



УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ
ГРАДЕЖЕН ФАКУЛТЕТ СКОПЈЕ



Фросина Александар Пановска Георгиевска

**УНАПРЕДУВАЊЕ НА УПРАВУВАЊЕТО СО
ВОДОСТОПАНСКИТЕ СИСТЕМИ СО ПРИМЕНА НА
МАШИНСКО УЧЕЊЕ**

Докторски труд

Скопје, 2026



УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ
ГРАДЕЖЕН ФАКУЛТЕТ СКОПЈЕ



Фросина Александар Пановска Георгиевска

УНАПРЕДУВАЊЕ НА УПРАВУВАЊЕТО СО ВОДОСТОПАНСКИТЕ СИСТЕМИ СО ПРИМЕНА НА МАШИНСКО УЧЕЊЕ

Докторски труд

Скопје, 2026

Докторанд:
ФРОСИНА АЛЕКСАНДАР ПАНОВСКА ГЕОРГИЕВСКА

Тема:
**УНАПРЕДУВАЊЕ НА УПРАВУВАЊЕТО СО ВОДОСТОПАНСКИТЕ СИСТЕМИ СО ПРИМЕНА НА
МАШИНСКО УЧЕЊЕ**
**ADVANCEMENT IN MANAGEMENT OF WATER RESOURCES SYSTEMS WITH APPLICATION
OF MACHINE LEARNING**

Ментор:
Проф. д-р СТЕВЧО МИТОВСКИ,
Универзитет „Св. Кирил и Методиј“ во Скопје, Градежен факултет Скопје

Комисија за одбрана на докторскиот труд:
Проф. д-р ЉУПЧО ПЕТКОВСКИ (претседател)
Универзитет „Св. Кирил и Методиј“ во Скопје, Градежен факултет Скопје

Проф. д-р МАРИЈАНА ЛАЗАРЕВСКА,
Универзитет „Св. Кирил и Методиј“ во Скопје, Градежен факултет Скопје

Проф. д-р ТИНА ДАШИЌ,
Универзитет во Белград, Градежен факултет, Р. Србија

Проф. д-р ФЕРНАНДО САЛАЗАР,
Меѓународен центар за нумерички методи во инженерството (The International Centre for Numerical Methods in Engineering – CIMNE, Madrid Branch), Политехнички универзитет во Каталонија, Барселона, Шпанија (The Universitat Politècnica de Catalunya · BarcelonaTech - UPC)

Вонр. проф. д-р МИЛИЦА МАРКОВИЌ,
Универзитет во Ниш, Факултет за градежништво и архитектура во Ниш, Р. Србија

Проф. д-р СТЕВЧО МИТОВСКИ,
Универзитет „Св. Кирил и Методиј“ во Скопје, Градежен факултет Скопје

Научна област:
ХИДРОТЕХНИКА

Датум на одбрана:
15.09.2026

УНАПРЕДУВАЊЕ НА УПРАВУВАЊЕТО СО ВОДОСТОПАНСКИТЕ СИСТЕМИ СО ПРИМЕНА НА МАШИНСКО УЧЕЊЕ

– А п с т р а к т –

Водните ресурси претставуваат едни од најзначајните природни ресурси, чија просторна и временска нерамномерна распределба ја наметнува потребата од нивно рационално, одржливо и интегрално управување. Во современите услови на зголемени потреби за водоснабдување, наводнување, производство на електрична енергија и заштита од штетното дејство на водите, хидросистемите со брани и акумулации имаат особено значајна улога. Истовремено, браните како капитални инженерски објекти бараат континуирано техничко набљудување, анализа на нивното однесување и навремено препознавање на можни промени кои можат да влијаат врз нивната стабилност и функционалност.

Предмет на оваа докторска дисертација е примената на методи од машинско учење во инженерството за брани, со фокус на два меѓусебно поврзани сегменти: управување со водостопански системи и анализа на долготрајното однесување на брани врз основа на податоци од техничко набљудување. Основната цел на истражувањето е да се испита применливоста, точноста и инженерската употребливост на различни алгоритми од машинско учење при моделирање на сложени нелинеарни процеси поврзани со работата на акумулациите, предвидување на хидролошки и експлоатациони параметри, како и анализа на параметри кои го опишуваат однесувањето на браните.

Во рамки на дисертацијата е изработен сопствен програмски код во Пајтон (анг. „Python“) средина, составен од повеќе функционални целини и 29 пресметковни ќелии, во кои се имплементирани девет алгоритми од машинско учење. Кодот овозможува внес, обработка и анализа на податоци, спроведување на вкрстена валидација, споредба на перформансите на моделите преку соодветни статистички показатели, избор на најсоодветен модел и примена на избраните модели за предвидување и инженерска интерпретација на резултатите. Како дел од истражувањето, кодот е тестиран и спореден со комерцијално достапниот софтвер „NeuralTools“, со цел оценување на неговата доверливост и практична применливост.

Истражувањето е организирано преку две главни студии на случај. Првата студија се однесува на примена на машинско учење во управување со хидросистемите „Стрежево“ и „Злетовица“, преку анализа и предвидување на испуштања од акумулација, производство на електрична енергија и варијација на водното ниво во акумулација. Втората студија се однесува на анализа на однесувањето на различни типови брани, при што се разгледуваат водни нивоа во пиезометри, филтрациони количества, вертикални слегнувања и радијални поместувања. Во анализите се користени реални податоци од хидросистеми и брани во Македонија и надвор од неа, меѓу кои хидросистемите „Стрежево“ и „Злетовица“, браната „Кнежево“ и бетонска лачна брана во Франција.

Резултатите од истражувањето покажуваат дека методите од машинско учење можат да претставуваат корисна дополнителна инженерска алатка за анализа, предвидување и поддршка на одлучувањето кај сложени хидротехнички системи. Нивната примена овозможува препознавање на зависности меѓу влезните и излезните параметри, надополнување на временски низи, споредба на различни модели и добивање прогностички информации кои можат да придонесат кон подобро управување со акумулациите и подобро разбирање на однесувањето на браните. Сепак, резултатите укажуваат и на потребата од внимателна инженерска интерпретација, особено во случаи со ограничен број податоци, мала варијабилност на целната променлива или појава на вредности надвор од опсегот на тренирачките податоци.

Научниот и практичниот придонес на дисертацијата се состои во развој на применлива методологија за имплементација на машинско учење во инженерството за брани, преку сопствен програмски код, анализа на реални студии на случај и споредба на повеќе алгоритми во различни инженерски услови. Добиените сознанија можат да придонесат кон унапредување на процесите на управување со водните ресурси, техничко набљудување на брани и процена на нивното долготрајно однесување, како и кон понатамошна примена на интелигентни пресметковни методи во хидротехничкото инженерство.

Клучни зборови: брани, акумулации, машинско учење, однесување на брани, управување со хидросистеми, метрички показатели

ADVANCEMENT IN MANAGEMENT OF WATER RESOURCES SYSTEMS WITH APPLICATION OF MACHINE LEARNING

- A b s t r a c t -

Water resources represent one of the most important natural resources, whose spatial and temporal uneven distribution imposes the need for their rational, sustainable, and integrated management. Under modern conditions of increasing demands for water supply, irrigation, electricity generation, and protection against the harmful effects of water, hydraulic systems with dams and reservoirs play an especially significant role. At the same time, dams, as major engineering structures, require continuous technical monitoring, analysis of their behavior, and timely identification of potential changes that may affect their stability and functionality.

The subject of this doctoral dissertation is the application of machine learning methods in dam engineering, with a focus on two interconnected segments: water resources systems management and analysis of the long-term behavior of dams based on technical monitoring data. The main objective of the research is to investigate the applicability, accuracy, and engineering usefulness of various machine learning algorithms in modeling complex nonlinear processes related to reservoir operation, prediction of hydrological and operational parameters, as well as analysis of parameters describing dam behavior.

Within the framework of the dissertation, a custom program code was developed in the Python environment, consisting of several functional units and 29 computational cells in which nine machine learning algorithms were implemented. The code enables data input, processing, and analysis, implementation of cross-validation procedures, comparison of model performances using appropriate statistical indicators, selection of the most suitable model, and application of the selected models for prediction and engineering interpretation of the results. As part of the research, the code was tested and compared with the commercially available software “NeuralTools” in order to evaluate its reliability and practical applicability.

The research is organized through two main case studies. The first study concerns the application of machine learning in the management of the “Streževo” and “Zletovica” hydraulic systems through the analysis and prediction of reservoir releases, hydropower generation, and reservoir water level variations. The second study focuses on the analysis of the behavior of different types of dams, including the investigation of water levels in piezometers, seepage quantities, vertical settlements, and radial displacements. The analyses were carried out using real data obtained from hydraulic systems and dams in North Macedonia and abroad, including the “Streževo” and “Zletovica” hydraulic systems, the “Kneževo” dam, and a concrete arch dam in France.

The research results indicate that machine learning methods can represent a useful supplementary engineering tool for analysis, prediction, and decision support in complex hydraulic engineering systems. Their application enables the identification of relationships between input and output parameters, completion of time series data, comparison of different models, and generation of prognostic information that may contribute to improved reservoir management and better understanding of dam behavior. Nevertheless, the results also indicate the need for careful engineering interpretation, especially in cases involving limited datasets, low variability of the target variable, or the occurrence of values outside the range of the training data.

The scientific and practical contribution of the dissertation lies in the development of an applicable methodology for implementing machine learning in dam engineering through custom program code, analysis of real case studies, and comparison of multiple algorithms under different engineering conditions. The obtained findings may contribute to the improvement of water resources management processes, technical monitoring of dams, and assessment of their long-term behavior, as well as to the further application of intelligent computational methods in hydraulic engineering.

Keywords: dams, reservoirs, machine learning, dam behavior, hydraulic systems management, performance metrics

Keywords: dams, reservoirs, machine learning, dam behavior, hydraulic systems management, performance metrics'

Изјавувам дека докторскиот труд е оригинален труд што го имам изработено самостојно.

Своерачен потпис на докторандот
(на електронската верзија потпис и скратеницата с.р.)

Изјавувам дека електронската верзија на докторскиот труд е идентична со отпечатениот докторски труд.

Потпис на авторот, с.р

Листа на слики

Слика 1. Централен попречен пресек на брана „Стрежево“.	- 21 -
Слика 2. Основа на брана „Кнежево“ со опремата за оскултација.	- 23 -
Слика 3. Централен попречен профил на брана „Кнежево“.	- 23 -
Слика 4. Проценето месечно производство на електрична енергија од „ХЕЦ Б“ (Aleksova Dosevska, 2025).	- 24 -
Слика 5. Основа на браната во Франција, со ознака на локацијата на висоците на браната во форма на свезди.	- 25 -
Слика 6. Преглед на централниот блок на бетонската лачна брана со позиција на висоците за мерење на радијалните поместувања на блокот.	- 25 -
Слика 7. Приказ на чекорите во развој на моделите во Python.	- 28 -
Слика 8. Разлика во тврдо и меко програмирање (Gupta & Kulkarni, 1993).	- 29 -
Слика 9. Структура на тренирање и прогноза кај алгоритмот „Случајна шума“.	- 32 -
Слика 10. Шематски приказ на класификациски проблем решен со алгоритмот KNN.	- 38 -
Слика 11. Шематски приказ на пресметката на целната варијабла кај BT моделите.	- 40 -
Слика 12. Шематски приказ на четирите слоеви кај GRNN моделот.	- 41 -
Слика 13. Графичка презентација на процесот на 5 - кратна вкрстена валидација (J. Martyr & R. Rogers, 2021).	- 44 -
Слика 14. Приказ на попречен пресек на брана „Кнежево“ со локација на пиезометрите FP_14 и FP_15 во фундаментот под браната.	- 120 -
Слика 15. Приказ на попречен пресек на брана „Кнежево“ со локација на пиезометрите „FP_2“ и „FP_3“ во фундаментот под браната.	- 134 -
Слика 16. Приказ на попречните пресеци на стационажа km 0+124, 0+149 и 0+174 по оска на круна на браната „Кнежево“ со позиција на функционалните келии за мерење на слегнувањата.	- 156 -

Листа на прилози

- Прилог 1. Графички приказ на користените вредности на предиктор варијаблата “CB3” и предиктор варијаблите “Zwl” и “T” во процесот на обучување на моделите со машинско учење за анализа на случај $CB3=f(Zwl,T)$ - 53 -
- Прилог 2. Графички приказ на предиктор варијаблите “Zwl” и “T” кои се користат во процесот на предвидување на целната варијабла “CB3” за анализа на случај $CB3=f(Zwl,T)$ - 53 -
- Прилог 3. Споредбен приказ на добиените резултати од моделите креирани со алгоритмот GRNN тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла “CB3” за анализа на случај $CB3=f(Zwl,T)$ - 54 -
- Прилог 4. Споредбен приказ на добиените резултати од моделите креирани со алгоритмот ANN тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла “CB3” за анализа на случај $CB3=f(Zwl,T)$ - 55 -
- Прилог 5. Графички приказ на пресметани разлики во предвидувањата помеѓу моделите во “NeuralTools” и Пајтон. - 55 -
- Прилог 6. Споредбен приказ на добиените резултати од четирите модели - два ANN и два GRNN модели тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла „CB3“ како и вредностите на предиктор варијаблите „Zwl“ и „T“, за анализа на случај $CB3=f(Zwl,T)$ - 56 -
- Прилог 7. График за варијаблите дотек „Q_inflow“ и истек „Q_discharge“ од акумулацијата, за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 60 -
- Прилог 8. График за варијаблата волумен на вода во акумулацијата „Reservoir_volume_10-6“ за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 60 -
- Прилог 9. График за варијаблата ниво на вода во акумулацијата „Zwl“ за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 60 -
- Прилог 10. Графички приказ на вредноста на R2 за секој анализиран алгоритам на сетот податоци за обучување со петкратна вкрстена валидација за анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 63 -
- Прилог 11. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 68 -
- Прилог 12. Приказ на MSE за секој обучен модел анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 69 -
- Прилог 13. Приказ на R2 за секој обучен модел анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 70 -
- Прилог 14. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 70 -
- Прилог 15. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 71 -
- Прилог 16. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF анализа на случај $Qdischarge = f(Zwl, ResVol, Qinflow)$ - 71 -
- Прилог 17. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT

анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 71 -
Прилог 18. Графици со резидуали за секој обучуван модел.....	- 72 -
Прилог 19. Графички приказ на предвидените вредности од моделите RF, KNN и BT во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флуктуацијата на водното ниво во акумулацијата „Zwl“.....	- 73 -
Прилог 20. Графички приказ на предвидените вредности од моделите SVR, DT и XGBoost во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флуктуацијата на водното ниво во акумулацијата „Zwl“.-	74
Прилог 21. Графички приказ на предвидените вредности од моделите GPR, GRNN и ANN во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флуктуацијата на водното ниво во акумулацијата „Zwl“.-	74
Прилог 22. 3Д график со приказ на пресметаните оптимални вредности на целната варијабла „Q_discharge“ во функција од трите предиктор варијабли „Zwl“, „Q_inflow“ и „Reservoir_volume_10-6“.....	- 75 -
Прилог 23. 3Д график со приказ на 20-те издвоени вредности на целната варијабла „Q_discharge“.	- 76 -
Прилог 24. Преглед на ковергентноста на моделот од Хаос оптимизација низ историскиот развој на итерациите.....	- 76 -
Прилог 25. График за варијаблите „P_kW“ - генерирана моќност, „Q_inflow“ - дотекот во акумулацијата и „Q_energy“ - енергетски обработениот проток за периодот од јануари, 2025 година до април, 2026 година, за приказ на историската низа на податоци која е земена предвид во процесот на обучување на моделите за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 78 -
Прилог 26. График за варијаблите „ResVol“ - волумен во акумулација и „Zwl“ - ниво на вода во акумулација за периодот од јануари, 2025 година до април, 2026 година, за приказ на историската низа на податоци која е земена предвид во процесот на обучување на моделите за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 78 -
Прилог 27. График со приказ на генерирана моќност во хидроцентралата „P_kW“, дотек во акумулацијата „Q_inflow“, енергетски обработен проток „Q_energy“ и проток кој се испушта низ темелниот испуст „Q_bottom outlet“, регистрирани во периодот на работа на хидроцентралата од 2025 до 2026 година за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 78 -
Прилог 28. График со приказ на предиктор варијаблата „Reservoir_volume_10-6“ кој ќе се применува во фазата на прогноза од обучените модели со машинско учење за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 79 -
Прилог 29. График со приказ на предиктор варијаблите „Q_inflow“ и „Q_discharge“ (или „Q_energy“) кои ќе се применуваат во фазата на прогноза од обучените модели со машинско учење за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 79 -
Прилог 30. График со приказ на предиктор варијаблата „Zwl“ која ќе се применува во фаза на прогноза од обучените модели со машинско учење за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 79 -
Прилог 31. Графички приказ на вредноста на R2 за секој анализиран алгоритам на сетот податоци за обучување со петкратна вкрстена валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 81 -
Прилог 32. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за	

анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 86 -
Прилог 33. Приказ на MSE за секој обучен модел за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 87 -
Прилог 34. Приказ на R2 за секој обучен модел за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 87 -
Прилог 35. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 88 -
Прилог 36. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 88 -
Прилог 37. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 89 -
Прилог 38. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 89 -
Прилог 39. Графици со резидуали за секој обучуван модел за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 90 -
Прилог 40. График со приказ на пресметаните вредности на целната варијабла „ P_{kW} “ пресметани со помош на сите девет обучени модели, приказ на предиктор варијаблите „ Q_{energy} “ и „ Q_{inflow} “	- 91 -
Прилог 41. График со приказ на пресметаните предвидени вредности на целната варијабла „ P_{kW} “ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на BT, DT и RF.	- 92 -
Прилог 42. График со приказ на пресметаните предвидени вредности на целната варијабла „ P_{kW} “ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на SVR, KNN и XGBoost.	- 92 -
Прилог 43. График со приказ на пресметаните предвидени вредности на целната варијабла „ P_{kW} “ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на ANN, GRNN и GPR.	- 93 -
Прилог 44. 3Д график со приказ на пресметаните оптимални вредности на целната варијабла P_{kW} во функција од четирите предиктор варијабли „ Zwl “, „ Q_{inflow} “, „ Q_{energy} “ и „ $Reservoir_volume_{10-6}$ “	- 94 -
Прилог 45. 3Д график со приказ на 20-те издвоени вредности на целната варијабла P_{kW}	- 94 -
Прилог 46. Преглед на ковергентноста на моделот од Хаос оптимизација низ историскиот развој на итерациите.	- 95 -
Прилог 47. Графички приказ на целната варијабла „ Zwl “ и предиктор варијаблите „ $Q_{hec_strezhevo}$ “, „ Q_{inflow} “, „ Q_{hec_mini} “ и „ $Q_{gl_dovoden_organ}$ “ со вредности кои се применуваат во процесот на обучување на моделите.	- 96 -
Прилог 48. Графички приказ на целната варијабла „ Zwl “ и предиктор варијаблата „ $Rain$ “ со вредности кои се применуваат во процесот на обучување на моделите.	- 96 -
Прилог 49. Графички приказ на целната варијабла „ Zwl “ и предиктор варијаблите „ $Q_{hec_strezhevo}$ “, „ Q_{inflow} “, „ Q_{hec_mini} “ и „ $Q_{gl_dovoden_organ}$ “ со вредности кои се применуваат во процесот на предвидување.	- 97 -
Прилог 50. Графички приказ на целната варијабла „ Zwl “ и предиктор варијаблата „ $Rain$ “ со вредности кои се применуваат во процесот на предвидување на целната варијабла. ...	- 97 -
Прилог 51. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за	

анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 104 -
Прилог 52. Приказ на R2 за секој обучен модел за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 105 -
Прилог 53. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 105 -
Прилог 54. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 106 -
Прилог 55. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 106 -
Прилог 56. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 106 -
Прилог 57. Графици со резидуали за секој обучуван модел за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 107 -
Прилог 58. Топлотна мапа со приказ на поврзаноста на предиктор варијаблите за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 108 -
Прилог 59. Топлотна мапа со приказ на сличноста во предвидувањата од различните модели тренирани на различни алгоритми за анализа на случај 1.3 – $Zwl = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	- 109 -
Прилог 60. Графички приказ на пресметаните разлики – e_i помеѓу предвидените и реалните вредности на целната варијабла, за секој трениран модел.....	- 110 -
Прилог 61. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со XGBoost, RF и BT, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.....	- 112 -
Прилог 62. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со ANN, GPR и SVR, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.....	- 112 -
Прилог 63. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со DT, GRNN и KNN, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.....	- 112 -
Прилог 64. 3Д визуелизација на пронајдените оптимални максимуми на целната функција $MAX F(Zwl): Zwl = f(Q_{inflow}, Rain_{10_6}, Q_{hec_mini}, Q_{hec_strezhevo}, Q_{gl_dovoden_kanal})$, со обојување на издвоени резултати.....	- 114 -
Прилог 65. Локација на издвоените 20 максимуми од оптимизацијата на целната варијабла Zwl во 3Д просторот на предиктор варијаблите „Q_inflow“, „Q_hec_strezhevo“ и „Rain_10-6“.....	- 115 -
Прилог 66. Конвергенција на оптимизацијата кон пронаоѓање на решение на функцијата $MAX F(Zwl): Zwl = f(Q_{inflow}, Rain_{10_6}, Q_{hec_mini}, Q_{hec_strezhevo}, Q_{gl_dovoden_kanal})$	- 115 -
Прилог 67. Приказ на вредностите на целните варијабли „FP_14“ и „FP_15“, како и предиктор варијаблата „Zwl“, користени во фазата на обучување на моделите.....	- 121 -
Прилог 68. Приказ на вредностите на предиктор варијаблата „Zwl“ за временскиот период за кој	

ќе се извршува предвидување на целните варијабли - нивоите во пиезометрите „FP_14“ и „FP_15“	- 122 -
Прилог 69. Пресметани резидуали по обучување на моделот за целната варијабла „FP_14“ со примена на 20% задржан сет на податоци за валидација на моделот.	- 124 -
Прилог 70. Пресметани резидуали по обучување на моделот за целната варијабла „FP_15“ со примена на 20% задржан сет на податоци за валидација на моделот.	- 124 -
Прилог 71. Важност на варијаблите пресметана за моделот RF за анализа на случај $FP_{14}=f(Zwl, dZw, p)$	- 125 -
Прилог 72. Пермутациска важност пресметана за моделот RF за анализа на случај $FP_{14}=f(Zwl, dZw, p)$	- 125 -
Прилог 73. График со приказ на предвидените вредности на целните варијабли „FP_14“ и „FP_15“ со моделите тренирани со алгоритмите RF и BT, соодветно.....	- 125 -
Прилог 74. График со приказ на вредностите на целната варијабла „V_notch_left“ и предиктор варијаблите „Zwl“, „FP_14“ и „FP_15“ кои се користат во процесот на обучување на моделите.	- 126 -
Прилог 75. График со приказ на вредностите на целната варијабла „V_notch_left“ и предиктор варијаблата „Q_inflow“ кои се користат во процесот на обучување на моделите.	- 127 -
Прилог 76. График со приказ на предиктор варијаблите „Zw“, „FP_14“, „FP_15“ и „Q_inflow“ за периодот на предвидување на целната варијабла „V_notch_left“.....	- 127 -
Прилог 77. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација анализа на случај $V_notch_left=f(Zwl, Q_inflow, FP_{14}, FP_{15})$	- 129 -
Прилог 78. График со приказ на предвидените вредности на целната варијабла 'V_notch_left' (график GRNN_preds), предиктор варијаблите „Zwl“, „FP_14“ и „FP_15“.....	- 130 -
Прилог 79. График со приказ на предвидените вредности на целната варијабла „V_notch_left“ (график GRNN_preds), и предиктор варијаблата „Q_inflow“	- 131 -
Прилог 80. График со приказ на регистрираните филтрациони протекувања „V_notch_left (observed)“ и пресметаните филтрациони истекувања (predicted) согласно предвидувањата од моделот обучен со GRNN.	- 131 -
Прилог 81. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“ без изостанување на екстремните измерени вредности кои значително отстапуваат од границите.....	- 133 -
Прилог 82. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“.....	- 133 -
Прилог 83. График со приказ на повисоките граници на предупредување „T_high“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“.....	- 133 -
Прилог 84. Приказ на вредностите на целните варијабли „FP_2“ и „FP_3“, како и предиктор варијаблата „Zwl“.	- 134 -
Прилог 85. Приказ на вредностите на предиктор варијаблата „Zwl“ за временскиот период за кој ќе се извршува предвидување на целните варијабли - нивоите во пиезометрите „FP_2“ и „FP_2“	- 135 -
Прилог 86. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за моделот $FP_3=f(Zwl)$	- 138 -

Прилог 87. График со приказ на предвидувањата на целните варијабли FP_2 и FP_3 во однос на предиктор варијаблата Zwl за разгледуваниот временски период од 2018 до 2025 година.....	138 -
Прилог 88. График со приказ на вредностите на целната варијабла „V_notch_right“ и предиктор варијаблите „FP_2“, „FP_3“ и „Zwl“ применети во процесот на обучување на моделите.	139 -
Прилог 89. График со приказ на вредностите на целната варијабла „V_notch_right“ и предиктор варијаблата „Q_inflow“ кои се применети во процесот на обучување на моделите.-	139 -
Прилог 90. График со приказ на вредностите на предиктор варијаблите „FP_2“, „FP_3“, „Zwl“ и „Q_inflow“ применети во процесот на предвидување на целната варијабла „V_notch_right“.	140 -
Прилог 91. График со приказ на предвидените вредности на целната варијабла „V_notch_right“, предиктор варијаблите „Zwl“, „FP_2“ и „FP_3“	145 -
Прилог 92. График со приказ на предвидените вредности на целната варијабла „V_notch_right“, и предиктор варијаблата „Q_inflow“	145 -
Прилог 93. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_right“, со приказ на измерените вредности за филтрацијата во разгледуваниот период „V_notch_right_measured“	147 -
Прилог 94. График со приказ на повисоките граници на предупредување „T_high“ на предвидените податоци за целната варијабла „V_notch_right“, со приказ на измерените вредности за филтрацијата во разгледуваниот период „V_notch_right_measured“	147 -
Прилог 95. График со приказ на вредностите на целната варијабла „V_notch_middle“ и предиктор варијаблата „Zwl“ применети во процесот на обучување на моделите.	148 -
Прилог 96. График со приказ на вредностите на целната варијабла „V_notch_middle“ и предиктор варијаблата „Q_inflow“ кои се применети во процесот на обучување на моделите.-	148 -
Прилог 97. График со приказ на вредностите на предиктор варијаблите „Zwl“ и „Q_inflow“ применети во процесот на предвидување на целната варијабла „V_notch_middle“.-	148 -
Прилог 98. Графички приказ на пресметаните предвидувања на целната варијабла „V_notch_middle“ со моделот трениран на GPR алгоритмот.	152 -
Прилог 99. Графички приказ на пресметаните предвидувања на целната варијабла „V_notch_middle“ со моделите тренирани на RF и SVR алгоритмите.	152 -
Прилог 100. Графички приказ на осреднетите вредности на целната варијабла „V_notch_middle“ и предиктор варијаблата „Zwl“	153 -
Прилог 101. Графички приказ на осреднетите вредности на целната варијабла „V_notch_middle“ и предиктор варијаблата „Q_inflow“	154 -
Прилог 102. График со приказ на коридорот со пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_middle“, со приказ и на измерените вредности на целната варијабла во разгледуваниот период.	155 -
Прилог 103. График со приказ на коридорот со повисоките граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_middle“, со приказ и на измерените вредности на целната варијабла во разгледуваниот период.	155 -
Прилог 104. Графички приказ на мерените слегнувања „SC“ во сите 10 келии кои се предмет на анализа за периодот од 2018 до 2024 година, заедно со мереното ниво во	

акумулацијата „Zwl“ во истите временски периоди на регистрирање на слегнувањата.	- 156 -
Прилог 105. Графички приказ на водното ниво во акумулацијата „Zwl“ за секој ден почнувајќи од 01.11.2018 година па се до 31.12.2025 година - податоци кои се користат во сите модели за слегнувања во функција на предиктор варијабли.	- 157 -
Прилог 106. Графички приказ на пресметаните предвидувања за $SC1=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC1, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 163 -
Прилог 107. Графички приказ на пресметаните предвидувања за $SC2=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC2_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 164 -
Прилог 108. Графички приказ на пресметаните предвидувања за $SC4=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC4_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 164 -
Прилог 109. Графички приказ на пресметаните предвидувања за $SC5=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC5_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 164 -
Прилог 110. Графички приказ на пресметаните предвидувања за $SC6=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC6_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 165 -
Прилог 111. Графички приказ на пресметаните предвидувања за $SC7=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC7_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 165 -
Прилог 112. Графички приказ на пресметаните предвидувања за $SC9=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC9_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 165 -
Прилог 113. Графички приказ на пресметаните предвидувања за $SC12=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC12_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 166 -
Прилог 114. Графички приказ на пресметаните предвидувања за $SC13=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC13_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 166 -
Прилог 115. Графички приказ на пресметаните предвидувања за $SC14=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC14_measured, и приказ на мереното водно ниво во акумулацијата - предиктор варијаблата „Zwl“.	- 166 -
Прилог 116. Пресметани предвидувања за слегнувањата во ќелиите “SC1”, “SC2”, “SC4”, “SC5”, “SC6”, “SC7”, “SC9”, “SC12”, “SC13” и “SC14” со дополнет приказ на предиктор варијаблата „Zwl“ - ниво на вода во акумулацијата.	- 167 -
Прилог 117. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерните ќелии „SC1“, „SC2“ и „SC4“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.	- 170 -
Прилог 118. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерните ќелии „SC5“, „SC6“ и „SC7“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.	- 170 -
Прилог 119. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во	

мерните ќелии „SC9“, „SC12“ и „SC13“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.....	- 171 -
Прилог 120. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерната ќелија „SC14“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.....	- 172 -
Прилог 121. Графички приказ на користените вредности на предиктор варијаблата „CB2“ и предиктор варијаблите „Zwl“ и „T“ во процесот на обучување на моделите со машинско учење за анализа на случај $CB2=f(Zwl,T)$	- 172 -
Прилог 122. Графички приказ на предиктор варијаблите „Zwl“ и „T“ кои се користат во процесот на предвидување на целната варијабла „CB2“ за анализа на случај $CB2=f(Zwl,T)$..	- 173 -
Прилог 123. График на важност на варијаблите (анг. „Feature importance“) во моделот со алгоритмот RF.....	- 177 -
Прилог 124. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со алгоритмот SVR.....	- 178 -
Прилог 125. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со алгоритмот RF.....	- 178 -
Прилог 126. Графици со резидуали за трите обучени модели.	- 179 -
Прилог 127. Графички приказ на пресметаните предвидувања на целната варијабла „CB2“ во функција од предиктор варијаблите „Zwl“ и „T“	- 180 -

Листа на табели

Табела 1. Конструктивни големина за брана „Стрежево“	- 20 -
Табела 2. Преглед на повеќе алгоритми за машинско учење и минималниот потребен број на податоци за тренирање на моделите.	- 30 -
Табела 3. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 61 -
Табела 4. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 64 -
Табела 5. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 65 -
Табела 6. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 66 -
Табела 7. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 66 -
Табела 8. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 67 -
Табела 9. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 68 -
Табела 10. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 68 -
Табела 11. Резултати за вредностите на метричките показатели по процесот на обучување во рамки на кој во секој модел е спроведена и петкратна вкрстена валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	- 69 -
Табела 12. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 80 -
Табела 13. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 82 -
Табела 14. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 83 -
Табела 15. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 83 -
Табела 16. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 84 -
Табела 17. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за анализа на случај $P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	- 85 -

Табела 18. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 85 -
Табела 19. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 86 -
Табела 20. Резултати за вредностите на метричките показатели по процесот на обучување во рамки на кој во секој модел е спроведена и петкратна вкрстена валидација за анализа на случај $P_{kW} = f(Zwl, ResVol, Qinflow, Qenergy)$	- 86 -
Табела 21. Резултати од прелиминарната петкратна вкрстена валидација изработена на деветте модели со машинско учење за $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 98 -
Табела 22. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 100 -
Табела 23. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 101 -
Табела 24. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 101 -
Табела 25. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 102 -
Табела 26. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 103 -
Табела 27. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 103 -
Табела 28. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за анализа на случај 1.3 – $Zwl = f(Qinflow, Rain, Qhec_strezhevo, Qmini_hec, Qgl_dovoden_organ)$	- 104 -
Табела 29. Нумерички вредности на максималните, минималните и просечните разлики помеѓу реалните вредности на целната варијабла и пресметаните вредности со тренираните модели.	- 111 -
Табела 30. Табеларен приказ на метричките показатели на најдобрите модели во анализите на случај дел од студија на случај 1.	- 116 -
Табела 31. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_14“.	- 122 -
Табела 32. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_15“.	- 123 -
Табела 33. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за моделот $V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15)$	- 128 -
Табела 34. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација	

анализа на случај $V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15)$	- 129 -
Табела 35. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_2“.....	- 135 -
Табела 36. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_3“.....	- 136 -
Табела 37. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $FP_2=f(Zwl)$	- 137 -
Табела 38. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за моделот $FP_3=f(Zwl)$	- 138 -
Табела 39. Резултати од метричките показатели од прелиминарната петкратна вкрстена валидација на моделите за $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$	- 140 -
Табела 40. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$	- 142 -
Табела 41. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за студија на случај $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$	- 142 -
Табела 42. Табеларен приказ на метричките показатели од изработената петкратна вкрстена валидација на студија на случај $V_notch_middle=f(Zwl, Q_inflow)$	- 149 -
Табела 43. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај $V_notch_middle=f(Zwl, Q_inflow)$. -	150
Табела 44. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за студија на случај $V_notch_middle=f(Zwl, Q_inflow)$	- 151 -
Табела 45. Максимални теоретски пресметани слегнувања за секоја ќелја која е предмет на анализа, во [cm].....	- 157 -
Табела 46. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC1=f(Zwl, t)$	- 157 -
Табела 47. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC2=f(Zwl, t)$	- 158 -
Табела 48. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC4=f(Zwl, t)$	- 158 -
Табела 49. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC5=f(Zwl, t)$	- 158 -
Табела 50. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC6=f(Zwl, t)$	- 158 -
Табела 51. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC7=f(Zwl, t)$	- 159 -
Табела 52. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC9=f(Zwl, t)$	- 159 -
Табела 53. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC12=f(Zwl, t)$	- 159 -

Табела 54. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC13=f(Zw1,t)$	- 160 -
Табела 55. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC14=f(Zw1,t)$	- 160 -
Табела 56. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за седумте студии на случај $SC1=f(Zw1,t)$, $SC2=f(Zw1,t)$, $SC5=f(Zw1,t)$, $SC7=f(Zw1,t)$, $SC9=f(Zw1,t)$, $SC12=f(Zw1,t)$ и $SC13=f(Zw1,t)$	- 162 -
Табела 57. Просечна вредност на метричките показатели од процесот на обучување на BT моделите по спроведена петкратна валидација, за студии на случај $SC4=f(Zw1,t)$, $SC6=f(Zw1,t)$ и $SC14=f(Zw1,t)$	- 163 -
Табела 58. Табеларен приказ на дел од резултатите од пресметката на екстраполираните слегнувања во десетте ќелии, со вредности на времето на достигнување на 70% од пресметаните теоретски слегнувања.	- 169 -
Табела 59. Табеларен приказ на дел од резултатите од пресметката на екстраполираните слегнувања во десетте ќелии, со вредности на времето на достигнување на 80% и 90% од пресметаните теоретски слегнувања.	- 169 -
Табела 60. Резултати од прелиминарна петкратна вкрстена валидација за анализа на случај 2.7 - $CB2=f(Zw1,T)$	- 173 -
Табела 61. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај $CB2=f(Zw1,T)$	- 175 -
Табела 62. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $CB2=f(Zw1,T)$	- 176 -
Табела 63. Табеларен приказ на метричките показатели од секој сет на петкратната вкрстена валидација во рамки на обучувањето на моделот со SVR, за анализа на случај $CB2=f(Zw1,T)$	- 176 -
Табела 64. Пресметани максимални и минимални радијални поместувања „CB2“ со секој од трите обучени модели со машинско учење за $CB2=f(Zw1,T)$	- 180 -
Табела 65. Табеларен приказ на метричките показатели на најдобрите модели во анализите на случај дел од студија на случај 2.	- 181 -

Листа на кратенки

CNN	Конволуциони невронски мрежи (анг. „Convolutional Neural Networks“)
LSTM	Невронска мрежа со долго – краткоторчна меморија (анг. „Long short-term memory“)
ML	Машинско учење (анг. „Machine Learning“)
ANN	Вештачки невронски мрежи (анг. „Artificial Neural Networks“)
RF	Случајна шума (анг. „Random Forest“)
SVR	Регресија со потпомогнати вектори (анг. „Support Vector Regressor“)
DT	Одлучувачки дрва (анг. „Decision Trees“)
KNN	К – најблиски соседи (анг. „K – nearest neighbors“)
BT	Поткрени дрва (анг. „Boosted trees“)
GRNN	Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“)
GPR	Гаусов процесен регресор (анг. „Gaussian Process Regressor“)
XGBoost	XGBoost алгоритам за градиентно поткревање
CGO	Хаос оптимизација (анг. „Chaos Game Optimization“)
R^2	Коефициент на детерминација
CV	Вкрстена валидација (анг. „Cross - validation“)
Zwl	Ниво на вода во акумулација
SC	Вертикални слегнувања во келија за слегнувања (анг. „Settlement cell“)
Q_{inflow}	Дотек во акумулација
Q_{energy}	Енергетски обработен проток
$Q_{discharge}$	Проток кој се испушта од акумулацијата
Q_{hec}	Проток кој се испушта од хидроцентрала
ResVol	Волумен на вода во акумулација
ХЕЦ	Хидроелектрична централа
FP	Пиезометар во основата под браната (анг. „Foundation Piezometer“)
CB	Поместувања на вертикален висок
T	Температура
t	Време
DamResML	Програмскиот код изработен во Пајтон средина
P_kW	Генерирана моќност
Rain	Регистриран дожд

СОДРЖИНА

Вовед.....	- 1 -
1. Предмет и цел на истражувањето	- 6 -
2. Преглед на досегашни истражувања во областа	- 8 -
2.1. Кибернетика и развој на програмските јазици.....	- 8 -
2.2. Досегашна примена на машинско учење во инженерство за брани и водостопански системи	- 9 -
3. Опис на анализираните хидросистеми со брани и акумулации.....	- 18 -
3.1. Хидросистем „Стрежево“	- 20 -
3.2. Хидросистем „Злетовица“	- 22 -
3.3. Лачна брана во Франција.....	- 24 -
4. Опис на креираниот код DamResML	- 26 -
4.1. Креирање на DamResML во Пајтон средина	- 26 -
4.2. Машинско учење.....	- 29 -
4.2.1. Подготовка на базата на податоци	- 30 -
4.2.2. Избор на алгоритми од машинско учење.....	- 31 -
4.2.3. Прелиминарна валидација на моделите	- 43 -
4.2.4. Избор на алгоритми за обучување на моделите	- 45 -
4.2.5. Предвидување на целната варијабла	- 45 -
4.3. Хаос оптимизација	- 46 -
4.3.1. Дефинирање на влезни податоци и модели	- 47 -
4.3.2. Дефинирање на граници и ограничувања	- 47 -
4.3.3. Точност на решенијата.....	- 47 -
4.3.4. Генерирање на нови решенија (CGO механизам)	- 47 -
4.3.5. Целна функција	- 47 -
4.3.6. Визуелизација и интерпретација.....	- 48 -
5. Тестирање и валидација на DamResML.....	- 49 -
5.1. „NeuralTools“	- 49 -
5.1.1. Досегашна примена на „NeuralTools“ во објавени истражувања	- 50 -
5.2. Споредба на резултати од модел во „NeuralTools“ и модел во DamResML	- 52 -
5.2.1. Подготовка на податоците за модел $CB3=f(Z_{wl}, T)$	- 52 -
5.2.2. Резултати од споредбата на предвидувањата	- 53 -
5.3. Сублимаат од спроведената споредба на резултатите	- 56 -
6. Студија на случај 1.....	- 58 -
6.1. Анализа на случај 1.1 – $Q_{\text{discharge}} = f(Z_{wl}, \text{ResVol}, Q_{\text{inflow}})$	- 59 -
6.1.1. Подготовка на податоците	- 59 -
6.1.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите	- 61 -
6.1.3. Обучување на моделите.....	- 63 -

6.1.4. Резултати од обучувањето на моделите	69 -
6.1.5. Предвидување на целната варијабла	72 -
6.1.6. Хаос оптимизација на целната варијабла	74 -
6.2. Анализа на случај 1.2 – $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$	77 -
6.2.1. Подготовка на податоците.....	77 -
6.2.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите	79 -
6.2.3. Обучување на моделите	81 -
6.2.4. Резултати од обучувањето на моделите.....	86 -
6.2.5. Предвидување на целната варијабла	91 -
6.2.6. Хаос оптимизација на целната варијабла	93 -
6.3. Анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$	95 -
6.3.1. Подготовка на податоците.....	95 -
6.3.2. Прелиминарна петкратна вкрстена валидација	97 -
6.3.3. Обучување на моделите	99 -
6.3.4. Резултати од обучувањето на моделите.....	105 -
6.3.5. Предвидување на целната варијабла	109 -
6.3.6. Хаос оптимизација на целната варијабла	113 -
6.4. Сублигат од спроведената студија на случај 1.....	116 -
6.4.1. Согледувања од добиените метрички показатели од студија на случај 1	116 -
6.4.2. Коментар на добиените резултати од студија на случај 1.....	116 -
7. Студија на случај 2.....	119 -
7.1. Анализа на случај 2.1 – $FP_{14}=f(Z_{wl})$ и $FP_{15}=f(Z_{wl})$	120 -
7.1.1. Подготовка на податоците	120 -
7.1.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите ...	122 -
7.1.3. Обучување на моделите	123 -
7.1.4. Предвидување на целната варијабла.....	125 -
7.2. Анализа на случај 2.2 – $V_{notch_left}=f(Z_{wl}, Q_{inflow}, FP_{14}, FP_{15})$	126 -
7.2.1. Подготовка на податоците	126 -
7.2.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите...	127 -
7.2.3. Обучување на моделот со GRNN.....	129 -
7.2.4. Предвидување на целната варијабла	130 -
7.2.5. Дефинирање на граници за предупредување (thresholds)	131 -
7.3. Анализа на случај 2.3 – $FP_{2}=f(Z_{wl})$ и $FP_{3}=f(Z_{wl})$	134 -
7.3.1. Подготовка на податоците	134 -
7.3.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите...	135 -
7.3.3. Обучување на моделите.....	137 -
7.3.4. Предвидување на целните варијабли $FP_{2}=f(Z_{wl})$ и $FP_{3}=f(Z_{wl})$	138 -
7.4. Анализа на случај 2.4 – $V_{notch_right}=f(Z_{wl}, Q_{inflow}, FP_{2}, FP_{3})$	139 -

7.4.1. Подготовка на податоците	139 -
7.4.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите ...	140 -
7.4.3. Обучување на моделите.....	141 -
7.4.4. Концепт на тежинско осреднување (анг. „weighted averaging“).....	143 -
7.4.5. Предвидување на целната варијабла	144 -
7.4.6. Дефинирање на граници за предупредување (thresholds)	146 -
7.5. Анализа на случај 2.5 – $V_notch_middle=f(ZwI, Q_inflow)$	147 -
7.5.1. Подготовка на податоците	147 -
7.5.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите ...	148 -
7.5.3. Обучување на моделите.....	150 -
7.5.4. Предвидување на целната варијабла со тежинско осреднување	152 -
7.5.5. Дефинирање на граници за предупредување (thresholds)	154 -
7.6. Анализа на случај 2.6 – $SC_i=f(ZwI,t)$	155 -
7.6.1. Подготовка на податоците	155 -
7.6.2. Прелиминарна петкратна вкрстена валидација.....	157 -
7.6.3. Обучување на моделите.....	161 -
7.6.4. Предвидување на целните варијабли $SC_i=f(ZwI,t)$	163 -
7.6.5. Екстраполација на слегнувањата.....	167 -
7.7. Анализа на случај 2.7 – $CB2=f(ZwI,T)$	172 -
7.7.1. Подготовка на податоците	172 -
7.7.2. Прелиминарна петкратна вкрстена валидација	173 -
7.7.3. Обучување на моделите	175 -
7.7.4. Резултати по обучувањето на моделите	177 -
7.7.5. Предвидување на целната варијабла.....	179 -
7.6. Сублимаг од спроведената студија на случај 2.....	181 -
7.6.1. Согледувања од добиените метрички показатели од студија на случај 2	181 -
7.6.2. Коментар на добиените резултати од студија на случај 2.....	182 -
8. Заклучоци и препораки за понатамошни истражувања	184 -
8.1. Општи заклучоци.....	184 -
8.2. Заклучоци од студија на случај 1 – примена на машинско учење во управување со водостопански системи	185 -
8.2.1. Научни согледувања од анализа на случај 1.1	186 -
8.2.2. Научни согледувања од анализа на случај 1.2	186 -
8.2.3. Научни согледувања од анализа на случај 1.3	187 -
8.2.4. Заклучоци од студија на случај 1	187 -
8.3. Заклучоци од студија на случај 2 – примена на машинско учење во анализа на однесувањето на брани.....	188 -
8.3.1. Научни согледувања од анализа на случај 2.1.....	188 -
8.3.2. Научни согледувања од анализа на случај 2.2	189 -

8.3.3. Научни согледувања од анализа на случај 2.3	- 190 -
8.3.4. Научни согледувања од анализа на случај 2.4	- 190 -
8.3.5. Научни согледување од анализа на случај 2.5	- 190 -
8.3.6. Научни согледувања од анализа на случај 2.6	- 191 -
8.3.7. Научни согледувања од анализа на случај 2.7	- 192 -
8.3.8. Заклучоци од студија на случај 2	- 192 -
8.4. Препораки за идни истражувања	- 193 -
Користена литература	- 195 -

ВОВЕД

Водата претставува национално богатство. Нејзиното ползување се базира на потребите на општеството и поединците, засновано на постулатите на добро од општ интерес и демократско располагање, со примена на концепт на интегралност и одржливост при управувањето со водните ресурси (Djordjevic, 1990; Votruba & Broža, 1989; Merrett, 1997). Таа претставува незаменлива материја за опстанокот на човекот, растителниот и животинскиот свет. Покрај тоа, се ползува за потребите на наводнување и како основна сировина во индустриските процеси, како и за производство на електрична енергија. Познато е дека над 70% од површината на планетата Земја е покриена со вода, но само околу 3% отпаѓаат на слатките води, кои може да се користат од страна на човечката популација (Agnew & Woodhuse, 2011). Од друга страна е притисокот кон водните ресурси, кој е значително зголемен по појавата на индустриската револуција од крајот на 18-тиот век, кога бројноста на населението на планетата Земја почнува експоненцијално да се зголемува. За споредба, во првите години од новата ера проценето е дека во светот живееле околу 200 милиони жители, додека во времето на индустриската револуција апроксимативно живееле 1 милијарда жители во светот. Денес, нешто повеќе од 2 века по индустриската револуција, во светот живеат над 8 милијарди жители (Roser & Ritchie, 2023). Зголемената бројност на населението влијае пропорционално на зголемување на потребите на вода за водоснабдување и наводнување, како и на зголемено производство на електрична енергија како обновлив извор на енергија. За споредба, од 1950 година до денес, потребите за вода за пиење се зголемени за повеќе од три пати, што предизвикува своевиден дисбаланс во однос на расположливите водни ресурси.

Значителна неповолност е фактот дека водните ресурси се просторно и временски нерамномерни распределени (Davie, 2008). Оттука, единствен начин за да се користи овој бесценет ресурс е истиот да се акумулира, што се постигнува со изградба на инженерски објекти – брани, со кои се овозможува формирање на водни акумулации (Tanchev, 2014; Novak, Moffat, Nalluri, & Narayanan, 2007; Chen S. H., 2015). Преку процес на управување со водните ресурси (акумулациите) се обезбедува задоволување на потребите за вода за различните намени (Loucks & Beek, 2017). Имено, со користење на водата од акумулациите човекот обезбедува вода за пиење за населените места, за наводнување на земојделски површини и енергетски го користи нејзиниот потенцијал за производство на електрична енергија од обновлив извор на енергија. Од друга страна, со формирањето на акумулациите, човекот се штити и од штетното дејство на водите, со намалување на ефектите од поплавен бран при појава на интензивни врнежи.

Низ светот, брани со акумулации се градат уште од времето на античките Египќани и Месопотамци (Schnitter, 1994; Smith, 1971). Сепак, факт е дека нашите претходници немале теоретски

познавања за примената на физиката и математиката, па изградбата на овие хидротехнички објекти се сведува на искусствено знаење. Браната Сад ел – Кафара (Schnitter, 1994; Smith, 1971), изградена во 2600 година пне во близина на градот Каиро во Египет, сèуште постои (во одреден обем, бидејќи дел од истата е разрушен поради поплави) и сведочи за инженерството на брани во време кога населението се борело со надојдените води од реката Нил. По 18-тиот век, со продлабочување на теоретските познавања од областа на хидрауликата, теоријата на конструкциите и геотехниката, изградбата на браните започнува нова ера кога изборот на материјали, усвојувањето на димензиите на браните и процената на нивната стабилност се базира на закони од физиката и од математиката.

Денес, вештачката интелигенција (која го вклучува и машинското учење) го наоѓа своето место во речиси секој аспект од инженерството (Flach, 2012; Marsland, 2009; Flood I. , Simulating the construction process using neural networks, 1990; Flood & Kartam, Neural networks in civil engineering II: Systems and application, 1994; Flood & Paul, 1996), соодветно - и во областа на градежното инженерство. Иако сèуште со извесна доза на резерва и внимателност, академскиот кадар и инженерската фела во областа на градежништвото го следат развојот на вештачката интелигенција со тенденција да најдат соодветна примена на овие напредни методи во различните области. Вештачките невронски мрежи се типичен пример за модерен интердисциплинарен пристап кој помага во решавање на различни инженерски проблеми кои не можат да се третираат со досега применуваното нумеричко моделирање и статистички методи. Невронските мрежи се способни да собираат, меморираат, анализираат, обработуваат и корелираат мноштво на податоци добиени од експериментални (лабораториски или теренски) истражувања, од техничко набљудување на брани и придружни објекти или од спроведени нумерички анализи (Sammur & Webb, 2011). Тие се илустрација на софистицирана техника на моделирање која може да се користи за решавање на многу сложени проблеми. „Тренираната“ невронска мрежа служи како аналитичка алатка за доверливи предикции на резултатите за серија на влезни податоци кои не биле вклучени во процесот на „учење“ на мрежата. Нивната оперативност е поедноставена и се одликува со соодветен степен на коректност. Вештачките невронски мрежи, заедно со фази логиката и генетските алгоритми, спаѓаат во групата на симболични методи на интелигентни пресметки и обработка на податоци кои функционираат според принципите на т.н. „меко“ пресметување (анг. „soft computing“) (Zadeh, 1993). Невронските мрежи се развиени како резултат на позитивните карактеристики на неколку различни истражувачки сегменти: обработка на податоци, невро-биологија и физика (Kecman, 2001). Истражувањата во областа на невронските мрежи покажале дека имаат солиден успех во предвидувањето на сериите на податоци со генерирање на меѓузависности помеѓу анализираните параметри (варијабли и целни параметри) и според тоа можат да се користат за создавање прогностички модели за различен вид на проблеми и тези во сферата на инженерството. Затоа, развиените модели за вештачки неврони само не потсетуваат на структурата на биолошките неврони и немаат претензии да ја копираат нивната вистинска состојба. Вештачкиот неврон ги прима влезните сигнали и ги генерира излезните сигнали (Kecman, 2001).

Водостопанските системи се состојат од брана (како клучен објект) со водна акумулација и придружни објекти, кои функционираат како целина и овозможуваат системот да ја исполни својата цел т.е. задоволување на потребите за вода за различните намени. Тука од првостепено значење се два сегменти: стабилноста на браната и придружните објекти и функционалноста на водостопанскиот систем. Иако навидум се работи за два различни сегменти, сепак тие се поврзани меѓу себе бидејќи првиот сегмент - стабилноста на браната и придружните објекти е предуслов за вториот сегмент - функционалноста на водостопанскиот систем (исполнување на целната функција т.е. целите и намените за кои е изграден системот). Со процесот на техничко набљудување на браните, кој е базиран на сознанијата за однесувањето на браната според креираниот нумерички модел и спроведената нумеричка анализа, се следат т.е. се врши мерење на поместувања на браната (внатрешни и надворешни), напрегањата, филтрационите протекувања и појава на зони на процедни води, ниво на вода во пиезометри (отворени и затворени) и појава на нестабилност (свечишта и деформаци, кои можат да бидат од локален карактер или пак од глобален, со што се загрозува сигурноста на браната и придружните објекти). Во рамки на техничкото набљудување се следат и климатски и хидрометеоролошки големини (атмосферски врнежи, температура, влажност на воздух, атмосферски притисок), кои имаат соодветно влијание

како врз однесувањето на браната од аспект на стабилност така и од ефективното и рационално управување со водниот ресурс (акумулацијата). Имено, управувањето со водните ресурси се базира на следење на влезните хидролошки големина (дотекувања) во акумулацијата, степенот на акумулирање и излезните големина (регулирани протекувања), преку процес на нивно континуирано мерење и интерпретација, со што се донесува иновирано управувачко правила и се подобрува оперативноста на водостопанскиот систем во идниот период (Bozorg, O. H.; ed., 2021; Biswas, 1975; Grafton & Hussey, 2011; Zeinalie, Bozorg-Haddad, & B., 2021; Wang, Yang, & Wang, 2016; Jain & Singh, 2003; Pidd, 2004). Според тоа, примената на напредни методи од машинско учење во наведените два клучни сегменти од инженерството за брани е императив во контекст на стекнување на нови сознанија, следење и подобрување на степенот на сигурност на браната како и подобрување на степенот на функционалност на водостопанскиот систем (генерирање на повисоки добивки од системот) преку генерирање на прогностички модел за однесувањето на истите во идниот период.

Браните со водните акумулации се капитални објекти, кои на денешен степен на развој се засноваат на концептот на интегралност и одржливост (Grigg, 2016). Според тоа, неминовно е континуирано техничко набљудување (оскултација) на состојбата и однесувањето на браните како и примена на напредни методи за анализа на одговорот конструкцијата при различните фази на оптоварување во еден здружен процес. Управувањето со акумулациите е од есенцијална важност во рамки на современиот начин на живот. Подобрувањето на процесот на управување со водниот ресурс, акумулиран во корисниот простор на акумулацијата е преку континуирано следење на клучните параметри (хидролошко дотекување, ниво на вода во акумулација и регулирани протекувања од акумулацијата) и анализа на степенот на задоволување на поставените цели т.е. критериумот за исполнување на потребите за вода по време, простор и количина за различните намени (Loucks & Van Beek, 1981; Mays, 2005). Примената на напредни методи, во насока на стекнување на нови сознанија за одговорот и однесувањето на браните при различни оптоварувања и надворешни влијанија како и во однос на подобрување на процесот на управувањето со водостопанските системи е од круцијално значење за вака сложените системи. Во предметната докторска дисертација е даден фокус и акцент на примена на методи од машинско учење во наведените два сегменти, преку процес на „учење и тренирање“ на алгоритмите од машинско учење и генерирање на образец за однесување на браната во идниот период и управувањето со водостопанските системи, што овозможува да се стекнат нови сознанија и информации, а во насока на подобрување на сигурноста и функционалноста на овие сложени системи.

Истражувањето кое е дел од предметната докторска дисертација се фокусира на примената на различни алгоритми од машинско учење во повеќе задачи од инженерството за брани, поделени во две различни студии на случај, со цел согледување на можностите на имплементација на овие паметни алгоритми во решавање на комплексни инженерски проблеми поврзани со управувањето и однесувањето на хидросистемите со брани и акумулации. Истражувањето е изработено со податоци од реално управување на повеќе брани во Македонија и надвор од неа, обработени во изворен код креиран во Пајтон (анг. „Python“) средина именуван како „DamResML“, кој е дел од докторската дисертација. Самиот DamResML код се состои од 29 ќелии со девет алгоритми за машинско учење кои се тестираат во сите студии на случај, а во интерпретација на крајните резултати се користат исклучиво моделите кои даваат најдобри резултати.

Содржински, докторската дисертација се состои од воведно поглавје без нумерација, седум нумерирани поглавја со презентација на целта на истражувањето, методологијата и резултатите, крајно поглавје како заклучок, и листа на користена литература.

Во првото поглавје насловено како „Предмет и цел на истражувањето“ е даден вовед во проблематиката која е предмет на обработка во рамки на докторската дисертација, како и основните цели и насоки на истражувањето. Во поглавјето е опишана потребата од примена на техники и алатки од машинското учење во инженерството на брани и управувањето со водостопански системи, при што образложени се мотивите за развој на сопствен програмски код

на програмскиот јазик Пајтон. Дополнително, накратко се опишани двете студии на случај преку кои се анализира применливоста на алгоритмите од машинското учење за моделирање и анализа на сложени инженерски процеси. Воедно, дефинирани се очекуваните научни и практични придонеси од истражувањето, со акцент на унапредување на методологиите за управување со водните ресурси и оценување на долготрајното однесување и сигурноста на браните.

Во второто поглавје од дисертацијата насловено како „Преглед на досегашни истражувања во областа“ образложен е преглед на досегашните истражувања поврзани со примената на машинското учење и техниките од вештачката интелигенција во хидротехниката, инженерството на брани и управувањето со водните ресурси. Анализирани се бројни научни трудови од областа на прогнозирање на хидролошки параметри, анализа на однесувањето на брани, управување со акумулации, процена на ризик од поплави, транспорт на нанос и моделирање на подземни води. Прегледот на литературата покажува дека современите ML алгоритми, како ANN, SVR, RF, XGBoost, LSTM и други интелигентни модели, имаат значителен потенцијал за моделирање на сложени нелинеарни процеси во хидротехничкото инженерство. Дополнително, преку анализата на постојните истражувања се согледуваат актуелните трендови, предности и ограничувања на примената на машинското учење во оваа област, што претставува основа за дефинирање на целите и методологијата на оваа дисертација.

Во третото поглавје насловено како „Опис на анализираните хидросистеми со брани и акумулации“ даден е опис на анализираните хидросистеми со брани и акумулации чии податоци се користат во рамки на истражувачкиот дел од докторската дисертација. Претставени се основните конструктивни, хидролошки и експлоатациони карактеристики на хидросистемите „Стрежево“ и „Злетовица“, како и на бетонска лачна брана во Франција управувана од компанијата „Électricité de France“.

Во четвртото поглавје насловено како „Опис на изработениот код DamResML“ даден е детален опис на изработениот код во Пајтон средината, ќелиите, чекорите на пресметка и алгоритмите од машинско учење вклучени во процесот. Станува збор за комплексен код составен од 29 ќелии и повеќе функционални целини, изработени со користење на постојни библиотеки за машинско учење.

Во петтото поглавје насловено како „Тестирање и валидација на DamResML“ презентирани се студии на случај со податоци од лачната брана во Франција, пресметана паралелно – со кодот во Пајтон и со комерцијално достапниот код – „NeuralTools“ кој е интегриран во Microsoft Excel средина. Паралелната пресметка е изработена со два алгоритми – GRNN и ANN, а потоа произнесени се заклучоци во однос на излезните резултати. Од анализите се заклучува дека „NeuralTools“ и креираниот код DamResML предвидуваат вредности со разгледуваните алгоритми кои речиси во целост се согласуваат.

Во шестото поглавје насловено како „Студија на случај 1“ презентирани се три студии на случај поврзани со управување на хидросистемите „Стрежево“ и „Злетовица“, каде со помош на машинско учење се предвидуваат и оптимизираат целните варијабли – испуштања на вода од акумулацијата Кнежево, производството на енергија од „ХЕЦ Б“ на хидросистемот „Злетовица“ и варијацијата на водното ниво во акумулација „Стрежево“.

Во седмото поглавје насловено како „Студија на случај 2“ презентирани се седум студии на случај поврзани со однесување на три различни типови на брани. Во првите пет анализи се разгледува предвидување на водни нивои во избрани пиезометри на брана „Кнежево“ чие однесување е во корелација со варијацијата на водното ниво во акумулацијата. Во овие студии на случај, се разгледува и филтрационото количество под низводното потпорно тело на брана „Кнежево“ со земање предвид на нивото на вода во акумулација како и мерените нивои во пиезометрите во близина на дренажните колектори. Во шестата анализа се разгледуваат вертикални поместувања на брана „Кнежево“ во однос на нивото на вода во акумулацијата и времето, изработени се предвидувања за надополнување на временските низи, а потоа извршена е пресметка на времето потребно за достигнување на теоретските максимални вредности за

вертикалните слегнувања. Во седмата анализа на случај се анализираат радијалните поместувања во централниот блок на лачната бетонска брана во Франција.

Како крајно поглавје од докторската дисертација, презентирани се заклучоци и препораки за понатамошни истражувања.

Дисертацијата завршува со прегледна референцирана листа на користена литература во која има над 120 книги, трудови и интернет содржини кои придонесоа во истражувањето.

1. ПРЕДМЕТ И ЦЕЛ НА ИСТРАЖУВАЊЕТО

Предмет на истражувањето во рамки на докторската дисертација е систематска примена на техники и алатки од областа на машинското учење за оцена на долготрајното однесување на брани, како и за унапредување на процесите на управување со водостопански системи, преку развој и имплементација на изворен програмски код на програмскиот јазик Пајтон (анг. „Python“). Примената на методите од машинското учење во рамките на докторската дисертација се разгледува преку две меѓусебно поврзани студии на случај.

Првата студија на случај е насочена кон анализа на процесот на управување со комплексен, повеќедимензионален и повеќенаменски водостопански систем, базирана на временски серии на податоци добиени од извршени теренски мерења.

Втората студија на случај опфаќа анализа на долготрајното однесување на браните со примена на обучени модели со машинско учење со податоци од оскултација на неколку различни типови на брани со цел воспоставување функционални и статистички зависности помеѓу релевантните параметри преку кои се следи и оценува однесувањето на браната во текот на времето.

Во рамки на истражувањето изработен е програмски код именуван како DamResML, кој во себе вклучува повеќе алгоритми од машинско учење. Имено, со примена на програмскиот јазик Пајтон изработен е програмскиот код DamResML во кој се спроведуваат анализите во рамки на истражувањето по претходно извршена верификација на неговата точност и функционалност. Практично, ова е иницијалниот чекор во истражувањето – креирање на функционален и верифициран код кој овозможува да се генерираат врски и зависности помеѓу разгледуваните параметри, со соодветен табеларен/графички приказ на соодветните коефициенти и критериуми за проценка на адекватноста и подобноста на различните модели. Кодот претставува вреден ресурс за истражувачкиот и инженерскиот кадар, како и за студентите по градежништво, кои ќе можат да ги користат овие напредни техники и соодветно применат во рамки на реалното инженерство.

Потребата од изработка на ваков комбиниран програмски код, кој вклучува повеќе алгоритми од машинско учење во инженерството за брани потекнува токму од моменталната актуелност на машинското учење во различните области на истражување, што недвосмислено создава можност и соодветно на тоа наметнува користење на овие алатки во наведената област. Како институција која има високообразовна, научноистражувачка и применувачка високостручна дејност, со морална и етичка должност, потребно е да ги истражime можностите но и границите на примена

на овие напредни методи и алатки во нашата област на истражување, и истата да ја пренесеме како знаење на сегашните и идните генерации на истражувачки и инженерски кадар од областа на градежништвото.

Ваквиот напреден пристап ќе придонесе за подобро и понапредно следење на сигурноста на браните и подобро управување со водните ресурси преку интегрирање на машинско учење на нов и практичен начин. Воедно, програмскиот код со отворен пристап и методологијата што може да се аплицира го прават погоден за адаптација и во други истражувачки проекти.

Целта на истражувањето е да се испита степенот, применливоста и начинот на интеграција на алатките од машинското учење во инженерството на брани, со фокус на нивниот потенцијал за унапредување на процесите на управување со водостопански системи и за подобрување на оцената на долготрајното однесување и безбедноста на браните.

Во рамки на првата студија на случај, анализата се заснова на влезни податоци што опфаќаат регистрирани временски серии на хидролошки дотекувања во акумулацијата, како и излезни големини кои се однесуваат на нивоата на водата во акумулацијата и регулираните протекувања, добиени од страна на оператори на реални хидросистеми. Овие податоци се специфицирани и обработени од аспект на исполнување на повеќе управувачки критериуми, вклучувајќи обезбедување на доверливо водоснабдување и наводнување, максимизација на производството на електрична енергија и задоволување на еколошките барања.

Преку соодветна спецификација на влезните варијабли и целните големини во рамки на развиениот програмски код, се овозможува генерирање на релевантни корелациски и функционални зависности, во согласност со однапред дефинирани критериуми за точност. Врз основа на воспоставените зависности се развива прогностички модел кој го опишува модусот на однесување на системот и овозможува анализа и потенцијална модификација на постојните управувачки правила, со цел постигнување на подобрени оперативни и системски перформанси.

Втората студија на случај е насочена кон анализа на долготрајното однесување на различни типови на брани преку примена на алатки од машинското учење, со цел унапредување на методологиите за следење и оценување на нивната сигурност. Влезните податоци опфаќаат излезни резултати од временски серии на мерени големини добиени од процесот на техничко набљудување (оскултација) на браната.

Со примена на пристап базиран на дефинирање на релевантни варијабли и целни големини, ќе се воспостават соодветни корелациски зависности со задоволителен степен на точност, врз основа на кои ќе се развие прогностички модел за опишување на модусот на однесување на браната. Овој модел ќе овозможи детекција на нерегуларни или атипични однесувања, како и поефикасно и порано следење на конструкцискиот одговор, со што ќе се придонесе кон стекнување на подлабоки сознанија, подоверлива и пообјективна оценка на долготрајното однесување и сигурноста на браната.

“Without data, you’re just another person with an opinion.”
W. Edwards Deming

2. ПРЕГЛЕД НА ДОСЕГАШНИ ИСТРАЖУВАЊА ВО ОБЛАСТА

2.1. Кибернетика и развој на програмските јазици

Развојот на системската анализа и современите методи за моделирање на комплексни водостопански системи е тесно поврзан со развојот на кибернетиката и пресметувачките науки. Кибернетиката претставува научна дисциплина која ги проучува принципите на управување, комуникација и повратна информација кај природните и техничките системи. Според речникот на Универзитетот Кембриџ, таа претставува научно проучување на преносот на информации во машините и електронските системи, споредувајќи ги овие процеси со функционирањето на човечкиот мозок и нервниот систем, додека според „Encyclopaedia Britannica“ кибернетиката се дефинира како теорија на контрола применлива кај комплексните системи. Заедничка карактеристика на овие дефиниции е концептот на повратна информација (анг. „feedback“), преку кој се овозможува следење на состојбата на системот и соодветно прилагодување на неговото однесување.

Потеклото на терминот произлегува од старогрчкиот збор *kybernetikos*, што означува „вештина на управување“ или „уметност на кормиларот“. Терминот во научна смисла прв го употребува Андре Мари Ампер во делото „Есеј за филозофијата на науките“ (фр. ‘Essai sur la philosophie des Sciences’) (Ampère, 1834), додека современите темели на кибернетиката ги поставува Норберт Винер со книгата „Cybernetics: Or Control and Communication in the Animal and the Machine“ (Wiener, 1948). Винер ја дефинира кибернетиката како наука за контрола и комуникација кај живите организми и машините, при што концептите на информациски тек, повратна врска и стабилност стануваат основа за моделирање и управување со сложени системи.

Во текот на понатамошниот развој се појавуваат повеќе интерпретации на поимот кибернетика. Ешби (Ashby, 1956) ја дефинира кибернетиката како „уметност на управувањето“, Куфињал (Couffignal, 1956) како „уметност на обезбедување ефикасна оперативност“, Миндел (Mindell, D. A., 2000) ја нагласува интеракцијата човек-машина заснована врз принципите на комуникација и контрола, додека Кофман (Kauffman, 2007) ја разгледува како наука за самоорганизирачки системи. Денес кибернетиката претставува интердисциплинарна научна област која ги интегрира

математиката, информатиката, автоматското управување, теоријата на системите и вештачката интелигенција.

Развојот на кибернетиката бил неразделно поврзан со развојот на пресметувачките машини и програмските јазици. Значајни придонеси дале Чарлс Бабаџ, кој го формулирал концептот на програмабилен компјутер (Babbage, 1834), Ада Лавлејс, автор на првиот алгоритам наменет за механички компјутер, Алан Тјуринг кој со концептот на универзалната машина ги поставил темелите на современата теориска информатика (Turing, 1936), како и Џон Фон Нојман, чија архитектура со зачувана програма претставува основа на современите компјутерски системи (Aspray, 1990). Дополнителен придонес дал Конрад Зусе со развојот на програмскиот јазик Планкалкул (анг. „Plankalkül“), кој се смета за првиот програмски јазик со високо ниво на апстракција.

Во педесеттите и шеесеттите години на минатиот век се појавуваат првите широко применувани програмски јазици, како FORTRAN, COBOL, ALGOL и LISP, кои овозможиле развој на научни, инженерски и деловни апликации (Richard Wexelblat, ed., 1981; Thomas Bergin, Jr.; Richard Gibson, Jr.; ed., 1996; Jermar, 1987; Bugliarello & Gunther, 1974). Подоцна следува развојот на Pascal, C, C++, Perl, Python, Java и JavaScript, со што значително се прошируваат можностите за развој на комплексни софтверски системи, нумерички модели и алгоритми за анализа на податоци.

Особено значење за современите инженерски дисциплини има програмскиот јазик Пајтон, кој благодарение на својата едноставност, отворен код и богат систем на библиотеки стана еден од најкористените јазици во областа на науката за податоци, машинското учење и вештачката интелигенција. Библиотеки како NumPy, Pandas, Scikit-learn, TensorFlow и PyTorch овозможуваат развој на напредни модели за анализа, оптимизација и предвидување, што денес претставува основа за интелигентно управување со сложени инженерски системи, вклучувајќи ги и водостопанските системи.

Токму ваквиот историски развој – од кибернетиката, преку теоријата на системите и програмските јазици, до современите методи на машинско учење – ја претставува технолошката основа врз која се развиваат современите модели за симулација, предвидување и оптимизација применети во оваа докторска дисертација.

2.2. Досегашна примена на машинско учење во инженерство за брани и водостопански системи

Процесот на учење од мерени податоци добиени преку реално управување или извршени експерименти како и пренесувањето на тоа знаење во аналитички модели припаѓа на математички тип на задачи така наречени „меко“ програмирање. Под „меко“ програмирање (анг. „soft computing“) се подразбираат моделите како невронските мрежи, потпомогнати вектор машини и фази логика (Kesman, 2001). За разлика од „тврдото“ програмирање (анг. „hard computing“) кое се потпира на точни алгоритми и прецизни пресметки, „мекото“ програмирање дозволува флексибилни, адаптирани модели со поголем степен на толеранција, што прави овие модели да бидат погодни за решавање на проблеми во практичното инженерство каде дефинирањето на егзактни решенија некогаш е отежнато. Како подрачје на „мекото“ програмирање се јавува вештачката интелигенција која претставува широко поле на различни системи и модели кои ја „имитираат“ човечката интелигенција. Во јадрото на вештачката интелигенција се јавува машинското учење и неговите потполиња, кое е присутно речиси еден век во научната сфера - уште од 1940 година (Sun, Yu, & Dong, 2023). Од почетокот на постоењето на пресметувачките машини, постоела идејата за тоа дали тие може да „учат“. Се сметало дека, доколку се успее во насока нивно „учење“ т.е. да се подобруваат преку искусствено и податочно знаење, ќе се постигне значителен технолошки исчекор (Mitchel, 1997).

Денес, напредните информатички алатки и технологија кои чинат сè помалку - истовремено надминувајќи ги перформансите на своите претходници - играат огромна улога во сè повеќе

присутната примена на вештачката интелигенција во секојдневниот живот (Ibrahim, 2016).

Машинското учење покажува негова примена со висока ефикасност и практичност во различни сфери, како препознавање на шеми, предвидување на низа временски податоци, дијагностика, контрола на процеси, оптимизација, предвидување на финансиски текови и многу повеќе (Kesman, 2001) што се потврдува со дефиницијата дека самиот израз „машинско учење“ се однесува на автоматска детекција на обрасци во податоците кои може да ни бидат од корист (Shalev-Shwartz & Ben-David, 2014). Поради проблемите и процесите поврзани со водните ресурси и нивното управување, интересно е да се истражи можната употреба на машинското учење во оваа област. Водата покрива околу 71% од земјината топка, но само 3,5% од вкупните води претставуваат слатки води (Water Science School, 2019). Дополнително, просторната и временската распределба на водните ресурси ретко се совпаѓаат со потребите на современото живеење. Ова наметнува сложено управување со оскудните водни ресурси чие користење станува сè поскапо, а побарувачката сè поголема.

Досега, објавени се бројни реферати со резултати од истражувања со успешна примена на машинското учење во решавање на водостопански проблеми и проблеми од областа на браните и акумулациите, како што се предвидување на потреби за вода, определување на веројатност на појава на води во сушни предели, анализа на однесувањето на браните, предвидување на нивои на вода во подземните аквифери, и многу други (A. Ahmed, Sayed, Abdoulhalik, Moutari, & Oyedele, 2024). Како што следи, даден е кус опис на повеќе избрани реферати со резултати кои ја докажуваат неприкосновената дводецениска континуирана примена на машинското учење во инженерството за брани и хидротехниката.

Белкаш и соработниците (Beljkas, Krulanovic, Rutesic, Lazarevska, & Zafirovski, 2025) ја анализираат примената на повеќе модели на машинско учење за предвидување на побарувачката на вода, со цел подобрување на управувањето со водоснабдителните системи и оптимизација на распределбата на водните ресурси. Во истражувањето се споредуваат повеќе современи алгоритми на вештачка интелигенција, при што нивните перформанси се оценуваат преку стандардни статистички индикатори за точност на предвидувањето. Авторите покажуваат дека моделите базирани на машинско учење успешно ги препознаваат сложените нелинеарни зависимости помеѓу историската потрошувачка, временските услови и останатите релевантни фактори кои влијаат врз побарувачката на вода. Резултатите укажуваат дека моделите на вештачка интелигенција обезбедуваат повисока точност и поголема робусност во однос на традиционалните методи за прогнозирање, што ги прави соодветна алатка за поддршка при оперативното управување и планирањето на водоснабдителните системи. Авторите заклучуваат дека интеграцијата на методите на машинско учење може значително да придонесе за поефикасно и поодржливо управување со водните ресурси, особено во услови на променлива побарувачка и климатски промени.

Хосеини, Нодушан, Барати и Шахејадари (Hosseini, Nodoushan, Barati, & Shahheydari, 2016) примениле адаптивен неуро-фази интерфејс систем (анг. „Adaptive Neuro-Fuzzy Inference System – ANFIS“) за определување на коефициентот на преливање кај преливник со лавиринтен облик. Во рамки на истражувањето биле анализирани различни геометриски и хидраулички параметри кои влијаат врз капацитетот и ефикасноста на преливникот, како што се висината на водата, геометријата на преливниот раб и односите помеѓу конструктивните димензии. Авторите покажале дека ANFIS моделот успешно ги комбинира предностите на невронските мрежи и фази логиката, овозможувајќи моделирање на сложени нелинеарни хидраулички процеси со висока точност. Резултатите покажале дека предложениот модел дава подобри предвидувања на коефициентот на преливање во споредба со класичните емпириски равенки и традиционалните методи, што укажува на потенцијалот на интелигентните модели за примена во хидротехничкото инженерство и проектирањето на хидраулички објекти.

Шахра, Ву и Романо (Shahra, Wu, & Romano, 2019) анализирале историски податоци од системи за дистрибуција на вода со примена на техники од машинско учење со цел подобрување на управувањето со водоснабдителните мрежи. Во рамки на истражувањето биле обработувани

податоци за потрошувачка, притисок и истекувања од системот, при што алгоритмите биле искористени за идентификација на обрасци во однесувањето на мрежата и за откривање на аномалии кои може да укажуваат на истекувања. Авторите покажале дека анализата на историски временски серии овозможува подобро предвидување на идната побарувачка за вода, поефикасна распределба на ресурсите и навремена детекција на загуби во системот. Студијата претставува значаен пример за примена на податочно управувани модели во современите системи за управување со водоснабдување.

Хуанг, Ченг, Јанг и Гуо (Huang., Cheng, Yang, & Guo, 2019) примениле конволуциони вештачки невронски мрежи (анг. „Convolutional Neural Networks“ или накратко CNN) за предвидување на нивото на водата во крајбрежни области изложени на влијание од бранови, плими и екстремни временски услови. Во рамки на истражувањето биле анализирани временски серии од хидролошки и метеоролошки параметри, при што CNN моделите биле искористени за автоматско извлекување на релевантни карактеристики од податоците. Авторите покажале дека конволуционите мрежи можат успешно да ги моделираат сложените временски и просторни зависимости кои влијаат на промените на нивото на водата. Резултатите укажале дека предложениот пристап обезбедува висока точност и стабилност при краткорочно предвидување на нивото на водата во крајбрежни региони.

Буи и соработниците (Bui, et al., 2020) предложиле нов пристап за проценка на подложноста на ненадејни поплави со примена на модели за длабоко учење (анг. „Deep Learning Neural Networks – DLNN“). Истражувањето било спроведено за планински регион во Виетнам изложен на чести тропски бури, при што биле анализирани повеќе фактори како што се врнежи, наклон на теренот, геологија, земјишна покривка и хидролошки карактеристики. Авторите развиле невронска мрежа со повеќе скриени слоеви која покажала висока способност за идентификација на подрачја со висок ризик од ненадејни поплави. Резултатите покажале дека моделите базирани на длабоко учење можат успешно да ги опфатат сложените нелинеарни зависимости помеѓу различните природни фактори и појавата на поплави, притоа обезбедувајќи повисока точност во однос на традиционалните методи.

Рен, Лиу, Ниу, Леи и Жанг (Ren, Liu, Niu, Lei, & Zhang, 2020) имплементирале интерпретабилен модел базиран на долго краткорочно памтење (анг. „Long Short-Term Memory – LSTM“) со механизам на внимание (анг. „attention mechanism“) за предвидување на поместувањата кај бетонски брани. Во студијата биле користени временски серии од оскултациски мерења, како што се ниво на вода во акумулацијата, температура и регистрирани поместувања на конструкцијата. Воведениот механизам овозможил идентификација на најзначајните влезни параметри кои влијаат врз однесувањето на браната, со што моделот станал не само точен, туку и интерпретабилен од инженерски аспект. Резултатите покажале дека предложениот модел има висока способност за предвидување на долготрајното однесување на бетонските брани и може да претставува корисна алатка за следење на нивната сигурност и навремено откривање на потенцијални аномалии.

Хосеини (Hosseiny, 2021) развил напреден модел базиран на модифицирана У – нет (анг. „U-Net“) архитектура, означена како ‘U-NetRiver’, за предвидување на длабочината и просторниот опфат на речните поплави. Во студијата биле користени податоци добиени од дводимензионални хидраулички симулации, при што како влезни параметри биле употребени дигитален модел на терен и вредности на протекување. Моделот овозможил автоматско препознавање на речната геометрија и поплавените површини, при што предвидувањата за длабочината на водата покажале значително подобрување во споредба со класичните хидраулички модели. Истражувањето покажало дека длабокото учење има потенцијал за значително забрзување на процесот на анализа и симулација на поплавни настани, со задржување на задоволителна точност на резултатите.

Насловот „Soft Computing and Machine Learning in Dam Engineering“, објавен во 2023 година како специјално издание на списанието Water од издавачката куќа MDPI (Special edition of 'Water', 2023), претставува значајна современа публикација посветена на примената на методите од

машинското учење и техниките на меко програмирање во инженерството на брани. Во рамки на ова специјално издание вклучени се повеќе научни трудови кои обработуваат различни аспекти од анализа, мониторинг и оценување на сигурноста и однесувањето на бетонски и насипни брани, со примена на современи алгоритми за машинско учење, длабоко учење и интелигентна обработка на податоци. Посебен акцент е ставен на користењето на временски серии од оскултациски мерења, како што се поместувања, филтрации, температури, нивоа на вода и други параметри релевантни за следење на конструкциското однесување на браните во текот на времето. Дел од трудовете се насочени кон развој на прогностички модели за предвидување на поместувања и деформации кај бетонски брани, додека други истражувања обработуваат методологии за детекција на аномалии, анализа на ризик и идентификација на потенцијално небезбедни состојби преку примена на невронски мрежи, потпомогнати векторски машини, ансамбл методи и LSTM модели. Во публикацијата се прикажува и значењето на интерпретабилноста на моделите, односно можноста инженерски да се објаснат воспоставените функционални зависности помеѓу влезните параметри и однесувањето на конструкцијата. Дополнително, повеќе трудови ја нагласуваат предноста на машинското учење во однос на класичните статистички и емпириски методи, особено при обработка на големи количини податоци и моделирање на сложени нелинеарни процеси. Со тоа, ова специјално издание претставува значаен научен придонес и релевантен преглед на современите трендови во примената на вештачката интелигенција и машинското учење во инженерството на брани и управувањето со нивната сигурност.

Салазар, Иразабал и Конде креирале веб-апликација базирана на засилени регресиони дрвја (анг. „Boosted Regression Trees“), развиена за анализа на податоци за следење на брани со можност за примена и во други проблемски задачи (Salazar, Irazabal, & Conde, 2024). Имено, во Меѓународниот центар за нумерички методи во инженерство (The International Centre for Numerical Methods in Engineering – CIMNE), изработен е код со алгоритмот „засилени регресиони дрвја“ (анг. ‘Boosted Regression Trees’) за примена при процена на конструкциското однесување на брани. Кодот е наречен ‘SoLdiEr’ (кратенка од SOLution for Dam behavior Interpretation and safety Evaluation with BRT) и е достапен на веб страна за слободно користење (<https://cimne.shinyapps.io/SOLDIER/>) (Salazar, Irazabal, & Conde, 2024). Наведената апликација ја третира проблематиката на долготрајно однесување на брана и детектирање и утврдување на нерегуларности (аномалии) во однесувањето на истата во рамки на разгледуван временски период.

Ахмед, Куршид, Алшамери и соработниците (Sayed Ahmed, Khursheed, Alshameri, Hamed, & Elshaarawy, 2026) разработиле модел базиран на машинско учење за процена на филтрацијата низ хетерогени земјени брани поставени на водопропустна основа. Во истражувањето биле анализирани сложени хидрогеолошки услови кои влијаат на појавата и интензитетот на филтрационите процеси, при што биле користени повеќе параметри поврзани со геометријата на браната, карактеристиките на материјалите, водопропустноста на подлогата и условите на филтрација. Авторите примениле повеќе модели од машинското учење за предвидување на количината на филтрација, а посебен акцент бил ставен на интерпретабилноста на моделите преку примена на SHAP (кратенка од англискиот назив „SHapley Additive Explanations“) анализа, со цел да се идентификува релативното влијание на секој влезен параметар врз резултатите од моделот. Резултатите покажале дека моделите од машинското учење можат со висока точност да ги моделираат сложените нелинеарни зависности присутни во процесите на филтрација кај земјени брани, при што пристапот овозможил подобро инженерско разбирање на механизмите кои влијаат врз сигурноста на браната. Авторите заклучиле дека интеграцијата на современи техники од машинско учење со интерпретабилни методологии претставува значаен чекор кон развој на поефикасни системи за анализа, следење и управување со сигурноста на земјените брани.

Киу, Су и соработниците (Qiu, Su, Fang, & Qi, 2026) развиле хибриден модел за предвидување на поместувањата кај бетонски брани, при што посебен акцент бил ставен на влијанието на температурната хистереза врз однесувањето на конструкцијата. Во истражувањето авторите анализирале дека температурните промени не влијаат моментално врз деформациите на браната, туку постои временско задоцнување поради акумулацијата и преносот на топлина низ бетонската маса. Со цел попрецизно моделирање на овој процес, бил развиен хибриден пристап кој

комбинира статистички методи и алгоритми од машинско учење, овозможувајќи подобро опфаќање на сложените нелинеарни зависимости помеѓу температурата, нивото на водата и регистрираните поместувања. Авторите вовеле дополнителни параметри поврзани со долгорочниот температурен ефект и температурното задоцнување, со што значително се подобрила точноста на предвидувањата во споредба со класичните модели за следење на однесувањето на бетонските брани. Резултатите покажале дека предложениот модел има висока способност за репродукција на реалното однесување на браната и може да претставува корисна алатка за системите за оскултација, анализа на сигурноста и рано откривање на потенцијални аномалии кај хидротехничките конструкции.

Марковиќ и соработниците (Markovic, Brankovic Markovic, Stosovic Andrejevic, & Zivkovic, 2021) предлагаат нов пристап за предвидување на порниот притисок во нефункционални мерни ќелии кај земјени брани со примена на вештачки невронски мрежи (анг. „Artificial Neural Networks“ или накратко – ANN). Истражувањето е мотивирано од честата појава на дефекти кај инструментите за мониторинг, при што недостигот од мерни податоци може значително да ја намали сигурноста при оценката на однесувањето на браната. Авторите развиваат ANN модел кој, врз основа на достапните податоци од останатите исправни мерни места и релевантните хидролошки и експлоатациони параметри, успешно ги реконструира вредностите на порниот притисок во неисправните сензори. Резултатите покажуваат висока точност на предвидувањето и потврдуваат дека моделот е способен да ги опише сложените нелинеарни врски помеѓу параметрите кои влијаат врз процесот на филтрација во телото на браната. Авторите заклучуваат дека примената на вештачките невронски мрежи претставува ефикасна алатка за поддршка на системите за мониторинг и за зголемување на сигурноста при оценката на состојбата на браните, особено во ситуации кога дел од инструментите се неисправни или недостасуваат мерни податоци.

Сноузи и соработниците (Snousy, 2025) развиле напреден хибриден модел базиран на вештачки невронски мрежи и оптимизациски техники за подобрување на предвидувањето на индексот на квалитет на подземни води во полу-сушни региони. Во истражувањето биле користени различни физичко-хемиски параметри на подземните води, како што се рН вредност, електропроводливост, концентрации на растворени соли и јони релевантни за наводнување, при што моделот бил обучен да воспостави функционални зависимости помеѓу измерените параметри и индексот на квалитет на водата. Авторите примениле оптимизациски пристап со цел подобрување на структурата и параметрите на невронската мрежа, што резултирало со значително подобрување на точноста на предвидувањата. Постигнатата вредност од $R^2 = 0.99$ при тестирањето укажува на исклучително висока согласност помеѓу предвидените и реално измерените вредности, што потврдува дека ваквите хибридни модели имаат значителен потенцијал за примена во управувањето со водните ресурси и оценувањето на погодноста на водите за наводнување, особено во региони со ограничени водни ресурси и изразени климатски влијанија.

Саџр и соработниците (Saqr, 2025) развиле напреден ансамбл-модел за машинско учење со цел предвидување на нивото на подземните води во сливот на реката Ергене во Турција, регион кој е под значителен притисок од интензивно земјоделство, индустриски активности и климатски промени. Во рамки на истражувањето биле користени долгорочни неделни временски серии на податоци за периодот од 1966 до 2023 година, што овозможило анализа на сложените и долгорочни промени во однесувањето на аквиферот. Авторите предложили интегриран ансамбл-пристап кој комбинира повеќе плитки и длабоки модели на машинско учење, при што нивната оптимизација била извршена со примена на алгоритмот ‘Coronavirus Herd Immunity Optimizer (CHIO)’, инспириран од механизмите на колективен имунитет. Ваквиот пристап овозможил подобра калибрација на параметрите и оптимизација на процесот на учење на моделите, со што значително се подобрила точноста на предвидувањата. Резултатите покажале исклучително висока успешност на моделот, со вредности од приближно $R^2 \approx 0.99$ и $RMSE \approx 0.5$ m, при што ансамбл-моделот ги надминал поединечните модели на машинско учење во однос на стабилноста и точноста на резултатите. Авторите заклучиле дека интеграцијата на повеќе техники од машинско учење со современи оптимизациски алгоритми претставува моќна алатка за моделирање на комплексни хидрогеолошки процеси и може да обезбеди значителна поддршка во процесите на

одржливо управување со подземните води и аквиферските системи во услови на растечки еколошки и антропогени влијанија.

Поголем број реферати се објавени во базите со литература со резултати од примена на алгоритми од машинско учење во прогноза на протоци во реки и транспорт на нанос во речни корита. Џејн (Jain S. K., 2001) применил вештачки невронски мрежи (анг. „Artificial Neural Networks“ или накратко ANN) за процена на концентрацијата на нанос на две мерни станици на реката Мисисипи. Во истражувањето биле користени податоци за нивото на водата, протокот и претходно измерените концентрации на седименти, при што невронска мрежа била обучена да воспостави зависност помеѓу хидролошките параметри и моменталната концентрација на седименти. Авторот покажал дека ANN моделите можат успешно да ги моделираат нелинеарните процеси поврзани со транспортот на седименти и да дадат попрецизни резултати од традиционалните методи базирани на емпириски криви на проток и ниво на вода. Истражувањето претставува едно од раните значајни истражувања за примена на невронски мрежи во хидрологијата и транспортот на седименти.

Наги и соработниците (Nagy, Watanabe, & Hirano, 2002) користеле ANN модел за предвидување на вкупната концентрација на седименти во водотеци врз основа на повеќе хидраулички и геометриски параметри. Во моделот биле вклучени параметри како тангенцијални напрегања, Фрудов број, Рејнолдсов број, наклон на дното, длабочина на тек и карактеристики на зрното на материјалот. Авторите извршиле анализа на чувствителност со цел идентификација на најзначајните влезни параметри, при што ANN моделот покажал подобри резултати во споредба со класичните равенки за транспорт на седименти. Истражувањето покажало дека невронските мрежи имаат значителен потенцијал за моделирање на сложените процеси на транспорт на седименти во природни водотеци.

Батачарја и соработниците (Bhattacharya, Price, & Solomatine, 2007) примениле ANN и модели на дрва (анг. „Model Trees“ или накратко – MT) за предвидување на брзината и вкупниот транспорт на седименти. Во рамки на истражувањето биле анализирани различни комбинации на влезни параметри со цел да се утврди кои податоци обезбедуваат најголема точност при моделирањето на транспортот на седименти. Резултатите покажале дека и ANN и MT моделите обезбедуваат значително подобри предвидувања во споредба со традиционалните емпириски равенки, особено кај комплексни хидраулички услови. Авторите заклучиле дека методите од машинското учење овозможуваат подобро опфаќање на нелинеарните односи присутни во процесите на транспорт на седименти.

Цимен (Cimen, 2008) користел потпомогнати векторски машини (анг. „Support Vector Machine“ или накратко - SVM) за предвидување на дневната концентрација на суспендирани седименти врз основа на дневниот проток на две реки во САД. Моделот бил обучен со податоци од повеќегодишни мерења, при што како единствен влезен параметар бил користен дневниот проток, а како излез – концентрацијата на седименти. Иако моделот покажал одредени ограничувања при екстремни вредности, резултатите покажале дека SVM пристапот дава подобри резултати во споредба со ANN, фази логика и ANFIS модели. Истражувањето укажало дека SVM претставува стабилна и ефикасна алатка за моделирање на хидролошки процеси со ограничен број влезни параметри.

Чау (Chau K. , 2006) применил т.н. метод за оптимизација со рој од честици (анг. „Particle Swarm Optimization“ или накратко PSO) за тренирање на ANN модел за предвидување на нивото на водата на реката Шинг Мун. Во истражувањето биле користени податоци од возводни мерни станици и историски мерења на набљудуваната станица, при што PSO алгоритмот покажал подобра способност за оптимизација на параметрите на невронската мрежа во споредба со класичниот алгоритам со повратна пропагација. Авторот заклучил дека PSO претставува ефикасна алтернатива за избегнување на локални минимуми во процесот на обука на ANN моделите. Чау (Chau K. , 2007) продолжил со развој на модифициран PSO алгоритам во повеќе чекори за обука на ANN модели за предвидување на нивото на водата во реално време. Во споредба со стандардниот PSO и пристапот со повратна пропагација (анг. „backpropagation“), предложениот модел постигнал

повисока точност и побрза конвергенција. Истражувањето покажало дека комбинацијата на интелигентни оптимизациски алгоритми и невронски мрежи има значителен потенцијал за примена во системите за хидролошко предвидување и управување со водни ресурси.

Сонг и соработниците (Song, et al., 2020) предлагаат модел за предвидување на буични поплави базиран на невронски мрежи со долгорочна краткотрајна меморија (анг. „Long short-term memory“ или накратко LSTM), со цел подобрување на точноста на краткорочните хидролошки прогнози во планински сливови. Авторите развиваат модел составен од повеќе едночечорни LSTM мрежи кои како влезни параметри ги користат измерените и прогнозираните врнежи, како и претходните вредности на протекот, овозможувајќи истовремено земање предвид на просторните и временските карактеристики на хидролошките процеси. Моделот е тестиран на сливот на реката Анхе во Кина, при што постигнува висока точност за временски хоризонти на прогнозирање од 1 до 10 часа. Особено добри резултати се добиени кај поголемите поплавни бранови, каде процентот на успешно предвидени настани надминува 94% за прогнози до 5 часа однапред. Дополнителната анализа покажува дека претходно измерениот протек има најголемо влијание врз краткорочните прогнози, додека влијанието постепено се намалува со зголемување на временскиот хоризонт. Авторите заклучуваат дека LSTM моделите претставуваат моќна алатка за моделирање на сложените нелинеарни врски помеѓу врнежите и протекот, овозможувајќи навремено и сигурно предвидување на буичните поплави и значително подобрување на системите за рано предупредување и управување со водните ресурси.

Хоанг Нам и соработниците (Nam Hoang, Tran Thu, Hoang Duc, & Le Van, 2025) ја испитуваат примената на невронски мрежи со долгорочна краткотрајна меморија (анг. „Long short-term memory“ или накратко LSTM) за дневно предвидување на протекот на реката на хидролошката станица Хоја Дует во Виетнам. Истражувањето е базирано на 49-годишна база на податоци (1975–2023), која ги опфаќа врнежите и протеците во сливот на реката Нган Сау. Авторите ја анализираат точноста на моделот за различни временски хоризонти на прогнозирање (1, 3, 5 и 7 дена) и различни должини на временскиот прозорец, при што најдобри резултати се добиени за еднодневна прогноза со временски прозорец од 90 дена, достигнувајќи Наш-Сутклиф (анг. „Nash-Sutcliffe“ или накратко NSE) коефициент од 0,864. Дополнително е утврдено дека моделот постигнува значително подобри резултати во дождовниот период отколку во сушниот, што укажува на поголема сложеност при моделирањето на ниските протоци. Авторите заклучуваат дека LSTM моделите претставуваат ефикасна и сигурна алатка за предвидување на речниот протек, способна успешно да ги моделира сложените временски и нелинеарни зависности во хидролошките серии, при што нивната примена може значително да придонесе кон подобро управување со водните ресурси, намалување на ризикот од поплави и подобрување на оперативното планирање на хидротехничките системи.

Богнер и Падрон (Bogner & Padron S., 2026) истражуваат нов пристап за подобрување на потсезонските хидролошки прогнози преку искористување на стохастичката природа на моделите базирани на длабоко учење. Авторите применуваат модел од типот „Трансформер на временска фузија“ (анг. „Temporal Fusion Transformer“ или накратко TFT) за истовремено предвидување на протекот и температурата на водата на 22 хидролошки станици во Швајцарија, при што се анализира влијанието на различни случајно иницијализирани модели (анг. „random seeds“) врз точноста и сигурноста на прогнозите. Наместо случајноста да се третира како недостаток, истражувањето покажува дека комбинирањето на повеќе модели обучени со различни почетни иницијализации овозможува добивање на поквалитетни веројатносни прогнози и намалување на неизвесноста на предвидувањата. Резултатите покажуваат подобрување на прогнозната способност во однос на поединечните модели, особено кај подолгите временски хоризонти, при што предложениот пристап обезбедува подобра калибрација и поголема робусност на прогнозите. Авторите заклучуваат дека вклучувањето на повеќе модели со длабоко учење претставува ефикасен начин за подобрување на хидролошките прогнози и управувањето со неизвесноста, што го прави овој пристап особено погоден за оперативни системи за управување со водните ресурси, заштита од поплави и поддршка при донесување одлуки во сложени хидролошки услови.

Кон крајот на 90-тите години на минатиот век, инженери и научници работеле во истражување на можната примена на техниките од машинско учење во предвидување на височината на морските бранови, како исклучителен феномен чие предвидување може да биде од огромно значење за реакција и спасување на загрозени приморски подрачја во услови на временски непогоди. Една од првите студии во оваа област е спроведена од Део и Наиду кои примениле вештачки невронски мрежи (анг. „Artificial Neural Networks“ или накратко ANN) во предвидување на бранови во реално време (Deo & Naidu, 1999). Како влезни податоци во нивните прогнозни модели биле само карактеристики на бранот. Добиените резултати биле поточно споредено со резултатите од класичните регресивни модели, со зголемена точност во пократките времиња на предвидување. Цаи и соработниците (Tsai, Lin, & Shen, 2002) работеле на предвидување на брановите врз основа на карактеристики на опсервирани бранови наместо со примена на податоци за ветерот, со примена на ANN модели. Резултатите покажале висока точност во предвидувањата со примена на податоци од опсервирани бранови во краток временски период.

За управувањето на водостопанските системи – брани со акумулации, објавени се поголем број на научни трудови со одлични резултати од примената на машинското учење. Вохерем, Одјејами, Угвоке и соработниците (Woherem, et al., 2026) развиле интегрирана рамка базирана на машинско учење за предвидување и управување со економските последици од прекуграничното испуштање на води од браните во Нигерија, со посебен акцент на влијанието на браната Лагдо врз сливот на реката Бенуе. Истражувањето било мотивирано од честите и разорни поплави во Нигерија, кои предизвикуваат значителни хуманитарни, инфраструктурни и економски последици, дополнително влошени од акумулацијата на седименти и намалувањето на ефективниот капацитет на акумулацијата. Авторите развиле мулти-задачен модел (анг. „Multi-Task Learning – MTL“) способен истовремено да предвидува интензитет на поплави, промени на нивото на водата и економски штети, при што неговите перформанси биле споредени со поединечни модели како Random Forest, XGBoost и Logistic Regression. Во моделот биле интегрирани хидролошки, метеоролошки и социоекономски податоци за периодот од 1982 до 2024 година, вклучувајќи дотекувања, испусти од брана, ниво на акумулација, врнежи, оштетувања на инфраструктура и влијание врз населението. Резултатите покажале висока точност на предложениот модел, со вредности од приближно $R^2 = 0.88$ за процена на економските влијанија и $MAE \approx 0.15\text{--}0.17$ за предвидување на нивото на водата, при што MTL рамката ги надминала поединечните модели во однос на стабилноста и точноста на резултатите. Авторите заклучиле дека ваквите интегрирани модели со машинско учење имаат значителен потенцијал за поддршка на системите за управување со поплави, рано предупредување и климатска отпорност, особено во региони изложени на комплексни хидролошки и антропогени влијанија.

Сенапати, Бера, Дута и соработниците (Senapati, Bera, Dutta, Pal, & Das, 2026) разработиле пристап базиран на машинско учење за оптимален избор на површини погодни за собирање на дождовница (Surface Rainwater Harvesting – SRWH) во полусуви региони, со цел поддршка на одржливото управување со водните ресурси. Истражувањето било спроведено во сливот на реката Дваракежвар, при што биле анализирани различни геоморфолошки, хидролошки, геолошки и климатски параметри, како што се наклон на теренот, тип на почва, густина на дренажна мрежа, користење на земјиште, врнежи и инфилтрациски карактеристики. Авторите примениле повеќе модели од машинско учење за идентификација и класификација на најпогодните локации за акумулирање на дождовница, при што моделите биле споредени според нивната точност и способност за генерализација. Резултатите покажале дека интеграцијата на машинско учење со GIS и просторна анализа овозможува значително подобро препознавање на погодни локации во споредба со традиционалните методи, што придонесува кон поефикасно управување со водните ресурси во региони изложени на воден дефицит и климатски промени. Авторите заклучиле дека ваквиот пристап претставува корисна алатка за планирање на системи за собирање и управување со дождовница, со потенцијална примена во различни хидролошки и климатски услови.

Модаб и Реза (Modab & Reza, 2025) примениле модели базирани на „привлекување внимание“ (анг. „attention-driven“) за длабоко учење за предвидување на нивото на водата во акумулација врз основа на мултиваријантни временски серии. Истражувањето било спроведено на браната Неса во

југоисточен Иран, при што биле анализирани повеќе хидролошки и временски зависни параметри со цел развој на модел способен за прецизно краткорочно предвидување на нивото на водата во акумулацијата. Авторот комбинирал современи архитектури за длабоко учење, вклучувајќи CNN-BiLSTM и ConvLSTM2D модели со механизам за „привлекување внимание“, со што се овозможило подобро препознавање на временските и нелинеарните зависности помеѓу влезните податоци. Резултатите покажале дека механизмот значително ја подобрува способноста на моделот за препознавање на најрелевантните информации во временските серии, што резултирало со висока точност и стабилност на предвидувањата. Авторот заклучил дека ваквите модели имаат значителен потенцијал за примена во управувањето со акумулации, контрола на поплави и планирање на водните ресурси, особено во услови на зголемена хидролошка неизвесност и климатски промени.

Бранковиќ и Марковиќ (Branković Marković & Marković, 2026) предлагаат современ пристап за повеќехоризонтно веројатносно предвидување на нивото на водата во акумулации со примена на длабоки невронски мрежи. Во истражувањето се развиваат модели кои, покрај историските хидролошки податоци, ги земаат предвид временскиот хоризонт на прогнозата и појавата на екстремни хидролошки настани, со цел да се подобри точноста и сигурноста на предвидувањата. Авторите применуваат напредни архитектури на длабоко учење за генерирање на веројатносни прогнози, со што покрај очекуваната вредност се добива и информација за неизвесноста на прогнозата. Резултатите покажуваат дека предложените модели постигнуваат значително подобра точност во споредба со конвенционалните пристапи, особено при подолги временски хоризонти и за време на екстремни хидролошки услови. Авторите заклучуваат дека интеграцијата на моделите со длабоко учење во системите за управување со акумулации овозможува навремено и посигурно донесување на оперативни одлуки, придонесувајќи за подобро управување со водните ресурси, намалување на ризикот од поплави и ефикасно користење на акумулациите.

Жанг, Јуе, Басирифард и соработниците (Zhang, Yue, Basirifard, Cao, & Yang) предложиле нов модел за длабоко учење од типот Mamba State Space Model, означен како Structured State Space for Sequence Modeling with Diagonal State Spaces (S4D), за симулација на испустите од акумулации и анализа на работата на брани во голем размер. Истражувањето претставува една од првите примени на ваков тип модели во областа на управувањето со водни ресурси и инженерството на брани. Авторите го тестирале моделот врз 441 брана во континенталниот дел на САД (CONUS), при што неговите перформанси биле споредени со широко користени модели како LSTM и XGBoost. Во анализата биле користени повеќе хидролошки, климатски и оперативни параметри кои влијаат врз одлуките за испуштање на вода од акумулациите, при што S4D моделот покажал подобра просторна и временска способност за репродукција на реалните оперативни правила и режими на работа на браните. Дополнително, авторите примениле SHAP анализа (кратенка од англ. „Shapley Additive Explanations“) со цел интерпретација на моделот и идентификација на најзначајните хидрометеоролошки фактори кои влијаат врз испустите од акумулациите. Резултатите покажале дека предложениот модел ги надминува LSTM и XGBoost кај најголем дел од анализираните акумулации, што укажува на значителниот потенцијал на современите state space deep learning архитектури за примена во управување со водни ресурси, оптимизација на работата на акумулации и анализа на сложени хидролошки системи.

Согласно наведените извори на истражувања спроведени во хидротехниката за решавање на најразлични проблеми со примена на машинско учење, може да се заклучи дека досегашната примена на современите техники од вештачката интелигенција доминираат во најновите истражувањата споредено со останатите традиционални истражувачки методи. Како инженери, кон овие нови техники треба да пристапиме внимателно и систематично, додека излезните големини и резултати треба да ги интерпретираме врз база на теоретско знаење, спроведени аналитички и нумерички пресметки и практично знаење. Досегашните истражувања ја покажуваат неприкосновената примена на алатки од машинско учење во оваа област, но сепак е потребен систематичен пристап со креирање на модели во кои се разгледуваат и користат долги и доверливи низи на податоци за јасно дефинирани и темелно проучени појави и студии на случај.

3. ОПИС НА АНАЛИЗИРАНИТЕ ХИДРОСИСТЕМИ СО БРАНИ И АКУМУЛАЦИИ

Голем број податоци од теренски мерења се користат во рамки на докторската дисертација како податоци добиени од страна на операторите на постојни функционални комплексни хидросистеми. Во ова поглавје накратко е даден опис за секој хидросистем со брана и акумулација чии податоци понатаму се применуваат во истражувачкиот дел на докторската дисертација. Станува збор за податоци кои се регистрираат како дел од редовната оскултација на браните и акумулациите.

Првата студија на случај опфаќа повеќе анализи со кодот од машинско учење кои се однесуваат на управување со хидросистемите. Станува збор за хидросистеми со повеќе намени чие исполнување на цели е во директна зависност со природните појави, пред сè врнежите и површинското истекување. Со примена на кодот со алгоритми од машинско учење, во рамки на докторската дисертација ќе се оптимизира управувањето на целните варијабли врз основа на познатите податоци за досегашното повеќегодишно управување со хидросистемите.

Втората студија на случај опфаќа анализа на однесувањето на предметните брани со примена на кодот со машинско учење. Анализата на однесувањето кај бетонските и насипните брани претставува неминовен дел во процесот на планирање на хидросистемите со брани и акумулации, додека во тек на експлоатација континуираното следење на стабилноста претставува дел од Законските обврски на управителите со хидросистемите во нашата земја. Согласно Член 198 од Законот за води (Службен весник бр. 87 од 15.07.2008 година), правните лица кои управуваат со браните што се определени како брани од посебно значење се должни да воспостават и да организираат техничко набљудување (оскултација) на браната, придружните објекти и акумулациите. Правното лице кое управува со браните е должно најмалку еднаш годишно да изработи посебен елаборат за анализа и оцена на стабилноста и функционалноста на браните (Собрание, Р. Македонија, 2008).

Во современата инженерска пракса, мониторингот и известувањето за состојбата на браните претставуваат суштински и законски регулирани активности во поголемиот дел од развиените земји, особено во Европа и во САД. Иако не постои единствена унифицирана регулатива на ниво на цела Европска Унија, земјите членки имаат воспоставено национални законски рамки кои го

уредуваат управувањето со безбедноста на браните, при што тие често се усогласени со пошироките принципи на Европската унија и директивите како што е Директивата за водите (анг. „Water Framework Directive“). Во рамки на овие законски системи, операторите на брани се должни да воспостават системи за континуиран мониторинг, кои вклучуваат инструментално следење (на пример со пиезометри, инклинометри и други уреди), редовни визуелни инспекции, како и изработка на периодични технички извештаи што се доставуваат до надлежните институции. Во многу европски држави, слично како и во нашата држава, во Франција, Италија и Шпанија овие обврски се дополнително формализирани преку прописи кои бараат изработка на годишни извештаи за состојбата на браната, спроведување на детални ревизии на безбедноста во определени временски интервали, како и подготовка на планови за итни случаи, со јасно дефинирани процедури за постапување при ризични настани. Овие регулативи ја опфаќаат целокупната животна фаза на браната – од проектирање и изградба, преку експлоатација, па сè до евентуално рушење на браната, со што мониторингот не е изолирана активност, туку континуиран процес на управување со ризик.

Во Соединетите Американски Држави, системот за регулирање на безбедноста на браните е уште поразгранет и се заснова на комбиниран пристап помеѓу федерални и државни институции. Клучна улога имаат организации како Федералната Регулаторна Комисија за Енергетика (анг. „Federal Energy Regulatory Commission“) и Инженерскиот корпус при армијата на САД (анг. „U.S. Army Corps of Engineers“), кои воспоставуваат строги барања за мониторинг, особено кај хидроенергетските објекти. Преку законски акти како Законот за развој на водните ресурси (анг. „Water Resources Development Act“) од 1986 година, се бара од државите да воспостават програми за безбедност на брани кои вклучуваат редовни инспекции (обично на секои неколку години), проценка на ризик, следење на однесувањето на конструкцијата и имаат обврска за известување. Во рамките на овие програми, операторите се должни да подготват т.н. планови за мониторинг и надзор (анг. „Dam Safety Surveillance and Monitoring Plans“), во кои детално се дефинираат типовите на мерни податоци, фреквенцијата на мерење и начинот на нивна анализа и известување. Дополнително, голем број сојузни држави имаат сопствени регулативи кои ги надополнуваат федералните барања, со што се обезбедува повеќеслоен систем на контрола и надзор. Сепак, важно е да се нагласи дека нивото на регулација може да варира во зависност од големината и ризикот на браната, при што помалите објекти во некои случаи подлежат на поограничени барања.

Генерално, во Европа и САД мониторингот и известувањето за браните не се сметаат за опционални активности, туку за основна законска и инженерска обврска насочена кон заштита на човечките животи, инфраструктурата и животната средина. Овие активности најчесто вклучуваат континуирано собирање и анализа на податоци, идентификација на потенцијални аномалии, како и благовремено известување до надлежните органи. Сепак, и покрај високото ниво на регулативна развиеност, во практиката сè уште доминираат класични, детерминистички пристапи базирани на гранични вредности и инженерско искуство, што отвора простор за унапредување преку примена на современи методи, како што се алгоритмите за машинско учење. Во таа насока, современите истражувања се насочени кон трансформација на традиционалниот мониторинг во интелигентни системи за предвидување и рана детекција на ризици, што претставува значаен чекор напред во управувањето со безбедноста на браните.

Во Кина, мониторингот и известувањето за состојбата на браните не само што се задолжителни, туку се регулирани преку централизиран и строго контролиран законски систем кој се смета за еден од најорганизираните во светот. Со оглед на огромниот број брани (најголем во глобални рамки) и значењето на водостопанските и хидроенергетските системи, Република Кина воспоставила силна институционална рамка, предводена од Ministry of Water Resources of China, која ја надгледува безбедноста, експлоатацијата и мониторингот на браните. Основниот законски документ што ја дефинира оваа област е Regulations on Safety Management of Dams, кој пропишува обврски за сопствениците и операторите на брани во однос на нивното одржување, следење и редовно известување.

Во рамки на овој систем, секоја брана е обврзана да има воспоставен систем за мониторинг кој

опфаќа инструментално следење на клучни параметри како што се деформации, порен притисок, филтрација и ниво на вода, како и редовни визуелни инспекции. Податоците од мониторингот мора систематски да се обработуваат и анализираат, а резултатите се доставуваат до надлежните институции преку стандардизирани извештаи. Покрај тоа, кинескиот систем вклучува и класификација на браните според нивниот ризик и значење, при што за високоризичните објекти важат построги барања за фреквенција на мониторинг, деталност на анализите и обем на известување. Законската регулатива исто така наложува периодични сеопфатни безбедносни ревизии, кои мора да се спроведуваат во точно определени временски интервали, како и изработка на планови за итни случаи.

Една од карактеристичните особености на кинескиот пристап е силната дигитализација и централизирано управување со податоците. Во последните години, значително се инвестира во развој на национални платформи за „паметен мониторинг“ (smart dam monitoring), каде што податоците од илјадници брани се собираат во реално време и се анализираат со напредни методи, вклучително и алгоритми за машинско учење. Овој пристап овозможува рана детекција на аномалии и подобро управување со ризикот на национално ниво. Оттука, може да се заклучи дека, слично како во Европа и САД, мониторингот и известувањето за браните во Кина се строго задолжителни и законски регулирани, но со дополнителен акцент на централизирана контрола и технолошка интеграција, што претставува значајна предност во однос на традиционалните, децентрализирани системи.

3.1. Хидросистем „Стрежево“

Хидросистемот „Стрежево“ претставува едноакмулационен повеќенаменски хидросистем лоциран во сливот на река Шемница, како дел од сливот на Црна река. Истиот е изграден во 1982 година, со примарна цел наводнување на 20200 ha од Пелагониското поле. Дополнително, водата од хидросистемот се користи за водоснабдување на градот Битола и околината, производство на електрична енергија, како и обезбедување на потребните количини на вода за потребите на РЕК Битола.

Хидросистемот „Стрежево“ се состои од следните објекти:

- 1) Алиментационен (собирен) канал,
- 2) Брана „Стрежево“ со придружни објекти,
- 3) Главен доводен канал,
- 4) Детална цевководна мрежа за наводнување,
- 5) Зафат и довод на вода за РЕК „Битола“,
- 6) Хидроелектрични централи.

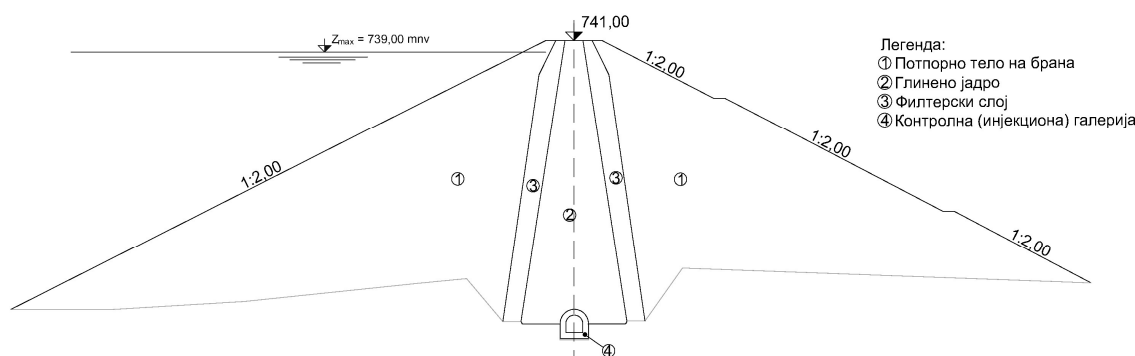
Податоците наведени во ова поглавје се преземени од Елаборатот за анализа и оценка на стабилноста и функционалноста на браната „Стрежево“ – Битола со придружните објекти и теренот околу акумулацијата за 2024 година (Кондински, 2025).

Браната „Стрежево“ е каменоземјана брана со централно глинено јадро, песочни филтерски слоеви, возводно и низводно потпорно тело од чакал и камени заштитни облоги на возводната и низводната страна. Браната е изградена на река Шемница, 24 km возводно од нејзиниот влив во Црна Река и на 1.5 km низводно од селото Стрежево. Со нејзината изградба се формира акумулационен простор со вкупен волумен од 119.23 милиони m³, од кои 108,87 милиони m³ се корисен волумен. Конструктивните карактеристики на брана „Стрежево“ се дадени во Табела 1.

Табела 1. Конструктивни големина за брана „Стрежево“

кота на круната на браната	741.00 mnv
кота на максимален водостој	739.00 mnv
кота на нормален водостој	737.50 mnv

кота на минимален водостој	694.00 mnv
висина на браната над речното корито	76 m
должина на круната на браната	632 m
ширина на круната на браната	10 m
максимална ширина во фундаментот	425 m
наклон на возводната косина	1 : 2 и 1 : 4.4
наклон на низводната косина	1 : 2
наклон на низводната баластна берма	1 : 10
наклон на косините на глиненото јадро	1 : 0.15
волумен на телото на браната	$4.219 \cdot 10^6 \text{ m}^3$
волумен на акумулацијата на кота на нормален водостој	$119.23 \cdot 10^6 \text{ m}^3$
корисен волумен	$108.87 \cdot 10^6 \text{ m}^3$
површина на водното огледало на кота на нормален водостој	432.86 ha
должина на водното огледало на кота на нормален водостој	6.25 km



Слика 1. Централен попречен пресек на брана „Стрежево“.

Собирниот канал од хидросистемот „Стрежево“ овозможува зафаќање на дел од водите на седум реки од Баба Планина: Кишавска, Граешка, Остречка, Злокуќанска, Стара Река, Киндерка и Драгор и нивно доведување до месноста Превалец, каде се внесуваат во сливот на реката Шемница. Од Собирниот канал се врши и дополнително снабдување со сирова вода на ЈП Водовод – Битола, а воедно претставува и алтернативен извор на сирова вода за потребите на термоелектраната Битола и Фабриката за шеќер во Битола.

Главниот доведен канал од хидросистемот „Стрежево“ има задача да ја транспортира и дистрибуира водата од акумулацијата „Стрежево“ до зафатните градби на главните цевководи од Деталната цевководна мрежа и цевководите за водоснабдување на РЕК Битола со сирова вода. Управувањето со течењето на водата низ каналот се врши со помош на методот на „динамичка регулација“, што претставува автоматизиран процес на далечинско управување од управувачки центар сместен во Дирекцијата на Системот.

Во склоп на техничкото решение на хидројазелот на браната за зафаќање на водите од акумулацијата, за евакуација на големите води, како и за празнење на самата акумулација изведени се поголем број на придружни објекти, чиј краток опис е даден во продолжение.

Зафатната градба (зафатна кула, доведен тунел, затворацница, дисипатор на водена енергија со дефлектор - бучница и прелив, и сигурносен испуст - брзоток со ски-скок) служи за зафаќање на водата од акумулација и доведување на истата до Главниот доведен канал, кој понатаму ја дистрибуира водата до низводните корисници. Лоцирана е на десниот брег од преградното место на браната, и предвидена е како длабински зафат со доведен тунел. Зафаќањето на водата е преку зафатната глава со груба решетка – кула зафат и преку доведен тунел со дијаметар 2.50 m со должина од 318.64 m, по што истата се доведува до ХЕЦ „Стрежево“ од каде се врши регулација на испуштањето на водата од акумулацијата. После ХЕЦ „Стрежево“, смирената вода се испушта во Главниот дводен канал. На 152.42 m од почетокот на доводниот тунел се наоѓа кула - затворацница во која е сместена фина решетка и табласт затворац со димензии 2.0/2.5 m. Оваа кула истовремено има улога и на водостан.

Преливниот орган овозможува евакуација на $Q_{\max} = 170 \text{ m}^3/\text{s}$, односно дел од максималните води со веројатност на појава 0.01% кои се пресметани како $400 \text{ m}^3/\text{s}$. Преливникот се наоѓа на левиот брег од преградното место на браната и се состои од вертикална преливна шахта со фиксен преливен раб, хоризонтален тунел, брзоток и слапиште. Преливната шахта има кота на круната 737.5 мнв, дијаметар во круната 14.0 m и должина на преливање 44.0 m, додека хоризонталниот тунел има дијаметар 3.30 m и должина 365 m кој завршува со брзоток со ширина 6 m и должина 255.2 m и слапиште за уништување на енергијата.

За евакуација на водата за време на градбата на браната, изведен е опточен тунел лоциран на десниот брег од преградното место. Истиот е со дијаметар 4.60 m и должина 476.80 m, а димензиониран е за пропуштање на водна количина од $150 \text{ m}^3/\text{s}$. После изградбата на браната опточниот тунел е пренаменет во темелен испуст, чија намена е празнење на акумулацијата како и испуштање на вода за биолошки загарантираниот минимум низводно од браната. Влезот во темелниот испуст е на кота 684.0 мнв каде е сместена груба решетка и помошен табласт затворац, а спојот со опточниот тунел е преку армирано-бетонски кос тунел со дијаметар 2.0 m и должина од 21.0 m. Во самиот опточен тунел почнувајќи од подолжната осовина на браната е поставена челична цевка со дијаметар 1500 mm и должина од 252.0 m, слободно ослонета на бетонски потпори. Почетокот на цевката е анкеран во бетонски чеп во опточниот тунел непосредно после кој е сместен сигурносен дисков затворац со дијаметар 1500 mm, а крајот е на самиот излез од опточниот тунел каде е сместен табласт регулационен затворац со дијаметар од 1500 mm.

Во иста вертикална рамнина со подолжната осовина на браната т.е. во темелот на глиненото јадро изведена е контролна (инјекциона) галерија, чија намена во текот на градењето била да овозможи изведување на работите на главната инјекциона завеса и контактното инјектирање, а во експлоатациониот период - контрола на однесувањето на браната и евентуални дополнителни инјекциони работи. Контролната галерија е изведена од од армиран бетон, со потковичаст попречен профил, со должина од 622.30 m.

Покрај опишаните придружни објекти, во рамки на хидросистемот „Стрежево“ функционираат и следните објекти:

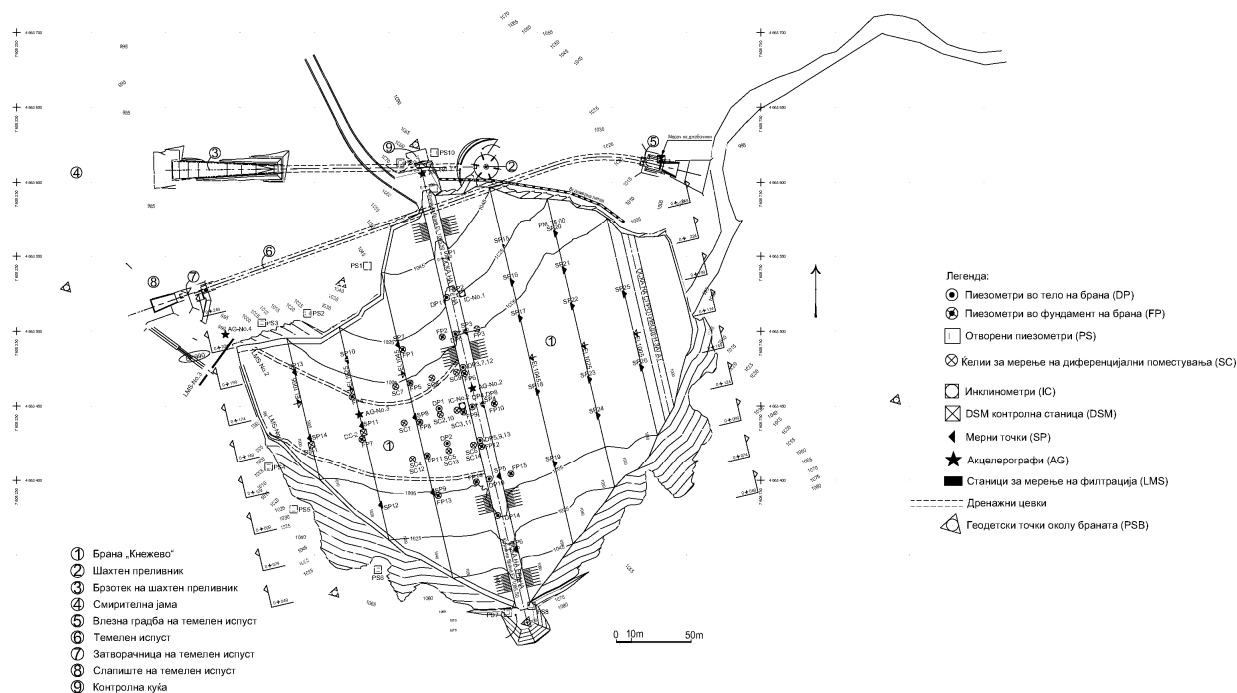
- командна зграда, во која е сместен персоналот за техничко набљудување на браната и дел од опремата за оскултација;
- мала метеоролошка станица;
- хемиско-бактериолошка лабораторија;
- рибник за калифорниска пастрмка
- експлоатациони патишта во должина од 2.52 km;
- електро-енергетски објекти: ХЕЦ „Стрежево“ и ХЕЦ „Експлоатационен минимум“, со кои е овозможено електро-енергетско искористување на водата за потребите на Главниот доведен канал и ХЕЦ „Биолошки минимум“, со која е овозможено електро-енергетско искористување на задолжителниот биолошки минимум кој се испушта во реката Шемница.

3.2. Хидросистем „Злетовица“

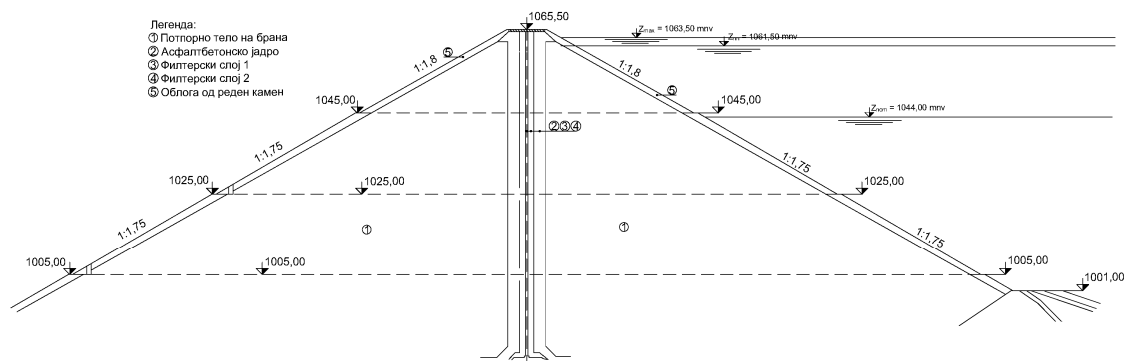
Хидросистемот „Злетовица“ претставува повеќенаменски систем за водоснабдување, наводнување и производство на електрична енергија лоциран во североисточниот дел на Р.С. Македонија во атарот на градовите Пробиштип и Кратово. Со работата на ХС „Злетовица“ се обезбедува вода за водоснабдување во општините Пробиштип, Штип, Свети Николе, Лозово и Карбинци. Хидросистемот се состои од брана „Кнежево“, преливен орган во десниот бок во форма на шахтен преливник со тунел, брзотек, ски отскок и смирителна јама проектирани на максимален проток од $Q_{\max} = 293 \text{ m}^3/\text{s}$, темелен испуст, прибранска хидроцентра и доводни цевководи за водоснабдување на општините Пробиштип, Штип, Свети Николе, Лозово и Карбинци (Dosevska Aleksova, 2023).

Реализацијата на ХС „Злетовица“ се одвива во повеќе фази:

- Фаза I – изградба на брана „Кнежево“ со придружни објекти, изградба на пристапен пат од Злетово до преградниот профил и изградба на доводни цевководи за водоснабдување до општините Пробиштип, Штип, Свети Николе, Лозово и Карбинци,
- Фаза II – изградба на систем за наводнување на површини во атарот на општината Пробиштип,
- Фаза III – изградба на мали хидроцентрали за производство на електрична енергија.



Слика 2. Основа на брана „Кнежево“ со опремата за оскултација.



Слика 3. Централен попречен профил на брана „Кнежево“.

Реализацијата на Фаза I е завршена во 2010 година со изградбата на каменонасипната брана „Кнежево“. Своевремено браната „Кнежево“ била единствена во Македонија од типот на каменонасипна брана со асфалтбетонско јадро. Максималната височина на браната над теренот изнесува 75m. Котата на круна на брана е 1065.50m, додека максималното водно ниво во акумулацијата е на кота 1063.50m. Нормалното работно ниво во акумулацијата е на кота 1061.50m. Во круната, браната е долга 290m, а широка 10m. Возводната и низводната косина е симетрични, со наклони 1:m = 1:1.8. Вкупниот волумен во акумулацијата е $23.5 \times 10^6 \text{ m}^3$.

На Слика 2 е прикажана основа на браната со инструментите за оскултација на браната, додека на Слика 3 даден е приказ на централниот, највисок попречен пресек на брана „Кнежево“.

Реализацијата на Фаза II е во тек, додека дел од Фаза III беше реализирано во изминатиот период со пуштање во употреба на прибранската хидроцентрала низводно од брана „Кнежево“.

именувана како „ХЕЦ Б“. Инсталираната хидроцентрала се наоѓа на цевковод DN500 кој е огранок (рачва) на темелниот испуст DN1200, а локациски се наоѓа веднаш низводно од темелната затворачница на брана „Кнежево“. Номиналниот нето пад на кој треба да работи хидроцентралата е $H_n = 54.10\text{m}$, со номинален проток од $Q_n = 0.350\text{ m}^3/\text{s}$ и инсталирана моќност од $P_{\text{ins}} = 200\text{ kW}$. Целта на оваа прибранска хидроцентрала е енергетски да се обработи протокот кој задолжително се испушта низводно од браната за обезбедување на биолошки загарантирано протекување. Проценето е дека на годишно ниво ќе се произведуваат околу 1320 MWh електрична енергија (Слика 4).



Слика 4. Проценето месечно производство на електрична енергија од „ХЕЦ Б“ (Aleksova Dosevska, 2025).

3.3. Лачна брана во Франција

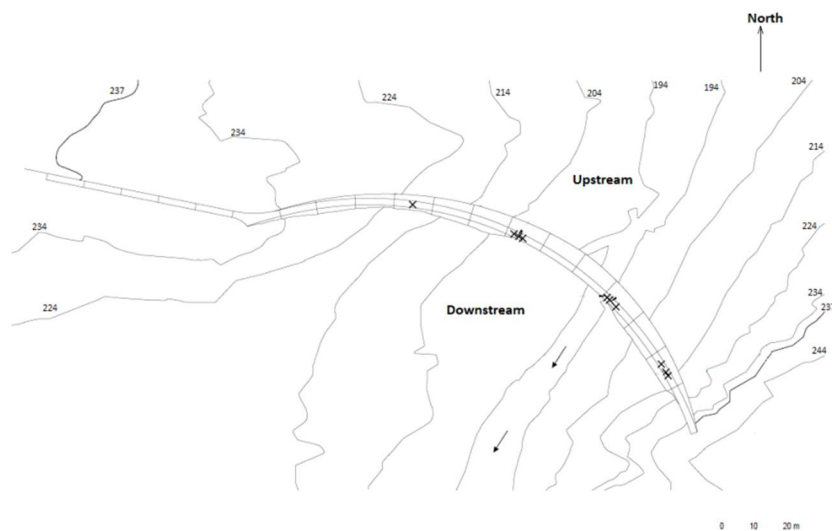
Трета брана која се разгледува во рамки на докторската дисертација е лачна брана од Франција со која управува јавната компанија „Électricité de France“ - или накратко EDF, која стопанисува со енергетските објекти во државата. Станува збор за бетонска лачна брана со височина од 45m, дебелина во круна на брана од 2m, дебелина во дното од браната од 6m, должина во круна на брана од 166m', работно водно ниво од 237m и ниво на круна на брана од 239m. Браната е изградена во периодот помеѓу 1957 и 1960 година, како тип на лачна брана со двојна закривеност, асиметрична поради обликот на речното корито.

Работното ниво во акумулацијата, круната на браната, како и сите понатамошни информации во врска со височините на браната не се мерени во однос на нивото на вода во морето туку се дадени во [m] мерени од една иста референтна височина.

Браната се наоѓа на речиси 2000mnnv, на непозната локација. Станува збор за студија на случај која е презентирана како тематска задача во рамки на 16-тата Интернационална работилница на нумерички анализи за брани (анг. „16th International Benchmark Workshop on Numerical Analysis of Dams“) во организација на Комитетот за компјутерски аспекти на анализи и проектирање на брани (анг. „Committee on Computational Aspects of Analysis and Design of Dams“) на Интернационалната комисија за големи брани (анг. „ICOLD - International Commission on Large Dams“), Тема А.

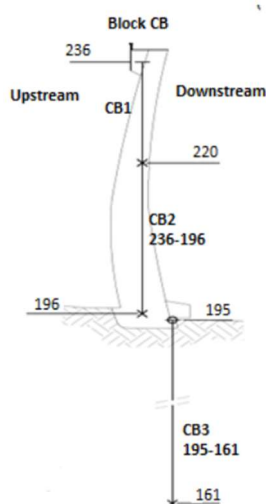
Предметната лачна брана е користена во студија на случај 2, каде се истражува стабилноста на браната преку анализа на мерените поместувања на браната и нивни предвидувања во иден период. Станува збор за бетонска лачна брана изработена од бетон во кој има цемент во количина од 300 kg/m³. Просечната јакост на притисок на бетонот изнесува 34 MPa (постигната по 90 дена зреење на бетонот).

Браната е опремена со голем број на инструменти за следење на нејзината стабилност и функционалност, вклучувајќи висоци, сензори за детектирање на дилатација на фугите, пиезометри и мерачи на процедурните води. Прибирањето на податоци за состојбата на браната се врши од првиот момент на полнење на акумулацијата. Низата на податоци прибрани со инструментите е веќе пречистена и обработена од страна на операторот, па како дел од тематската задача дадени се валидирани низи на податоци.



Слика 5. Основа на браната во Франција, со ознака на локацијата на висоците на браната во форма на ѕвезди.

Како дел од студија на случај 2, од тематските расположливи податоци од оскултацијата на бетонската брана ќе се користат податоците за радијални поместувања на браната измерени од висоците во централниот блок на браната (Слика 6).



Слика 6. Преглед на централниот блок на бетонската лачна брана со позиција на висоците за мерење на радијалните поместувања на блокот.

Во анализите кои се дел од студија на случај 2 од докторската дисертација, стабилноста на конкретната брана ќе се анализира со пресметка на радијалните поместувања во централниот блок „CB2“ како целна варијабла, во функција од водното ниво во акумулацијата „Zwl“ кое се регистрира на дневно ниво и температурата на воздухот „T“ која е дадена како интерполиран график за разгледуваната брана бидејќи на локација на самата брана не се вршат мерења на температурата.

Дополнително, во поглавјето „5. Тестирање и валидација на DamResML“ ќе се користат податоци од радијалните поместувања во централниот блок „CB3“ како целна варијабла, во функција од водното ниво во акумулацијата „Zwl“ кое се регистрира на дневно ниво и температурата на воздухот „T“. Со овие податоци ќе се изврши споредба на добиените резултати од предвидувањата пресметани со DamResML кодот и комерцијалниот софтвер „Neural Tools“ на Palisade.

“Machine intelligence is the last invention that humanity will ever need to make.”
Nick Bostrom

4. ОПИС НА КРЕИРАНИОТ КОД DamResML

4.1. Креирање на DamResML во Пајтон средина

Сите алгоритми за машинско учење кои се применуваат во истражувачкиот дел на докторската дисертација се изработени со помош на програмскиот јазик Пајтон, во апликацијата Jupyter lab. Се работи за код кој е создаден како комбинација од различни слободно достапни датотеки како numpy, pandas, warnings, sklearn, xgboost, matplotlib.pyplot и други. Самиот код се состои од вкупно 29 ќелии, секоја со различна задача поврзана со тренирање, тестирање, објава на резултати, прогноза на таргет варијаблата и оптимизација на истата.

Кодот, во понатамошниот текст ќе се повзума со акронимот ‘DamResML’ – акроним кој во себе ги содржи кратенките од англиските називи на поимите *брана* (анг. „Dam“), *акумулација* (анг. „Reservoir“) и *машинско учење* (анг. „Machine Learning“).

Во самиот код вклучени се следните алгоритми за машинско учење: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“), (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“), (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“), (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Кодот е концепиран во неколку чекори:

1. Чекор 1 – ќелии за внес на податоци. Во ова поглавје од кодот, истиот се повикува на табеларни надворешни датотеки со низа на податоци од интерес. Во надворешната датотека важен е распоредот на колоните на податоците. Првата колона секогаш треба да ги содржи податоците од целната таргет варијабла, додека останатите колони во датотеката претставуваат предиктор варијабли. Првиот ред во датотеката претставува текстуален ред во кој се пишуваат кратките ознаки на секој параметар, без разлика дали е целен или е предиктор.

2. Чекор 2 – ќелии за спроведување на преткратната вкрстена валидација на податоците за прелиминарна оцена на моделите. Во овој дел од кодот, за сите алгоритми се спроведува прекратна вкрстена валидација на податоците (анг. „5 fold cross validation“) за да се согледаат

метричките параметри на моделите со прелиминарен тренинг и предвидување на податоците со кои располагаме. Преткратната вкрстена валидација претставува процес во кој целиот сет на податоци се дели во 5 подсетови - четири се применуваат за тренирање на моделот по соодветниот алгоритам додека петтиот служи како тест сет. Ова се повторува пет пати – сè додека секој сет не стане и тест и тренинг сет. По завршување на овој процес на петкратно тестирање на податоците, кодот објавува просечни резултати за MSE, RMSE, MAE и R^2 параметрите. По нив, прелиминарно се оценува подобноста на применетите алгоритми за понатмошно обучување на моделите.

3. Чекор 3 – петкратна вкрстена валидација за секој модел со калибрирани хиперпараметри за подлабока анализа на секој модел. По оваа детална петкратна вкрстена валидација во која по потреба се извршува и фино подесување на хиперпараметрите на секој алгоритам, се донесува одлука кои сè од алгоритмите понатаму ќе се користат за тренирање на моделите. Се работи за вкупно девет модели, и иако целта на овој чекор е дел од моделите да се исфрлат – особено оние со неповолни метрички показатели, во секоја студија на случај се извршува тренирање со сите девет алгоритми. По завршување на овој процес на петкратно тестирање на податоците, кодот објавува просечни резултати за MSE, RMSE, MAE и R^2 параметрите за секој модел.

4. Чекор 4 – поделба на сетот на влезни податоци на 20% задржан сет (анг. „holdout“) и 80% сет за тренирање. Задржаниот сет се издвојува од сетот за тренирање со цел да се изврши предвидување на 20% од влезниот сет на податоци по завршување на тренингот на моделите. Со резултатите од ова предвидување на краток дел од сетот на податоци, кодот прикажува резултати за предвидувањата како важност на варијаблите (анг. „Feature importance“), пермутациска важност (анг. „permutation importance“) и графици со резидуали. Без издвојување на задржаниот сет, кодот нема да може да ги издвои посочените излези бидејќи ќе нема сет на податоци со кои ќе може да изврши кратко предвидување.

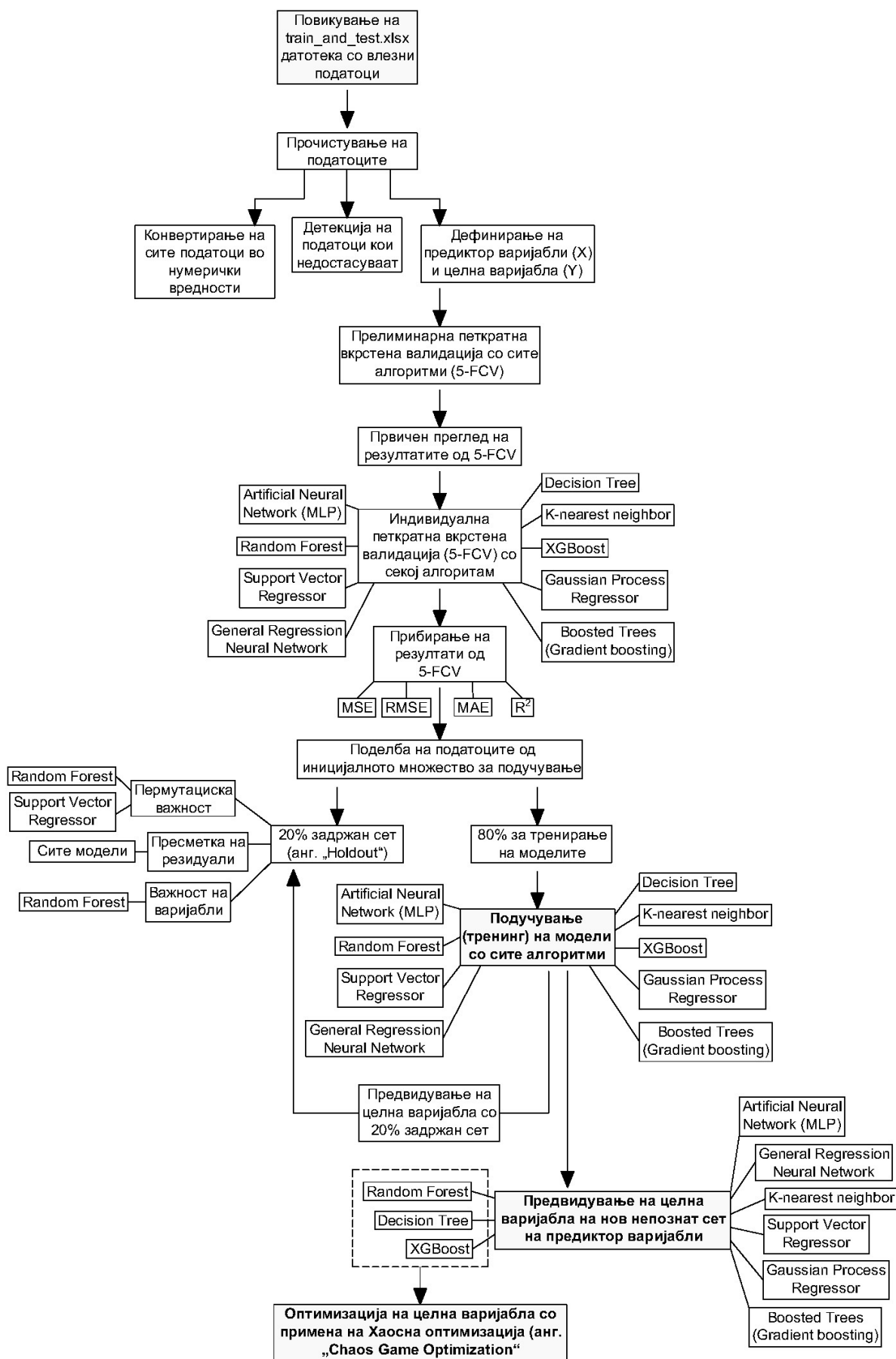
5. Чекор 5 – тренинг на моделите по секој алгоритам за машинско учење. Во ова поглавје се извршува обучување на девет модели со девет различни алгоритми за машинско учење, со примена на 80% од податоците од сетот за обучување кој е внесен во кодот во првото поглавје.

6. Чекор 6 – прелиминарни метрички параметри за издвоениот задржан сет на влезни податоци кој е исклучен од процесот од петкратната вкрстена валидација, со графички приказ на излезните резултати од метриките. Овој сет е издвоен од процесот на петкратна валидација, со цел истиот да се искористи како сет за предвидување од страна на тренираните модели и приказ на прелиминарни резултати. Станува збор за краток сет кој се применува пред сè за приказ на резидуали (остатоци од грешки) во предвидувањата со тренираните модели.

7. Чекор 7 – извршување на прогноза на таргет варијаблата врз основа на изработените тренирани модели. Во ова поглавје, модераторот повикува надворешна датотека со предиктор варијаблите, а кодот извршува прогноза врз основа на стекнатото знаење од претходното поглавје.

8. Чекор 8 – Хаос оптимизација – извршување на оптимизација на таргет варијаблата врз основа на резултатите од прогнозата за истата од претходното поглавје.

За јасен преглед на текот на преземање на влезните податоци, извршување авкрстената валидација, тренинг процесот, предвидувањето и оптимизацијата на целната варијабла, изработен е графикон, прикажан во Слика 7.



Слика 7. Приказ на чекорите во развој на моделите во Python.

4.2. Машинско учење

Уште во почетоците на создавањето на пресметувачките машини, инженерите се прашувале дали машината може да учи. Доколку би можеле да ја програмираме машината да учи, т.е. да се подобрува со тек на време врз основа на стекнатото знаење, исходот би бил неверојатен. Машините тогаш би се користеле за најразлични цели, како во медицината на пример при донесување на одлуки за комбинација на лекови за лекување на најразлични болести, или во домовите каде компјутерите би ја оптимизирале потрошувачката на енергија врз основа на потребите на домаќинството. Доколку успееме да ги натераме компјутерите да учат, би се отвориле голем број можности во сите сфери на хуманото функционирање, а единиците кои би ги користеле би добиле и предност во однос на својата конкуренција со зголемување на ефикасноста во своето работење (Mitchel, 1997). Во моментот, сеуште немаме доволно познавање за тоа како да ги натераме пресметувачките машини да учат, но различни алгоритми се развиени кои ефикасно се справуваат со одредени задачи кои бараат учење. Во процесот на копање на податоци (анг. „data mining“) голем број на алгоритми од машинско учење се применуваат за да се откријат корисни информации од големи комерцијални бази на податоци како на пример бази од медицински центри, банки, и други. Очигледно, со збогатување на нашите познавања кон работата и функционирањето на пресметувачките машини, ќе станува неизбежна примената на машинското учење во компјутерските науки и технологија (Mitchel, 1997).

Машинското учење претставува мултидисциплинарно подрачје кое опфаќа повеќе полиња како вештачката интелигенција, веројатност и статистика, теоријата на пресметувачката комплексност, филозофија, психологија, невробиологија и други. Машинското учење е дел од т.н. меко програмирање (анг. „soft computing“), кое има цел да пронајде решенија на проблеми кои се тешки или нерешливи со примена на традиционалните „тврди“ методи на пресметка (Ibrahim, 2016). Мекото програмирање се состои од различни техники кои реплицираат донесување одлуки за одредени проблеми кои наликуваат на природното хумано донесување одлуки (Amin Hariri-ARdebili, Salazar, Pourkamali-Anaraki, Mazza, & Mata, 2023). Мекото програмирање се базира на парцијална вистина, несигурност и апроксимација во решавањето на сложени проблеми. Главниот принцип во мекото програмирање е користењето на толеранцијата на неточноста, несигурноста и парцијалната вистина за да се постигнат резултати поблиски до реалноста (Zadeh, 1993). На Слика 8 може да се воочи груб опис на разликите во пристапот и размислувањето кај тврдото програмирање и мекото програмирање (Gupta & Kulkarni, 1993). Тврдото програмирање се базира на точни математички пресметки за решавање на конкретни проблеми, за разлика од мекото програмирање кое се потпира на генетски алгоритми, создадени да донесуваат одлуки на начин како што тоа го прави нервниот систем кај човекот – хаотично, врз основа на претходно искуство.



Слика 8. Разлика во тврдо и меко програмирање (Gupta & Kulkarni, 1993).

Бидејќи досега многупати се спомна терминот „алгоритам“, редно е да се даде кратко појаснување што се подразбира под овој термин. Едноставна дефиниција за „алгоритам“ е: алгоритам претставува системски сет на функции кои се извршуваат на одредена низа на податоци, налик на процедура (Paul Muller & Massaron, 2021). Во суштина, машинското учење се потпира на алгоритми за да анализира обемни низи на податоци. Во моментот, ни најдобрите алгоритми не можат да размислуваат, чувствуваат, да покажат самосвест или да искажат слободна волја. Единствено што може да направат е да извршат предикциски анализи многу побрзо од било

кој човек. Тоа е клучната предност на моделите со машинско учење – може да го направи секојдневното функционирање на човекот поефикасно споредено со примена на други методи во решавање на истата задача (Paul Muller & Massaron, 2021). Оттука, не треба да се мешаат термините „учење“ и „интелегенција“. Пресметувачките машини, со помош на алгоритмите од машинско учење, стекнуваат знаење за одредената задача, но немаат способност на свесност нити пак за донесување на одлуки надвор од стекнатото знаење.

При изработка на моделите со машинско учење, чекорите во кои се креира моделот може да се опишат во следниот редослед: (1) Прибирање на податоци и подготовка на базата на податоци со јасна поделба на таргет и предиктор варијабли, (2) Избор на алгоритми од машинско учење, (3) Валидација на влезните податоци, (4) Избор на алгоритми за тренирање на моделите, (5) Предвидување на таргет варијаблата врз основа на низи на податоци од предиктор варијаблите. Како што следи, даден е краток опис на секој чекор од изработка на моделите, со детален опис на применетите алгоритми.

4.2.1. Подготовка на базата на податоци

База на податоци може да биде било кој сет на податоци кој содржи информации од било каква природа. Веројатно најважно е да имаме интерес да ги обработуваме базите на податоци и да извлечеме некако „знаење“ од истите. Во овој иницијален чекор од моделирањето, треба да се напомене дека базите на податоци најчесто содржат голем број на вредности кои се хаотични, без очигледен ред и тренд.

Секој сет на податоци, уште во оваа иницијална фаза на проучување на податоците, се дели на таргет или целна варијабла, и предиктор варијабли. Целна варијабла е варијаблата за која ќе се изработи модел и за која ќе се прават предвидувања од финално тренираните модели. Предиктор варијабли се сите останати податоци во сетот кои модераторот смета дека се во корелација со таргет варијаблата. Сите варијабли може да бидат опишани како нумерички или описни варијабли, но моделите со машинско учење најдобро ги разбираат нумеричките варијабли, особено регресионите модели (Paul Muller & Massaron, 2021). Нумерички вредности имаат варијаблите кои по својата природа се бројки, додека описните варијабли може да содржат текстуална содржина, знаци и други компоненти.

Моделите со машинско учење, интуитивно, подобро „учат“ кога овие бази се обемни, со голем број на податоци. Во Табела 2 може да се согледаат примери од различни алгоритми и минималниот потребен број на податоци за тренирање на моделите (Alwosheel, van Cranenburgh, & G. Chorus, 2018), (Blum & Langley, 1997).

Табела 2. Преглед на повеќе алгоритми за машинско учење и минималниот потребен број на податоци за тренирање на моделите.

Алгоритам	Минимален број на податоци потребен да се создаде модел	Посакуван број на податоци за да се создаде модел
Линеарна регресија (анг. „Linear Regression“)	30 – 50	100+
Одлучувачко дрво (анг. „Decision Tree“)	80 -100	200+
Случајна шума (анг. „Random Forest“)	100 – 150	300+
Гаусов процесен регресор (анг. „Gaussian Process Regression“)	30 – 100	<1000
Регресија со потпомогнати вектори (анг. „Support Vector Regression“)	20 – 120	200+

Големината на базата на податоци значително влијае врз пристрасноста на моделот (анг. „Bias“) и варијансата на моделот (анг. „Variance“). Пристрасноста претставува систематска грешка на моделот која настанува која моделот е премногу едноставен за да ја опише вистинската зависност помеѓу предиктор варијаблите. Високи вредности за пристрасноста води кон незадоволителни вредности на таргет варијаблата дури и во процесот на тренирање на моделот, како и на потценување на предвидените вредности на целната варијабла (анг. „underfitting“). Варијансата, од друга страна, ја претставува чувствителноста на моделот кон промените во сетот на податоци. Висока вредност на варијансата води кон преценување на предвидените вредности, а во процесот на тренирање може да покаже одлични резултати и да го наведе модераторот кон погрешни заклучоци.

Во машинското учење не постои универзално дефинирана минимална големина на множествоо на податоци, но постојат повеќе насоки и емпириски истражувања кои обезбедуваат корисни референтни вредности. Често користено правило укажува дека бројот на примероци за тренирање треба да биде најмалку 10 пати поголем од бројот на предиктор варијабли. Понапредни истражувања посочуваат дека големината на множеството на податоци треба да се движи помеѓу 10 и 100 пати од бројот на варијаблите, во зависност од сложеноста на моделот. Современите истражувања исто така покажуваат дека малите множества (на пример, помалку од 300 примероци) имаат тенденција да даваат преоптимистички резултати поради претерување во предвидувањата (анг. „overfitting“), особено кај сложени модели како што се невронските мрежи. Затоа, потребната големина на податочниот сет зависи од конкретниот проблем и е под влијание на фактори како што се димензионалноста на карактеристиките, сложеноста на моделот и варијабилноста на податоците.

4.2.2. Избор на алгоритми од машинско учење

Во оваа фаза од моделирањето, модераторот прави избор кои сè алгоритми ќе ги користи за тренирање, валидација и предвидување во понатамошните чекори. Изборот е огромен – постојат над 30 различни алгоритми кои активно се користат и тоа во презентирани резултати од истражувања од областа на хидротехниката (Liu, Zhou, & Yang, 2024). Во зависност од природата на проблемот, се одбира модел за решавање на регресиони задачи, класификациски задачи или кластер задачи. Во конкретниот истражувачки модел применет во рамки на докторската дисертација, изработен е код во Пајтон во кој вклучени се следните алгоритми за машинско учење: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Како што следи, презентирани е кус опис на секој алгоритам за машинско учење кој е имплементиран во Пајтон кодот.

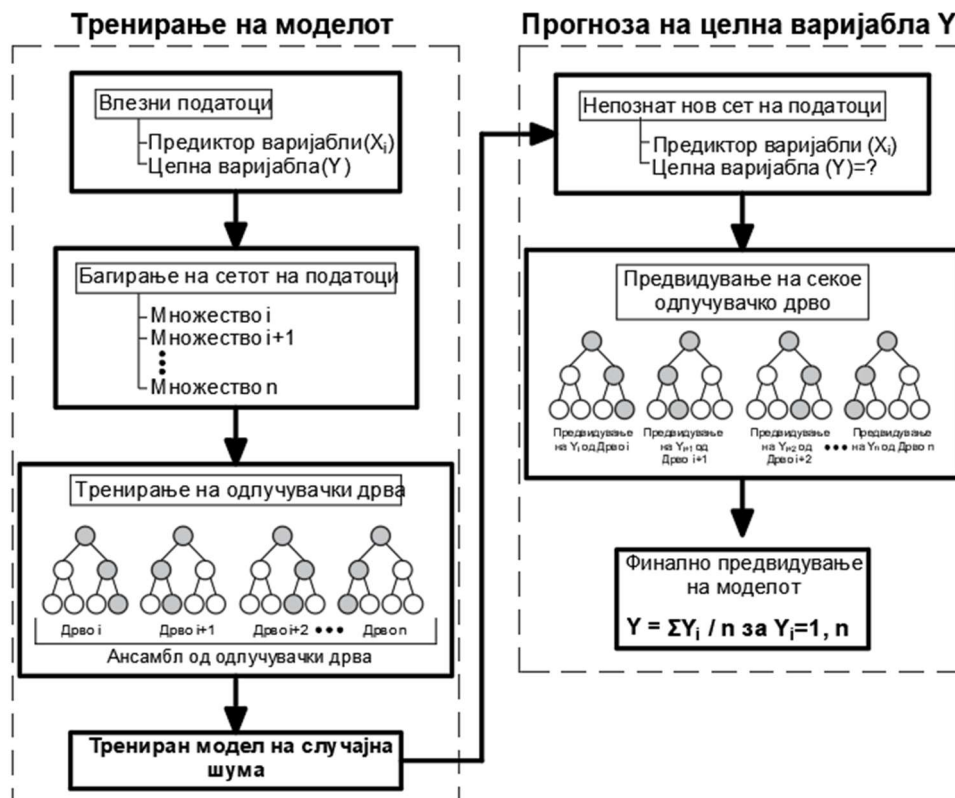
4.2.2.1. Случајна шума (Random Forest)

Алгоритмот „Случајна шума“ (анг. „Random Forest“) или накратко RF претставува алгоритам од повеќе дрва на одлучување (анг. „Decision trees“) (Breiman, Random Forests, 2001). Станува збор за ансамбл од повеќе дрва на одлучување, кои се создаваат врз основа на багирање (анг. „Bagging – Bootstrap Aggregating“) на сетот на податоци за учење на моделот. Багирање претставува ансамбл-метод во машинското учење при кој повеќе модели се обучуваат врз основа на случајно избрани примероци од податочното множество, а нивните предвидувања се агрегираат за да се добие конечниот резултат (Breiman, Bagging Predictors, 1996). Под „агрегација“ се подразбира прибирање на резултатите од предвидувањето на секое одлучувачко дрво, при што финалниот прогнозен резултат за целната варијабла Y е секогаш просечна, средна вредност од предвидувања на сите одлучувачки дрва, како што е прикажана во структурата на тренинг и прогноза на Слика 9.

Принципот на функционирање на алгоритмот Случајна шума е појаснет како што следи. Нека алгоритмот Случајна шума се дефинира како:

$$f(x, b_k), k = 1 \dots n \quad (4.2.1)$$

Во равенка (4.2.1), $f(x, b_k)$ претставува класификатор со структура на дрво, каде x претставува множеството на влезни податоци за тренирање на моделот, b_k се множества на примероци одбрани од множеството x и секое од множествата k го одредува формирањето на посебно дрво на одлучување. Секое множество b_k има иста димензија како и множеството на влезни податоци x . Формирањето на меѓусебно релативно независни дрва е клучот на овој алгоритам бидејќи на тој начин дрвата се заштитуваат едни со други од своите индивидуални грешки во класификацијата (Markovska, 2020).



Слика 9. Структура на тренирање и прогноза кај алгоритмот „Случајна шума“.

Основниот принцип на алгоритмот на „Случајна шума“ се базира на случајно одбирање на примероци од оригиналното множество за тренирање и формирање на b_k множества, а одбирањето на примероците се прави со замена на варијаблите. Оваа постапка се нарекува багирање (анг. „Bagging – Bootstrap Aggregating“) на сетот на податоци за учење на моделот. За да се појасни постапката на багирање, даден е пример како што следи.

Нека е дадено оригинално множество x за тренирање на моделот со „Случајна шума“, со варијабли $x = \{1, 2, 3, 4, 5, 6\}$. Со багирање, може да се формираат повеќе множества, на пример множество $b_1 = \{1, 2, 3, 4, 4\}$ за тренирање на одлучувачко дрво T_1 , и множество $b_2 = \{1, 2, 2, 3, 4, 4\}$ за тренирање на одлучувачко дрво T_2 . Новоформираните множества имаат иста димензија како оригиналното множество, но варијаблите не се идентични. Сите примероци од оригиналното множество се вклучуваат во новите множества, но истите подлежат на замена, со што дел од нив се повторуваат, а дел недостасуваат. На овој начин се формираат случајни и меѓусебно независни множества за обучување на дрвата во шумата, со што се воведува случајност во алгоритмот.

Врз сложеноста на моделот на „Случајна шума“ влијаат следните параметри:

(1) Бројот на одлучувачки дрва ($n_estimators$),

- (2) Длабочината на секое дрво (`max_depth`),
- (3) Бројот на примероци во секој лист од одлучувачките дрва (`min_samples_split`), и
- (4) Број на различни варијабли кои се земаат предвид во секој јазол (`max_features`).

Бројот на одлучувачки дрва n_T директно влијае врз длабочината на моделот и неговата точност, но не секогаш голем број на дрва значи и многу точен модел. Во Пајтон кодот, оваа карактеристика се внесува преку функцијата `n_estimators` и истата се варира во три инстанции – за вредности $n_T = 50, 100$ и 150 . Моделот формира шума од 50, 100 и 150 дрва, а потоа донесува одлука за најдобриот број на дрва од анализираните со помош на вгнездена трикратна внатрешна валидација (анг. „GridSearchCV“).

Длабочината на секое дрво T_{depth} расте сè до ограничената вредност преку функцијата `max_depth` во `DamResML` кодот. Во конкретните анализи, длабочината се тестира за три инстанции – нула, 10 и 20. Помала длабочина на дрвото може да доведе кон едноставен модел со ниска варијанса и зголемена генерализација на предвидувањата. Од друга страна, голема длабочина на дрвото го усложнува моделот, ја зголемува варијансата и може да предуслови модел кој ги препознава нелинеарните врски помеѓу варијаблите.

Бројот на примероци во секој лист n_{leaf} од одлучувачките дрва T_i се ограничува со функцијата `min_samples_split` преку кој на моделот му се посочува со колку примероци да се подели секој јазол, т.е. колку примероци да има во секој лист од дрвата. Функцијата, во `DamResML` кодот се анализира во три инстанции – за вредности 2, 5 и 10. За помали вредности на бројот на примероци во секој лист, се формираат многу длабоки дрва и постои ризик за претерување во предвидувањата. Поголемиот број на примероци во листовите може, пак, да доведе кон поедноставни дрва и зголемена генерализација на предвидувањата.

Бројот на варијабли m , од вкупниот број на варијабли M , кои се земаат предвид во секој јазол се контролира преку функцијата `max_features` која се тестира во три инстанции, за вредности $m = \begin{cases} \sqrt{M} \\ \log_2 M \\ M \end{cases}$.

Сите варијанси на параметрите n_T, T_{depth}, n_{leaf} и m се комбинираат за тренирање на модели со алгоритмот за „Случајна шума“, а со примена на внатрешна трикратна вкрстена валидација се врши оптимизација на вредностите на истите. Со оптимизирани вредности на параметрите се тренира финален модел на кој дополнително се врши надворешна петкратна валидација. Финалниот трениран модел се применува во подоцнежната фаза на предвидување на целната варијабла.

4.2.2.2. Вештачка невронска мрежа (Artificial Neural Network)

Вештачките невронски мрежи (анг. „Artificial Neural Networks“) или накратко ANN претставуваат една од најзначајните и најшироко применуваните техники во областа на машинското учење, инспирирани од начинот на функционирање на човечкиот мозок (Gurpey, 1997). Тие се составени од меѓусебно поврзани елементарни единици наречени „неврони“, кои колективно овозможуваат моделирање на сложени нелинеарни зависимости помеѓу влезните и излезните податоци. Благодарение на нивната способност за учење од податоци, ANN се применуваат во различни инженерски дисциплини, вклучувајќи хидротехника, каде што се користат за предвидување на проток, ниво на вода, производство на хидроенергија и анализа на однесување на брани.

Една вештачка невронска мрежа се состои од три основни типови слоеви:

- Влезен слој (анг. „Input layer“) – ги прима влезните податоци (на пример: доток, ниво на акумулација, испуст, итн.)
- Скриени слоеви (анг. „Hidden layers“) – вршат трансформација на податоците преку низа од математички операции

→ Излезен слој (анг. „Output layer“) – го дава конечниот резултат (на пример: предвидена моќност или проток)

Секој неврон во мрежата прима влезни вредности, ги множи со соодветни тежини (анг. „weights“), додава пристрасност (анг. „bias“) и ја пропушта сумата низ активациска функција.

Работата на еден неврон може да се претстави со следната равенка:

$$y = f(\sum_{i=1}^n w_i x_i + b) \quad (4.2.2)$$

, каде што:

- x_i се влезните вредности
- w_i се тежинските коефициенти
- b е пристрасност (поместување)
- f е активациска функција (ReLU, sigmoid, tanh итн.)

Оваа функција овозможува мрежата да моделира сложени нелинеарни процеси, што е особено важно во инженерски проблеми како што се и проблемите поврзани со водостопанското управување.

Процесот на тренирање на Вештачките невронски мрежи се одвива во неколку чекори:

1. Пропагирање напред (анг. „Forward propagation“). Во оваа фаза, влезните податоци се пропуштаат низ мрежата, слој по слој, сè до добивање на излез.

2. Пресметка на грешка (анг. „Loss function“). Во овој чекор се пресметува разликата помеѓу предвидениот излез и вистинската вредност, најчесто преку MSE (анг. „Mean Squared Error“) и MAE (анг. „Mean Absolute Error“),

3. Пропагирање назад (анг. „Backpropagation“). Во овој чекор, грешката се враќа назад низ мрежата, при што се пресметуваат градиенти за секоја тежина,

4. Ажурирање на тежини. Во овој чекор, тежинските коефициенти се ажурираат со цел да се намали грешката, најчесто преку алгоритам како Постепено спуштање (анг. „Gradient Descent“) или Оптимизаторот Адам (анг. „Adam optimizer“).

Процесот на тренирање се повторува повеќе пати (епохи) сè додека моделот не научи оптимални параметри. Комплексноста на невронската мрежа зависи од следните параметри (1) бројот на скриени слоеви n_{layer} , (2) бројот на неврони во секој слој $n_{neurons}$, (3) типот на активациската функција f_{type} , (4) брзината на учење l_{rate} , и (5) бројот на епохите n_{epochs} . Поголема мрежа може да ги опфати и нелинеарните зависности помеѓу влезните варијабли, но кај истата постои ризик од претерување на предвидувањата.

Во конкретниот DamResML код во кој е внесен алгоритмот за „Вештачки невронски мрежи“, во однос на параметрите, внесени се следните вредности:

1) Број на скриени слоеви n_{layer} и бројот на неврони во секој слој $n_{neurons}$ се внесени како `hidden_layer_sizes=(100,100)`. Со ова, внесени се 2 скриени слоеви со 100 неврони во секој од нив.

2) Типот на активациска функција е усвоен како `solver="adam"` што значи дека се применува Оптимизаторот Адам (анг. „Adam optimizer“) како активациска функција.

(3) Брзината на учење l_{rate} е усвоена како `learning_rate_init=0,001`, вообичаена за оптимизаторот Адам. Брзината на учење l_{rate} влијае на тежинските коефициенти, и тоа:

$$w_{i+1} = w_i - l_{rate} \cdot g \quad (4.2.3)$$

, каде:

w_{i+1} – тежински коефициент во претходен чекор на учење,

w_i – тежински коефициент во сегашен чекор на учење,

l_{rate} – брзина на учење

g – градиент на грешка во однос на тежински коефициент, $g = \frac{\Delta Loss}{\Delta w}$

На овој начин тежинските коефициенти континуирано се корегираат во секој чекор на учење на моделот.

Откако согласно алгоритмот се формира модел кој стекнал знаење врз основа на влезните податоци, моделот се применува во предвидување на целната варијабла со внесување на нови влезни податоци. Во процесот на предвидување се извршува единствено пропагацијата напред (анг. „Forward propagation“) и нема ажурирање на тежинските коефициенти.

4.2.2.3. Регресија со потпомогнати вектори (Support Vector Regressor)

Регресијата со потпомогнати вектори (анг. „Support Vector Regressor“) или накратко SVR претставува надгледуван алгоритам за машинско учење кој произлегува од теоријата на Потпомогнати векторски машини (анг. „Support Vector Machines“) или накратко SVM, развиена од Владимир Вапник (Cortes & Vapnik, 1995). Додека SVM првично е наменет за класификација, SVR е негова адаптација за решавање на регресиски проблеми, односно предвидување на континуирани вредности.

Основната идеја на SVR не е да се минимизира грешката за секој поединечен примерок, туку да се пронајде функција која најдобро ги опишува податоците во рамки на однапред дефинирана толеранција. Оваа толеранција се дефинира преку параметарот ε (анг. „epsilon“), кој ја претставува дозволената грешка околу регресиската функција. Со други зборови, SVR се обидува да најде функција $f(x)$ која отстапува од реалните вредности најмногу за ε , при што сите грешки помали од ε не се казнуваат.

Математички, задачата на SVR се формулира како оптимизациски проблем, каде целта е да се минимизира комплексноста на моделот (односно да се обезбеди „рамна“ функција), истовремено дозволувајќи одредени отстапувања од реалните вредности. Ова се постигнува преку минимизација на нормата на тежинскиот вектор, што резултира со добра генерализација на моделот врз нови, непознати податоци.

Еден од клучните концепти кај SVR е употребата на таканаречени „потпорни вектори“ (анг. „support vectors“). Тоа се податоци кои лежат на границите на ε - интервалот или надвор од него и директно влијаат врз дефинирањето на регресиската функција. За разлика од класичните регресиски методи, кај SVR само мал дел од потпорните вектори т.е. од податоците ја дефинираат конечната функција, што го прави моделот ефикасен и отпорен на шум (Smola & Schölkopf, 2004).

Дополнително, SVR користи т.н. техника „kernel trick“, односно техника која овозможува трансформација на податоците во повисокодимензионален простор, каде што линеарната регресија станува возможна и за нелинеарни зависимости. Најчесто користени кернел функции се (M. Bishop, 2006):

- линеарен кернел,
- полиномијален кернел,
- радијално-базирана функција (анг. „RBF kernel“),
- сигмоиден кернел.

Особено RBF кернелот е широко применет поради неговата способност да моделира комплексни нелинеарни врски, што е од големо значење во инженерските проблеми, вклучително и во хидротехниката и управувањето со водостопански системи.

SVR содржи неколку важни хиперпараметри кои значително влијаат на перформансите на моделот (M. Bishop, 2006). Најзначајни се:

- 1) параметарот C , кој го контролира балансот помеѓу сложеноста на моделот и дозволената грешка,
- 2) ε (epsilon), кој ја дефинира ширината на зоната на толеранција,
- 3) γ (gamma), кој ја контролира влијанието на поединечните податоци кај RBF кернелот.

Висока вредност на параметарот C резултира со модел кој се обидува да ги минимизира грешките по секоја цена, што може да доведе до претерување во предвидувањето (overfitting). Од друга страна, помала вредност на C овозможува поголема генерализација, но со потенцијално поголема грешка. Слично, параметарот γ влијае на тоа колку „локално“ или „глобално“ моделот ја интерпретира зависноста помеѓу влезните и излезните податоци (Hastie, Tibshirani, & Friedman, 2009).

Една од главните предности на SVR е неговата способност да работи ефикасно со податочни сетови со висока димензионалност и релативно мал број на примероци. Ова го прави погоден за инженерски апликации каде податоците може да бидат ограничени, но комплексни. Дополнително, SVR покажува добра отпорност на претерување во предвидувањата, особено кога е правилно калибриран.

Сепак, алгоритмот има и одредени ограничувања. Изборот на соодветен кернел и оптималните хиперпараметри може да биде сложен процес и често бара употреба на техники како што се вкрстена валидација и оптимизација преку мрежно пребарување. Исто така, SVR може да биде пресметковно скап при големи податочни сетови, бидејќи времето на тренирање расте со бројот на примероци.

Во контекст на ова истражување, SVR е применет како еден од основните модели за предвидување на целните променливи во водостопански системи, при што неговата способност за моделирање на нелинеарни зависности и добра генерализација претставуваат значајна предност во споредба со класичните регресиски методи.

4.2.2.4. Одлучувачки дрва (Decision Trees)

Одлучувачките дрва (анг. „Decision Trees“) или накратко DT претставуваат еден од најинтуитивните и широко користени алгоритми во машинското учење, применливи како за класификација, така и за регресија. Во контекст на регресиски проблеми, DT се користи за предвидување на континуирани вредности преку рекурзивно делење на податочниот простор на помали и похомогени подгрупи (Hastie, Tibshirani, & Friedman, 2009).

Основната идеја на алгоритмот е да се изгради структура во форма на дрво во која секој внатрешен јазол претставува услов (тест) врз одредена влезна променлива, секоја гранка претставува резултат од тој услов, а секој лист (анг. „leaf node“) содржи предвидена вредност. Процесот започнува од коренскиот јазол (анг. „root node“), каде целиот податочен сет се дели врз основа на критериум кој најдобро ја намалува варијабилноста на целната променлива.

Кај регресиските дрва, најчесто користен критериум за делење е минимизација на средноквадратната грешка (анг. „Mean Squared Error – MSE“) или варијансата во рамките на секој јазол. Алгоритмот избира поделба која резултира со најголемо намалување на варијансата, односно создава подгрупи во кои вредностите на целната променлива се што е можно поуниформни (James, Witten, Hastie, & Tibshirani, 2021).

Процесот на градење на дрвото се одвива рекурзивно и е познат како алчен алгоритам (анг. „greedy algorithm“), бидејќи во секој чекор се избира локално оптималната поделба без гаранција за глобално оптимално решение. Дрвото продолжува да се разгранува сè додека не се исполни некој од условите за запирање, како што се:

- 1) достигнување на максимална длабочина (`max_depth`),
- 2) минимален број на примероци во јазол (`min_samples_split`),
- 3) или минимален број на примероци во лист (`min_samples_leaf`).

Крајниот резултат е структура која може лесно да се интерпретира, бидејќи патот од коренот до листот претставува низа од логички услови кои водат до одредена предвидена вредност.

Една од главните предности на одлучувачките дрва е нивната висока интерпретабилност. За разлика од моделите „црна кутија“ (анг. „black box“) како што се невронските мрежи, DT овозможуваат јасно разбирање на влијанието на поединечните влезни параметри врз резултатот.

Ова е особено важно во инженерските апликации, каде интерпретацијата на резултатите има практично значење.

Дополнително, алгоритмот не бара нормализација или скалирање на податоците и може да работи со податоци од различни типови (нумерички и категоријални). Тој исто така е релативно отпорен на нелинеарности и може да моделира комплексни односи помеѓу променливите.

Сепак, DT имаат и значајни недостатоци. Еден од главните проблеми е нивната склоност кон претерување во предвидувањата (анг. „overfitting“), особено кога дрвото е длабоко и премногу комплексно. Во такви случаи, моделот може да ја „научи“ шумата во податоците наместо вистинската зависност, што резултира со слаба генерализација врз нови податоци.

За да се намали овој ефект, се користат техники како што се:

- 1) ограничување на длабочината на дрвото,
- 2) поставување минимален број на примероци по лист, или
- 3) примена на техники за кастрење на дрвото (анг. „Pruning“), со што се отстрануваат гранки кои не придонесуваат значајно за подобрување на моделот.

Исто така, DT се нестабилни во смисла дека мали промени во податочниот сет може да резултираат со значително различна структура на дрвото. Оваа слабост е една од причините зошто често се користат ансамбл методи како RF и Gradient Boosting, кои комбинираат повеќе дрва за да се постигне подобра стабилност и точност.

Во математичка смисла, предвидувањето кај регресиските дрва се врши како просечна вредност на целната променлива за сите примероци кои припаѓаат на даден лист. Ова значи дека моделот создава парцијално константна апроксимација на вистинската функција (анг. „Piecewise constant“), што може да биде ограничување при моделирање на многу мазни зависимости (Breiman, Friedman, Olshen, & J. Stone, 1984).

Во контекст на ова истражување, DT е применет како базичен модел за споредба со понапредни алгоритми. Неговата едноставност, брзина и интерпретабилност го прават погоден за иницијална анализа на податоците и идентификација на релевантни влезни параметри. Сепак, поради неговата склоност кон претерување во предвидувањата, резултатите од овој модел се внимателно споредени со ансамбл методи и други алгоритми со подобра генерализација.

4.2.2.5. К – најблизок сосед (K – nearest neighbor)

Алгоритмот К – најблизок сосед (анг. „K-Nearest Neighbors“) или накратко KNN претставува еден од наједноставните, но истовремено и многу ефикасни алгоритми од областа на машинското учење. Тој припаѓа во групата на алгоритми со надгледано учење, што значи дека за негово обучување се користат податоци кај кои се познати влезните параметри и соодветните излезни вредности (Altman, 1992). KNN може да се применува и за класификација и за регресија, а во рамките на ова истражување е користена неговата регресиска варијанта, позната како K-Nearest Neighbors Regressor, со цел предвидување на континуирани вредности.

Основната идеја на алгоритмот е дека сличните случаи во минатото најчесто имаат слични резултати. Кога ќе се внесе нов примерок за предвидување, алгоритмот не гради класичен математички модел како кај другите методи, туку ги бара К најблиски соседи од веќе постоечките обучувачки податоци (Cover & Hart, 1967). Еден пример за класификациски проблем шематски претставен на график е даден на Слика 10.

Близината помеѓу соседите најчесто се определува преку математичка мерка за растојание, како што е Евклидово растојание, кое се пресметува според равенката:

$$d_{x,y} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.2.4)$$

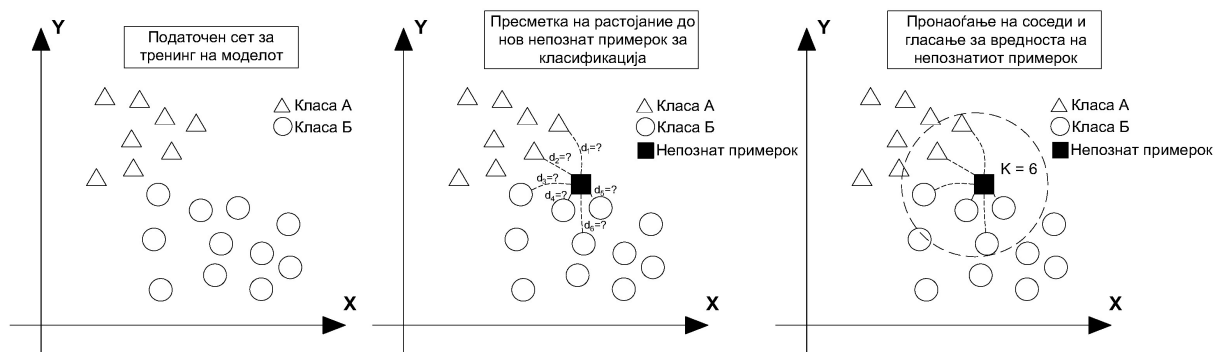
,каде што x_i и y_i се две точки во n-димензионален простор, а $d_{x,y}$ е нивното меѓусебно растојание (Cover & Hart, 1967).

Растојанието може да се пресметува и по други математички релации, како на пример пресметка на растојанието по Менхетан (анг. „Manhattan Distance“) или пак по Минковски (анг. „Minkowski Distance“). Пресметката на растојанието по Менхетан е прикажано во равенката (5) додека по Минковски – во равенката (6) (Cover & Hart, 1967).

$$d_{x,y} = |x_1 - x_2| + |y_1 - y_2| \quad (4.2.5)$$

$$d_{x,y} = \sqrt[p]{(x_1 - x_2)^p + (y_1 - y_2)^p} \quad (4.2.6)$$

Кај регресиските проблеми, конечната предвидена вредност се добива како просек од вредностите на најблиските соседи (Hastie, Tibshirani, & Friedman, 2009). На пример, доколку за одредена состојба на акумулација се бара предвидување на произведена мокност, алгоритмот ги пронаоѓа најсличните историски состојби според параметри како водостој, доток, зафатена количина вода и други влезни величини, а потоа ја пресметува просечната вредност на излезот.



Слика 10. Шематски приказ на класификациски проблем решен со алгоритмот KNN.

Најважен параметар кај KNN е бројот K, односно бројот на соседи кои ќе се земат предвид. Ако K е премногу мало (на пример K=1), моделот станува чувствителен на шум и случајни варијации во податоците. Ако K е премногу големо, моделот станува прегруб и може да изгуби способност за детекција на локални трендови. Поради тоа, изборот на оптимална вредност на K најчесто се врши експериментално преку вкрстена валидација.

Една од најголемите предности на KNN алгоритмот е неговата едноставност и фактот што не бара сложен процес на обучување. Дополнително, алгоритмот добро функционира кај нелинеарни зависимости, бидејќи одлуката ја базира на локални сличности наместо на глобална математичка функција. Ова го прави погоден за инженерски проблеми кај кои зависностите меѓу параметрите не се секогаш линеарни и лесно дефинирани (Han, Kamber, & Pei, 2012).

Од друга страна, KNN има и одредени недостатоци. Алгоритмот може да биде побавен кај големи множества на податоци, бидејќи за секое ново предвидување мора да се пресмета растојание до сите обучувачки примероци. Исто така, тој е чувствителен на размерот на влезните податоци, па затоа е неопходна претходна стандардизација или нормализација на променливите (Han, Kamber, & Pei, 2012).

4.2.2.6. Поткрени дрва (Boosted Trees)

Алгоритмот „Поткрени дрва“ (анг. „Boosted Trees“) или накратко БТ претставува современ ансамбл метод од областа на машинското учење, кој се базира на комбинирање на повеќе поединечни модели на регресиони дрва со цел добивање на поточен и постабилен конечен модел (Hastie, Tibshirani, & Friedman, 2009). Основната идеја на поткрениот пристап е секој следен модел да ги коригира грешките направени од претходните модели, при што постепено се подобрува способноста за предвидување (Shao & Deng, 2018). Во регресиски задачи, како што е случајот во ова истражување, БТ се користи за предвидување на континуирани вредности како што се водостој, проток, произведена енергија или други параметри поврзани со управување на водостопански системи.

За разлика од класичните единечни DT модели, кај кои едно дрво ја носи целата одлука, кај

ансамбл моделот BT се создава серија од многу дрва со помала сложеност. Првиот модел се обучува врз основните податоци, а секое следно дрво се фокусира на примероците кај кои претходниот модел направил најголема грешка. На тој начин, секој нов модел претставува корекција на претходниот, а конечната прогноза се добива како збир или тежински збир од сите поединечни модели.

Најчесто користен пристап е т.н. постепено поткревање (анг. „Gradient Boosting“), каде процесот на обучување се заснова на минимизација на функција на грешка преку градиентен пристап (Chen & Guestrin, 2016). Во секоја итерација се додава ново дрво кое го моделира преостанатиот резидуал (грешка) од претходниот чекор. Процесот на растење со учење на моделот може да се претстави со следните равенки (Wang, Huang, Wang, & Li, 2020):

$$\begin{aligned} y_0 &= 0 \\ y_1 &= v \cdot f_1(x; \theta_1) = y_0 + v \cdot f_1(x; \theta_1) \\ y_2 &= v \cdot \sum_{i=1}^2 f_i(x; \theta_i) = y_1 + v \cdot f_2(x; \theta_2) \\ &\dots \\ y_T &= v \cdot \sum_{i=1}^T f_i(x; \theta_i) = y_{(T-1)} + v \cdot f_T(x; \theta_T) \end{aligned} \quad (4.2.7)$$

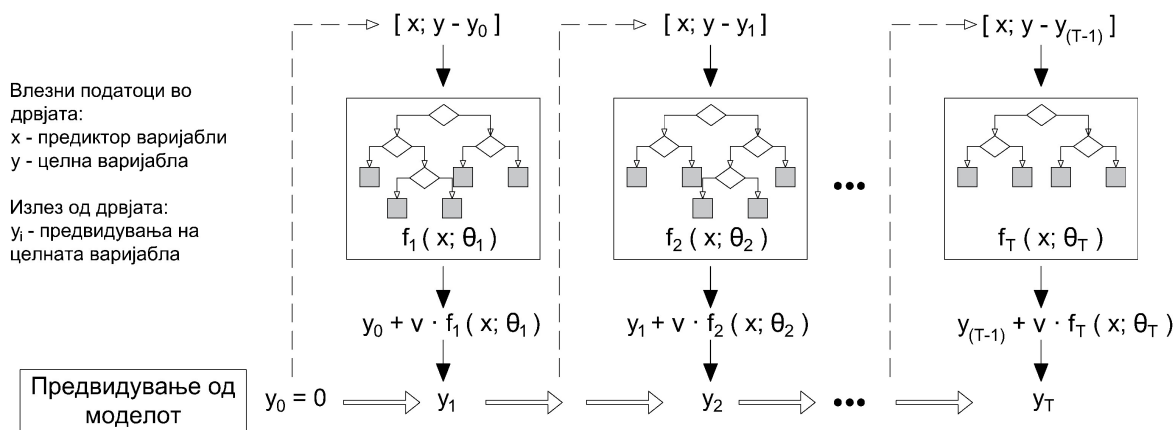
, каде „T“ е бројот на одлучувачки поткренати дрва, θ_i – структура на i-тото поткренато дрво со вклучен број на јазли и лисја на дрвото, v – градиент на учење кој се движи помеѓу $0 < v < 1$ кој влијае на придонесот на поединечните поткренати дрва во процесот на целосниот модел, y_T – предвидување на целната варијабла со користење на претходните i-ти поткренати дрвја, и f_i – излез од i-тото поткренато дрво без имплементиран градиент на учење кој ги користи предиктор варијаблите x за да ги апроксимира резидуалите $y - y_{i-1}$ од дрвната структура θ_i . Шематски приказ на пресметката на целната варијабла од моделот е прикажан на Слика 11.

Една од најголемите предности на BT алгоритмот е високата точност при моделирање на сложени и нелинеарни зависимости. Бидејќи секое ново дрво ги поправа слабостите на претходните, моделот има способност да открива скриени релации во податоците и да дава многу прецизни предвидувања. Дополнително, алгоритмот добро работи и кога постојат различни типови на влезни променливи, без потреба од строги претпоставки за нивната распределба.

Сепак, алгоритмот има и одредени ограничувања. Доколку бројот на дрва е преголем или параметрите не се правилно избрани, може да дојде до претерување во предвидувањата, односно моделот премногу добро да ги научи обучувачките податоци, а послабо да генерализира на нови случаи. Поради тоа, важни параметри се бројот на дрва ($n_estimators$), длабочината на дрвата (max_depth) и стапката на учење ($learning_rate$), кои треба внимателно да се оптимизираат.

Во рамките на ова истражување, алгоритмот „Поткренати дрва“ покажа едни од најдобрите резултати според повеќе статистички индикатори како што се MSE, RMSE, MAE и R^2 , што укажува на висока точност и стабилност. Ова потврдува дека ансамбл методите базирани на алгоритам а поткревање се исклучително погодни за предвидување и оптимизација кај сложени водостопански системи.

Поради својата способност за моделирање на нелинеарни процеси, висока точност и добра робустност, BT претставува еден од најмоќните алгоритми во современото машинско учење и има широка примена во инженерството, енергетиката и управувањето со природни ресурси.



Слика 11. Шематски приказ на пресметката на целната варијабла кај БТ моделите.

4.2.2.7. Генерализирана регресиона невронска мрежа (General Regression Neural Network)

Алгоритмот „Генерализирана регресиона невронска мрежа“ (анг. „General Regression Neural Network“) или накратко GRNN претставува специјализиран тип на вештачка невронска мрежа, развиен за решавање на регресиски проблеми, односно за предвидување на континуирани нумерички вредности. Овој модел е предложен од Donald F. Specht во 1991 година и претставува варијација на архитектурата на Радијалната Базна Невронска Мрежа (анг. „Radial Basis Neural Network“) (Specht, 1991). GRNN е особено погоден за инженерски задачи кај кои постојат сложени и нелинеарни врски помеѓу влезните и излезните параметри (Izonin, K. Kazantzi, Tkachenko, & Mitoulis, 2025).

За разлика од класичните вештачки невронски мрежи, кај кои е потребен итеративен процес на обучување преку алгоритам со повратна пропација (анг. „backpropagation“), GRNN има многу побрз процес на учење. Тој директно ги користи постоечките обучувачки податоци за формирање на моделот и не бара долготрајна оптимизација на тежинските коефициенти (Haykin, 1994). Ова го прави исклучително ефикасен кај проблеми каде е потребна брза имплементација и стабилни резултати.

Архитектурата на GRNN обично се состои од четири слоја:

- 1) Слој со влезни параметри (анг. „Input Layer“) – ги презема влезните параметри;
- 2) Скриен слој (анг. „Pattern Layer“) – секој обучувачки примерок претставува еден неврон;
- 3) Собирен слој (анг. „Summation Layer“) – ги собира пондерираните резултати;
- 4) Слој со излезни резултати (анг. „Output Layer“) – ја дава конечната предвидена вредност.

Основниот принцип на работа се базира на проценка на густина на веројатност и користење на Гаус кернел функција (анг. „Gaussian kernel function“) за пресметка на тежините на секој примерок. Шематски приказ за процесот на предвидување на целна варијабла со модел базиран на GRNN е прикажан на Слика 12. За секој нов влезен примерок, алгоритмот ја пресметува сличноста со сите обучувачки примероци преку пресметка на Евклидово растојание. Примероците кои се поблиски добиваат поголема тежина, а подалечните помала. Евклидовото растојание се пресметува како:

$$E_{k,i} = \sqrt{\sum_{j=1}^n (x_{k,j} - x_{i,j})^2} \quad (4.2.8)$$

Гаусовата кернел функција за пресметка на тежината на i-тиот примерок се пресметува како:

$$G_{k,i} = \exp\left(-\frac{(E_{k,i})^2}{\sigma^2}\right) \quad (4.2.9)$$

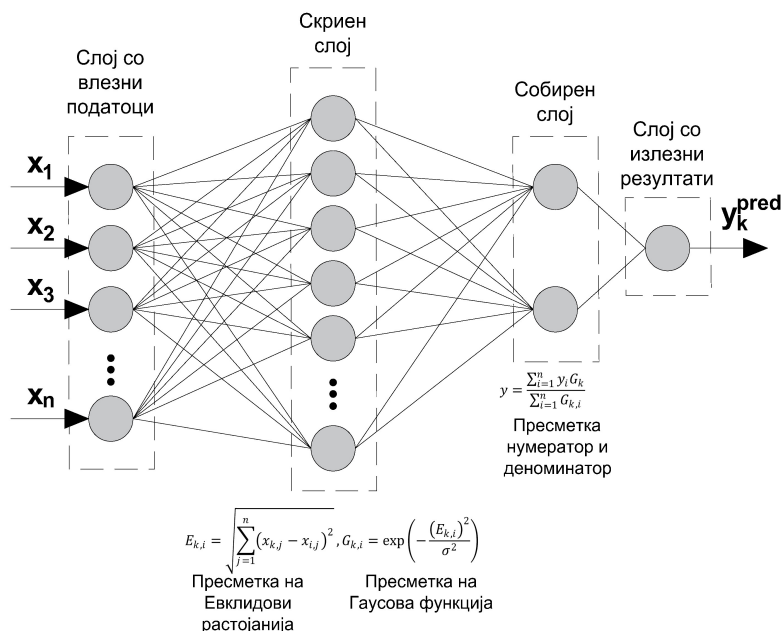
каде $G_{k,i}$ претставува тежината на i-тиот примерок, $E_{k,i}$ претставува растојанието до новиот примерок, а σ е параметар за ширината на Гаусовата кернел функција познат како проширувачки

или мазен параметар (анг. „spread or smoothing parameter“).

Конечната предвидена вредност се добива како пондериран просек од сите излезни вредности:

$$y = \frac{\sum_{i=1}^n y_i G_{k,i}}{\sum_{i=1}^n G_{k,i}} \quad (4.2.10)$$

Ова значи дека GRNN ги користи сите достапни историски податоци, но најголемо влијание имаат примероците кои се најслични на новата состојба.



Слика 12. Шематски приказ на четири слоеви кај GRNN моделот.

Една од најголемите предности на GRNN е способноста за моделирање на сложени нелинеарни зависимости без потреба од голем број параметри за подесување. Алгоритмот е робустен, стабилен и често покажува многу добри резултати кај помали и средни множества на податоци (Izonin, K. Kazantzi, Tkachenko, & Mitoulis, 2025). Дополнително, бидејќи нема класично долго обучување, времето за имплементација е релативно кратко.

Сепак, GRNN има и одредени недостатоци. Бидејќи секој обучувачки примерок се чува во мрежата, мемориските барања можат да бидат поголеми кај многу големи бази на податоци. Исто така, изборот на параметарот σ е критичен – премала вредност може да доведе до преосетливост и претерување во предвидувањата, а преголема – до прегрубо измазнување на резултатите.

4.2.2.8. Гаусов процесен регресор (Gaussian Process Regressor)

Алгоритмот „Гаусов процес регресор“ (анг. „Gaussian Process Regressor“) или накратко GPR, претставува напреден непараметарски метод од областа на машинското учење кој се користи за решавање на регресиски проблеми, односно за предвидување на континуирани нумерички вредности. За разлика од класичните регресиски модели кои претпоставуваат однапред дефинирана функционална форма, GPR користи веројатносен пристап, при што се моделира распределбата на можните функции кои можат да ги опишат податоците. Ова му овозможува на алгоритмот не само да даде предвидена вредност, туку и да ја процени несигурноста на предвидувањето, што претставува значајна предност во примена на овој модел во решавање на инженерски комплексни проблеми.

GPR е математички дефиниран како множество на случајни променливи кај кои секој конечен подмножество има заедничка Гаусова распределба. Во практична смисла, тоа значи дека секоја точка во просторот на податоци е поврзана со останатите точки преку функција на коваријанса, позната како кернел функција (анг. „kernel function“). Оваа функција ја дефинира сличноста помеѓу различни влезни примероци и претставува основа на моделот.

Формално, GPR може да се запише како:

$$f(x) \sim GP(m(x), k(x, x')) \quad (4.2.11)$$

, каде што $m(x)$ е средната функција, а $k(x, x')$ е коваријансната функција или кернел (анг. „kernel“) која ја опишува зависноста меѓу точките x и x' .

Најчесто користена коваријансна функција кај GPR е радијалната базна функција (анг. „Radial Basis Function“) или накратко RBF, познат и како Гаусов кернел (анг. „Gaussian kernel“):

$$k(x, x') = \sigma_f^2 \cdot \exp\left(-\frac{\|x - x'\|^2}{2 \cdot l^2}\right) \quad (4.2.12)$$

, каде што l е параметар кој ја контролира мазноста на функцијата, а σ_f^2 ја определува варијансата. Овој кернел овозможува моделирање на мазни и континуирани нелинеарни врски помеѓу предиктор и целните податоци.

Една од најголемите предности на GPR е тоа што покрај предвидената вредност, моделот дава и стандардна девијација или доверлив интервал за секое предвидување. Ова е особено важно кај хидротехнички и водостопански системи, каде несигурноста на прогнозите може да има директно влијание врз донесување на оперативни одлуки, управување со акумулации и безбедност на објекти.

Дополнително, GPR е многу ефикасен кај мали и средни множества на податоци, каде често постигнува многу висока точност. Бидејќи алгоритмот е флексибилен и нелинеарен, тој успешно моделира сложени физички процеси кај кои класичните линеарни методи не се доволни.

Сепак, GPR има и одредени ограничувања. Главен недостаток е високата пресметковна сложеност, бидејќи при обучување е потребна инверзија на коваријансна матрица, што го прави алгоритмот помалку погоден за многу големи множества на податоци. Исто така, изборот на соодветен кернел и неговите параметри има значително влијание врз резултатите и бара внимателна калибрација.

Во рамките на ова истражување, GPR беше применет за моделирање на параметри поврзани со управување на водостопански системи и однесување на брани. Иако алгоритмот теоретски има голем потенцијал поради можноста за проценка на несигурност, добиените резултати зависат од структурата на податоците, изборот на кернел функцијата и бројот на достапни примероци.

4.2.2.9. XGBoost алгоритам за градиентно поткревање (XGBoost)

Алгоритмот „XGBoost“ (анг. „Extreme Gradient Boosting“) претставува еден од најмоќните и најшироко применувани алгоритми од областа на машинското учење, особено кај задачи поврзани со регресија, класификација и рангирање (Chen & Guestrin, 2016). Тој претставува унапредена и оптимизирана верзија на класичниот „Gradient Boosting“ метод, развиена со цел да обезбеди повисока точност, побрзо обучување и подобра способност за генерализација. Поради овие карактеристики, XGBoost често постигнува врвни резултати на научни натпревари при решавање на комплексни инженерски проблеми.

Основниот принцип на работа на XGBoost се базира на ансамбл пристап, при што повеќе поединечни модели со одлучувачки дрва се комбинираат во еден силен предиктивен модел. Наместо сите дрва да се обучуваат независно, тие се додаваат секвенцијално, така што секое ново дрво ги коригира грешките направени од претходните дрва. На овој начин моделот постепено ја подобрува точноста.

Во регресиски проблеми, конечната предвидена вредност се добива како збир од придонесите на сите дрва:

$$y_i = \sum_{j=1}^n f_j(x_i) \quad (4.2.13)$$

, каде што f_j е поединечно одлучувачко дрво, а n е вкупниот број дрва во ансамблот.

Главната разлика помеѓу класичниот Gradient Boosting и XGBoost е тоа што кај XGBoost

процесот на оптимизација е понапреден и вклучува регуларизација, со што се намалува ризикот од претерување во предвидувањата. Целната функција може да се претстави како:

$$F = \sum l(y, y_i) + \sum \theta(f_j) \quad (4.2.14)$$

, каде првиот дел l ја претставува грешката на моделот, а вториот дел θ претставува казна за премногу сложени дрва. Со ова се добива подобар баланс помеѓу точност и стабилност.

Една од најголемите предности на XGBoost е неговата способност за работа со нелинеарни и комплексни зависности, што го прави исклучително погоден за инженерски проблеми. Алгоритмот автоматски може да открива интеракции меѓу променливите и да моделира процеси кои не можат лесно да се опишат со класични математички равенки.

Дополнително, XGBoost има бројни практични предности:

- 1) многу брзо обучување благодарение на паралелна обработка,
- 2) можност за пресметка во услови на вредности кои недостасуваат во множеството,
- 3) висока точност кај структурирани табеларни податоци,
- 4) вграден механизам за контрола на претерување во предвидувањата;
- 5) можност за добивање на податоци за важност на параметрите (анг. „feature importance“).

Во рамките на ова истражување, XGBoost беше применет за моделирање и предвидување на параметри поврзани со управување на водостопански системи, како што се водостoj, доток, испусти и произведена енергија. Резултатите покажаа висока точност и стабилност, што потврдува дека алгоритмот успешно ги препознава сложените релации во податоците.

Сепак, XGBoost има и одредени недостатоци. Поради големиот број параметри, како што се `n_estimators`, `max_depth`, `learning_rate`, `subsample` и други, потребна е внимателна калибрација за постигнување оптимални резултати. Доколку параметрите не се соодветно избрани, може да дојде до претерување во предвидувањата или непотребно долго време на обучување.

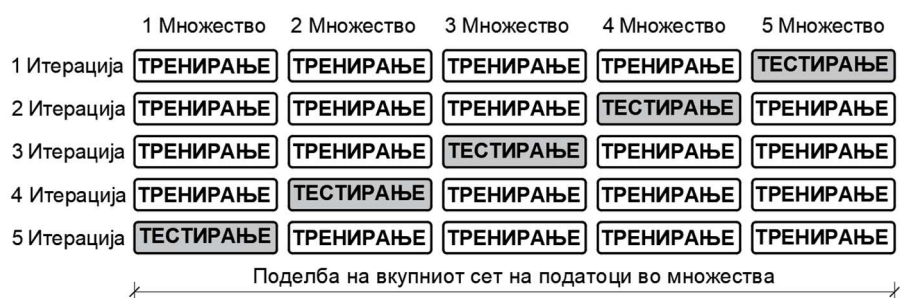
Поради својата висока ефикасност, робустност и способност за моделирање на сложени нелинеарни процеси, XGBoost денес претставува еден од најдобрите алгоритми во машинското учење и има широка примена во финансии, медицина, енергетика, хидротехника и управување со природни ресурси (Chen & Guestrin, 2016). Во инженерски апликации, особено кај анализи со повеќе влезни параметри и комплексна динамика, XGBoost претставува исклучително силна алатка за предвидување и оптимизација.

4.2.3. Прелиминарна валидација на моделите

Различни методи може да се применат за валидација на моделите врз основа на квалитетот на базите на влезни податоци. Често користен е методот на поделба на сетот на дел за тренирање и дел за тестирање (анг. „train and test split“). Со овој метод, делот од сетот за тренирање вообичаено е од 70 – 75%, додека делот од сетот за тестирање е вообичаено 25 – 30%. За да се добијат значајни резултати од овој пристап кон валидација на квалитетот на податоците, потребно е сетот да е добро „измешан“, да нема подредени вредности на таргет варијаблата и да нема очигледна зависност помеѓу таргет предиктор варијаблите. Доколку сетот е подреден, може да се очекува од идниот модел да прави претерани предвидувања (анг. „overestimation“) или да ги потценува предвидувањата (анг. „underestimation“) споредено со реални очекувани вредности на таргет варијаблата. Доброто „мешање“ на базата на податоци многу често изостанува, особено кога станува збор за сетови со помал број на податоци (Paul Muller & Massaron, 2021). Очигледно, доколку станува збор за сет на податоци со мал број на вредности, тренираниот модел понатаму има големи шанси да стане пристрасен кон претерување или потценување на вредностите на таргет варијаблата.

За ова да се надмине, се применува k – кратна вкрстена валидација (анг. „ k – fold cross validation“).

K – кратната вкрстена валидација е математичка функција која го дели сетот на податоци на множества со еднаков број на случајно избрани податоци од вкупниот сет. Потоа, секое множество се користи како сет за тестирање, додека останатите множества се користат како сетови за тренирање. Секоја итерација користи различни множества за тестирање, со што математичката функција пресметува проценка на грешката што се јавува во процесот на тренирање и тестирање. Процесот започнува со тренирање на k-1 множества, и тестирање на единственото останато множество. Понатаму, во следната итерација, друго множество се одбира да биде тест множество, а преостанатите k-1 се користат за тренирање. Овој процес се одвива сè додека сите множества не се искористат по еднаш како тест множества. Често применети се 5 – кратната вкрстена валидација каде вкупниот сет се дели на 5 множества, и 10 – кратната вкрстена валидација каде вкупниот сет се дели на 10 множества (J. Martyr & R. Rogers, 2021). Шематски приказ на процесот на валидација на вкупниот сет податоци со 5 – кратна вкрстена валидација е прикажан на Слика 13.



Слика 13. Графичка презентација на процесот на 5 - кратна вкрстена валидација (J. Martyr & R. Rogers, 2021).

K – вкрстената валидација не дава трениран модел. Таа тренира одреден број на множества од вкупниот сет, потоа тренираниот модел го тестира на изоставеното множество податоци, а на крај како резултат дава метрички показатели по кои се оценува квалитетот на податоците и креираните модели. Метрички показатели кои се користат во овој процес, најчесто се: (1) Коефициентот на детерминација, означен како R^2 , (2) Средната апсолутна грешка или накратко MAE (анг. „Mean Absolute Error - MAE“), (3) Средна квадратна грешка или накратко MSE (анг. „Mean Square Error - MSE“) и (4) Корен од средната квадратна грешка или накратко RMSE (анг. „Root of Mean Square Error“ – RMSE). Вообичаено, при оценка на метричките показатели, се гледа просекот од сите итерации во процесот на K – вкрстена валидација. Како што следи, даден е кус опис на секој метрички показател, неговата математичка интерпретација и значењето на добиените вредности.

4.2.3.1. Коефициент на детерминација, R^2

Коефициентот на детерминација, означен како R^2 , претставува мерка за тоа колкав дел од варијансата на зависната променлива е објаснета со регресискиот модел. Неговата вредност се движи во интервалот од 0 до 1, при што повисоки вредности укажуваат на подобра усогласеност помеѓу моделот и набљудуваните податоци. Вредност на R^2 блиска до 1 означува дека моделот успешно ја опишува зависноста помеѓу влезните и излезните променливи, додека пониски вредности укажуваат на ограничена објаснителна моќ на моделот. Математичката равенка со која се пресметува R^2 е прикажана во равенка 4.2.15. Во равенката, ' y_i ' ја претставува реалната вредност на таргет варијаблата, ' y_i' ' претставува предвидувањето за таргет варијаблата, ' y_{aver} ' е средна вредност од реалните вредности на таргет варијаблата, и ' n ' е бројот на вредности на таргет варијаблата.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i')^2}{\sum_{i=1}^n (y_i - y_{aver})^2} \quad (4.2.15)$$

4.2.3.2. Средна апсолутна грешка, MAE

Средната апсолутна грешка (MAE) ја претставува просечната вредност на апсолутните разлики помеѓу предвидените и реално набљудуваните вредности. Оваа метрика обезбедува интуитивна интерпретација на грешката, бидејќи е изразена во истите единици како и зависната променлива. Помали вредности на MAE укажуваат на поголема точност на моделот и помало просечно

отстапување на предвидувањата од реалните податоци. Математичката равенка со која се пресметува MAE е прикажана во равенка 4.2.16. Во равенката, 'y_i' ја претставува реалната вредност на таргет варијаблата, 'y_i' претставува предвидувањето за таргет варијаблата, и 'n' е бројот на вредности на таргет варијаблата.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (4.2.16)$$

4.2.3.3. Средна квадратна грешка, MSE

Средната квадратна грешка (MSE) претставува просек на квадратите на разликите помеѓу предвидените и набљудуваните вредности. Со квадрирање на грешките, оваа метрика им дава поголема тежина на поголемите отстапувања, што ја прави особено чувствителна на појава на екстремни грешки. Поради изразување во квадратни единици на мерената величина, интерпретацијата на MSE е помалку интуитивна, но е многу корисна за споредба и оптимизација на модели. Математичката равенка со која се пресметува MSE е прикажана во равенка 4.2.17. Во равенката, 'y_i' ја претставува реалната вредност на таргет варијаблата, 'y_i' претставува предвидувањето за таргет варијаблата, и 'n' е бројот на вредности на таргет варијаблата.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (4.2.17)$$

4.2.3.4. Корен на средната квадратна грешка, RMSE

Коренот од средната квадратна грешка (RMSE) се добива како квадратен корен од MSE и претставува мерка за просечното отстапување на предвидените вредности од реално набљудуваните. Како и MAE, RMSE е изразена во истите единици како и зависната променлива, но за разлика од неа, е почувствителна на поголеми грешки поради квадратниот третман во основата на пресметката. Помали вредности на RMSE укажуваат на подобра точност и стабилност на моделот. Математичката равенка со која се пресметува RMSE е прикажана во равенка 4. Во равенката, 'y_i' ја претставува реалната вредност на таргет варијаблата, 'y_i' претставува предвидувањето за таргет варијаблата, и 'n' е бројот на вредности на таргет варијаблата.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (4.2.18)$$

4.2.4. Избор на алгоритми за обучување на моделите

Алгоритмите за тренирање на моделите се дефинирани во самиот код во Python, и претставуваат дел од мноштвото моделирани алгоритми во самиот код. Моделирани се следните алгоритми: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Одбраните алгоритми се проценети како алгоритми чии модели даваат висока точност во фазата на валидација на податоците, уште пред тренирањето на моделите. Точноста е проценета преку согледување на вредностите од извршената петкратна валидација за секој алгоритам во фазата на прелиминарна валидација на секој модел со сетот на влезни податоци.

4.2.5. Предвидување на целната варијабла

Предвидувањето на целната варијабла се извршува по обучување на деветте модели со машинско учење кои се тренирани во претходниот чекор. Предвидувањето е извршува со повикување на надворешна .xls датотека која содржи податоци за предиктор варијаблите, а целната варијабла е исклучена т.е. непозната во предвидувањето. Како резултат на процесот на предвидување се добива податочен сет со предвидени вредности на целната варијабла, изнесени надвор од DamResML кодот во .xls датотека.

4.3. Хаос оптимизација

По завршување на процесот на предвидување, во дел од моделите применета е хаос оптимизација (анг. „Chaos Game Optimization“) или накратко CGO за оптимизирање на предиктор варијаблите кон минимизирање или максимизирање на целната варијабла. Хаос оптимизација се применува во моделите каде има повеќе од две предиктор варијабли на целната варијабла.

Хаос оптимизацијата претставува стохастичка оптимизациска метода која се базира на принципите на теоријата на хаос и нелинеарната динамика. За разлика од класичните оптимизациски алгоритми кои користат случајни или детерминистички пристапи за пребарување на дозволени решенија (Djordjevic, 1990), хаос оптимизацијата користи хаотични секвенци за да обезбеди подобро покривање на просторот на можни решенија и да избегне заглавување во локални оптимуми.

Основата на овој метод лежи во користењето на хаотични мапи, како што е логистичката мапа, која генерира детерминистички, но непредвидливи и чувствителни на почетни услови секвенци. Овие карактеристики овозможуваат хаотичните секвенци да се однесуваат слично на случајни броеви, но со подобри својства во однос на рамномерната распределба и покриеноста на пребарувачкиот простор.

Една од најчесто користените хаотични функции е логистичката мапа, дефинирана како:

$$x_{k+1} = r \cdot x_k \cdot (1 - x_k) \quad (4.3.1)$$

,каде параметарот r обично се зема во интервалот (3.57, 4), со што се обезбедува хаотично однесување на системот. Генерираната секвенца x_k се користи за создавање кандидат-решенија во процесот на оптимизација.

Процесот на хаос оптимизација обично се состои од неколку чекори. Прво, се дефинира почетна вредност и се генерира хаотична секвенца. Потоа, оваа секвенца се мапира во просторот на можни решенија (на пример, во рамките на дозволените граници на предиктор варијаблите). За секое генерирано решение се пресметува вредноста на целната функција. Алгоритмот итеративно генерира нови решенија и го задржува најдоброто пронајдено решение според дефинираниот критериум (минимизација или максимизација).

Една од клучните предности на хаос оптимизацијата е нејзината способност за глобално пребарување. Поради хаотичната природа на генерираните секвенци, алгоритмот има поголема веројатност да ги истражи различните региони од просторот на решенија, за разлика од класичните методи кои можат да се заглават во локални оптимуми. Ова е особено важно кај сложени инженерски проблеми со повеќе локални екстреми.

Дополнително, хаос оптимизацијата е релативно едноставна за имплементација и не бара сложени математички пресметки како градиенти или хесијани, што ја прави погодна за нелинеарни и недиференцијабилни функции. Ова е значајна предност во практични апликации каде аналитичката форма на функцијата не е позната или е тешко да се изведе.

Сепак, методот има и одредени ограничувања. Еден од главните предизвици е изборот на соодветна хаотична мапа и параметри, бидејќи различни мапи може да резултираат со различна ефикасност на пребарувањето. Исто така, како и кај другите стохастички методи, не постои гаранција дека ќе се пронајде глобалниот оптимум, туку само приближно решение.

Во современите истражувања, хаос оптимизацијата често се комбинира со други оптимизациски техники, како што се генетски алгоритми, Particle Swarm Optimization (PSO) или градиентни методи, со цел да се подобри конвергенцијата и стабилноста на решението. Овие хибридни пристапи се покажуваат како особено ефективни во сложени инженерски системи, каде постојат повеќе ограничувања и нелинеарни зависности помеѓу варијаблите.

Ќелијата со методологијата на Хаос оптимизацијата е последна во DamResML кодот. Процесот на оптимизација во самата ќелија, чекор по чекор, е појаснет како што следи.

4.3.1. Дефинирање на влезни податоци и модели

Во почетната фаза се дефинираат влезните променливи т.е. предиктор варијаблите (анг. „feature variables“), како што се кота на водно огледало (Zwl), волумен на акумулација, дотек во акумулација (Q_inflow) и други – во согласност со студијата на случај која е предмет на разгледување. Дополнително, се одбира еден или повеќе претходно обучени модели (на пр. Random Forest, XGBoost, Gradient Boosting), кои служат како сурогат модели на реалниот систем. Резултатите од предвидувањето со овие модели се користат како влезен податочен сет во Хаос оптимизацијата.

4.3.2. Дефинирање на граници и ограничувања

Границите на пребарување (анг. „search space“) се определуваат врз основа на историските податоци од податочниот сет генериран во процесот на предвидување на целната варијабла. За да се исклучат екстремните вредности од податочниот сет, во моделот се користат 1% и 99% квантил од податоците (се исклучуваат податоците со вредност под 1% од податочниот сет и над 99% од податочниот сет). Дополнително, во моделот е оставен простор да се внесат и инженерски ограничувања, со што се обезбедува реална оптимизација во физички оправдан опсег.

4.3.3. Точност на решенијата

За да се спречат нереални или физички невозможни решенија, се воведува казнена функција (анг. „penalty function“) базирана на т.н. растојание од историските податоци. Секоја нова генерирана точка се нормализира и се пресметува нејзиното растојание до најблиската реална точка од базата на податоци. Доколку ова растојание ја надмине дефинираната граница, се додава казна во функцијата на цел, со што таквите решенија се елиминираат од оптималниот избор.

4.3.4. Генерирање на нови решенија (CGO механизам)

Алгоритмот CGO започнува со иницијална популација од случајно генерирани решенија во рамките на дефинираните граници. Во секоја итерација, секоја единка од популацијата се ажурира преку комбинација од три компоненти:

- глобално најдобро решение (gbest),
- средна вредност на популацијата,
- случајно избрана друга точка од популацијата.

Овие компоненти се комбинираат со тежински фактори добиени од хаотична логистичка мапа, која обезбедува нелинеарна и стохастичка динамика на пребарување. Овој пристап овозможува подобро истражување на просторот на решенија и избегнување на локални оптимуми.

Дополнително, се воведува мала случајна варијација (jitter) со цел да се одржи диверзитетот на популацијата.

4.3.5. Целна функција

Целната функција (анг. „Objective function“) има задача да ја максимизира предвидената вредност на целната променлива. Бидејќи алгоритмот минимизира, се користи негативна вредност од предвидувањето. Во функцијата се вклучени:

- предвидената вредност од моделот,
- казна за нереалистични решенија,
- казна за излезни вредности надвор од дозволения опсег,
- (опционално) казна за несогласување меѓу повеќе модели (стандардна девијација).
- Евалуација и селекција на решенија

Во текот на итерациите, се евалуираат илјадници потенцијални решенија. Се задржуваат сите испитани точки, по што се врши:

- филтрирање на физички прифатливи решенија,
- отстранување на дупликати,

- сортирање според највисока предвидена вредност и најмало растојание до реални податоци.

Од сите кандидати се издвојуваат најдобрите 20 сценарија (Тор 20), кои претставуваат оптимални работни состојби на системот.

4.3.6. Визуелизација и интерпретација

Конечните резултати се прикажуваат табеларно и графички (3D дијаграм), при што се споредуваат со историските податоци. Овие сценарија не се директно земени од податоците, туку се резултат на оптимизациски процес кој комбинира машинско учење и стохастичка оптимизација.

“Prediction is very difficult, especially if it’s about the future.”
Niels Bohr

5. ТЕСТИРАЊЕ И ВАЛИДАЦИЈА НА DamResML

За тестирање и валидација на изработениот код DamResML, во рамки на докторската дисертација изработено е паралелно предвидување со модели од машинско учење изработени со DamResML кодот кој е дел од дисертацијата, и комерцијално достапен софтвер за предвидување на база на модели со машинско учење – „NeuralTools“.

5.1. „NeuralTools“

„NeuralTools“ претставува комерцијален софтверски додаток (анг. „add-in“) за Microsoft Excel, развиен од компанијата „Palisade“, денес дел од компанијата „Lumivero“. Софтверот е наменет за анализа на податоци, препознавање на шеми и предвидување на вредности со примена на вештачки невронски мрежи (анг. „Artificial Neural Networks“ – ANN). „NeuralTools“ овозможува директна интеграција на алгоритмите за машинско учење во Excel околина, без потреба од познавање на програмски јазици, што го прави погоден за инженерски и деловни анализи каде податоците веќе постојат во табеларна форма.

Основниот принцип на работа на „NeuralTools“ се базира на процес на обучување на вештачка невронска мрежа преку постоечки податоци. Во Excel табелата најпрво се дефинираат влезните (предиктор) варијабли и излезната (целна) варијабла. Потоа, софтверот автоматски врши процес на обучување на невронската мрежа со користење на историските податоци. За време на процесот на обучување, невронската мрежа постепено ги прилагодува тежинските коефициенти помеѓу невроните со цел да се минимизира грешката помеѓу предвидените и реалните вредности (Palisade Corporation, 2013).

„NeuralTools“ користи повеќеслојни вештачки невронски мрежи (анг. „Multilayer Perceptron“ – MLP), кои се состојат од:

- Влезен слој (input layer),
- Еден или повеќе скриени слоеви (hidden layers),
- Излезен слој (output layer).

Во влезниот слој се внесуваат вредностите на предиктор варијаблите, како на пример ниво на вода во акумулација, дотек, волумен во акумулација или температура. Потоа информацијата се пренесува низ скриените слоеви, каде секој неврон врши математичка трансформација преку

активациски функции и тежински коефициенти. На крај, во излезниот слој се добива предвидената вредност на целната варијабла.

Процесот на обучување во „NeuralTools“ најчесто се базира на алгоритам со повратна пропација (анг. „backpropagation“), каде грешката помеѓу реалните и предвидените вредности се враќа наназад низ мрежата со цел корекција на тежините. На тој начин мрежата постепено „учи“ да ја опишува зависноста помеѓу влезните и излезните параметри.

Една од предностите на „NeuralTools“ е што софтверот автоматски:

- врши поделба на податоците на тренинг, тест и validation сетови,
- ја избира архитектурата на мрежата,
- ги прилагодува параметрите на обучување,
- врши оптимизација на моделот,
- пресметува статистички показатели за успешноста на моделот.

По завршување на процесот на обучување, „NeuralTools“ генерира повеќе резултати и статистички анализи, како:

- графици на реални и предвидени вредности,
- резидуали (остатоци),
- важност на влезните варијабли,
- корелации,
- табели со метрички показатели.

Вообичаено, софтверот ги прикажува:

- MSE (Mean Squared Error),
- RMSE (Root Mean Squared Error),
- MAE (Mean Absolute Error),
- Процентуални грешки,
- корелационски коефициенти помеѓу реални и предвидени вредности.

Во споредба со DamResML моделите развиени во ова истражување, „NeuralTools“ претставува затворен систем (анг. „black-box“) бидејќи корисникот има ограничена контрола врз деталната архитектура на моделот и математичките чекори во процесот на обучување. Од друга страна, кодот DamResML во Пајтон овозможува целосна контрола врз:

- изборот на алгоритам,
- архитектурата на мрежата,
- хиперпараметрите,
- начинот на валидација,
- preprocessing методите,
- статистичките анализи,
- визуелизацијата на резултатите.

Поради тоа, споредбата помеѓу резултатите добиени со DamResML и „NeuralTools“ претставува корисен пристап за верификација на добиените модели. Доколку резултатите од двата софтверски пристапи покажуваат слични трендови и слични метрички показатели, тоа претставува дополнителна потврда за стабилноста и доверливоста на развиените модели со машинско учење.

5.1.1. Досегашна примена на „NeuralTools“ во објавени истражувања

„NeuralTools“ претставува доста често користена алатка за предвидувања во голем број на објавени истражувања, во инженерски планирања, прогнозни анализи, производствени процеси и управување со ризици. Според креаторот на софтверот, повеќе од 150 000 корисници во светот го користат „NeuralTools“ во своето секојдневие, а објавени се над 100 реферати со резултати добиени токму со примена на оваа алатка. Како што следи даден е кус опис на неколку објавени истражувања од областа на инженерството, биологијата, градежништвото и хидротехниката каде истражувачите во својата работа го користеле токму „NeuralTools“ и експлицитно го навеле како

користена алатка.

- **Предвидување на цени на материјали во градежништвото** (Shiha, M. Dorra, & Nassar, 2020)

Во предметното истражување авторите применуваат модели со вештачки невронски мрежи за предвидување на цените на градежните материјали во Египет, со користење на макроекономски индикатори како влезни параметри. Во анализата се користени „NeuralTools“, Excel и Пајтон околина со цел развој и споредба на ANN модели за прогнозирање на цените на челик и цемент. „NeuralTools“ е применет за обучување на невронските мрежи, анализа на зависностите помеѓу економските параметри и генерирање на предвидувања за идните движења на цените. Авторите заклучуваат дека ANN моделите имаат висока способност за моделирање на сложени нелинеарни врски помеѓу економските параметри и цените на градежните материјали, а „NeuralTools“ се покажал како практична алатка за брзо развивање и тестирање на ANN модели.

- **Предвидување на параметри поврзани со транспорт на нанос** (Atieh, Mehlretter, Gharabaghi, & R., 2015)

Во наредното истражување се разгледува примена на вештачки невронски мрежи за предвидување на параметри поврзани со транспорт на нанос и седиментни криви во сливови без директни мерења. Целта на истражувањето е развој на ANN модели кои можат да ја опишат врската помеѓу хидролошките параметри и транспортот на нанос. „NeuralTools“ е применет за обучување на ANN модели и анализа на зависностите помеѓу хидролошките променливи. Истражувањето покажува дека невронските мрежи успешно ги моделираат сложените нелинеарни процеси во хидрологијата и овозможуваат проценка на параметри за сливови каде нема доволно мерени податоци.

- **Оптимизирање на производство со примена на ANN** (Casalino, Facchini, Mortello, & Mummolo, 2016)

Во ова истражување авторите применуваат модели со вештачки невронски мрежи за оптимизација на производствени процеси. Во трудот експлицитно е наведено дека ANN моделите се развиени со користење на „NeuralTools“ софтверскиот додаток за Microsoft Excel од компанијата Palisade. „NeuralTools“ е употребен за обучување на повеќеслојни невронски мрежи, анализа на влијанието на производствените параметри и оптимизација на производствениот процес. Истражувањето покажува дека ANN моделите успешно можат да ги моделираат сложените индустриски процеси и да обезбедат оптимални комбинации на параметри за подобрување на квалитетот и ефикасноста на производството.

- **Модел со невронски мрежи за термална деактивација на салмонела кај кокошки** (Oscar, 2017)

Во овој труд се применуваат вештачки невронски мрежи за моделирање на процесот на термичка деактивација на бактеријата салмонела. Авторите експлицитно наведуваат дека „NeuralTools“ е користен за развој на повеќеслојна невронска мрежа со напредување кон напред (анг. „feedforward“). Софтверот е применет за обучување на моделите, тестирање на точноста на предвидувањата и анализа на влијанието на температурните параметри врз процесот на деактивација. Резултатите покажуваат дека ANN моделите развиени со „NeuralTools“ имаат висока точност и можат успешно да се применат во моделирање на сложени биолошки и термички процеси.

- **Предвидувања на електропребенење** (анг. „electrospinning“) со примена на вештачки невронски мрежи (Brooks & Tucker, 2015)

Во ова истражување „NeuralTools“ е применет за обучување на ANN модели за предвидување на параметри поврзани со процесот на електропребенење при производство на материјали. Авторите наведуваат дека моделите биле обучувани со Palisade „NeuralTools“ додаток за Microsoft Excel. Софтверот бил употребен за анализа на зависностите помеѓу влезните производствени параметри и карактеристиките на добиените материјали. Истражувањето покажува дека невронските мрежи

успешно можат да предвидуваат сложени нелинеарни производствени процеси и да помогнат во оптимизација на технолошките параметри.

→ **Предвидување на цени на јагленот** (Matyjaszek, Riesgo Fernandez, Krzemien, Wodarski, & Fidalgo Valverde, 2019)

Во овој труд авторите вршат споредба помеѓу класични временски серии базирани на ARIMA модели и модели со вештачки невронски мрежи за прогнозирање на цените на коксен јаглен. Во истражувањето експлицитно е наведено дека „NeuralTools“ 7.5 бил користен за симулација и обучување на невронските мрежи. Софтверот е применет за анализа на временските серии, генерирање на предвидувања и споредба на точноста помеѓу ANN и ARIMA моделите. Резултатите покажуваат дека невронските мрежи имаат подобра способност за моделирање на нелинеарните промени во цените на јагленот во споредба со класичните статистички методи.

→ **Примена на вештачки невронски мрежи во управување со градежни проекти за проценка на ризик во градењето** (Mossalam & Arafa, 2018)

Во ова истражување „NeuralTools“ е применет во областа на градежното инженерство и управување со ризици во градежни проекти. Авторите експлицитно наведуваат дека Palisade Neural Network Add-on бил избран како алатка за развој на ANN моделите. „NeuralTools“ бил користен за анализа на факторите кои влијаат врз ризикот во градежните проекти, обучување на невронските мрежи и проценка на веројатноста за појава на ризици. Истражувањето покажува дека ANN моделите успешно можат да се применат за анализа на сложени инженерски и организациски проблеми каде постојат повеќе меѓусебно поврзани фактори.

5.2. Споредба на резултати од модел во „NeuralTools“ и модел во DamResML

За анализа на точноста на барем дел од моделираните алгоритми за машинско учење во DamResML кодот кој е дел од оваа докторска дисертација, одбран е сет на податоци со кој ќе се тренира и подучува модел во комерцијално достапниот „NeuralTools“ и модел во DamResML кодот. Тука треба да се напомене дека во „NeuralTools“ може да се изберат два типа на алгоритми за обучување на моделот: (1) Генерална регресиона невронска мрежа (анг. „General Regression Neural Network“) или накратко GRNN, и (3) Вештачка невронска мрежа на база на повеќеслоен перцептрон перцептор (анг. „Artificial Neural Network based on Multilayer perceptron regressor“) или накратко ANN. DamResML кодот, од друга страна, овозможува избор на повеќе алгоритми со машинско учење, меѓу кои и напоменатите – GRNN и ANN. Оттука, сетот на податоци кој ќе се разгледува ќе се обучува на овие два алгоритми со помош на „NeuralTools“ и DamResML кодот, по што ќе се направи споредба на добиените резултати.

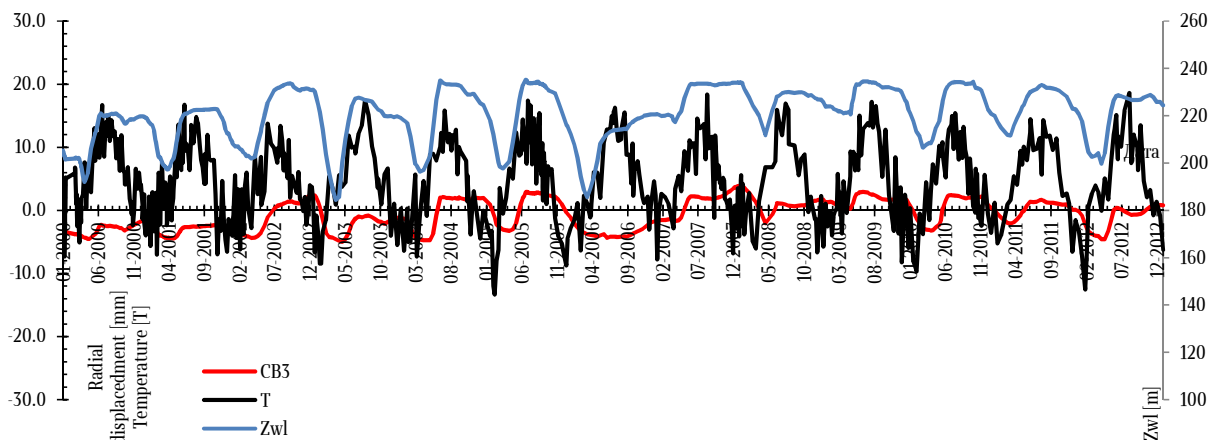
5.2.1. Подготовка на податоците за модел $CB3=f(Zwl,T)$

Сетот на податоци кој ќе се користи во оваа фаза на валидација на кодот претставува низа на регистрирани радијални поместувања во централниот блок на лачната брана во Франција накратко именувани како „CB3“, ниво на водата во акумулацијата „Zwl“ и амбиенталната температура пресметана во корелација со регистрирана температура на далечна метеоролошка станица „T“.

Целта на анализата на случај ќе биде пресметка на радијалните поместувања „CB3“ за периодот помеѓу 27.01.2013 година до 18.12.2017 година кога истите не се познати, со помош на поучени модели со машинско учење кои ќе се обучуваат на сет на мерени податоци на целната и предиктор варијаблите од периодот 19.01.2000 година па сè до 27.01.2013 година.

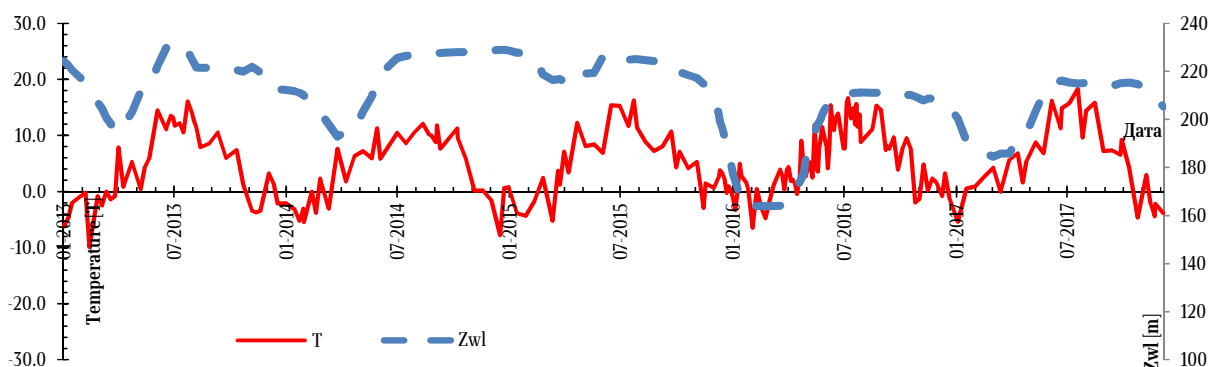
Сетот на податоци на целната варијаблата „CB3“ и предиктор варијаблите „Zwl“ и „T“ кои ќе се применуваат во процесот на обучување на моделите, се состои од вкупно 699 вредности за периодот помеѓу 19.01.2000 година па сè до 27.01.2013 година. Мерењата се на дневно ниво, но

истите се извршувани во неправилни временски интервали т.е. не се извршувани секојдневно. Вредностите на целната варијабла “CB3” и предиктор варијаблите “Zwl” и “T” користени во процесот на обучување на моделите со машинско учење се прикажани во Прилог 1.



Прилог 1. Графички приказ на користените вредности на предиктор варијаблата “CB3” и предиктор варијаблите “Zwl” и “T” во процесот на обучување на моделите со машинско учење за анализа на случај $CB3=f(Zwl,T)$.

Сетот на податоци на предиктор варијаблите “Zwl” и “T” кои ќе се користат во периодот на предвидување на целната варијабла “CB3” се состои од вкупно 226 вредности за периодот помеѓу 27.01.2013 година па сè до 18.12.2017 година, исто така регистрирани во неправилни временски интервали, на дневно ниво. За овој период, целната варијабла “CB3” е непозната. Истите се прикажани графички во Прилог 2.



Прилог 2. Графички приказ на предиктор варијаблите “Zwl” и “T” кои се користат во процесот на предвидување на целната варијабла “CB3” за анализа на случај $CB3=f(Zwl,T)$.

Целната варијабла – радијалните поместувања “CB3” во централниот блок, во анализите е во [mm]. Предиктор варијаблата “Zwl” – ниво на вода во акумулацијата е во [m], додека амбиенталната температура “T” е во мерна единица [K].

Целната варијабла “CB3” може да има позитивни и негативни вредности како резултат на насоката на напредување на радијалните поместувања – возводно или низводно. Така, позитивните вредности на радијалните поместувања значат поместувања во низводен правец, додека негативните – во возводен правец.

5.2.2. Резултати од споредбата на предвидувањата

По извршеното обучување на моделите со машинско учење на база на алгоритмите GRNN и ANN во “NeuralTools” и DamResML кодот креиран како дел од докторската дисертација, пресметани се предвидувања на целната варијабла “CB3” – радијални поместувања во централниот блок на анализираната лачна брана во функција на предиктор варијаблите “Zwl” и “T”.

Како што следи, дадено е кратко појаснување на добиените резултати.

Моделот со алгоритмот за учење GRNN изработен во “NeuralTools” во текот на обучувањето има вредност на $RMSE = 0.545$, додека при тестирањето на 20% од влезниот податочен сет за обучување има вредност на $RMSE = 0.6589$. Вредноста на коефициентот на детерминација R^2 е $R^2 = 0.8509$.

Моделот со алгоритмот GRNN во DamResML кодот е изработен такашто за тестирање на обучениот модел внатрешно кодот прави петкратна вкрстена валидација на хиперпараметрите за да определи која нивна комбинација во анализираните сетови од вкрстената валидација ќе даде најдобри метрички показатели за понатаму да биде проследен како обучен финален модел. Вредноста на коефициентот на детерминација R^2 како средна вредност од петкратната вкрстена валидација изнесува $R^2 = 0.934$, додека средната вредност на $RMSE$ изнесува $RMSE = 0.682$. На графикот во Прилог 3 може да се согледа дека предвидените вредности на моделот со “NeuralTools” и DamResML меѓусебно се следат, освен за вредности во месец март 2016 година кога двата алгоритми предвидуваат вредности на целната варијабла кои делуваат нелогично – вредностите нагло се зголемуваат од -4.54mm на -0.322mm во временски период од 14 дена. Вредностите сами по себе не делуваат нелогични, но прегледани на историскиот график се забележува дека овој скок во вредностите е нелогичен поради соседните вредности пред и по скокот кои се движат околу -4.00mm . Во овој период каде се јавува скокот во предвидените вредности, моделот се соочува со комбинација на предиктор варијаблите која претходно не била презентирана во процесот на обучување.

Со цел нумеричко согледување на разликите во предвидувањата од двата модели каде се применува истиот алгоритам GRNN, пресметани се разлики “e” во предвидувањата како:

$$e = \text{displ}_{\text{DamResML}} - \text{displ}_{\text{NeuralTools}} = \dots [\text{mm}] \quad 5.2.2.1$$

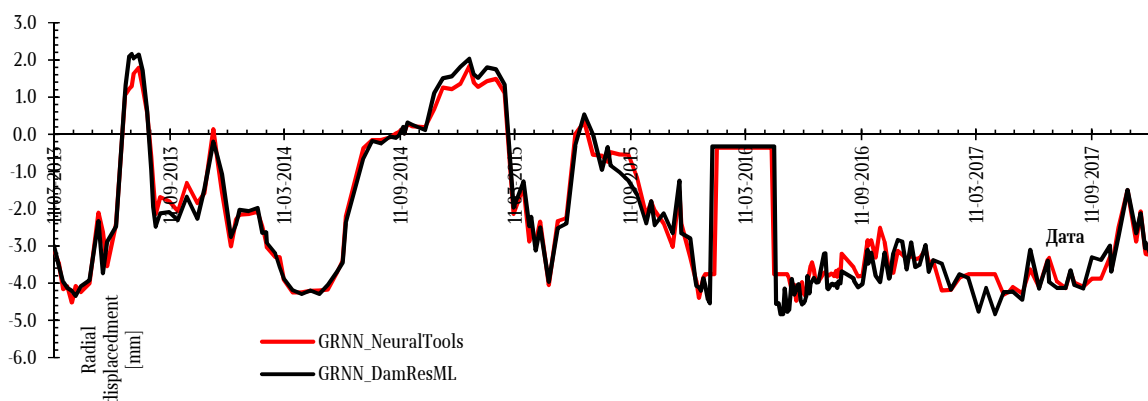
,каде:

e – разлика во предвидувањата во [mm],

$\text{displ}_{\text{DamResML}}$ – пресметани предвидувања на целната варијабла “CB3” со примена на DamResML,

$\text{displ}_{\text{NeuralTools}}$ – пресметани предвидувања на целната варијабла “CB3” со примена на NeuralTools,

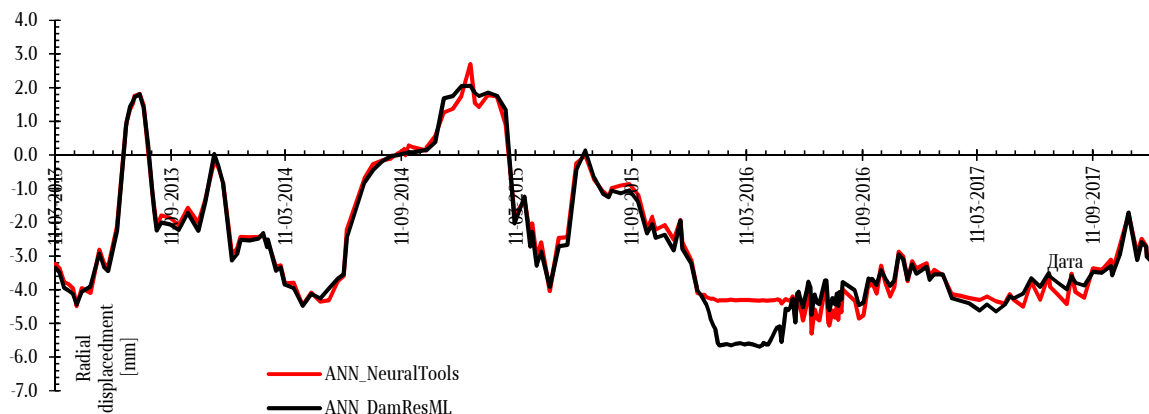
Историската низа на пресметани разлики е прикажана во Прилог 5, додека екстремите на разликите во предвидувањата од двата модели се движат во доста тесен опсег, од $e_{\min} = -1.462\text{mm}$ до $e_{\max} = 3.437\text{mm}$.



Прилог 3. Споредбен приказ на добиените резултати од моделите креирани со алгоритмот GRNN тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла “CB3” за анализа на случај $CB3=f(Zw1,T)$.

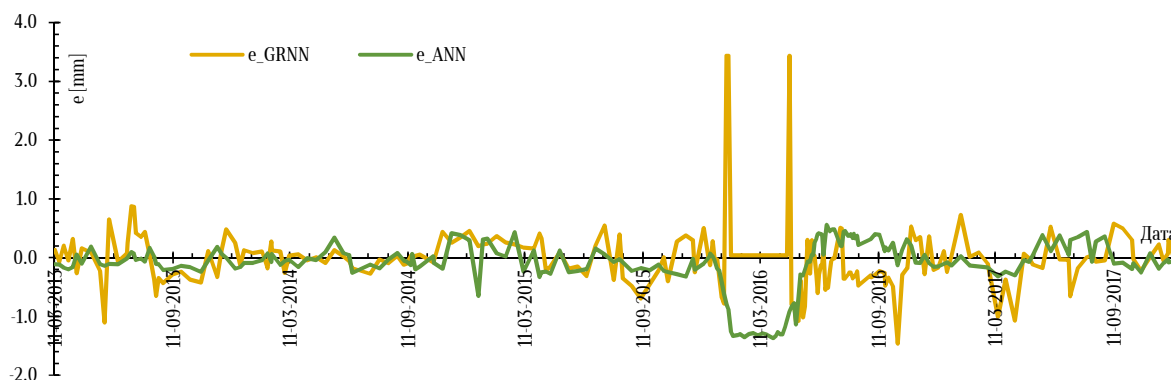
Откако беа презентирани резултатите добиени со GRNN алгоритмот во двата модели креирани во “NeuralTools” и DamResML, како што следи дадено е појаснување на резултатите добиени при споредба на предвидувањата од моделите тренирани на алгоритмот ANN, прикажани на графиконите во Прилог 4. Со споредба на моделите базирани на алгоритмот ANN (MLP) изработени во “NeuralTools” и во DamResML кодот, се согледуваат слични вредности за меѓусебните разлики на предвидувањата (Прилог 5) како кај моделите обучени на алгоритмот GRNN. Моделот обучен со

ANN во “NeuralTools” има вредност $RMSE = 0.6762$ во фазата на обучување, и вредност $RMSE = 0.6004$ во фаза на тестирање на моделот, со $R^2 = 0.8425$. Кај обучениот модел базиран на ANN во DamResML кодот, просечната вредност на $RMSE$ изнесува $RMSE = 0.6601$ и $R^2 = 0.9388$.



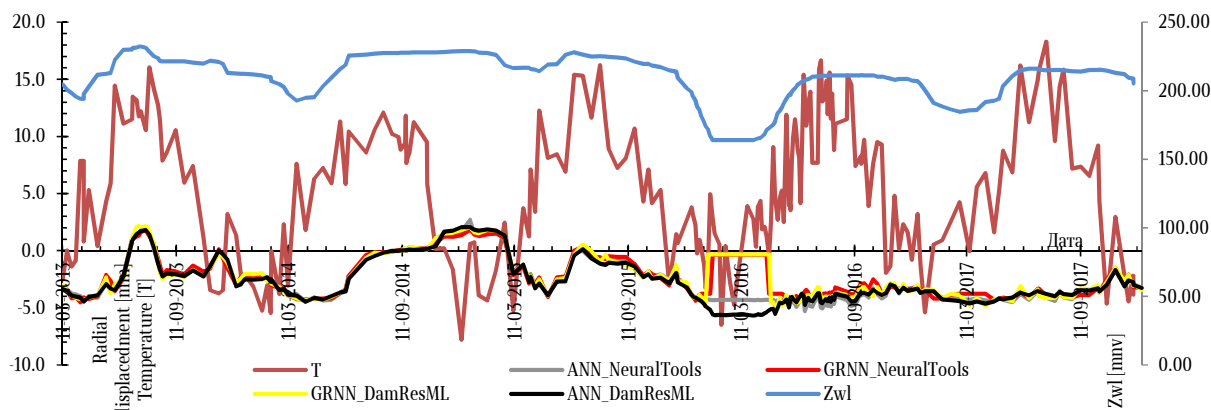
Прилог 4. Споредбен приказ на добиените резултати од моделите креирани со алгоритмот ANN тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла “CB3” за анализа на случај $CB3=f(Zw1,T)$.

Разликите во предвидувањата помеѓу двата модели се движат генерално околу хоризонталната оска и се блиски до нула (Прилог 5), со осцилации од $e_{min} = -1.370mm$ до $e_{max} = 0.559mm$. Разликите во предвидувањата со ANN алгоритмите се пресметани согласно равенството 5.2.2.1.



Прилог 5. Графички приказ на пресметани разлики во предвидувањата помеѓу моделите во “NeuralTools” и Пајтон.

Во периодот во месец мај, 2016 година, кога предвидувањата пресметани со GRNN моделите значително отскокнуваа од соседните вредности, кај моделите базирани на ANN алгоритмот едноставно ги нема овие отскокувања. Ова се должи на математичката пресметка во позадина врз основа на која алгоритмите ги прават осреднетите пресметки на целната варијабла во услови на еклатантни комбинации на вредности на предиктор варијаблите. За да се направи увид на вредностите на предиктор варијаблите во овој интересен период на разгледување на добиените податоци, даден е комбиниран приказ на графициите од предвидувањата од сите четири пресметки, заедно со вредности на предиктор варијаблите “T” и “Zw1” за овој период на предвидување на целната варијабла, прикажани во Прилог 6. На графикот може да се воочи дека во овој период – мај, 2016 година, нивото во акумулацијата подолг временски период се задржува на најниско ниво (околу вредност 164mmv), со осцилации на температурата од $0^{\circ}C$ до $-5^{\circ}C$. Со споредба на податочниот сет кој се користи во процесот на обучување на моделите прикажан во Прилог 1, може да се воочи дека најниското тренирано ниво на предиктор варијаблата “Zw1” е некаде 184mmv во услови на ниска амбиентална температура од с.с.а $3^{\circ}C$. Оттука, може да се потврди дека комбинацијата на предиктор варијаблите во мај, 2016 година навистина не е претставен во моделот во фаза на обучување и скокот во предвидените вредности кај GRNN моделите се должи токму на непознатиот сет на предиктор варијабли.



Прилог 6. Споредбен приказ на добиените резултати од четирите модели - два ANN и два GRNN модели тренирани во “NeuralTools” и DamResML кодот со приказ на пресметаните вредности на целната варијабла „CB3“ како и вредностите на предиктор варијаблите „Zwl“ и „T“, за анализа на случај CB3=f(Zwl,T).

5.3. Сублимат од спроведената споредба на резултатите

Во рамки на Поглавје 5 беше извршена валидација на DamResML кодот развиен во оваа докторска дисертација преку споредба на резултатите добиени со комерцијалниот софтверски пакет “NeuralTools”. Целта на анализата беше да се утврди дали развиениот код овозможува стабилни, реални и доверливи предвидувања споредливи со резултатите добиени од широко користена комерцијална алатка за моделирање со вештачки невронски мрежи.

Како студија на случај која се применува за изработка на паралелните предвидувања, користен е сетот на податоци за радијални поместувања во централниот блок на лачната брана во Франција “CB3”, нивото на водата во акумулацијата “Zwl” и амбиенталната температура “T”. Извршената анализа покажа дека моделите базирани на алгоритмите GRNN и ANN, обучени во DamResML кодот и во “NeuralTools”, генерално продуцираат слични трендови и слични вредности на предвидувањата на радијалните поместувања на лачната брана. Ова претставува значајна потврда дека развиениот DamResML код успешно ги моделира зависностите помеѓу нивото на водата во акумулацијата, амбиенталната температура и радијалните поместувања во централниот блок на браната.

Споредбата на метричките показатели дополнително ја потврди успешноста на развиените DamResML модели. Кај GRNN моделот, DamResML кодот постигна повисока вредност на коефициентот на детерминација $R^2 = 0.934$ во споредба со $R^2 = 0.8509$ кај моделот во “NeuralTools”, додека вредностите на RMSE беа меѓусебно блиски. Сличен тренд беше забележан и кај ANN моделите, каде DamResML кодот повторно постигна повисока вредност на $R^2 = 0.9388$ во споредба со $R^2 = 0.8425$ кај моделот во “NeuralTools”.

По согледување на резултатите од предвидувањата, кај GRNN моделите се јави отскокнување во вредностите во мај, 2016 година, по што се скокот во вредноста се провлекува како константна вредност сè до втората половина на месец април, 2016 година. Појавата на речиси константни предвидени вредности во текот на дел од 2016 година не е поврзана со конкретната софтверска имплементација, бидејќи истото однесување се јавува и кај GRNN моделот креиран во NeuralTools и кај GRNN моделот развиен во DamResML. Ваквиот резултат посочува дека причината произлегува од самата природа на GRNN алгоритмот и од распределбата на влезните податоци за анализираниот период. Во услови кога комбинациите на нивото на водата во акумулацијата и температурата не овозможуваат доволно јасно разграничување на промените во радијалното поместување, GRNN моделот тежнее кон измазнување на предвидувањата и нивно приближување кон средна пондерирана вредност. За разлика од тоа, ANN моделите, и во NeuralTools и во DamResML, покажуваат поголема способност за следење на променливоста на целната променлива во истиот период.

Од анализата може да се заклучи дека “NeuralTools” претставува практична и релативно едноставна алатка за ANN и GRNN моделирање во Excel околина, особено за корисници кои немаат напредно програмерско знаење. Сепак, неговата затворена (анг. „black-box“) природа и ограничената контрола врз архитектурата, методите за предпроцесирање и процесот на валидација претставуваат ограничување при развој на сложени инженерски модели. Од друга страна, DamResML кодот развиен во рамки на оваа докторска дисертација овозможува целосна контрола врз процесот на обучување, изборот на алгоритми, хиперпараметрите, методите на вкрстена валидација и анализата на резултатите, што овозможува развој на пофлексибилни и постабилни модели.

Резултатите од тестирањето и валидацијата потврдуваат дека развиениот DamResML код успешно може да се применува за анализа на сложени инженерски задачи и предвидување на однесувањето на браните како хидротехнички објекти. Дополнително, споредбата со “NeuralTools” претставува независна потврда за точноста, стабилноста и доверливоста на развиените модели со машинско учење во DamResML кодот.

6. СТУДИЈА НА СЛУЧАЈ 1

Во Поглавје 6, прикажани се детални описи и резултати на истражувањата од примената на моделите со машинско учење во управување со хидросистемите „Злетовица“ и „Кнежево“. Со примена на обучени модели со машинско учење, во анализираните студии на случај се согледуваат влијанијата на различните операторски одлуки врз однесувањето на хидросистемите, со акцент на една избрана целна варијабла.

Анализата на случај 1.1 се фокусира во предикција на протокот кој се испушта низводно од акумулација „Кнежево“ кој е клучен параметар во управувањето со хидросистемот бидејќи низводното испуштање од браната се зафаќа долж река Злетовица за различни потреби на населението – наводнување и водоснабдување. Дополнително, овој проток до неодамна се испуштал единствено за задоволување на потребите на водокорисниците, како и за одржување на биотопот и биоценозата во водотекот со пропуштање на количини на вода еднакви на загарантираниот биолошки минимум, но во последната година операторот на хидросистемот „Злетовица“ изгради мала хидроцентрала „ХЕЦ Б“ низ која енергетски ќе се обработува протокот кој се испушта низводно од браната. Со проценката на овој проток врз основа на работните услови во хидросистемот дефинирани преку ниво на вода во акумулација, дотек во акумулација и акумулиран волумен во акумулација, операторот во иднина може да го планира производството на енергија. Дополнително, во втората анализа на случај 1.2 изработен е предикционен модел за произведена моќност во хидроцентралата „ХЕЦ Б“ врз основа на досега прибраните податоци за производството на моќност во хидроцентралата.

Третата анализа на случај 1.3 е фокусирана кон анализа на нивото на водата во акумулацијата „Стрежево“ земајќи ги предвид познатите дотекувања и истекувања од акумулацијата. Оваа акумулација има стратешко регионално значење, а водните ресурси кои се слеваат во акумулацијата придонесуваат во водоснабдување на градот Битола и околината, наводнување на Битолско поле, како и подобрување на енергетската ефикасност на термоблоковите на РЕК „Битола“. Покрај набројаните корисници на водата од акумулацијата, операторот успешно управува со повеќе хидроцентрали. Со земање предвид на сите познати дотекувања и истекувања од акумулацијата, изработени се модели со машинско учење кои се користат во предвидување на целната варијабла „Zwl“, а понатаму и оптимизација на истата во екстремни услови на зголемена и намалена побарувачка на вода. Резултатите се прикажани како што следи, а на крајот од Поглавје

6 даден е кус сублимат на спроведените анализи.

6.1. Анализа на случај 1.1 – $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$

6.1.1. Подготовка на податоците

Во првата анализа на случај, вклучени се следните параметри од набљудувањето на брана „Кнежево“:

- 1) Ниво на вода во акумулација „Zwl“ во [mnv] прикажан во Прилог 9,
- 2) Волумен на вода во акумулацијата „Reservoir_volume_10-6“ во [10^6m^3] прикажан во Прилог 8,
- 3) Дотек во акумулацијата „Q_inflow“ во [m^3/s] прикажан во Прилог 7, и
- 4) Истек од акумулацијата „Q_discharge“ во [m^3/s] прикажан во Прилог 7.

Во оваа анализа на случај, целна варијабла е истекувањето од акумулацијата „Q_discharge“, додека останатите варијабли се предиктор варијабли. Во процесот на обучување, вклучени се вкупно 2241 податок за сите пет варијабли. Во процесот на предвидување, каде целната варијабла е исклучена, вклучени се вкупно 365 податоци за четирите предиктор варијабли. Целта на оваа анализа на случај е проценка на протокот кој се испушта низводно од акумулацијата „Кнежево“ за периодот од 1ви јануари 2025 година па сè до 31 декември 2025 година, и споредба на добиените резултати до мерените испуштени протекувања. Проценката на истекувањето од акумулацијата „Кнежево“ е клучен параметар во управувањето на овој хидросистем бидејќи со испуштањето на низводното протекување се обезбедуваат количини на потреба вода за низводните корисници – водоснабдување и наводнување на градот Кратово, Штип и Св. Николе.

Податоците користени во моделите се регистрирани како дневни мерени податоци, почнувајќи од ноември, 2019 година па сè до декември, 2025 година. Одредени денови нема информација за податоците поради прекин со снабдување на електрична енергија.

Нивото на вода во акумулација „Zwl“ операторот на хидросистемот секојдневно го регистрира во 7:30 часот наутро со помош на мерна летва поставена во десниот бок во непосредна близина на браната.

Податоците за дотек во акумулацијата „Q_inflow“ се пресметани податоци, истите не се мерени регистрирани податоци. Имено, операторот со хидросистемот врши секојдневно мерење на водното ниво во акумулацијата, а дополнително ја познава и кривата на површини и волумени на акумулацијата, па со познавање на овие два податоци се пресметува волуменот во акумулацијата „Reservoir_volume_10-6“ за истиот ден во 10^6m^3 . Понатаму, операторот врши мерење на истекувањата низ темелниот испуст и преливниот орган, со што познат е истекот од акумулацијата. Оттука, дотекот во акумулацијата е пресметан како:

$$Q_{\text{inflow}} = (V_i - V_{i-1}) \cdot 10^6 \cdot (24 \cdot 60 \cdot 60) + Q_{\text{discharge}} = \dots \left[\frac{\text{m}^3}{\text{s}} \right] \quad (6.1.1)$$

, каде:

Q_{inflow} – проценет дотек во акумулацијата во [m^3/s],

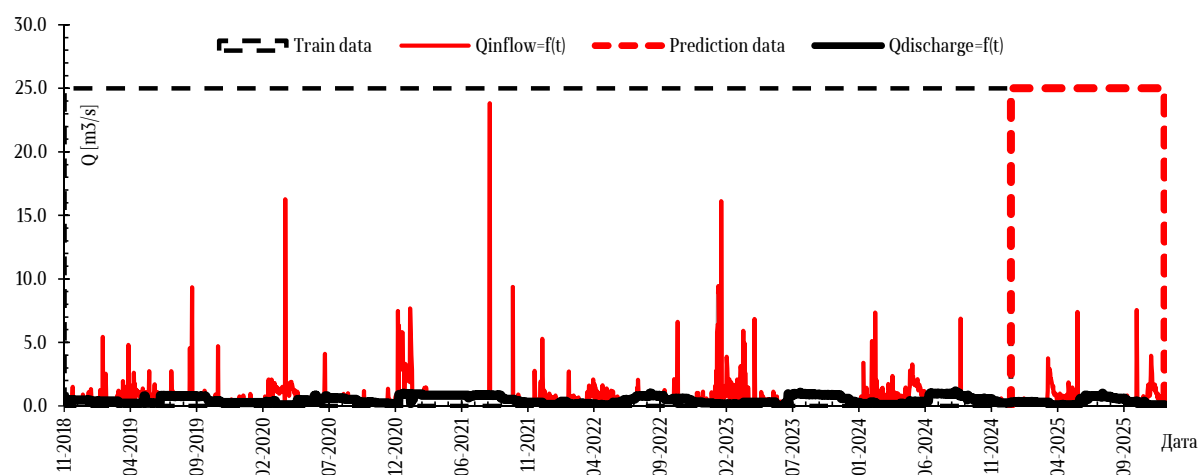
V_i – волумен на вода во акумулацијата во i – ден, во [10^6m^3],

V_{i-1} – волумен на вода во акумулацијата во $(i-1)$ – ден, во [10^6m^3],

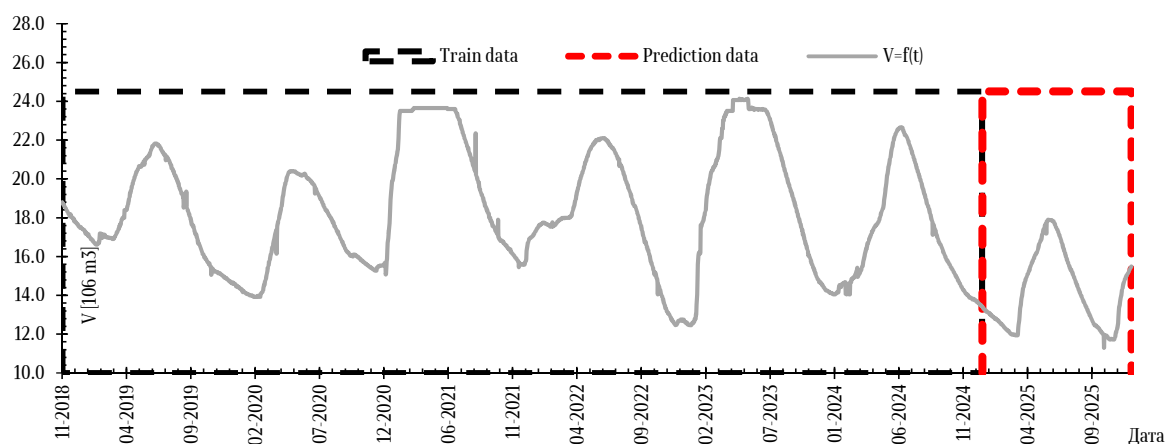
$Q_{\text{discharge}}$ – регистрирано истекување од акумулацијата во [m^3/s].

Тука треба да се напомене дека операторот на хидросистемот ја регистрира и дисконтинуирано врши мерења на филтрацијата во лев бок, десен бок и низ телото на браната. Истите се ред големина од $0.01 \text{ m}^3/\text{s}$ до $0.105 \text{ m}^3/\text{s}$ и не се регистрираат секојдневно туку во одредени неправилни временски интервали, поради што овој податок не е внесен во равенка (4) во пресметката на проценетиот дотек во акумулацијата. Операторот не врши мерења на испарувањето од

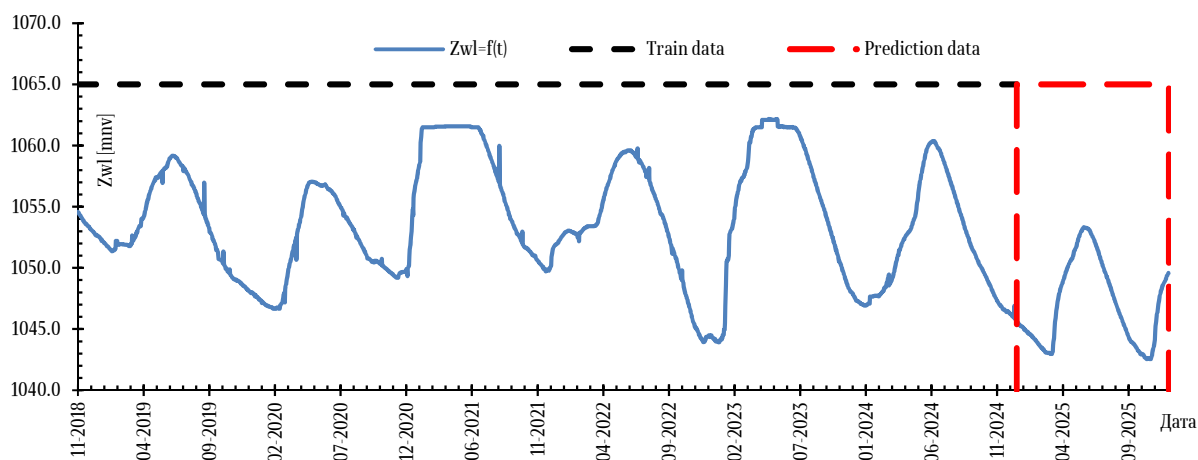
акумулацијата, па поради тоа и тој податок е исклучен во проценката на дотекот во акумулацијата.



Прилог 7. График за варијаблите дотек „Q_inflow“ и истек „Q_discharge“ од акумулацијата, за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.



Прилог 8. График за варијаблата волумен на вода во акумулацијата „Reservoir_volume_10-6“ за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.



Прилог 9. График за варијаблата ниво на вода во акумулацијата „Zwl“ за временските интервали на податоци кои се земени предвид во процесот на обучување на моделот и процесот на предвидување на целната варијабла за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

6.1.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Перцепсија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Целната варијабла е користена во нејзината оригинална форма, без примена на нормализација или стандардизација. Од друга страна, предиктор варијаблите беа стандардизирани во рамки на секој сет од петкратната валидација, и тоа само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R² прикажани во Табела 3.

Табела 3. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

5-fold CV					
No.	Algorithm	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	Random Forest	0.025	0.157	0.076	0.690
2	XGBoost	0.028	0.168	0.099	0.647
3	Boosted Trees	0.029	0.169	0.102	0.642
4	Artificial Neural Networks	0.032	0.178	0.115	0.604
5	K-Nearest Neighbors	0.033	0.181	0.097	0.591
6	Decision Tree	0.037	0.192	0.074	0.537
7	Support Vector Regressor	0.044	0.211	0.143	0.444
8	Gaussian Process	0.060	0.244	0.184	0.253
9	GRNN	0.061	0.247	0.207	0.240
max value:		0.061	0.247	0.207	0.690
min value:		0.025	0.157	0.074	0.240

Спроведената компаративна анализа на изработените модели со машинско учење врз основа на средноквадратната грешка - MSE покажува на хиерархија во нивната способност за предвидување. Добиените вредности на MSE се движат во интервал од 0.025 до 0.061, што посочува на умерена варијација во точноста на предвидувањето помеѓу разгледуваните модели. Најдобри резултати се постигнати со RF моделот (MSE = 0.025), по што следуваат XGBoost (0.028) и Boosted Trees (0.029). Конзистентната супериорност на ансамбл-методите базирани на дрвја укажува на нивната висока способност за моделирање на нелинеарни зависности и сложени интеракции помеѓу променливите во анализираниот сет на податоци. Средно ниво на перформанси е забележано кај вештачките невронски мрежи (MSE = 0.032) и K-Nearest Neighbors (0.033), што покажува дека овие модели успешно го апроксимираат однесувањето на системот, но со пониска прецизност во однос на ансамбл-моделите. Пониски перформанси се евидентирани кај DT (0.037) и SVR (0.044), што може да се поврзе со ограничувања во генерализацијата и чувствителност на структурата на податоците. Најслаби резултати се добиени со GPR (0.060) и GRNN (0.061), што укажува дека овие методи се помалку соодветни за анализираниот проблем, веројатно поради чувствителност на шум и потешкотии при моделирање на сложени, високодимензионални односи. И покрај сето погоре наведено, релативно тесниот опсег на вредности на MSE укажува дека изборот на модел има одредено влијание врз перформансите, но

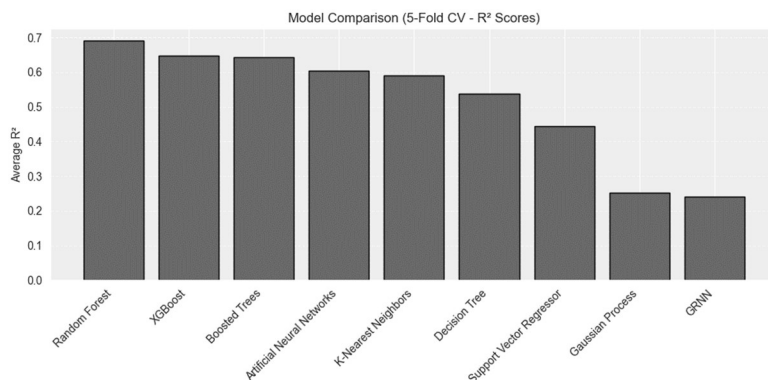
разликите помеѓу најдобрите модели се релативно мали. Ова сугерира дека понатамошно подобрување на точноста на предвидување може да се постигне првенствено преку подобрување на квалитетот на податоците и дополнување на предиктор варијаблите.

Анализата на резултатите врз основа на коренот на средноквадратната грешка – RMSE покажува конзистентен тренд со резултатите добиени од MSE, при што вредностите на RMSE се движат во интервал од $0.157 \text{ m}^3/\text{s}$ до $0.247 \text{ m}^3/\text{s}$. Овој опсег посочува на умерена разлика во точноста на предвидување помеѓу различните модели. Најдобри перформанси се постигнати со Random Forest (RMSE = 0.157), по што следуваат XGBoost (0.168) и Boosted Trees (0.169). Овие резултати дополнително ја потврдуваат супериорноста на ансамбл-методите базирани на одлучувачки дрвја, кои покажуваат висока способност за моделирање на сложени и нелинеарни процеси. Моделите Artificial Neural Networks (RMSE = 0.178) и K-Nearest Neighbors (0.181) покажуваат солидни, но нешто послаби резултати, што укажува на нивна способност за апроксимација на системот, но со ограничена прецизност во однос на водечките модели. Пониска точност е забележана кај Decision Tree (0.192) и Support Vector Regressor (0.211), што укажува на нивни ограничувања во генерализацијата и адаптацијата на комплексната структура на податоците. Највисоки вредности на RMSE се добиени кај Gaussian Process (0.244) и GRNN (0.247), што укажува дека овие модели имаат најголеми отстапувања од реалните вредности и се најмалку погодни за анализираниот проблем. Општо земено, разликата помеѓу најдобриот и најслабиот модел изнесува околу $0.09 \text{ m}^3/\text{s}$, што укажува дека иако изборот на модел има значајно влијание, разликите помеѓу најдобрите модели се релативно мали.

Анализата на резултатите врз основа на средната апсолутна грешка – MAE покажува дека вредностите се движат во интервал од 0.074 до 0.207, што укажува на значајни разлики во точноста на предвидување помеѓу применетите модели. Најниска вредност на MAE е постигната со Decision Tree (MAE = 0.074), по што следува Random Forest (0.076), додека XGBoost (0.099) и Boosted Trees (0.102) исто така покажуваат добри перформанси. Ниските вредности на MAE кај овие модели укажуваат дека тие обезбедуваат најмало просечно апсолутно отстапување од реалните вредности. Моделите K-Nearest Neighbors (0.097) и Artificial Neural Networks (0.115) покажуваат умерени резултати, што укажува дека тие генерално добро ги следат трендовите во податоците, но со нешто поголеми просечни отстапувања во споредба со најдобрите модели. Поголеми вредности на MAE се забележуваат кај Support Vector Regressor (0.143), што укажува на послаба точност во просечна смисла. Највисоки вредности се добиени кај Gaussian Process (0.184) и GRNN (0.207), што укажува дека овие модели имаат најголеми просечни апсолутни грешки и најслаба способност за точно предвидување. Интересно е да се забележи дека Decision Tree, иако не е најдобар според MSE и RMSE, постигнува најниска MAE, што укажува дека моделот има мали просечни грешки, но повремено прави поголеми отстапувања (што се рефлектира во повисок MSE и RMSE). Ова укажува на нестабилност во предвидувањата и потенцијално присуство на екстремни грешки.

Анализата на резултатите врз основа на коефициентот на детерминација - R^2 покажува дека вредностите се движат во интервал од 0.240 до 0.690, што укажува на значајни разлики во способноста на моделите да ја објаснат варијабилноста на целната променлива. Највисока вредност на R^2 е постигната со Random Forest ($R^2 = 0.690$), што укажува дека овој модел објаснува приближно 69% од варијансата на податоците. Следуваат XGBoost (0.647) и Boosted Trees (0.642), кои исто така покажуваат високо ниво на објаснувачка моќ. Овие резултати ја потврдуваат ефикасноста на ансамбл-методите базирани на дрвја во моделирање на сложени и нелинеарни процеси. Моделите Artificial Neural Networks ($R^2 = 0.604$) и K-Nearest Neighbors (0.591) постигнуваат умерени вредности, што укажува дека успеваат да објаснат значителен дел од варијабилноста, но со помала прецизност во однос на водечките модели. Понишки вредности се забележуваат кај Decision Tree (0.537) и Support Vector Regressor (0.444), што укажува на ограничена способност за генерализација и недоволно моделирање на комплексните односи во податоците. Најниски вредности на R^2 се добиени кај Gaussian Process (0.253) и GRNN (0.240), што значи дека овие модели објаснуваат само околу 24–25% од варијансата, односно имаат слаба предиктивна моќ за дадениот проблем. Општо земено, резултатите од R^2 анализата ја потврдуваат супериорноста на ансамбл-моделите и нивната способност да ја репрезентираат структурата на податоците. Сепак,

вредностите под 0.7 кај сите модели укажуваат дека значителен дел од варијабилноста останува необјаснет, што сугерира потреба од дополнително подобрување преку подобар избор на влезни променливи, поквалитетни податоци и подлабока анализа на физичките процеси кои го дефинираат системот. Графички приказ на пресметаниот просечен R^2 за секој алгоритам е прикажан на Прилог 10.



Прилог 10. Графички приказ на вредноста на R^2 за секој анализиран алгоритам на сетот податоци за обучување со петкратна вкрстена валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Вкупната анализа врз основа на MSE, RMSE, MAE и R^2 покажува дека ансамбл-моделите базирани на одлучувачки дрвја (Random Forest, XGBoost и Boosted Trees) обезбедуваат најдобра и најконзистентна предиктивна точност. Овие модели постигнуваат најниски грешки и највисоки вредности на R^2 , што укажува на нивна супериорна способност за моделирање на нелинеарни и сложени односи во податоците.

Останатите модели покажуваат умерени до послаби перформанси, при што разликите помеѓу најдобрите алгоритми се релативно мали. Ова укажува дека понатамошното подобрување на точноста не зависи исклучиво од изборот на модел, туку пред сè од квалитетот на податоците, изборот на релевантни влезни променливи и разбирањето на физичките процеси што го дефинираат системот.

6.1.3. Обучување на моделите

Во процесот на обучување на моделите, по извршената петкратна валидација, донесена е одлука сите девет алгоритми да се користат во обучување на девет различни модели. Накратко секој модел е опишан како што следи, заедно во хиперпараметрите внесени во DamResML.

6.1.3.1. Модел на Вештачки невронски мрежи (ANN)

Прв обучен модел е моделот на Вештачки невронски мрежи или накратко ANN (анг. „Artificial Neural Network“). Во конкретниот DamResML код во кој е внесен алгоритмот за Вештачки невронски мрежи, во однос на параметрите, внесени се следните вредности:

1) Број на скриени слоеви n_{layer} и бројот на неврони во секој слој n_{neurons} се внесени како `hidden_layer_sizes=(100,100)`. Со ова, внесени се 2 скриени слоеви со 100 неврони во секој од нив.

2) Типот на активациона функција е усвоен како `solver="adam"` што значи дека се применува Оптимизаторот Адам (анг. „Adam optimizer“) како активациона функција.

(3) Брзината на учење l_{rate} е усвоена како `learning_rate_unit=0.001`.

Процесот на обучување траеше 2s и 613ms, со просечна вредност на $\text{MSE} = 0.0333$, $\text{RMSE} = 0.182$, $\text{MAE} = 0.119$ и $R^2 = 0.585$.

6.1.3.2. Модел на Регресија со потпомогнати вектори (SVR)

Вториот обучен модел е моделот на Регресија со потпомогнати вектори или накратко SVR (анг. „Support Vector Regressor“). Во дадениот код е имплементиран модел на Support Vector Regression (SVR), кој претставува метод базиран на поддржувачки вектори за решавање на регресиони

проблеми. Основната идеја на SVR е да се пронајде функција која најдобро ги апроксимира податоците, при што грешките помали од одреден праг (ϵ) не се казнуваат.

Моделот е дефиниран со следните параметри:

1) `kernel = "rbf"` - Се користи радијална базисна функција (Radial Basis Function), која овозможува моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) `C = 100` - Регуларизациски параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на C значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување (overfitting).

3) `epsilon = 0.1` - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

За евалуација на моделот е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Средната вредност на метричките показатели по обучувањето на моделот изнесуваат: Avg MSE = 0.0461, Avg RMSE = 0.2146, Avg MAE = 0.1456, Avg R^2 = 0.4232. Процесот на обучување траеше 3s и 571ms.

6.1.3.3. Модел на Случајна шума (RF)

Третиот обучен модел е моделот на Random Forest регресија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја. Основната идеја е да се изгради голем број дрвја врз различни подсети од податоците и нивните предвидувања да се комбинираат (се вадат просечни резултати), со цел да се подобри точноста и стабилноста на моделот.

За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators (50, 100, 150)` - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth (None, 10, 20)` - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split (2, 5, 10)` - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features ("sqrt", "log2", None)` - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата „None“ сите предиктор варијабли се вклучуваат во сите јазли, додека „sqrt“ и „log2“ земаат предвид само дел од предиктор варијаблите.

За добивање на објективна проценка на перформансите применета вгнездена n-кратна валидација, и тоа:

- Внатрешна (анг. „inner“) трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку т.н. „GridSearchCV“.
- Надворешна (анг. „outer“) петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Табела 4. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	max_depth	max_features	min_samples_split	n_estimators
5	10	None	2	150
4	20	None	5	100

Fold	max_depth	max_features	min_samples_split	n_estimators
3	20	None	2	150
2	20	None	10	50
1	20	None	2	150

Финалното обучување е направено согласно хиперпараметрите од третиот сет од петкратната надворешна валидација, со вредности: 'max_depth': 20, 'max_features': None, 'min_samples_split': 2, 'n_estimators': 150. Процесот на обучување траеше 52s и 802ms.

6.1.3.4. Модел Одлучувачко дрво (Decision Tree)

Четвртиот обучувачки модел е моделот на Decision Tree Regressor, кој претставува едноставен, но моќен алгоритам базиран на хиерархиско делење на податоците. Моделот функционира така што рекурзивно ги дели податоците на подгрупи врз основа на вредностите на предиктор варијаблите, при што на секој чекор се одбира поделба која најмногу ја намалува грешката (најчесто MSE).

Во овој случај, моделот е дефиниран со основните параметри (со random_state = 42 за репродуктивност), без дополнителна регуларизација, што значи дека дрвото може да расте до максимална сложеност и да ги следи податоците детално.

За евалуација на хиперпараметрите е применета петкратна вкрстена валидација, при што податоците се делат на пет сетови, а моделот се обучува и тестира пет пати со различни комбинации. Овој пристап овозможува стабилна и доверлива проценка на перформансите.

Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста на моделот. Бидејќи DT не бара скалирање на влезните податоци, моделот се применува директно врз оригиналните вредности на секоја варијабла. Вредностите за метриките се прикажани во Табела 5.

Табела 5. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај Q_{discharge} = f(Z_{wl}, ResVol, Q_{inflow}).

Fold	MSE	RMSE	MAE	R2
1	0.043	0.208	0.082	0.467
2	0.036	0.191	0.071	0.544
3	0.028	0.166	0.064	0.640
4	0.037	0.191	0.076	0.532
5	0.042	0.205	0.080	0.502

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 0.0371, Avg RMSE: 0.1922, Avg MAE: 0.0743, Avg R²: 0.5370. Процесот на обучување траеше 354ms.

6.1.3.5. Модел К – најблизок сосед (K – nearest neighbor)

Како петти трениран модел е моделот со K-Nearest Neighbors (KNN) регресија, кој претставува метод базиран на сличност помеѓу податоците. Основната идеја е дека вредноста на целната променлива за нов примерок се одредува како просек од вредностите на неговите k најблиски соседи во просторот на влезните променливи.

Моделот е дефиниран со параметар n_neighbors = 5, што значи дека за секое предвидување се земаат предвид 5 најблиски примероци од податоците за обучување.

За проценка на перформансите е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста на моделот. Конкретните вредности за моделот се дадени во Табела 6. Треба да се напомене дека KNN не гради експлицитен модел (анг. „lazy learning“), туку ги користи сите тренинг податоци при предвидување, што може да влијае на пресметковната ефикасност кај поголеми податочни множества.

Табела 6. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	MSE	RMSE	MAE	R2
1	0.034	0.185	0.098	0.577
2	0.035	0.187	0.102	0.561
3	0.027	0.165	0.091	0.642
4	0.031	0.176	0.093	0.607
5	0.036	0.191	0.103	0.566

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 0.0328, Avg RMSE: 0.1809, Avg MAE: 0.0972, Avg R²: 0.5906. Процесот на обучување траеше 233ms.

6.1.3.6. Модел со XGBoost (XGBoost Regressor)

Шестиот трениран модел е моделот на XGBoost Regressor, кој претставува напреден ансамбл-метод базиран на градиентно бустирање (анг. „gradient boosting“) на одлучувачки дрвја. Основната идеја е дека моделот се гради постепено, при што секое ново дрво ја коригира грешката направена од претходните дрвја, со цел да се постигне висока предиктивна точност.

Моделот е дефиниран со следните клучни параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но може да ја зголеми сложеноста на моделот.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое ново дрво. Помали вредности водат до постабилно учење, но бараат повеќе дрвја.

3) `objective = "reg:squarederror"` - Ја дефинира функцијата на загуба, при што се минимизира средноквадратната грешка (MSE).

4) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи XGBoost е базиран на дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета стандардизација на карактеристиките.

За евалуација на моделот применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста и објаснувачката моќ на моделот. Пресметаните вредности на метриките од моделот се прикажани во Табела 7.

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 0.026039, Avg RMSE: 0.1608, Avg MAE: 0.0844, Avg R²: 0.6759. Процесот на обучување траеше 443ms.

Табела 7. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	MSE	RMSE	MAE	R2
1	0.030	0.172	0.088	0.635
2	0.028	0.167	0.086	0.653
3	0.019	0.139	0.078	0.748
4	0.023	0.152	0.079	0.705
5	0.030	0.174	0.091	0.639

6.1.3.7. Модел со Гаусов Процесен Регресор (GPR)

Седмиот модел за обучување се базира на алгоритмот Гаусов Процесен Регресор (анг. „Gaussian Process Regression“). Тој претставува нелинеарен, пробабилистички алгоритам за машинско учење за регресиони проблемски задачи. Наместо да гради експлицитна функција, GPR ја моделира зависноста како распределба над можни функции, при што предвидувањата се базираат на коваријантна структура дефинирана преку кернел функција.

Во моделот е користен комбиниран кернел – т.н. „ConstantKernel × RBF“ (анг. „Radial Basis

Function“). Овој кернел овозможува моделирање на мазни нелинеарни односи помеѓу променливите, при што RBF ја контролира т.н. „должината на влијание“ (анг. „length scale“), а „ConstantKernel“ ја скалира вредноста на податоците.

Клучни параметри на моделот се:

1) `normalize_y = True` - Ја нормализира целната варијабла, што значително ја подобрува стабилноста на моделот.

2) `n_restarts_optimizer = 5` - Овозможува повеќекратно оптимизирање на параметрите на кернел функцијата, со цел да се избегне локален минимум и да се добие подобро вклопување.

Поради тоа што GPR е чувствителен на растојанија, применета е StandardScaler трансформација во рамки на методологијата, со што сите податоци од сетот за обучување се стандардизираат во секој сет од петкратната вкрстена валидација. Перформансите се оценуваат преку MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста. Иако GPR има теоретски висока флексибилност и способност за моделирање на сложени соодноси помеѓу различните варијабли, во практични услови може да биде чувствителен на шум и ограничен при работа со поголеми и покомплексни податочни множества.

Вредностите за метричките показатели од процесот на петкратна вкрстена валидација се прикажани во Табела 8.

Табела 8. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	MSE	RMSE	MAE	R2
1	0.150	0.388	0.195	-0.847
2	0.093	0.304	0.172	-0.156
3	0.051	0.225	0.159	0.335
4	0.512	0.716	0.193	-5.539
5	0.060	0.245	0.170	0.287

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 0.1731, Avg RMSE: 0.3755, Avg MAE: 0.1778, Avg R^2 : -1.1840. Процесот на обучување траеше 11m и 31s.

6.1.3.8. Модел со Поткренати Регресиони Дрва (BT)

Осмиот модел за обучување се базира на алгоритмот на поткренати регресиони дрва (анг. „Boosted Regression Trees“), односно т.н. „Gradient Boosting Regressor“, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја. Основната идеја е дека секое ново дрво се обучува да ја коригира грешката направена од претходните модели, со што постепено се подобрува точноста на предвидувањето.

Моделот е дефиниран со следните параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на моделот е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 9. BT како бустинг метод, е особено ефикасен во моделирање на сложени нелинеарни односи и често постигнува висока предиктивна точност.

Табела 9. Вредности на метричките показатели од процесот на обучување на ВТ моделот по спроведена петкратна валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	MSE	RMSE	MAE	R2
1	0.030	0.174	0.102	0.628
2	0.030	0.173	0.103	0.626
3	0.022	0.149	0.095	0.710
4	0.029	0.169	0.101	0.635
5	0.033	0.181	0.109	0.612

Просечните вредности на метричките показатели, се: Avg MSE: 0.0287, Avg RMSE: 0.1691, Avg MAE: 0.1021, Avg R²: 0.6421. Потребното време за обучување беше 660ms.

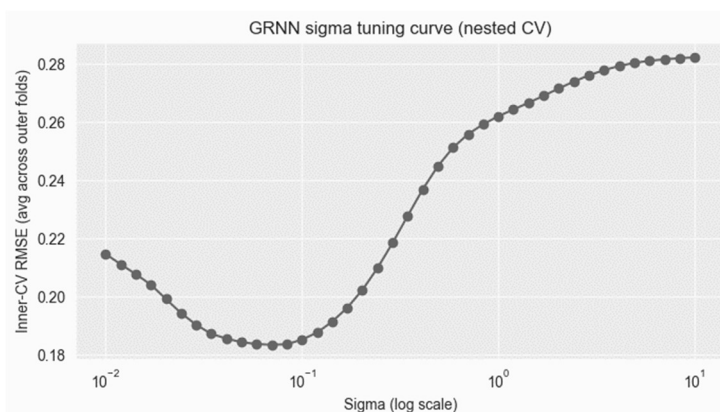
6.1.3.9. Модел со Генерализирана Регресиона Невронска Мрежа (GRNN)

Последниот модел е обучуван на алгоритмот на Генерализирана Регресиона Невронска Мрежа (анг. „General Regression Neural Network“). Тој претставува нелинеарен, ненадгледуван метод базиран на радијални базни функции. GRNN не гради експлицитна параметарска функција, туку предвидувањата ги врши преку тежинско осреднување на сите обучувани примероци, а тежинските коефициенти зависат помеѓу растојанието помеѓу примероците. Основен параметар на моделот е σ – сигма. Овој параметар ја контролира ширина на Гаус функцијата, односно степенот на „влијание“ на соседните точки. Помали вредности водат до локално чувствителен модел (ризик од т.н. „overfitting“), додека поголеми вредности создаваат помазни предвидувања (ризик од т.н. „underfitting“).

Во кодот е применета вгнездена трикратна валидација за избор на најдобра вредност на параметарот σ – сигма. Се тестираат вредности во рангот $\sigma = (0.01 - 10)$. Критериумот е да се избере вредност за σ што ќе резултира со минимален RMSE, а во конкретниот случај графикот на варијација на RMSE во зависност од параметарот σ по трикратната вкрстена валидација е прикажан во Прилог 11. Понатаму, со изборот на соодветна вредност за параметарот σ и дополнително се прави петкратна надворешна вкрстена валидација на податоците.

Табела 10. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Fold	Best Sigma	Fold RMSE
1	0.049	0.186
2	0.070	0.172
3	0.084	0.176
4	0.084	0.185
5	0.059	0.172



Прилог 11. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Од графикот во Прилог 11 може да се воочи дека минимумот на $\text{RMSE} = 0.183$ се добива за вредност на параметарот $\sigma \approx 0.05 \div 0.1$.

6.1.4. Резултати од обучувањето на моделите

По обучување на девет различни модели базирани на девет различни алгоритми, како што следи презентирани се сумарни резултати од сите модели за претходната фаза во форма на табели и графици.

Како првичен преглед на резултатите, во Табела 11 прикажани се вредности на метричките показатели добиени во процесите на петкратна вкрстена валидација во секој обучен модел.

Табела 11. Резултати за вредностите на метричките показатели по процесот на обучување во рамки на кој во секој модел е спроведена и петкратна вкрстена валидација анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Model	MSE	RMSE	MAE	R2
GPR	0.173	0.376	0.178	-1.184
SVR	0.046	0.215	0.146	0.423
Decision Tree	0.037	0.192	0.074	0.537
ANN	0.033	0.182	0.119	0.585
KNN	0.033	0.181	0.097	0.591
GRNN	0.032	0.178	0.105	0.601
Boosted Trees	0.029	0.169	0.102	0.642
XGBoost	0.026	0.161	0.084	0.676
Random Forest	0.025	0.157	0.078	0.690
min:	0.025	0.157	0.074	-1.184
max:	0.173	0.376	0.178	0.690

Прикажаните резултати во Табела 11 покажуваат јасна рангираност на моделите според нивната предиктивна точност, при што вредностите на MSE, RMSE, MAE и R^2 се конзистентни индикатори за перформансите.

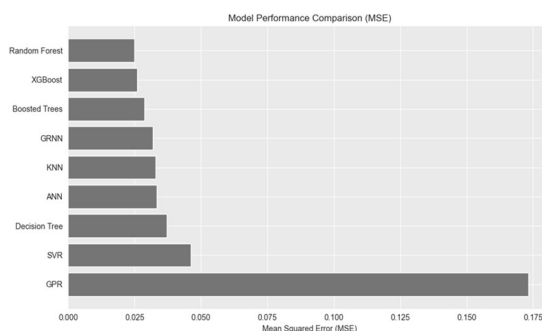
Најдобри резултати се постигнати со Random Forest (MSE = 0.025, RMSE = 0.157, R^2 = 0.690), по што следуваат XGBoost и Boosted Trees, кои покажуваат многу слични и конкурентни перформанси. Ова јасно укажува дека ансамбл-моделите базирани на одлучувачки дрвја се најсоодветни за моделирање на анализираниот систем, поради нивната способност да ги опфатат сложените нелинеарни односи и интеракции помеѓу променливите.

Моделите GRNN, KNN и ANN покажуваат умерени резултати, при што GRNN постигнува нешто подобра вредност на MSE, но со понизок R^2 во однос на ансамбл-моделите. Ова укажува дека овие методи делумно го опишуваат системот, но не ја достигнуваат стабилноста и точноста на водечките модели.

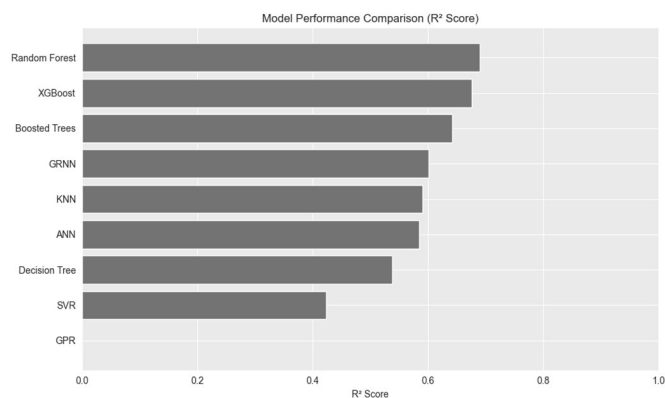
Decision Tree моделот покажува релативно ниска MAE (0.074), но повисоки MSE и RMSE, што укажува на присуство на поголеми поединечни грешки, односно нестабилност во предвидувањата.

SVR има значително пониски перформанси (R^2 = 0.423), што укажува на ограничена способност за моделирање на сложените нелинеарни зависимости во податоците.

Најслаби резултати се забележуваат кај Gaussian Process Regression (GPR), кој има негативна вредност на R^2 (-1.184), што укажува дека моделот не само што не ја објаснува варијансата на податоците, туку е полош и од едноставна средна вредност како предиктор.



Прилог 12. Приказ на MSE за секој обучен модел анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

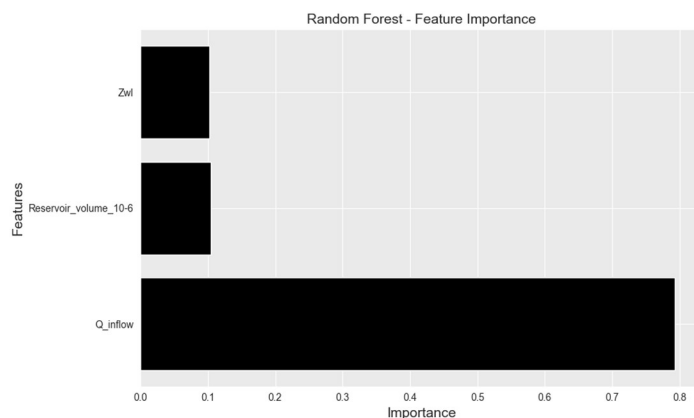


Прилог 13. Приказ на R^2 за секој обучен модел анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

За моделот обучуван со RF, DamResML кодот е програмиран да прикаже график на важност на варијаблите (анг. „Feature importance“) во моделот.

Важноста на варијаблите во моделот покажува колкаво е влијанието на предиктор варијаблите во предвидувањето на целната варијабла.

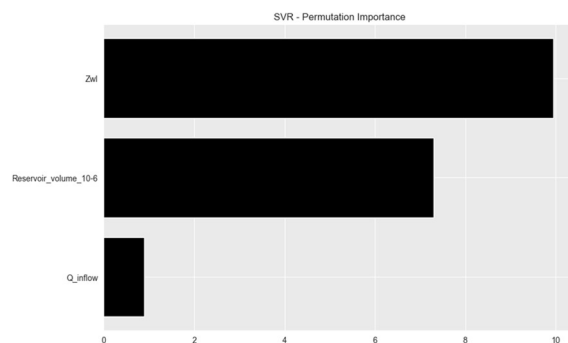
Графикот со важноста на варијаблите за моделот обучуван со RF е даден во Прилог 14, каде може да се воочи дека најголемо влијание во моделот има варијаблата „Q_inflow“ која се однесува на дотекот во акумулацијата со вредност 0.8, додека варијаблите „Reservoir_Volume_10-6“ и „Zwl“ имаат влијание од 0.1, соодветно. Доминанцијата на варијаблата „Q_inflow“ покажува дека истата има значително влијание врз предвидувањето на целната варијабла. Влијанието врз предвидувањето на целната варијабла од страна на останатите две предиктор варијабли е значително помало. Ова е очекуван резултат поради физичката поврзаност на целната варијабла – истек од акумулацијата, со дотекувањето во акумулацијата „Q_inflow“. Корелацијата е очекувана поради концепцијата на функционирање на хидросистемот „Злетовица“ – сите регулирани истекувања се ослободуваат преку темелниот испуст на браната, а бидејќи во акумулацијата се одржува релативно високо ниво, истекувањето низ темелниот испуст е тесно поврзано со дотекот во акумулацијата.



Прилог 14. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

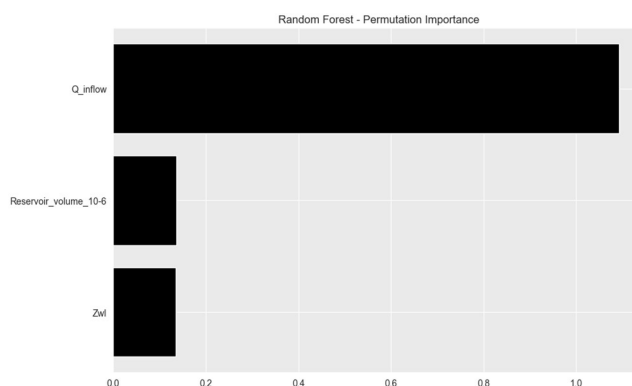
Дополнително, во DamResML кодот програмирано е извлекувањето на пермутациската важност (анг. „Permutation importance“) на моделите обучувани со SVR, RF и DT. Пермутациската важност претставува метод за оценување на значењето на предиктор варијаблите врз основа на нивното влијание врз перформансите на моделот. Основната идеја се состои во тоа што вредностите на една променлива случајно се мешаат (пермутираат), со што се нарушува нејзината врска со целната варијабла. Потоа се мери промената во точноста на моделот. Доколку пермутацијата на одредена променлива доведе до значително зголемување на грешката (на пример RMSE), тоа укажува дека моделот во голема мера се потпира на таа променлива. Напротив, ако промената е мала, тогаш нејзиното влијание е ограничено.

Овој пристап овозможува подиректна и пообјективна проценка на важноста на променливите, бидејќи се базира на реалното влијание врз предиктивната способност на моделот, за разлика од внатрешните мерки на важност кај одредени алгоритми.



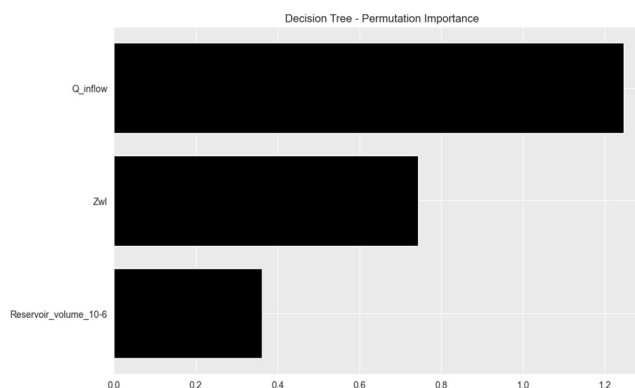
Прилог 15. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Во Прилог 15 прикажан е график на пермутациска важност на предиктор варијаблите за моделот обучуван со SVR. Може да се воочи дека варијаблата „Zwl“ – ниво на вода во акумулација, има доминантна важност во споредба со останатите две варијабли „Reservoir_volume_10-6“ и „Q_inflow“. За разлика од моделот обучуван со SVR, моделите обучувани со RF и DT покажуваат висока пермутациска важност на параметарот „Q_inflow“ како доминантен, а останатите - „Zwl“ и „Reservoir_volume_10-6“ се инфериорни (Прилог 16 и Прилог 17).



Прилог 16. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

Разликите во пермутациската важност може да се објаснат преку различните математички врски врз основа на кои различните алгоритми стекнуваат знаење. Оттука, секој алгоритам различно ги вреднува предиктор варијаблите и нивната важност во процесот на обучување.

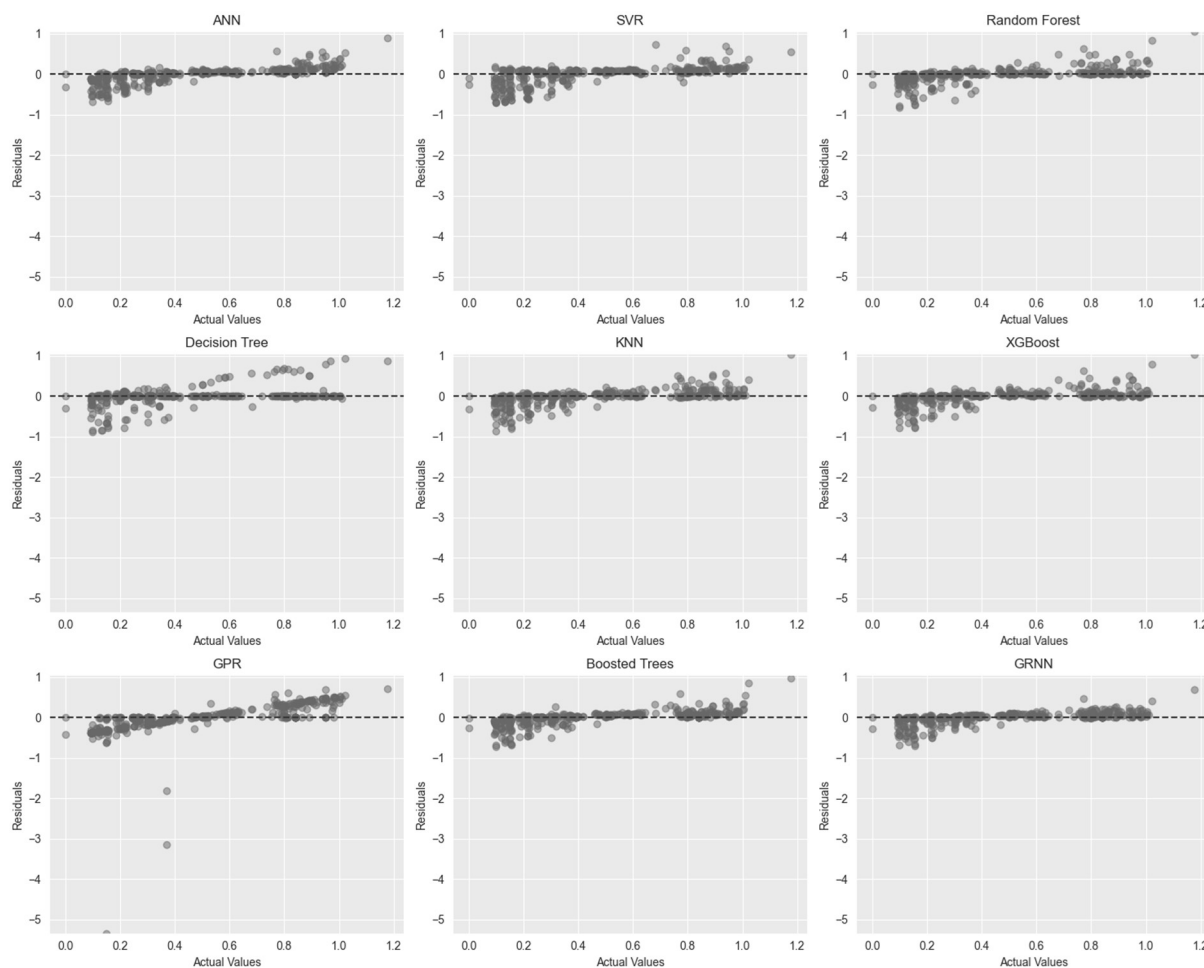


Прилог 17. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT анализа на случај $Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$.

За секој обучуван модел, во DamResML кодот програмирано е да се даде график на резидуали

(остатоци) – грешки во предвидувањата од секој модел. Сумарен график кој ги опфаќа сите 9 модели е прикажан во Прилог 18. Графиците со резидуали покажуваат дека, иако повеќето модели генерираат резидуали кои се приближно центрирани околу нула, присутен е јасен систематски тренд. Кај пониските вредности на целната варијабла доминираат негативни резидуали, додека кај повисоките вредности се забележуваат позитивни резидуали, што укажува на систематско преценување на ниските и потценување на високите вредности.

Ова однесување посочува на присуство на нелинеарности и сложени соодности помеѓу варијаблите кои не се целосно опфатени од моделите. Најдобри резултати покажуваат ансамбл-моделите (Random Forest, XGBoost и Boosted Trees), кои имаат најмала дисперзија на резидуалите, додека Gaussian Process моделот покажува значителна нестабилност и екстремни отстапувања. Дополнително, присутната варијација во ширината на резидуалите укажува на хетероскедастичност, односно зависност на грешката од вредноста на целната променлива.



Прилог 18. Графици со резидуали за секој обучуван модел.

6.1.5. Предвидување на целната варијабла

Целна варијабла во оваа студија на случај е истекувањето низ темелниот испуст „Q_discharge“. Со помош на обучени модели од машинско учење, се бара предвидување на целната варијабла со познавање на предиктор варијаблите:

- 1) Ниво на вода во акумулација „Zw1“ во [mnv] прикажан во Прилог 9,
- 2) Волумен на вода во акумулацијата „Reservoir_volume_10-6“ во [10^6m^3] прикажан во Прилог 8,
- и
- 3) Дотек во акумулацијата „Q_inflow“ во [m^3/s] прикажан во Прилог 7.

Обучени се вкупно девет модели со машинско учење базирани на следните алгоритми: (1)

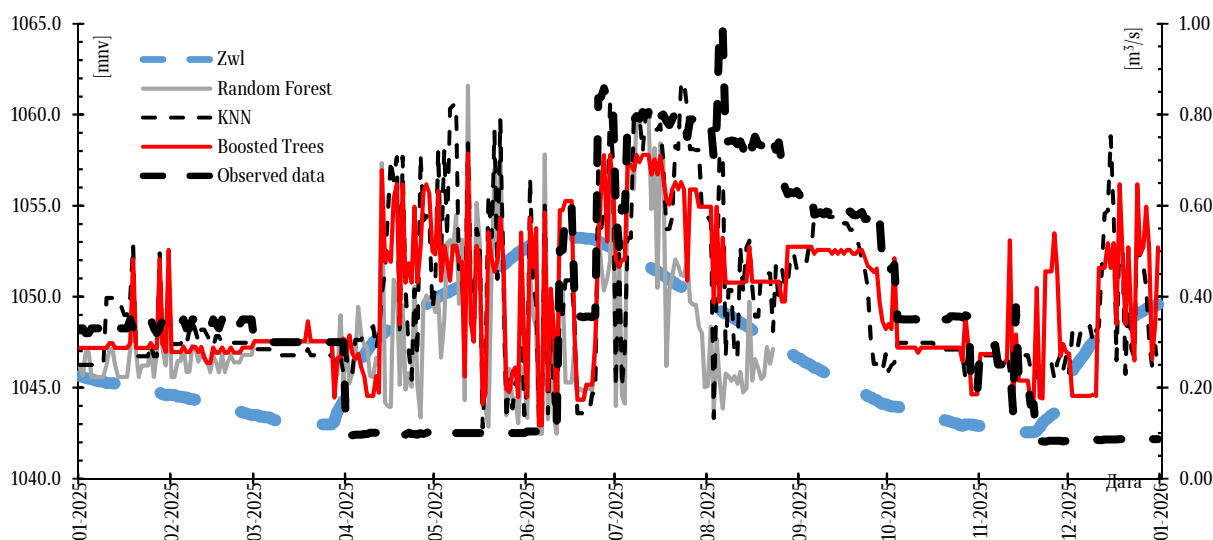
Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Перцепсија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Процесот на предвидување се изработува за 365 вредности на предиктор варијаблите, кои временски се однесуваат на периодот од 1^{ви} јануари 2025 година до 31 декември 2025 година. За овој период, позната е вредноста на целната варијабла, но истата се исклучува од процесот на предвидување, а понатаму се прави компарација на предвидувањата со она што е пресметано со моделите. Предиктор варијаблите се внесуваат во DamResML кодот преку датотеката „predictor_variable_for_prediction.xls“ и истата се повикува во ќелијата „Prediction“ по завршување на процесот на обучување на деветте модели.

По спроведениот процес на предвидување на целната варијабла од сите обучувани модели, предвидувањата од ANN моделот отстапуваат најмалку $0.002 \text{ m}^3/\text{s}$ и најмногу $0.382 \text{ m}^3/\text{s}$ од реалната вредност на целната варијабла. Кај SVR, овие вредности се $0.0000 \text{ m}^3/\text{s}$ и $0.677 \text{ m}^3/\text{s}$, респективно. Кај RF, минималното отстапување од реалната вредност е $0.0000 \text{ m}^3/\text{s}$, а максималното - $0.534 \text{ m}^3/\text{s}$. Во моделот DT, минималното и максималното отстапување е $0.009 \text{ m}^3/\text{s}$ и $0.620 \text{ m}^3/\text{s}$. Кај KNN тренираниот модел, минималната и максималната разлика во предвидувањето на целната варијабла е $0.003 \text{ m}^3/\text{s}$ и $0.542 \text{ m}^3/\text{s}$. Кај XGBoost – $0.001 \text{ m}^3/\text{s}$ и $0.549 \text{ m}^3/\text{s}$ соодветно. Кај GPR моделот овие вредности се $0.002 \text{ m}^3/\text{s}$ и $0.549 \text{ m}^3/\text{s}$. Кај GPR моделот, разликите се $0.002 \text{ m}^3/\text{s}$ и $0.433 \text{ m}^3/\text{s}$. Кај BT моделот, овие вредности се $0.001 \text{ m}^3/\text{s}$ и $0.384 \text{ m}^3/\text{s}$. Во последниот GRNN модел, разликите се движат од $0.001 \text{ m}^3/\text{s}$ до $0.478 \text{ m}^3/\text{s}$.

Разликите во предвидувањата споредени со реалните вредности на целната варијабла може да се согледаат на графиконите прикажани во Прилог 19, Прилог 20 и Прилог 21.

На Прилог 19 прикажана е споредба на предвидувањата направени со моделите Random Forest, K-Nearest Neighbour и Boosted Trees со реалните вредности на целната варијабла прикажани со графикот „Observed data“ и варијацијата на водното ниво во акумулацијата „Zwl“. Од графикот може да се воочи дека моделите се потпираат на водното ниво како предиктор варијабла и не успеваат во целост да ја предвидат целната варијабла во периодот помеѓу април, 2025 и јуни, 2025. Наместо да предвидуваат пониски вредности на целната варијабла, моделите предвидуваат зголемени испуштања кои одговараат на состојбата со нивото во акумулацијата.



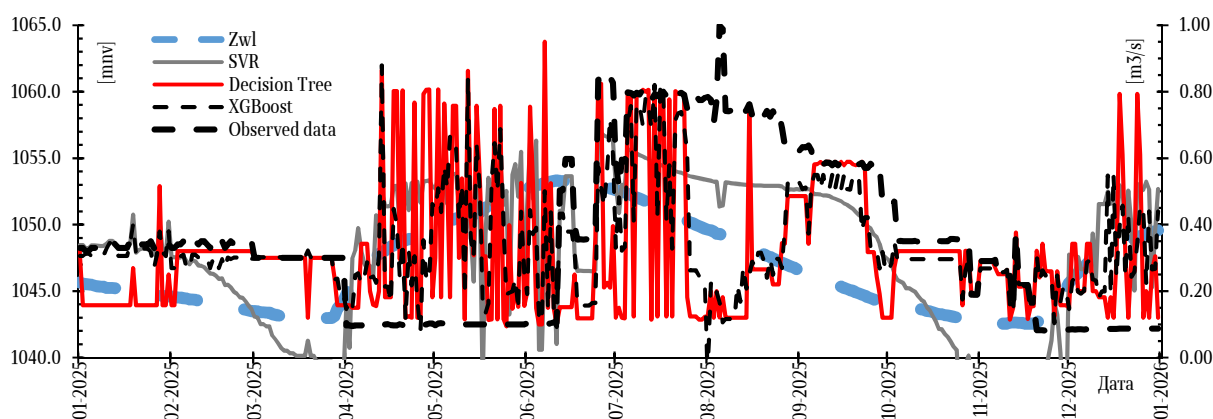
Прилог 19. Графички приказ на предвидените вредности од моделите RF, KNN и BT во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флукуацијата на водното ниво во акумулацијата „Zwl“.

Слични опсервации може да се воочат и на останатите шест модели Support Vector Regressor, Decision Tree, XGBoost, Gaussian Process Regressor, General Regression Neural Network и Artificial Neural Networks.

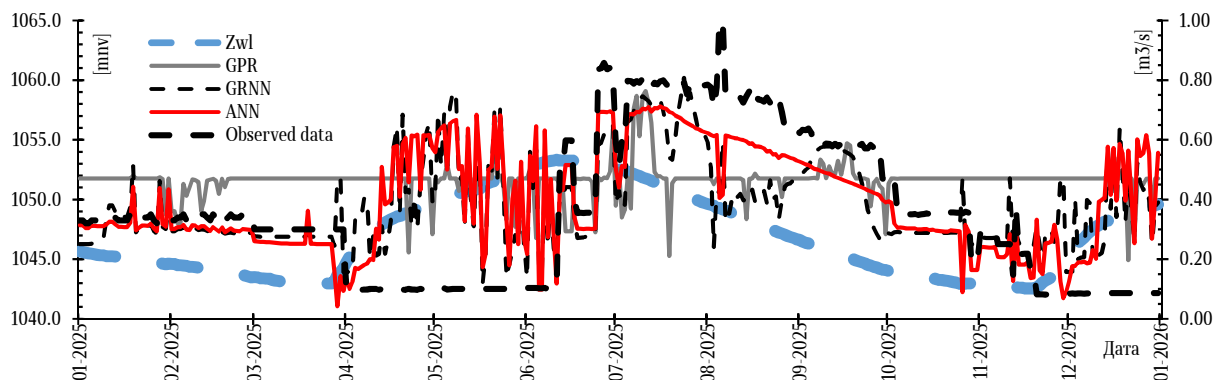
Моделот на SVR во одреден период предвидува и негативни вредности на целната варијабла, што е сосема погрешно. Тоа може да се согледа на Прилог 20 каде предвидувањата од SVR моделот опаѓаат и под нулата во периодот од март, 2025 до април, 2025, како и во периодот од ноември, 2025 до декември, 2025 година.

Моделот на GPR предвидува речиси константна вредност од $0.45 \text{ m}^3/\text{s}$ во целиот анализиран период на предвидувања, што може да се согледа во Прилог 21.

Оттука, може да се заклучи дека грешката во предвидувањата на моделите се движи од $0.000 \text{ m}^3/\text{s}$ до $0.677 \text{ m}^3/\text{s}$.



Прилог 20. Графички приказ на предвидените вредности од моделите SVR, DT и XGBoost во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флукуацијата на водното ниво во акумулацијата „Zwl“.



Прилог 21. Графички приказ на предвидените вредности од моделите GPR, GRNN и ANN во споредба со реалните вредности на целната варијабла „Q_discharge“ прикажани со графикот „Observed data“ и флукуацијата на водното ниво во акумулацијата „Zwl“.

6.1.6. Хаос оптимизација на целната варијабла

Како последен чекор од процесот на подучување на моделите и предвидување на целната варијабла, применета е хаос оптимизација со цел оптимизирање на целната варијабла. Во конкретниот случај, целната варијабла е „Q_discharge“ т.е. протокот кој се испушта низ темелниот испуст на браната „Кнежево“.

Хаос оптимизацијата се применува за максимизација на вредноста на предвидувањата на целната варијабла „Q_discharge“ од моделите RF, DT и XGBoost како модели со најдобри метрички показатели во процесот на подучување и предвидување на целната варијабла. Станува збор за вкупно 365 податоци на предиктор и таргет варијаблата.

Во самата Хаос оптимизација изоставени се ограничувањата на предиктор варијаблите. Самата функција на Хаос оптимизација содржи неколку клучни параметри кои влијаат на однесувањето, стабилноста и точноста на оптимизацијата.

Така, најпрво е дефинирана големината на популацијата преку линијата на код `pop_size = 80`, која го ограничува бројот на кандидати кои алгоритмот ги разгледува во секоја итерација. Во конкретниот случај, моделот анализира 80 различни комбинации на предиктор и целна варијабла во секоја итерација.

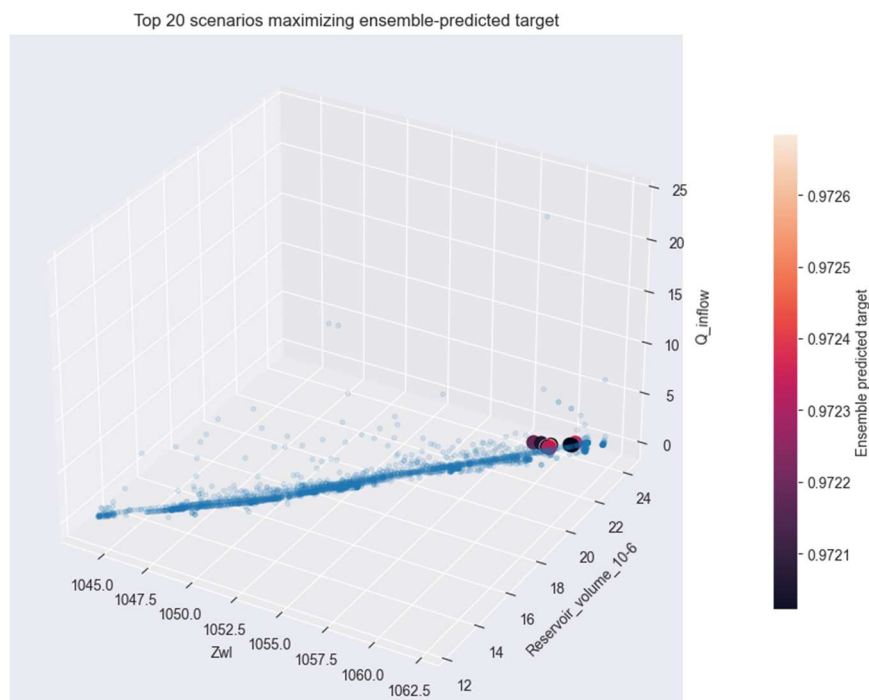
Бројот на итерации е втората големина од редот на клучните параметри во Хаос оптимизацијата. Станува збор за параметар кој во кодот е дефиниран преку линијата на код `iters = 300`, а истиот го ограничува бројот на циклуси на ажурирање на популацијата.

Во конкретниот код, алгоритмот прави 300 итерации, т.е. разгледува $80 \times 300 = 24\,000$ сценарија.

Понатаму следи дефинирањето на параметарот за ограничување на иницијализацијата на случајност, внесен како линија на код `seed = 42`. Овој параметар ја контролира почетната состојба на генераторот на случајни броеви.

Следно, дефинирани се граници на пребарување кои го ограничуваат просторот на оптимизација и спречуваат нереални решенија. Тие се дефинирани врз основа на статистичка анализа на податочниот сет, и исклучени се сите податоци кои припаѓаат во границите под 1% и над 99% од вредноста на варијаблата, т.е. се исклучуваат најмалите и најголемите вредности од податочниот сет, сè со цел пребарување на оптимумот во реален опсег без екстремни вредности.

Дефинирањето на овие параметри во напоменатите рамки овозможува билас на моделот за истражување и искористување на податоците со цел обезбедување на што е можно поголема популација, доволно итерации и фиксна вредност на иницијализацијата на случајност кои придонесуваат кон значително истражен податочен сет и стабилна конвергенција на оптимизацијата. На овој начин се обезбедува точност на решенијата.



Прилог 22. 3Д график со приказ на пресметаните оптимални вредности на целната варијабла „Q_discharge“ во функција од трите предиктор варијабли „Zwl“, „Q_inflow“ и „Reservoir_volume_10-6“.

За да стане појасно кои комбинации на влезните предиктор варијабли придонесуваат кон оптимум на целната варијабла, моделот продуцира 3D графикон со приказ на 20 оптимални

вредности на целната варијабла во XYZ координатен систем (Прилог 22). Самиот графикон ги прикажува и останатите вредности на целната варијабла од податочниот сет, но ги акцентира издвоените 20 оптимални вредности – во случајност максимални вредности на целната варијабла „Q_discharge“. Во Прилог 23 е прикажан 3Д графикон кој ги вклучува единствено оние издвоени оптимизирани 20 вредности на целната варијабла, додека на XYZ оските соодветно може да се отчитаат вредностите за предиктор варијаблите.

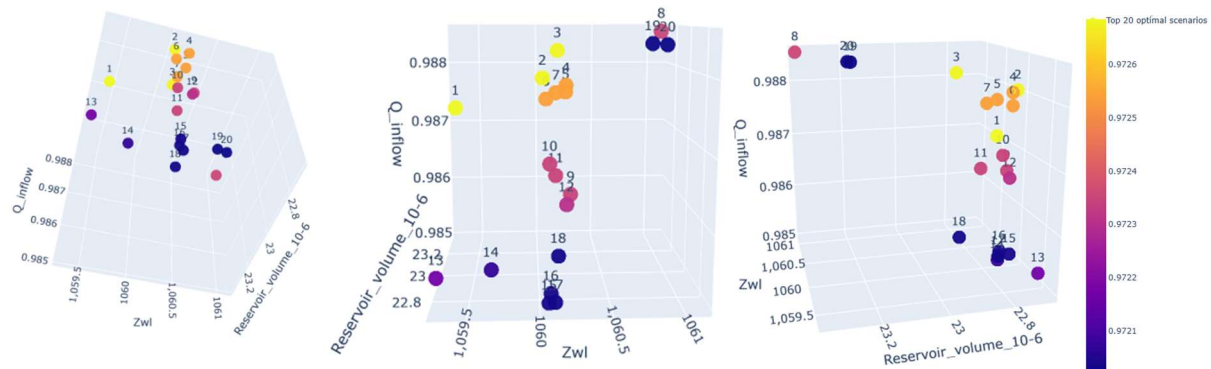
Од Прилог 23 директно може да се согледа дека поголемите вредности на целната варијабла се блиски со поголемите вредности за дотекот во акумулацијата „Q_inflow“.

Оптимумот кој Хаос оптимизацијата го пронајде се конкретниот податочен сет кој претставува комбинација на предвидени вредности на целната варијабла од различните поучени модели со машинско учење, е:

$$\text{MAX } Q_{\text{discharge}} = 0.9726 \text{ [m}^3/\text{s]}$$

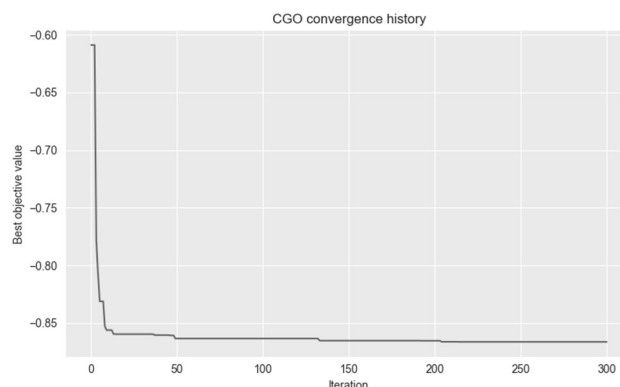
, во функција од предиктор варијаблите:

- 1) Ниво на вода во акумулација „Zwl“: 1059.418 [mnv],
- 2) Волумен на вода во акумулација „Reservoir_volume_10-6“: 22.851 [$\times 10^6 \text{ m}^3$],
- 3) Дотек на вода во акумулација „Q_inflow“: 0.9875 [m^3/s]



Прилог 23. 3Д график со приказ на 20-те издвоени вредности на целната варијабла „Q_discharge“.

За приказ на еволуцијата на најдобрата вредност на целната функција во зависност од бројот на итерации во Хаос оптимизацијата, прикажан е график на конвергентноста на моделот (Прилог 24).



Прилог 24. Преглед на ковергентноста на моделот од Хаос оптимизација низ историскиот развој на итерациите.

Во графиконот на Прилог 24, на апсисната оска е прикажан бројот на итерации, додека на ординатата е прикажана најдобрата постигната вредност на целната функција во секој чекор од оптимизацијата.

Во првите неколку итерации се забележува нагло намалување на вредноста на целната функција (од приближно -0.61 до околу -0.85). Овој стрмен пад укажува на ефикасно иницијално

истражување на просторот на решенија и способност на алгоритмот брзо да идентификува региони со значително подобри решенија, како и на добро дефинирани граници на пребарување и соодветно иницијализирана популација. Ова однесување е карактеристично за оптимизациски алгоритми кои успешно ја користат стохастичката компонента (во овој случај хаотичната логистичка мапа) за глобално пребарување.

Во средната фаза се прецизира решението (итерации 10–50), каде се забележува намалување на целната функција, но со значително помал интензитет. Се забележуваат мали, постепени подобрувања (од приближно -0.85 до -0.86), што укажуваат на транзиција од глобална експлорација кон локална експлоатација, фино подесување на веќе идентификуваниот оптимален регион, и стабилност на алгоритмот без значителни осцилации. Оваа фаза е клучна бидејќи алгоритмот ја подобрува точноста на решението без да ја наруши стабилноста.

Во завршната фаза или т.н. фаза на конвергенција (итерации 50–300) графикот покажува речиси хоризонтална линија (по 50та итерација), со минимални или занемарливи подобрувања (околу -0.865). Ова јасно укажува дека алгоритмот достигнал стабилно решение, не се пронаоѓаат подобри комбинации на влезни параметри, постигната е конвергенција кон глобален или квази-глобален оптимум. Недостигот на осцилации и стабилноста на кривата во оваа фаза укажуваат на добра дефинираност на функцијата на цел, отсуство на значителни локални минимални кои би го нарушиле процесот, и сигурност во добиеното оптимално решение.

Со оглед дека целната функција е дефинирана како негативна вредност од предвидувањето (со додадени казни), намалувањето на вредноста на графикот значи зголемување на предвидената вредност на целната променлива и подобрување на оптималноста на решението. Оттука, финалната вредност (-0.865) претставува најдобрата комбинација на влезни параметри која го максимизира излезот, при почитување на сите ограничувања.

6.2. Анализа на случај 1.2 – $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$

6.2.1. Подготовка на податоците

Во предметната анализа на случај ќе се разгледува генерирање на моќност низ прибранската хидроцентрала инсталирана низводно од брана „Кнежево“ именувана како „ХЕЦ Б“, која е ставена во употреба во 2025 година. Хидроцентралата е со инсталирана моќност од $P_{ins} = 200 \text{ kW}$ и инсталиран проток $Q_{ins} = 300 \text{ l/s}$. Него падот на кој хидроцентралата ќе работи се движи во границите $H_n = (39,10 - 73,60) \text{ m}$ (Aleksova Dosevska, 2025).

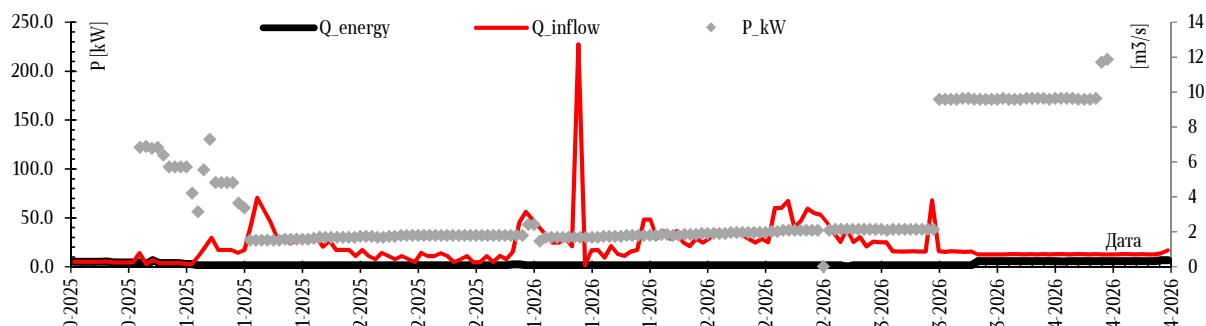
Бидејќи за самата работа на хидроцентралата имаме податоци само од периодот кога истата била во функција во дел од 2025 и дел од 2026 година (168 податок регистриран од страна на операторот), истите ќе се користат за процесот на обучување на моделите. За прогноза на производството на енергија ќе се користат податоци на предиктор варијаблите од периодот од 2018 до 2024 година.

Како целна варијабла во анализите ќе се разгледува генерираната моќност во хидроцентралата „ P_{kW} “, додека предиктор варијаблите ќе бидат:

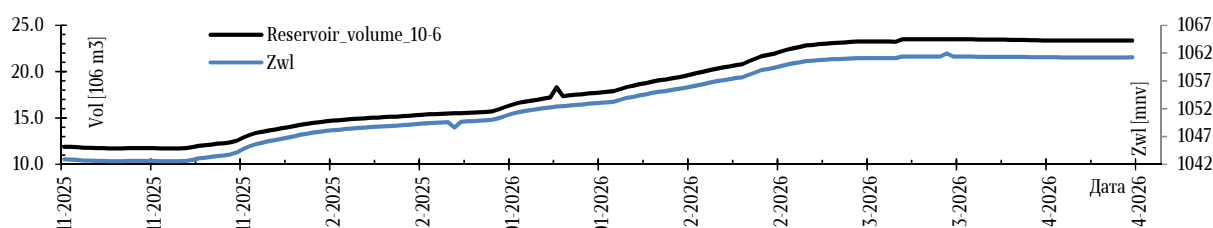
- 1) Нивото на вода во акумулацијата, „ Z_{wl} “ во [mnv],
- 2) Волуменот на вода акумулиран во акумулацијата, „ $Reservoir_volume_{10-6}$ “ во [10^6 m^3],
- 3) Дотек во акумулацијата, „ Q_{inflow} “ во [m^3/s], и
- 4) Истек низ темелен испуст кој ќе се изедначи со протокот кој енергетски ќе се обработува, „ $Q_{energy} = Q_{discharge}$ “ во [m^3/s].

За обучување на моделите, како што беше напоменато и претходно, ќе се користат мал број на податоци кои се регистрирани во периодот кога хидроцентралата е ставена во функција во 2025 година. Станува збор за 168 податок за целната варијабла „ P_{kW} “, и предиктор варијаблите „ Z_{wl} “, „ $Reservoir_volume_{10-6}$ “, „ Q_{inflow} “ и „ Q_{energy} “.

Податоците применети за предиктор варијаблите и целната варијабла за периодот на обучување на моделите се прикажани во форма на графикони во Прилог 25 и Прилог 26.

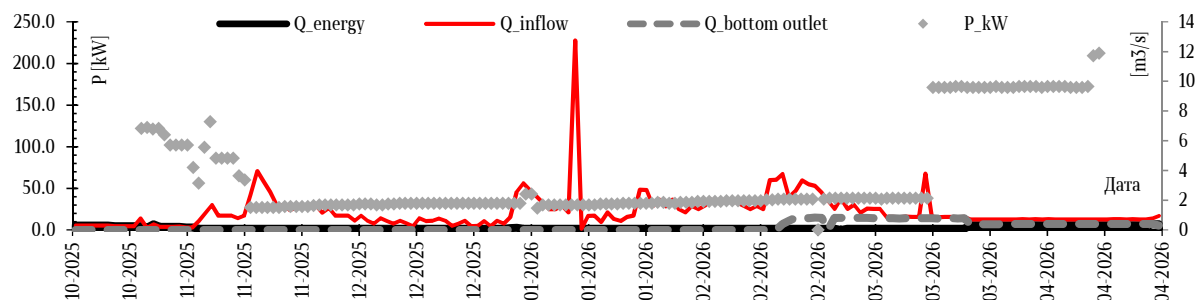


Прилог 25. График за варијаблите „P_kW“ - генерирана моќност, „Q_inflow“ - дотекот во акумулацијата и „Q_energy“ - енергетски обработениот проток за периодот од јануари, 2025 година до април, 2026 година, за приказ на историската низа на податоци која е земена предвид во процесот на обучување на моделите за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.



Прилог 26. График за варијаблите „ResVol“ - волумен во акумулација и „Zwl“ - ниво на вода во акумулација за периодот од јануари, 2025 година до април, 2026 година, за приказ на историската низа на податоци која е земена предвид во процесот на обучување на моделите за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

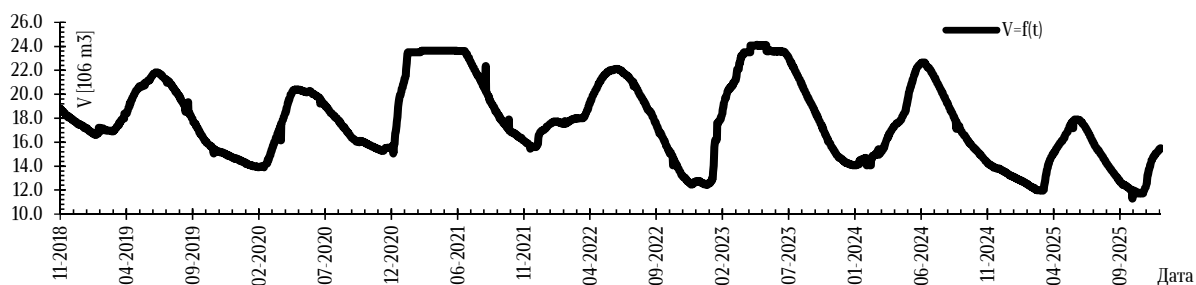
Тука треба да се напомене дека протокот кој енергетски се обработува и протокот кој се испушта низ темелниот испуст, во периодот на стапување на хидроцентралата во употреба, се различни протоци. Имено, хидроцентралата е поставена на главниот доведен цевковод за наводнување, а темелниот испуст е засебен цевковод. Во период на маловодие, протокот кој се испушта како биолошки минимум низводно од браната во целост енергетски се обработува. Спротивно, во случај на зголемени дотекувања, темелниот испуст дополнително се отвора, и покрај хидроцентралата операторот испушта одредени количини на вода низ темелниот испуст. Хидроцентралата е пуштена во функција релативно краток временски период кој воедно е и тест период за нејзиното работење. Поради тоа, не може да се донесе конкретен заклучок за донесувањето на одлука на операторот за дополнително отворање на темелниот испуст дури и во случај кога дотекувањата во хидроцентралата се во рамки на инсталираниот капацитет на хидроцентралата. На графиконот прикажан во Прилог 27 може да се забележи дека во периодот од 15.02.2026 до 20.03.2026 паралелно со производството на електрична енергија низ хидроцентралата операторот испушта и протек низ темелниот испуст кој енергетски не се обработува.



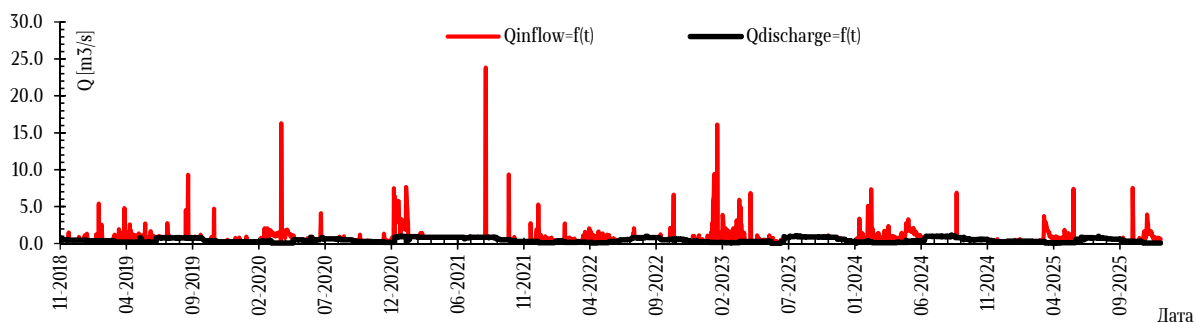
Прилог 27. График со приказ на генерирана моќност во хидроцентралата „P_kW“, дотек во акумулацијата „Q_inflow“, енергетски обработен проток „Q_energy“ и протек кој се испушта низ темелниот испуст „Q_bottom outlet“, регистрирани во периодот на работа на хидроцентралата од 2025 до 2026 година за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

За прогноза на производството на моќност ќе се користат податоци за предиктор варијаблите од периодот од 01.11.2018 година до 31.12.2024 година, или вкупно 2241 податок за секоја предиктор варијабла. Целта е да се согледа колкаво е производството на моќност во хидроцентралата во случај истата да работи во наредните години со идентичен режим како разгледуваните 5 години, со примена на модели од машинско учење.

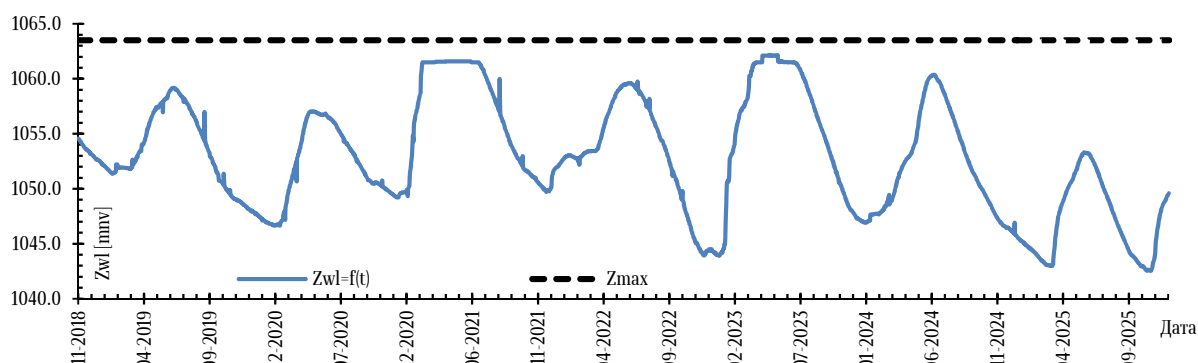
Вредноста на предиктор варијаблите применети во фазата на прогноза со обучените модели со машинско учење се прикажани во форма на графикони на Прилозите Прилог 28, Прилог 29 и Прилог 30.



Прилог 28. График со приказ на предиктор варијаблата „Reservoir_volume_10-6“ кој ќе се применува во фазата на прогноза од обучените модели со машинско учење за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.



Прилог 29. График со приказ на предиктор варијаблите „Q_inflow“ и „Q_discharge“ (или „Q_energy“) кои ќе се применуваат во фазата на прогноза од обучените модели со машинско учење за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.



Прилог 30. График со приказ на предиктор варијаблата „Zwl“ која ќе се применува во фаза на прогноза од обучените модели со машинско учење за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

6.2.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со

потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Целната варијабла е користена во нејзината оригинална форма, без примена на нормализација или стандардизација. Од друга страна, предиктор варијаблите беа стандардизирани во рамки на секој сет од петкратната валидација, но само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 12.

Табела 12. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за анализа на случај $P_{kW} = f(Z_{w1}, ResVol, Q_{inflow}, Q_{energy})$.

5-fold CV					
No.	Algorithm	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	Boosted Trees	49.013	5.948	2.044	0.981
2	Artificial Neural Networks	41.731	6.052	2.918	0.984
3	K-Nearest Neighbors	51.456	6.932	2.716	0.981
4	Random Forest	60.237	7.284	2.496	0.977
5	Decision Tree	68.084	7.528	2.695	0.975
6	GRNN	64.397	7.634	3.254	0.976
7	XGBoost	67.516	8.003	3.433	0.977
8	Support Vector Regressor	70.379	8.194	3.243	0.975
9	Gaussian Process	124.184	8.958	3.090	0.952
	max value:	124.184	8.958	3.433	0.984
	min value:	41.731	5.948	2.044	0.952

Резултатите добиени со примена на петкратната вкрстена валидација (анг. „5-fold cross validation“) овозможуваат објективна и доверлива споредба на перформансите на разгледуваните алгоритми за машинско учење, бидејќи секој модел е обучуван и тестиран на повеќе различни поделби од податочниот сет. Притоа, оценувањето е извршено преку четири клучни метрики: MSE (анг. „Mean Squared Error“), RMSE (анг. „Root Mean Squared Error“), MAE (анг. „Mean Absolute Error“) и R^2 (анг. „Coefficient of Determination“).

Во однос на метриката MSE, која ја претставува средната квадратична грешка и дополнително ги казнува поголемите отстапувања, најдобар резултат покажува моделот Artificial Neural Networks (ANN) со вредност 41.731, што укажува на најмало просечно квадратно отстапување меѓу реалните и предвидените вредности. Ова значи дека ANN моделот најуспешно ги минимизира поголемите грешки во предвидувањето. Следат Boosted Trees со 49.013 и K-Nearest Neighbors (KNN) со 51.456, што исто така претставува многу добар резултат. Највисока вредност на MSE има Gaussian Process со 124.184, што укажува на значително послаба стабилност и поголеми грешки во споредба со останатите модели.

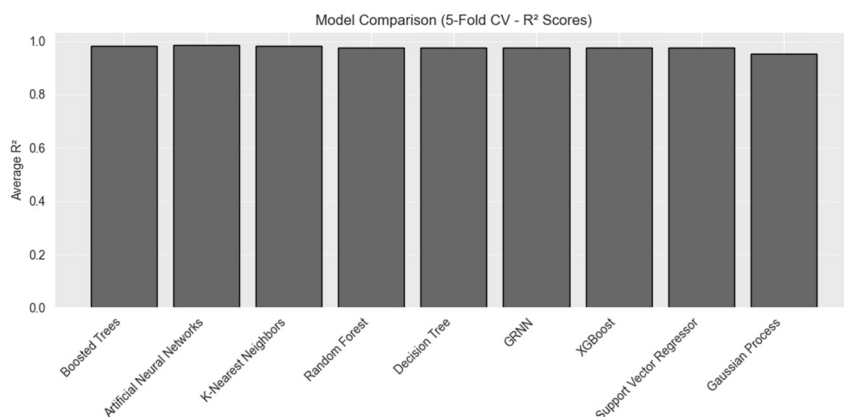
Кај RMSE, кој е корен од MSE и има иста мерна единица како целната променлива, најдобар резултат повторно има Boosted Trees со 5.948, што значи дека просечното отстапување на предвидувањата е најмало кај овој модел. Веднаш потоа следат ANN со 6.052 и KNN со 6.932. Вредностите под 7 претставуваат многу добра предиктивна точност. Најслаб резултат повторно има Gaussian Process со 8.958, што покажува дека овој модел има најголемо просечно растојание меѓу реалните и предвидените вредности.

Метриката MAE, која ја мери просечната апсолутна грешка без дополнително нагласување на екстремните грешки, покажува дека најпрецизен модел е Boosted Trees со вредност 2.044. Овој резултат укажува дека просечното апсолутно отстапување кај овој модел е најниско. Следуваат

Random Forest со 2.496, Decision Tree со 2.695 и KNN со 2.716. Иако ANN има најнизок MSE, неговиот MAE = 2.918 покажува дека моделот има нешто поголеми просечни отстапувања, но ретко прави големи грешки. Највисок MAE има XGBoost со 3.433, што значи дека во просек овој модел има најголемо апсолутно отстапување.

Во однос на R^2 коефициентот, кој го покажува процентот на објаснета варијанса во податоците, најдобар резултат има Artificial Neural Networks со 0.984, што значи дека моделот успешно објаснува околу 98.4% од варијацијата на целната променлива. Ова е одличен резултат и потврдува висока способност за генерализација. Следуваат Boosted Trees и KNN со 0.981, како и Random Forest и XGBoost со 0.977. Најнизок R^2 повторно има Gaussian Process со 0.952, што иако е понизок од останатите, сè уште претставува релативно добра точност. Графичкиот приказ на R^2 е даден во Прилог 31.

Генерално, може да се заклучи дека сите модели покажуваат високо ниво на успешност, бидејќи сите R^2 вредности се над 0.95, што е индикатор за квалитетни предикции. Сепак, како најуспешни модели се издвојуваат Artificial Neural Networks и Boosted Trees. ANN има најдобри вредности за MSE и R^2 , додека Boosted Trees покажува најдобри резултати за RMSE и MAE. Ова укажува дека двата модели се најсоодветни за конкретниот проблем, при што изборот меѓу нив зависи од тоа дали приоритет е минимизирање на големи грешки (ANN) или најмало просечно отстапување (Boosted Trees).



Прилог 31. Графички приказ на вредноста на R^2 за секој анализиран алгоритам на сетот податоци за обучување со петкратна вкрстена валидација за анализа на случај $P_{kW} = f(Z_{w1}, ResVol, Q_{inflow}, Q_{energy})$.

6.2.3. Обучување на моделите

Во процесот на обучување на моделите, по извршената петкратна валидација, донесена е одлука сите девет алгоритми да се користат во обучување на девет различни модели. Накратко секој модел е опишан како што следи, заедно во хиперпараметрите внесени во кодот во Python.

6.2.3.1. Модел на Вештачки невронски мрежи (ANN)

Прв обучен модел е моделот на Вештачки невронски мрежи или накратко ANN (анг. „Artificial Neural Network“). Во конкретниот DamResML код во кој е внесен алгоритмот за Вештачки невронски мрежи, во однос на параметрите, внесени се следните вредности:

- 1) Број на скриени слоеви n_{layer} и бројот на неврони во секој слој $n_{neurons}$ се внесени како `hidden_layer_sizes=(100,100)`. Со ова, внесени се 2 скриени слоеви со 100 неврони во секој од нив.
- 2) Типот на активациска функција е усвоен како `solver="adam"` што значи дека се применува Оптимизаторот Адам (анг. „Adam optimizer“) како активациска функција.
- (3) Брзината на учење l_{rate} е усвоена како `learning_rate_init=0.001`.

Процесот на обучување траеше 518ms, со просечна вредност на MSE = 152.00, RMSE = 11.418, MAE = 6.638 и $R^2 = 0.948$.

6.2.3.2. Модел на Регресија со потпомогнати вектори (SVR)

Вториот обучен модел е моделот на Регресија со потпомогнати вектори или накратко SVR (анг. „Support Vector Regressor“).

Моделот е дефиниран со следните параметри:

1) `kernel = "rbf"` - Се користи радијална базисна функција (Radial Basis Function), која овозможува моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) `C = 100` - Регуларизациски параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на `C` значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување (анг. „overfitting“).

3) `epsilon = 0.1` - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

За евалуација на моделот е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Средната вредност на метричките показатели по обучувањето на моделот изнесуваат: Avg MSE = 57.11, Avg RMSE = 7.306, Avg MAE = 2.265, Avg R^2 = 0.979. Процесот на обучување траеше 70ms.

6.2.3.3. Модел на Случајна шума (RF)

Третиот обучен модел е моделот на Random Forest регресија. За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators (50, 100, 150)` - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth (None, 10, 20)` - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split (2, 5, 10)` - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features ("sqrt", "log2", None)` - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијаблите.

За добивање на објективна проценка на перформансите применета вгнездена n -кратна валидација, и тоа:

- Внатрешна трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку т.н. „GridSearchCV“.
- Надворешна петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Овој пристап обезбедува непристрасна проценка на перформансите и ја намалува веројатноста за пренагласено прилагодување на моделот кон податоците. Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R^2 прикажани во Табела 13, што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување.

Табела 13. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај

$$P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy}).$$

Fold	max_depth	max_features	min_samples_split	n_estimators
5	None	None	2	150
4	None	None	5	150
3	None	None	2	50
2	None	None	2	150
1	None	None	5	150

Финалното обучување е направено согласно хиперпараметрите од петтиот сет од петкратната надворешна валидација (Табела 13), со вредности: 'max_depth': None, 'max_features': None, 'min_samples_split': 2, 'n_estimators': 50. Процесот на обучување траеше 26s и 559ms.

6.2.3.4. Модел Одлучувачко дрво (Decision Tree)

Четвртиот обучувачки модел е моделот на Decision Tree Regressor. Во овој случај, моделот е дефиниран со основните параметри (со random_state = 42 за репродуктивност), без дополнителна регуларизација, што значи дека дрвото може да расте до максимална сложеност и да ги следи податоците детално.

За евалуација на хиперпараметрите е применета петкратна вкрстена валидација, при што податоците се делат на пет сетови, а моделот се обучува и тестира пет пати со различни комбинации. Овој пристап овозможува стабилна и доверлива проценка на перформансите.

Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста на моделот. Бидејќи Decision Tree не бара скалирање на влезните податоци, моделот се применува директно врз оригиналните вредности на секоја варијабла. Вредностите за метриците се прикажани во Табела 14.

Табела 14. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај $P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Fold	MSE	RMSE	MAE	R2
1	96.125	9.804	4.074	0.962
2	80.941	8.997	2.882	0.969
3	91.353	9.558	2.941	0.963
4	71.303	8.444	3.182	0.980
5	0.697	0.835	0.394	1.000

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 68.083, Avg RMSE: 7.528, Avg MAE: 2.694, Avg R²: 0.975. Процесот на обучување траеше 24ms.

6.2.3.5. Модел K – најблизок сосед (K – nearest neighbor)

Како петти трениран модел е моделот со K-Nearest Neighbors (KNN) регресија, кој претставува метод базиран на сличност помеѓу податоците.

Моделот е дефиниран со параметар n_neighbors = 5, што значи дека за секое предвидување се земаат предвид 5 најблиски примероци од податоците за обучување.

За проценка на перформансите е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста на моделот. Конкретните вредности за моделот се дадени во Табела 15. Треба да се напомене дека KNN не гради експлицитен модел (lazy learning), туку ги користи сите тренинг податоци при предвидување, што може да влијае на пресметковната ефикасност кај поголеми податочни множества.

Табела 15. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај $P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Fold	MSE	RMSE	MAE	R2
1	51.299	7.162	3.106	0.980
2	29.465	5.428	2.394	0.989

Fold	MSE	RMSE	MAE	R2
3	108.434	10.413	3.594	0.956
4	30.994	5.567	2.570	0.991
5	37.091	6.090	1.915	0.990

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 0.0328, Avg RMSE: 0.1809, Avg MAE: 0.0972, Avg R²: 0.5906. Процесот на обучување траеше 233ms.

6.2.3.6. Модел со XGBoost (XGBoost Regressor)

Шестиот трениран модел е моделот на XGBoost Regressor. Моделот е дефиниран со следните клучни параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но може да ја зголеми сложеноста на моделот.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое ново дрво. Помали вредности водат до постабилно учење, но бараат повеќе дрвја.

3) `objective = "reg:squarederror"` - Ја дефинира функцијата на загуба, при што се минимизира средноквадратната грешка (MSE).

4) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи XGBoost е базиран на дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета стандардизација на карактеристиките.

За евалуација на моделот применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста и објаснувачката моќ на моделот. Пресметаните вредности на метриките од моделот се прикажани во Табела 16. Просечните вредности на метричките показатели изнесуваат: Avg MSE: 80.205, Avg RMSE: 8.461, Avg MAE: 2.754, Avg R²: 0.969. Процесот на обучување траеше 233ms.

Табела 16. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Fold	MSE	RMSE	MAE	R2
1	73.637	8.581	3.383	0.971
2	100.834	10.042	3.026	0.962
3	168.880	12.995	3.688	0.932
4	23.379	4.835	1.909	0.994
5	34.296	5.856	1.766	0.990

6.2.3.7. Модел со Гаусов Процесен Регресор (GPR)

Седмиот модел за обучување се базира на алгоритмот Гаусов Процесен Регресор (анг. „Gaussian Process Regression“). Во моделот е користен комбиниран кернел - ConstantKernel × RBF (Radial Basis Function). Овој кернел овозможува моделирање на мазни (smooth) нелинеарни односи помеѓу променливите, при што RBF ја контролира „должината на влијание“ (length scale), а ConstantKernel ја скалира вредноста на податоците.

Клучни параметри на моделот се:

1) `normalize_y = True` - Ја нормализира целната варијабла, што значително ја подобрува стабилноста на моделот.

2) `n_restarts_optimizer = 5` - Овозможува повеќекратно оптимизирање на параметрите на кернел функцијата, со цел да се избегне локален минимум и да се добие подобро вклопување.

Вредностите за метричките показатели од процесот на петкратна вкрстена валидација се прикажани во Табела 17.

Табела 17. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Fold	MSE	RMSE	MAE	R2
1	908.038	30.134	26.681	0.644
2	758.392	27.539	23.386	0.713
3	1382.147	37.177	27.742	0.445
4	999.030	31.607	25.537	0.724
5	1150.850	33.924	21.980	0.679

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 1039.69, Avg RMSE: 32.076, Avg MAE: 25.065, Avg R²: 0.641. Процесот на обучување траеше 698ms.

6.2.3.8. Модел со Поткрени Репресиони Дрвја (BT)

Осмиот модел за обучување се базира на алгоритмот на Boosted Regression Trees, односно Gradient Boosting Regressor, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја.

Моделот е дефиниран со следните параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Погolem број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на моделот е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 18. Gradient Boosting, како бустинг метод, е особено ефикасен во моделирање на сложени нелинеарни односи и често постигнува висока предиктивна точност.

Табела 18. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Fold	MSE	RMSE	MAE	R2
1	32.309	5.684	2.561	0.987
2	89.687	9.470	2.421	0.966
3	111.967	10.581	3.434	0.955
4	10.525	3.244	1.332	0.997
5	0.577	0.759	0.473	1.000

Просечните вредности на метричките показатели, се: Avg MSE: 49.0129, Avg RMSE: 5.947, Avg MAE: 2.044, Avg R²: 0.981. Потребното време за обучување беше 101ms.

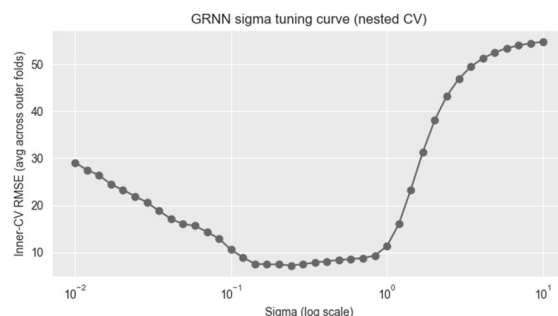
6.2.3.9. Модел со Генерализирана Репресиона Невронска Мрежа (GRNN)

Последниот модел е обучуван на алгоритмот на Генерализирана Репресиона Невронска Мрежа (анг. „General Regression Neural Network“).

Во кодот е применета вгнездена трикратна валидација за избор на најдобра вредност на параметарот σ – сигма. Се тестираат вредности во рангот $\sigma = (0.01 - 10)$. Критериумот е да се избере вредност за σ што ќе резултира со минимален RMSE, а во конкретниот случај графикот на варијација на RMSE во зависност од параметарот σ по трикратната вкрстена валидација е прикажан во Прилог 32. Понатаму, со изборот на соодветна вредност за параметарот σ и дополнително се прави петкратна надворешна вкрстена валидација на податоците.

Табела 19. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за анализа на случај P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy}).

Fold	Best Sigma	Fold RMSE
1	0.242446	8.417178
2	0.142510	0.976167
3	0.242446	13.232246
4	0.142510	9.107570
5	0.242446	4.357449



Прилог 32. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за анализа на случај P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy}).

Од графикот во Прилог 32 може да се воочи дека минимумот на RMSE = 4.357 се добива за вредност на параметарот $\sigma \approx 0.242$.

6.2.4. Резултати од обучувањето на моделите

По обучување на девет различни модели базирани на девет различни алгоритми, како што следи презентирани се сумарни резултати од сите модели за претходната фаза во форма на табели и графици.

Како првичен преглед на резултатите, во Табела 20 прикажани се вредности на метричките показатели добиени во процесите на петкратна вкрстена валидација во секој обучен модел.

Табела 20. Резултати за вредностите на метричките показатели по процесот на обучување во рамки на кој во секој модел е спроведена и петкратна вкрстена валидација за анализа на случај P_kW = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy}).

Model	MSE	RMSE	MAE	R2
GPR	1039.691	32.076	25.065	0.641
ANN	152.000	11.419	6.639	0.949
XGBoost	80.205	8.462	2.755	0.970
GRNN	69.766	7.218	2.832	0.972
Decision Tree	68.084	7.528	2.695	0.975
Random Forest	64.728	7.669	2.562	0.976
SVR	57.112	7.306	2.265	0.980
Boosted Trees	49.013	5.948	2.044	0.981
KNN	51.456	6.932	2.716	0.981

Резултатите прикажани во Табела 20 даваат јасна споредба на перформансите на девет различни алгоритми за машинско учење преку четири стандардни метрики: MSE, RMSE, MAE и R². Овие индикатори овозможуваат детална проценка на точноста, стабилноста и способноста за генерализација на секој модел при предвидување на целната варијабла.

Во однос на MSE, кој ја мери средната квадратична грешка и ги нагласува поголемите отстапувања, најдобар резултат има Boosted Trees со вредност 49.013, што покажува дека овој модел најуспешно ги минимизира големите грешки. Многу блиску се и KNN со 51.456 и SVR со 57.112, што укажува на висока точност. Random Forest со 64.728 и Decision Tree со 68.084 исто така покажуваат добри резултати. Од друга страна, ANN има MSE од 152.0, што е значително повисоко од најдобрите модели, додека убедливо најслаб резултат има Gaussian Process Regression (GPR) со 1039.691, што укажува на сериозни нестабилности и присуство на големи грешки во

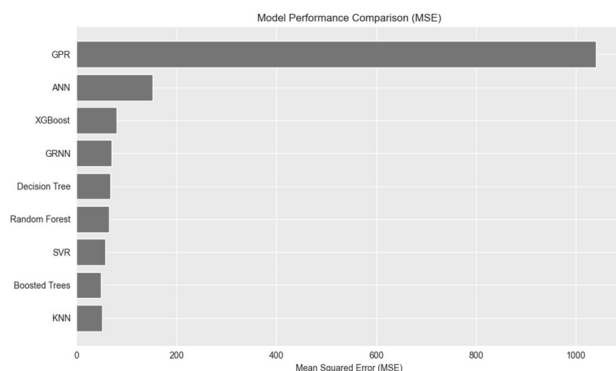
предвидувањата.

Кај RMSE, кој претставува корен од MSE и е полесен за интерпретација бидејќи е изразен во иста единица како целната променлива во [kW], повторно најдобар резултат има Boosted Trees со 5.948. Тоа значи дека просечното реално отстапување на моделот е најниско. Следуваат KNN со 6.932, GRNN со 7.218, SVR со 7.306, како и Decision Tree и Random Forest со вредности околу 7.5. ANN има RMSE од 11.419, што покажува дека моделот има поголеми осцилации во точноста, додека GPR со 32.076 повторно значително заостанува зад сите останати модели.

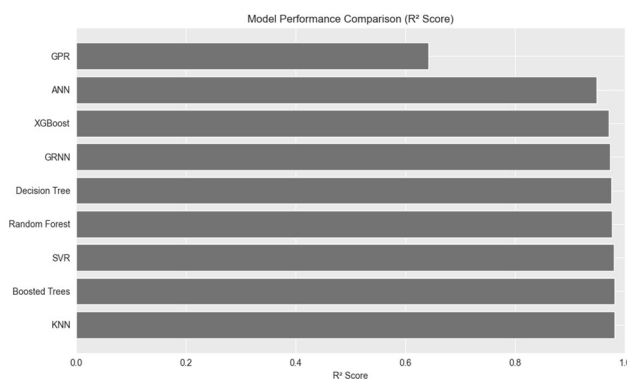
Во однос на MAE, кој ја мери просечната апсолутна грешка и не е толку чувствителен на екстремни вредности, најдобар модел повторно е Boosted Trees со 2.044, што значи дека во просек неговите предвидувања најмалку отстапуваат од реалните вредности. Следуваат SVR со 2.265, Random Forest со 2.562, Decision Tree со 2.695, KNN со 2.716 и GRNN со 2.832. Иако ANN има солиден R^2 , неговиот MAE од 6.639 покажува дека во просек има поголеми апсолутни грешки. Најслаб повторно е GPR со 25.065, што укажува на сериозно неточни предвидувања.

Кај R^2 коефициентот, кој го покажува процентот на варијација што моделот успева да ја објасни, најдобри резултати имаат Boosted Trees и KNN со 0.981, што значи дека објаснуваат околу 98.1% од варијансата на податоците. Многу блиску се SVR со 0.980, Random Forest со 0.976 и Decision Tree со 0.975. Овие вредности потврдуваат одлична способност за генерализација. ANN има $R^2 = 0.949$, што иако е добар резултат, е понизок од водечките модели. Најнизок резултат повторно има GPR со 0.641, што значи дека овој модел објаснува само 64.1% од варијацијата и не е соодветен за конкретниот проблем.

Генерално, може да се заклучи дека Boosted Trees е најдобриот модел во оваа анализа, бидејќи има најниски вредности за MSE, RMSE и MAE, како и највисок R^2 . Одлични резултати покажуваат и KNN, SVR, Random Forest и Decision Tree, кои претставуваат стабилни и сигурни модели. GRNN исто така покажува добра конкурентност. ANN има умерени резултати, додека GPR значително заостанува и не е препорачлив за ова множество податоци. Овие резултати потврдуваат дека ансамбл методите и локално-базирани модели покажуваат најголема ефикасност за анализираниот инженерски систем.



Прилог 33. Приказ на MSE за секој обучен модел за анализа на случај $P_{kW} = f(Z_{w1}, ResVol, Q_{inflow}, Q_{energy})$.

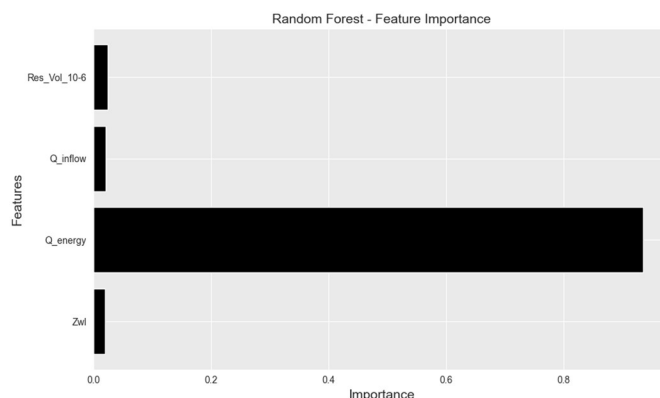


Прилог 34. Приказ на R^2 за секој обучен модел за анализа на случај $P_{kW} = f(Z_{w1}, ResVol, Q_{inflow}, Q_{energy})$.

За моделот обучуван со RF, Пајтон кодот е програмиран да прикаже график на важност на варијаблите (анг. „Feature importance“) во моделот.

Графикот со важноста на варијаблите за моделот обучуван со RF е даден во Прилог 35, каде може да се воочи дека најголемо влијание во моделот има варијаблата „Q_energy“ – енергетски обработен проток со вредност над 0.9, додека останатите предиктор варијабли „Res_Vol_10-6“, „Q_inflow“ и „Zwl“ имаат важност помала од 0.2. Пресметаната важност на варијаблата „Q_energy“ е очекувана со оглед на тоа дека производството на моќност е во директна физичка корелација со енергетскиот обработен проток т.е. моќноста правопрпорционално зависи од протокот, согласно равенството:

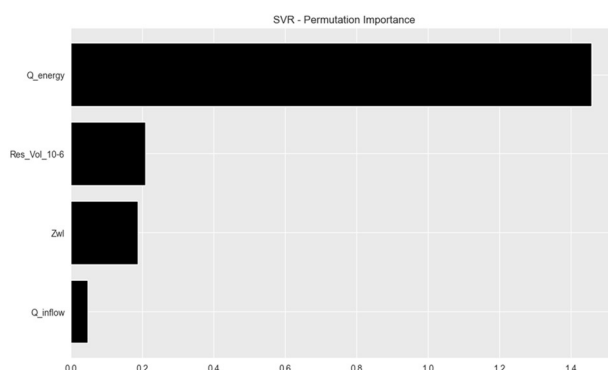
$$P = 9.81 \cdot \rho_w \cdot Q_{energy} \cdot H_n = \dots [W] \quad (6.2.1)$$



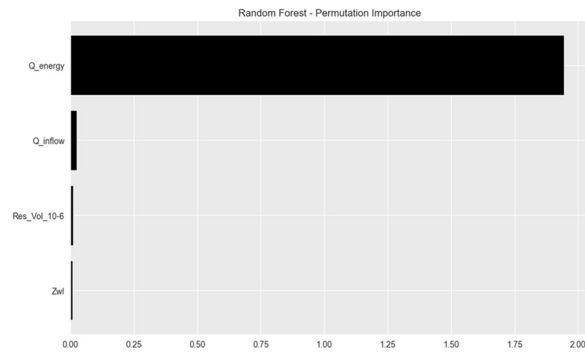
Прилог 35. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

Дополнително, во DamResML кодот програмирано е извлекувањето на пермутациската важност (анг. „Permutation importance“) на моделите обучувани со SVR, RF и DT. Пермутациската важност претставува метод за оценување на значењето на предиктор варијаблите врз основа на нивното влијание врз перформансите на моделот.

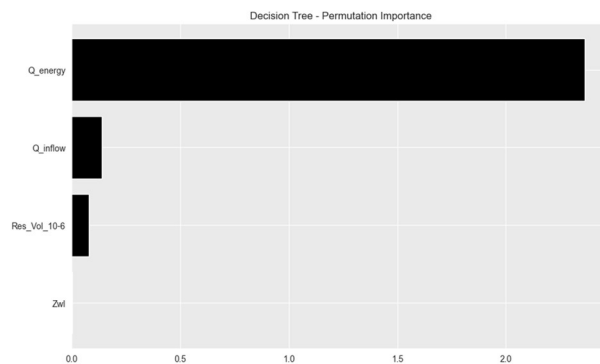
Во Прилог 36, Прилог 37 и Прилог 38 даден е приказ на пресметаната пермутациска важност за напоменатите модели. Се согледува дека кај сите три модели најголема пермутациска важност има предиктор варијаблата „Q_energy“, додека останатите „Res_vol_10-6“, „Zwl“ и „Q_inflow“ се во инфериорна позиција споредено со истата. Кај секој модел вредноста на пермутациската важност е различна за доминантната варијабла „Q_energy“, па така, кај моделот обучуван со SVR вредноста е 1.42, кај моделот обучуван со RF вредноста е 1.9 и кај моделот обучуван со DT вредноста на пермутациската важност на „Q_energy“ е 2.75.



Прилог 36. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.



Прилог 37. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.



Прилог 38. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

За секој обучуван модел, во DamResML кодот програмирано е да се прикаже график на резидуали (остатоци) – грешки во предвидувањата од секој модел. Сумарен график кој ги опфаќа сите 9 модели е прикажан во Прилог 39. Резидуалите се пресметуваат врз основа на кратко предвидување со обучените модели, спроведено на т.н. задржан сет на податоци (анг. „holdout“) кој не бил претходно вклучен во процесот на обучување на моделите, па така истиот е непознат за моделот. Ова издвојување на задржан сет настрана од сетот за обучување кодот го прави уште во првата ќелија од моделот кодот.

Прикажаните графици ја претставуваат резидуалната анализа за девет различни модели на машинско учење. Резидуалите ја претставуваат разликата помеѓу реалните и предвидените вредности, односно грешката на моделот. Во идеален случај, точките треба да бидат случајно распоредени околу хоризонталната нулта линија, без јасен тренд, со мала вертикална дисперзија. Таквата појава укажува дека моделот е стабилен, непристрасен и успешно ја опишува зависноста во податоците.

Кај моделот обучуван со ANN, најголем дел од резидуалите се концентрирани околу нулата кај пониските вредности, што покажува добра точност во тој домен. Сепак, кај неколку средни и повисоки вредности на предиктор варијаблата, се забележуваат поголеми негативни и позитивни отстапувања, што укажува дека моделот понекогаш потценува или преценува поединечни случаи. Генерално, ANN покажува добра, но не целосно конзистентна стабилност.

Кај моделот обучуван со SVR, точките се компактно групирани околу нултата линија, кај најголемиот дел од примероците. Постои еден поизразен негативен резидуал кај средна вредност, но во целина моделот покажува добра рамнотежа и мала пристрасност. Ова е показател за солидна генерализација.

Random Forest покажува добар распоред на резидуали околу нулата, но со неколку поголеми негативни отстапувања кај средните вредности. Тоа значи дека моделот во одредени случаи има тенденција на преценување на целната варијабла. Сепак, најголемиот број точки се блиску до нула, што укажува на висока точност на моделот.

Кај Decision Tree се забележува поголема дисперзија на резидуалите во споредба со Random Forest. Постојат неколку значајни позитивни резидуали, што укажува дека моделот понекогаш потценува реални вредности. Ова е очекувано бидејќи единечно дрво е почувствително на шум и варијации во податоците.

Моделот KNN покажува релативно добра распределба, но со повеќе изразени негативни резидуали кај средните вредности. Тоа значи дека во одредени случаи моделот ја преценува целната варијабла. Бидејќи KNN зависи од локални соседи, ваквите осцилации се чести кога податоците не се рамномерно распределени.

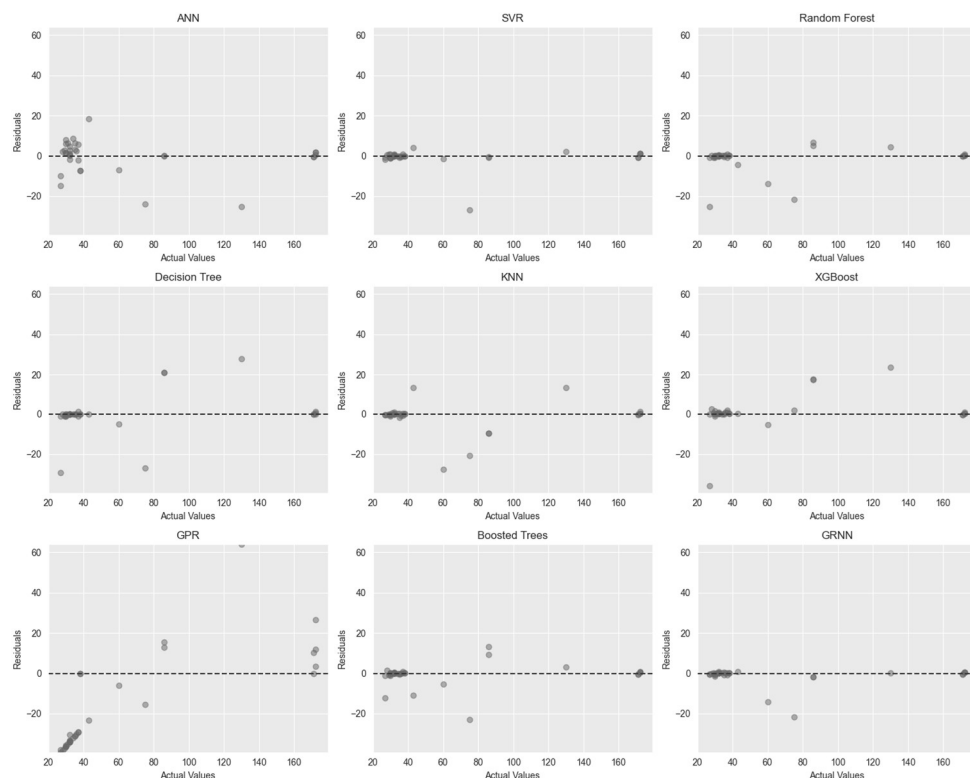
Кај XGBoost се забележува добра концентрација околу нулата за поголемиот дел од податоците, но има неколку изолирани точки со поголеми позитивни и негативни грешки. Генерално, моделот покажува добра точност и стабилност, со мала дисперзија кај дел од податоците.

Моделот GPR (Gaussian Process Regression) има најнеповолен резидуален распоред. Се забележуваат многу големи позитивни и негативни отстапувања, особено кај повисоките реални вредности. Тоа укажува на нестабилност и слаба способност за генерализација. Резидуалите не се случајно распоредени, туку покажуваат систематски проблеми во предвидувањето.

Кај Boosted Trees резидуалите се концентрирани блиску до нулата, со релативно мала дисперзија. Иако постојат неколку изолирани отстапувања, вкупната слика е многу добра. Ова потврдува дека ансамбл пристапот успешно ги намалува грешките и ја подобрува точноста.

Моделот GRNN покажува една од најдобрите распределби на резидуали. Најголем дел од точките се блиску до нултата линија, со мал број умерени негативни отстапувања. Ова укажува на добра стабилност и способност за следење на трендовите во податоците.

Генерално, најдобри резидуални карактеристики покажуваат Boosted Trees, SVR, Random Forest и GRNN, бидејќи точките се компактно распоредени околу нулата со мала дисперзија. ANN, KNN и XGBoost имаат добри резултати со одредени поединечни отстапувања. Decision Tree е умерено стабилен, додека GPR покажува најслаби резултати. Оваа анализа е во согласност со претходно добиените нумерички метрики, каде ансамбл моделите и стабилните нелинеарни алгоритми покажуваат највисока успешност во предвидување на вредноста на целната варијабла.



Прилог 39. Графици со резидуали за секој обучуван модел за анализа на случај $P_{kW} = f(Z_{wl}, ResVol, Q_{inflow}, Q_{energy})$.

6.2.5. Предвидување на целната варијабла

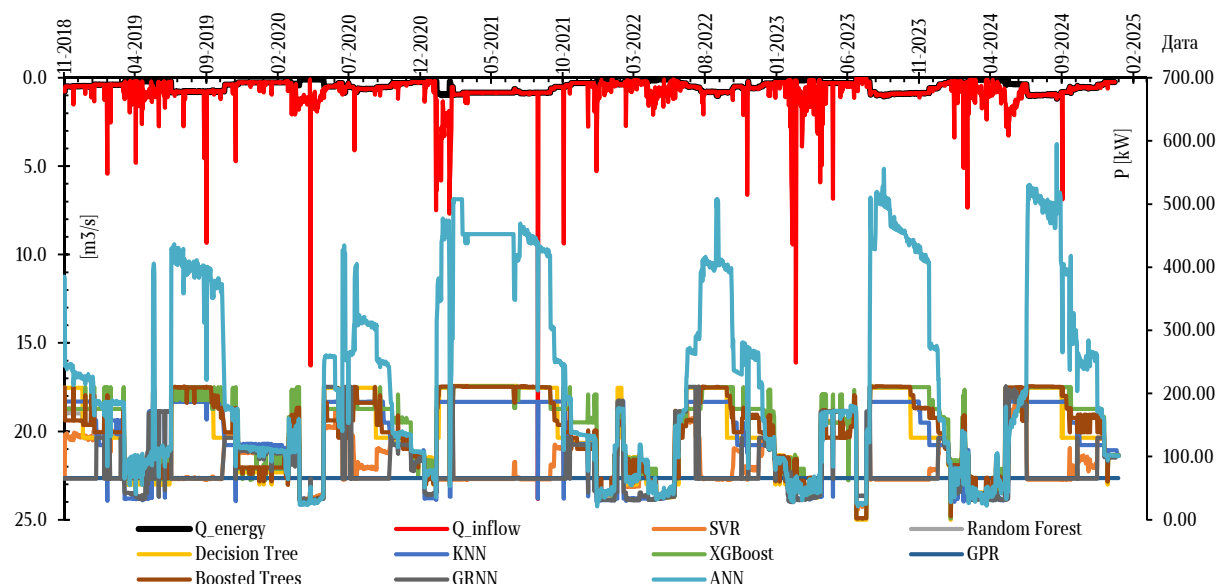
Целна варијабла во следната студија на случај е генерираната моќност „P_kW“ во прибранската хидроцентрала на браната „Кнежево“. Со помош на обучени модели од машинско учење, се бара предвидување на целната варијабла со познавање на следните предиктор варијабли:

- 1) Ниво на вода во акумулација „Zwl“ во [mnv] прикажан во Прилог 30,
- 2) Волумен на вода во акумулацијата „Reservoir_volume_10-6“ во [10^6m^3] прикажан во Прилог 28,
- 3) Дотек во акумулацијата „Q_inflow“ во [m^3/s] прикажан во Прилог 29,
- 4) Истекувања од акумулацијата низ темелниот испуст кои се изедначуваат со протокот кој енергетски би се обработувал, „Q_discharge“ = „Q_energy“ во [m^3/s] прикажан во Прилог 29.

Обучени се вкупно девет модели со машинско учење базирани на следните алгоритми: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Процесот на предвидување се извршува за 2241 познати вредности на предиктор варијаблите, кои временски се однесуваат на периодот на прибирање на податоците од 1ви ноември 2018 година до 31 декември 2024 година. За овој период нема вредност за целната варијабла со оглед на тоа дека хидроцентралата е пуштена во работа во 2025 година, но има вредностите на предиктор варијаблите согласно прибраните податоци од страна на операторот. Предиктор варијаблите се внесуваат во DamResML кодот преку датотеката „predictor_variable_for_prediction.xls“ и истата се повикува во ќелијата „Prediction“ по завршување на процесот на обучување на деветте модели.

Резултатите од предвидувањата се прикажани во прилозите кои следат. Сумарните предвидувања се прикажани во Прилог 40, заедно со предиктор варијаблите „Q_energy“ и „Q_inflow“.

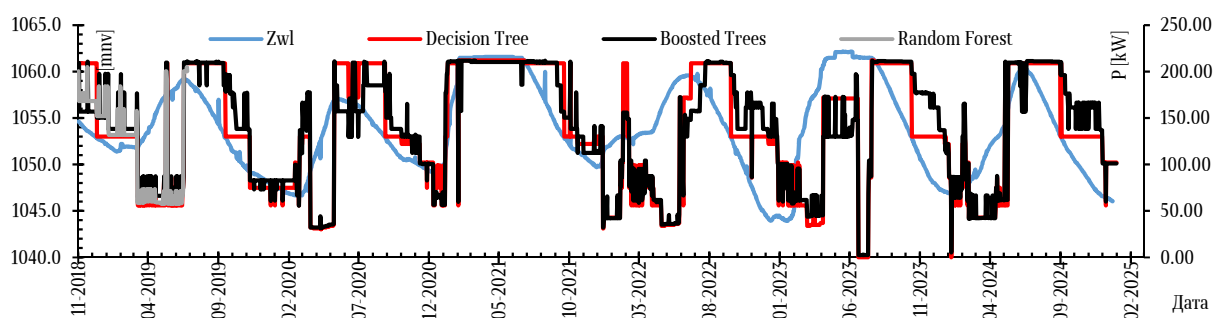


Прилог 40. График со приказ на пресметаните вредности на целната варијабла „P_kW“ пресметани со помош на сите девет обучени модели, приказ на предиктор варијаблите „Q_energy“ и „Q_inflow“.

Во Прилог 41 прикажани се предвидувањата за целната варијабла „P_kW“ во ординатната оска и тоа предвидувања од моделите тренирани со алгоритмите BT, DT и RF. Дополнително, во графикот прикажана е и варијацијата на водното ниво во акумулацијата „Zwl“ за периодот на предвидување, при што може да се воочи дека предвидувањата за целната варијабла во најголем

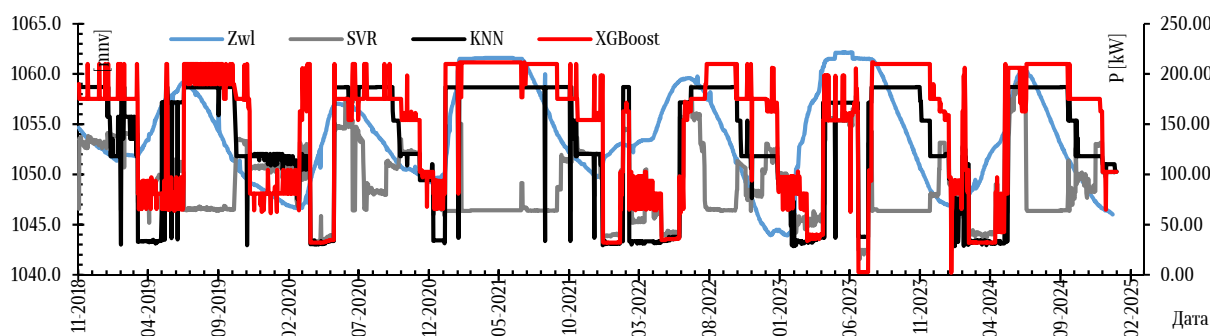
дел ја следат флукуацијата на водното ниво во акумулацијата; поголеми вредности за целната варијабла се пресметуваат за периодите со високи нивоа на водно ниво во акумулацијата, додека пониски вредности за целната варијабла се предвидуваат за периодите со пониски нивоа на водно ниво во акумулацијата, со одредени исклучоци. Нивото во акумулацијата за периодот на предвидување варира во границите помеѓу $Z_{wl} = 1043.92\text{m}$ и $Z_{wl} = 1062.18\text{m}$. Максималното ниво во акумулацијата е $Z_{\max} = 1063.50\text{m}$, со што може да се заклучи дека нивото во акумулацијата за разгледуваниот период се одржува релативно високо, во коридор од 19.59m под максималното водно ниво.

Согласно предвидувањата од моделот со BT, на годишно ниво хидроцентралата генерира моќност од $P_{\text{god}} = 50.59 \text{ MW/god}$ или вкупно $P_{\text{vk}} = 310.62 \text{ MW}$ за разгледуваниот период. Согласно предвидувањата од моделот со DT, генерираната моќност во хидроцентралата изнесува $P_{\text{god}} = 50.00 \text{ MW/god}$ на годишно ниво, или вкупно $P_{\text{vk}} = 306.98 \text{ MW}$ за разгледуваниот период. Слични се предвидувањата и со моделот со RF, каде на годишно ниво се предвидува генерирање на моќност во вредност од $P_{\text{god}} = 49.20 \text{ MW/god}$ или вкупно $P_{\text{vk}} = 302.05 \text{ MW}$ за разгледуваниот период на предвидувања.



Прилог 41. График со приказ на пресметаните предвидени вредности на целната варијабла „P_kW“ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на BT, DT и RF.

Во Прилог 42 прикажани се предвидувањата на целната варијабла „P_kW“ за разгледуваниот период на предвидувања согласно тренираните модели базирани на SVR, KNN и XGBoost. Идентично на Прилог 41, и овде е прикажана варијацијата на водното ниво во акумулацијата, при што може да се заклучи дека генерирањето на моќност како целна варијабла исто така ја следи флукуацијата на водното ниво во акумулацијата, слично на предвидувањата од моделите BT, DT и RF. Исклучок претставува генерирањето на моќност во месец јуни, 2023 година, кога може да се забележи дека при високо водно ниво во акумулацијата производството на моќност драстично опаѓа. Овој феномен се забележува во двата прилози - Прилог 41 и Прилог 42. Дополнително, на повеќе наврати се генерира висока вредност на моќноста во период на пониски водни нивоа во акумулацијата. Овие појави се резултат на влијанието на останатите предиктор варијабли врз предвидувањето на целната варијабла, особено „Q_energy“ – енергетски обработениот проток, што може да се забележи на графиконот во Прилог 40 каде во ординатната оска прикажани се предвидувањата од сите девет тренирани модели, како и вредностите на предиктор варијаблите „Q_energy“ и „Q_inflow“.

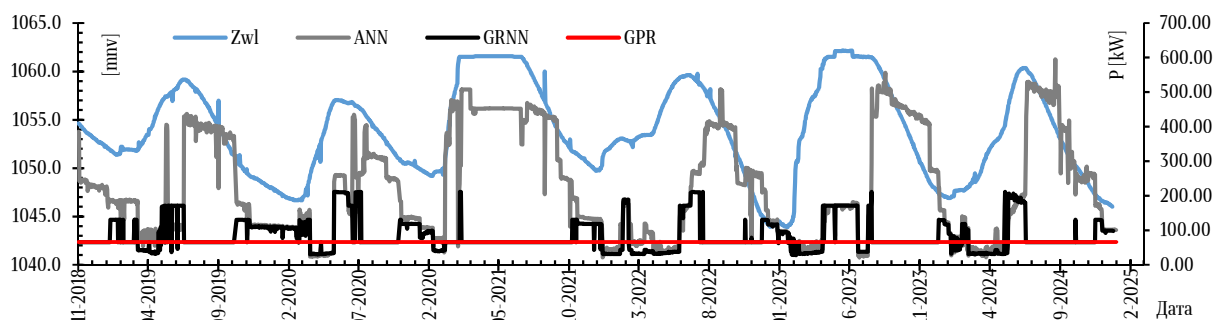


Прилог 42. График со приказ на пресметаните предвидени вредности на целната варијабла „P_kW“ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на SVR, KNN и XGBoost.

Согласно предвидувањата со моделите тренирани на SVR, KNN и XGBoost, произведената моќност со моделот на SVR е $P_{\text{god}} = 32.73 \text{ MW/god}$ или вкупна моќност за целиот период на пресметување од $P_{\text{vk}} = 200.93 \text{ MW}$. Кај моделот трениран со KNN, произведената моќност е пресметана како $P_{\text{god}} = 48.26 \text{ MW/god}$ на годишно ниво, или вкупно $P_{\text{vk}} = 296.33 \text{ MW}$. Со моделот трениран на XGBoost, пресметаните вредности за произведената моќност како целна варијабла, изнесуваат $P_{\text{god}} = 53.80 \text{ MW/god}$ на годишно ниво, или вкупно $P_{\text{vk}} = 330.34 \text{ MW}$.

Во графикот во Прилог 43 прикажани се предвидувањата од тренираните модели со ANN, GRNN и GPR заедно со предиктор варијаблата „Zwl“. Интересни се резултатите од моделот со GPR, каде може да се согледа дека низ целиот период на предвидувања моделот предвидува само една вредност на генерираната моќност – $P = 66.13 \text{ kW/den}$ или $P_{\text{god}} = 24.14 \text{ MW/god}$, т.е. $P_{\text{vk}} = 148.19 \text{ MW}$ за целиот разгледуван период. Ова е најниската пресметана вредност на предвидувањата од сите девет модели. Од друга страна, тренираниот модел со ANN дава највисоки вредности на предвидувањата од сите девет модели, со предвидено годишно производство од $P_{\text{god}} = 83.39 \text{ MW/god}$ или вкупно производство на моќност за целиот разгледуван период од $P_{\text{vk}} = 511.98 \text{ MW}$. Моделот трениран со алгоритмот GRNN пресметува вредности на предиктор варијаблата кои се пониски споредено со останатите модели, со пресметано годишно производство од $P_{\text{god}} = 30.54 \text{ MW/god}$ или вкупно производство на моќност од $P_{\text{vk}} = 187.52 \text{ MW}$ за целиот период на предвидувања.

Како резиме на предвидувањата, за разгледуваниот период од шест години, најниски вредности за генерираната моќност во хидроцентралата „P_kW“ предвидува моделот трениран со алгоритмот GPR, со вредност од само $P_{\text{god}} = 24.14 \text{ MW/god}$. Највисоки вредности за предвидувањата пресметува моделот трениран со ANN, со вредност од $P_{\text{god}} = 83.39 \text{ MW/god}$. Моделите базирани на дрва (DT, BT и RF) пресметуваат слични вредности за целната варијабла, блиску до $P_{\text{god}} = 50 \text{ MW/god}$. Оттука може да се заклучи дека на годишно ниво од прибранската хидроцентрала на брана „Кнежево“ може да се очекуваат од $P_{\text{god}} = (24+83) \text{ GW/god}$ произведена моќност во зависност од хидролошката состојба во сливот, нивото на водата во акумулацијата, акумулираниот волумен на вода во акумулацијата и енергетски испуштениот проток низ хидроцентралата.

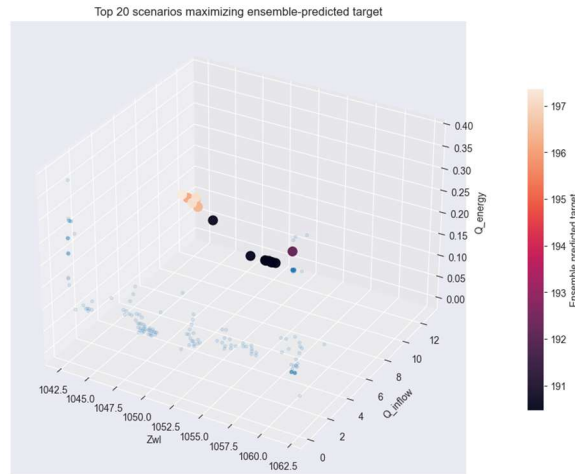


Прилог 43. График со приказ на пресметаните предвидени вредности на целната варијабла „P_kW“ - генерирање на моќност во хидроцентралата, пресметани со помош на обучените модели на ANN, GRNN и GPR.

6.2.6. Хаос оптимизација на целната варијабла

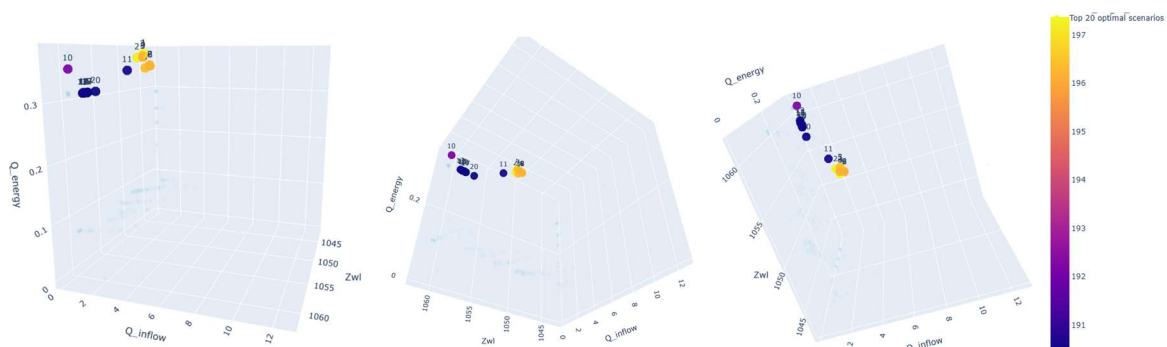
Како последен чекор од процесот на подучување на моделите и предвидување на целната варијабла, применета е хаос оптимизација со цел оптимизирање на целната варијабла. Во конкретниот случај, целната варијабла е „P_kW“ т.е. генерирана моќност од „ХЕЦ Б“ – прибранската хидроцентрала на брана „Кнежево“.

Хаос оптимизацијата се применува за максимизација на вредноста на предвидувањата на целната варијабла „P_kW“ со користење на сетот на податоци добиени како предвидувања од моделите RF, DT и XGBoost издвоени како модели со најдобри метрички показатели во процесот на подучување и предвидување на целната варијабла. Станува збор за вкупно 2241 податоци на предиктор и таргет варијаблата. Тука треба да се напомене дека во хаос оптимизацијата внесено е ограничување единствено на вредноста на целната варијабла и тоа во опсег од $P_{\text{kW}} = 0$ од $P_{\text{kW}} = 200$ со цел да се опфатат сите реални можни вредности на генерираната моќност.



Прилог 44. 3Д график со приказ на пресметаните оптимални вредности на целната варијабла P_kW во функција од четирите предиктор варијабли „Zwl“, „Q_inflow“, „Q_energy“ и „Reservoir_volume_10-6“.

За да стане појасно кои комбинации на влезните предиктор варијабли придонесуваат кон максимизација на целната варијабла, моделот продуцира 3D графикон со приказ на 20 оптимални вредности на целната варијабла во XYZ координатен систем (Прилог 44). Самиот графикон ги прикажува и останатите вредности на целната варијабла од податочниот сет, но ги акцентира издвоените 20 оптимални вредности – во случајот максимални вредности на целната варијабла „P_kW“. Во Прилог 45 е прикажан 3Д графикон кој ги вклучува единствено оние издвоени оптимизирани 20 вредности на целната варијабла, додека на XYZ оските соодветно може да се отчитаат вредностите за предиктор варијаблите. Бидејќи на 3Д графиконот може да се прикажат три предиктор варијабли, за репрезентација избрани се „Q_energy“, „Q_inflow“ и „Zwl“.



Прилог 45. 3Д график со приказ на 20-те издвоени вредности на целната варијабла P_kW.

Од Прилог 45 директно може да се согледа дека поголемите вредности на целната варијабла се блиски со поголемите вредности за нивото во акумулацијата „Zwl“ и енергетски обработениот проток „Q_energy“. Максимумот кој Хаос оптимизацијата го пронајде врз основа на конкретниот податочен сет кој претставува комбинација на предвидени вредности на целната варијабла од различните поучени модели со машинско учење, е:

$$\text{MAX P_kW} = 197.375 \text{ [kW]}$$

,во функција од предиктор варијаблите:

- 1) Ниво на вода во акумулација „Zwl“: 1051.013 [mnv],
- 2) Волумен на вода во акумулација „Reservoir_volume_10-6“: 23.488 [$\times 10^6 \text{ m}^3$],
- 3) Дотек на вода во акумулација „Q_inflow“: 1.897 [m^3/s],
- 4) Енергетски обработен проток „Q_energy“: 0.376 [m^3/s].

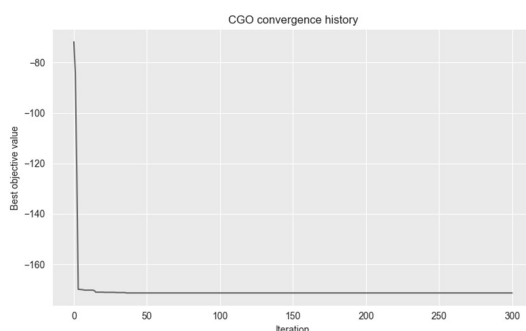
За приказ на еволуцијата на најдобрата вредност на целната функција во зависност од бројот на итерации во Хаос оптимизацијата, прикажан е график на конвергентноста на моделот (Прилог 46). Во графиконот на Прилог 46, на апцисната оска е прикажан бројот на итерации, додека на

ординатата е прикажана најдобрата постигната вредност на целната функција во секој чекор од оптимизацијата.

Во првите неколку итерации се забележува нагло намалување на вредноста на целната функција (од приближно -70 до околу -175). Овој стрмен пад укажува на ефикасно иницијално истражување на просторот на решенија и способност на алгоритмот брзо да идентификува региони со значително подобри решенија, како и на добро дефинирани граници на пребарување и соодветно иницијализирана популација. Ова однесување е карактеристично за оптимизациски алгоритми кои успешно ја користат стохастичката компонента (во овој случај хаотичната логистичка мапа) за глобално пребарување.

Во средната фаза се прецизира решението (итерации 10–50), каде се забележува намалување на целната функција, но со значително помал интензитет. Се забележуваат мали, постепени подобрувања (од приближно -175 до -170), што укажуваат на транзиција од глобална експлорација кон локална експлоатација, фино подесување (fine-tuning) на веќе идентификуваниот оптимален регион, и стабилност на алгоритмот без значителни осцилации. Оваа фаза е клучна бидејќи алгоритмот ја подобрува точноста на решението без да ја наруши стабилноста.

Во завршната фаза или т.н. фаза на конвергенција (итерации 50–300) графикот покажува речиси хоризонтална линија (по 50та итерација), со минимални или занемарливи подобрувања (околу -170). Ова јасно укажува дека алгоритмот достигнал стабилно решение, не се пронаоѓаат подобри комбинации на влезни параметри, постигната е конвергенција кон глобален или квази-глобален оптимум.



Прилог 46. Преглед на ковергентноста на моделот од Хаос оптимизација низ историскиот развој на итерациите.

Недостигот на осцилации и стабилноста на кривата во оваа фаза укажуваат на добра дефинираност на целната функција, отсуство на значителни локални минимални кои би го нарушиле процесот, и сигурност во добиеното оптимално решение. Со оглед дека целната функција е дефинирана како негативна вредност од предвидувањето (со додадени казни), намалувањето на вредноста на графикот значи зголемување на предвидената вредност на целната променлива и подобрување на оптималноста на решението. Оттука, финалната вредност (-170) претставува најдобрата комбинација на влезни параметри која го максимизира излезот, при почитување на воспоставените ограничувања.

6.3. Анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$

6.3.1. Подготовка на податоците

Наредната студија на случај се изработува со податоци од управување со хидросистемот „Стрежево“. Бидејќи станува збор за комплексен, повеќефункционален систем со различни водокорисници, одржувањето на водно ниво во акумулацијата кое ќе обезбеди максимално задоволување на потребите на водокорисниците претставува клучна задача на операторите. Водното ниво во акумулацијата зависи од дотекот во акумулацијата и врнежите на површината на акумулацијата кои ја полнат акумулацијата од една страна, како и испуштањето на вода низ

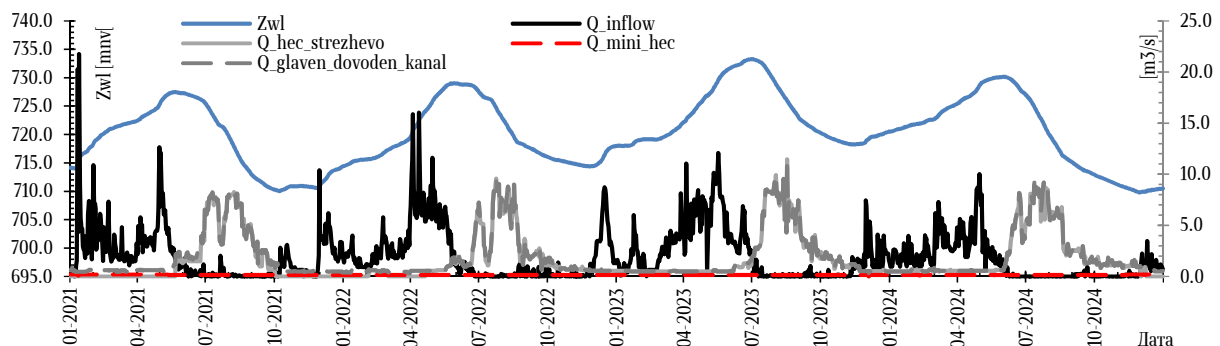
хидроцентралите и главниот доводен орган – од друга страна. Напоменатите параметри секојдневно се регистрираат и следат од страна на операторот, и истите беа ставени на располагање како податоци за изработка на докторската дисертација.

Анализата на случај се фокусира во (1) предвидување на целната варијабла „Zwl“ – ниво на вода во акумулацијата преку (2) анализа на влијанието на предиктор варијаблите „Rain“ – паднат дожд на површината на акумулацијата, „Q_hec_strezhevo“ – енергетски обработен проток низ низводната хидроцентрала „Стрежево“, „Q_mini_hec“ – енергетски обработен проток низ малата ХЕЦ за биолошки минимум, и „Q_gl_dovoden_organ“ – проток кој се испушта во главниот доводен канал кон хидромелиоративниот систем „Стрежево“ за наводнување на пелагониското поле.

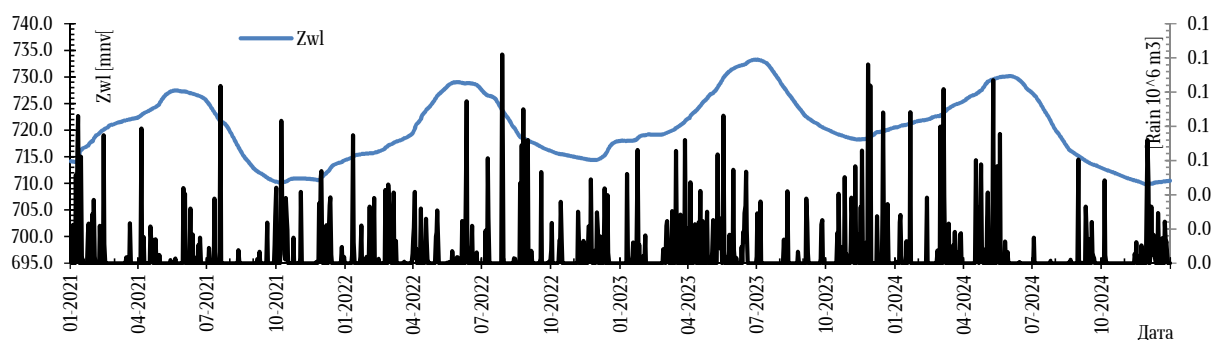
Вредностите на предиктор варијаблите кои се применуваат во процесот на обучување на моделите се преземени од операторот за временски период 2021 – 2024 година, регистрирани секојдневно. Станува збор за вкупно 1461 податок за секоја предиктор варијабла и целната варијабла „Zwl“.

Предиктор варијабли, се:

- 1) „Q_inflow“ – регистриран дотек во акумулацијата во $[m^3/s]$ прикажан во Прилог 47,
- 2) „Q_hec_strezhevo“ – регистриран енергетски обработен проток низ ХЕЦ „Стрежево“ во $[m^3/s]$ прикажан во Прилог 47,
- 3) „Q_hec_mini“ – регистриран енергетски обработен проток низ мини ХЕЦ за биолошки минимум, во $[m^3/s]$ прикажан во Прилог 47,
- 4) „Q_gl_dovoden_organ“ – регистрирано испуштање на вода низ главниот канал за ХМС „Стрежево“ во $[m^3/s]$ прикажан во Прилог 47, и
- 5) „Rain“ – паднат дожд на површината на акумулацијата регистриран во $[10^6 m^3]$ прикажан во Прилог 48.



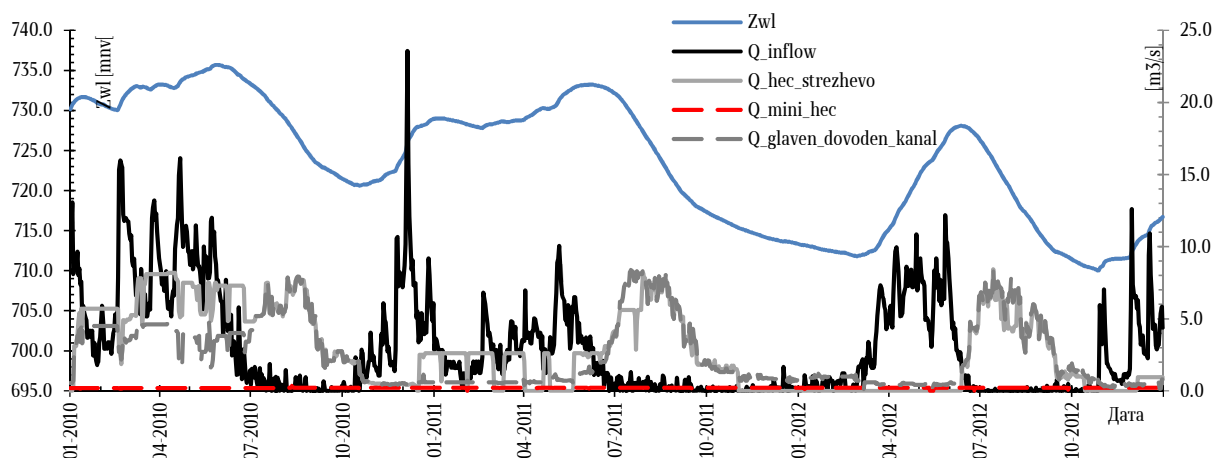
Прилог 47. Графички приказ на целната варијабла „Zwl“ и предиктор варијаблите „Q_hec_strezhevo“, „Q_inflow“, „Q_hec_mini“ и „Q_gl_dovoden_organ“ со вредности кои се применуваат во процесот на обучување на моделите.



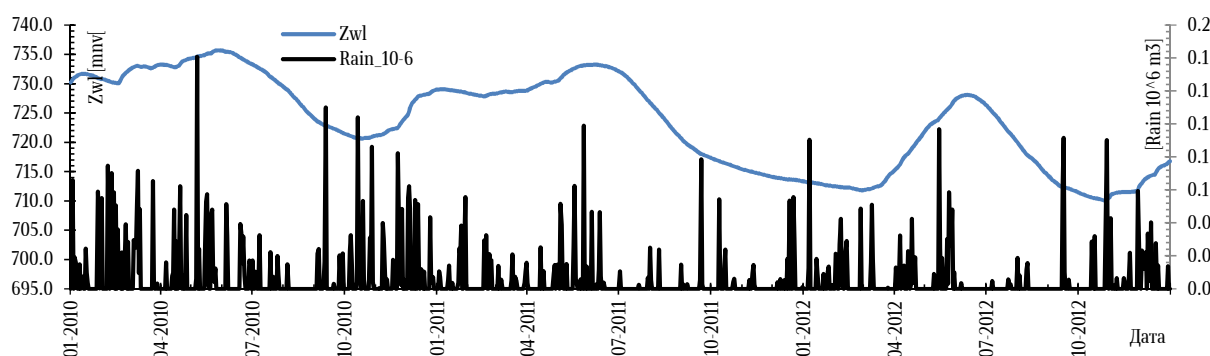
Прилог 48. Графички приказ на целната варијабла „Zwl“ и предиктор варијаблата „Rain“ со вредности кои се применуваат во процесот на обучување на моделите.

Периодот на предвидување на целната варијабла опфаќа 1091 податоци за секоја предиктор варијабла. Станува збор за регистрирани вредности на оперативните параметри за периодот од

2010 – 2012 година, каде покрај предиктор варијаблите, се води регистрација и на целната варијабла, па предвидувањата ќе се споредат и со реалните вредности од управувањето на системот. Вредностите на предиктор варијаблите како и целната варијабла за периодот на предвидување се прикажани во граfiците во Прилог 49 и Прилог 50.



Прилог 49. Графички приказ на целната варијабла „Zwl“ и предиктор варијаблите „Q_inflow“, „Q_hec_strezhevo“, „Q_mini_hec“ и „Q_glaven_dovoden_kanal“ со вредности кои се применуваат во процесот на предвидување.



Прилог 50. Графички приказ на целната варијабла „Zwl“ и предиктор варијаблата „Rain“ со вредности кои се применуваат во процесот на предвидување на целната варијабла.

6.3.2. Прелиминарна петкратна вкрстена валидација

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткрени регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Целната варијабла е користена во нејзината оригинална форма, без примена на нормализација или стандардизација. Од друга страна, предиктор варијаблите беа стандардизирани во рамки на секој сет од петкратната валидација, освен за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R² прикажани во Табела 21.

Табела 21. Резултати од прелиминарната петкратна вкрстена валидација изработена на деветте модели со машинско учење за $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Model	MSE	RMSE	MAE	R2
GPR	199831.168	293.297	26.360	-5519.245
ANN	155.997	10.640	4.718	-3.241
Decision Tree	25.892	5.082	3.136	0.303
GRNN	22.716	4.764	3.511	0.385
KNN	20.805	4.560	3.334	0.438
SVR	20.619	4.541	3.518	0.444
Boosted Trees	17.089	4.133	3.217	0.539
XGBoost	16.612	4.075	2.985	0.552
Random Forest	15.217	3.900	2.841	0.589
min:	15.217	3.900	2.841	-5519.245
max:	199831.168	293.297	26.360	0.589

Резултатите од петкратната вкрстена валидација од анализата за предвидување на нивото на водата во акумулацијата „Zwl“ врз основа на влезните параметри „Qinflow“, „Rain“, „Qhec_strezhevo“, „Qmini_hec“ и „Q_gl_dovoden_organ“ покажуваат дека кај поголемиот дел од моделите се добиени позитивни вредности на коефициентот на детерминација R^2 , што укажува на постоење на јасна зависност помеѓу предиктор и целната варијабла.

Моделите базирани на ансамбл методи, како што се Random Forest и XGBoost, покажуваат најдобри перформанси, со најниски вредности на RMSE (3.900 и 4.075) и највисоки вредности на R^2 (0.589 и 0.552). Ова укажува дека овие модели најуспешно ја препознаваат нелинеарната зависност помеѓу хидролошките дотекувања, истекувања од акумулацијата и нивото на водата во акумулацијата. Слично однесување покажува и моделот Boosted Trees со $R^2=0.539$, што дополнително ја потврдува предноста на ансамбл пристапите кај вакви комплексни системи.

Моделите Support Vector Regression и K-Nearest Neighbors, како и GRNN, покажуваат умерено добри резултати, со R^2 вредности во опсег од 0.38 до 0.44. Ова укажува дека и покрај различната природа на алгоритмите, тие успеваат да ја апроксимираат основната динамика на системот, но со нешто пониска точност во споредба со ансамбл моделите.

Од друга страна, Decision Tree моделот покажува ограничена предиктивна способност ($R^2=0.303$), што може да се објасни со неговата едноставна структура и тенденција кон пренапасување на тренинг податоците. Моделот Artificial Neural Networks покажува нестабилни резултати со негативна вредност на R^2 , што укажува дека за дадениот број на податоци и нивната структура, овој модел не е соодветно конфигуриран или бара поголем обем на податоци за да постигне стабилни резултати.

Особено изразено отстапување се забележува кај Gaussian Process Regression, каде се добива екстремно висока вредност на грешка и силно негативен R^2 . Ова укажува на нестабилност на моделот во конкретниот случај, најверојатно поради чувствителноста на GPR на скалата на податоците и изборот на кернел функцијата.

И покрај значајните разлики во перформансите, сите модели ќе бидат задржани и користени во понатамошната фаза на обучување. Ваквиот пристап овозможува поширока анализа на однесувањето на системот и обезбедува основа за примена на ансамбл техники, при што се комбинираат предвидувањата од повеќе модели со цел да се добие поробусна и постабилна проценка. Дополнително, задржувањето на сите модели овозможува споредбена анализа на различни алгоритамски пристапи и нивната применливост во моделирањето на хидролошките процеси. Добиените резултати укажуваат дека нивото на водата во акумулацијата може успешно да се моделира со помош на машинско учење, што е во согласност со физичката природа на процесот, бидејќи оваа варијабла директно зависи од балансот помеѓу дотекот и истекот на водата.

6.3.3. Обучување на моделите

За обучување на моделите, и покрај релативно полошите резултати од петкратната вкрстена валидација за GPR моделот, тренирани се девет модели со сите девет алгоритми за да се согледаат резултатите од метричките показатели од претходниот чекор во фазата на проценка на предвидувањата. Накратко секој модел е опишан како што следи, заедно во хиперпараметрите внесени во кодот DamResML.

6.3.3.1. Модел на Вештачки невронски мрежи (ANN)

Прв обучен модел е моделот на Вештачки невронски мрежи или накратко ANN (анг. „Artificial Neural Network“). Во конкретниот DamResML код во кој е внесен алгоритмот за Вештачки невронски мрежи, во однос на параметрите, внесени се следните вредности:

1) Број на скриени слоеви n_{layer} и бројот на неврони во секој слој $n_{neurons}$ се внесени како `hidden_layer_sizes=(100,100)`. Со ова, внесени се 2 скриени слоеви со 100 неврони во секој од нив.

2) Типот на активациска функција е усвоен како `solver="adam"` што значи дека се применува Оптимизаторот Адам (анг. „Adam optimizer“) како активациска функција.

(3) Брзината на учење l_{rate} е усвоена како `learning_rate_init=0.001`.

Процесот на обучување траеше 7s и 777ms, со просечна вредност на $MSE = 155.997$, $RMSE = 10.639$, $MAE = 4.718$ и $R^2 = -3.240$. Резултатите добиени за моделот на вештачки невронски мрежи (ANN – MLP) укажуваат на слаба предиктивна способност, што се потврдува со негативната вредност на коефициентот на детерминација $R^2 = -3.24$. Иако вредностите на $RMSE$ (10.64) и MAE (4.72) не се екстремно високи, негативниот R^2 покажува дека моделот не успева да ја опише зависноста помеѓу влезните и излезните променливи и дава полоши резултати од едноставна апроксимација со средна вредност. Ова најверојатно се должи на ограничениот број на податоци и комплексноста на моделот, при што ANN бара поголем обем на податоци и внимателна калибрација за да постигне стабилни и сигурни резултати.

6.3.3.2. Модел на Регресија со потпомогнати вектори (SVR)

Вториот обучен модел е моделот на Регресија со потпомогнати вектори или накратко SVR (анг. „Support Vector Regressor“).

Моделот е дефиниран со следните параметри:

1) `kernel = "rbf"` - Се користи радијална базисна функција (Radial Basis Function), која овозможува моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) `C = 100` - Регуларизациски параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на C значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување.

3) `epsilon = 0.1` - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

За евалуација на моделот е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриките MSE , $RMSE$, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Средната вредност на метричките показатели по обучувањето на моделот изнесуваат: Avg $MSE = 20.619$, Avg $RMSE = 4.540$, Avg $MAE = 3.518$, Avg $R^2 = 0.443$. Процесот на обучување траеше 182ms. Резултатите добиени за моделот на Support Vector Regression укажуваат на умерено добра предиктивна способност, со позитивна вредност на коефициентот на детерминација $R^2 = 0.444$. Вредностите на $RMSE$ (4.54) и MAE (3.52) покажуваат релативно ниско ниво на грешка, што укажува дека моделот успешно ја апроксимира основната зависност помеѓу влезните параметри и нивото на водата во акумулацијата. Иако точноста не е на највисоко ниво во споредба со најдобрите модели,

резултатите потврдуваат дека SVR претставува стабилен и робустен пристап за моделирање на нелинеарните зависности од конкретната анализа на случај.

6.3.3.3. Модел на Случајна шума (RF)

Третиот обучен модел е моделот на Random Forest регресија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја.

За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators` (50, 100, 150) - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth` (None, 10, 20) - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split` (2, 5, 10) - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features` ("sqrt", "log2", None) - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијаблите.

За добивање на објективна проценка на перформансите применета вгнездена n -кратна валидација, и тоа:

- Внатрешна трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку т.н. „GridSearchCV“.
- Надворешна петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Овој пристап обезбедува непристрасна проценка на перформансите и ја намалува веројатноста за пренагласено прилагодување на моделот кон податоците. Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување. Во Табела 22 даден е табеларен приказ на најдобрите хиперпараметри од надворешната петкратна вкрстена валидација.

Табела 22. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	max_depth	max_features	min_samples_split	n_estimators
5	10	sqrt	2	100
4	20	sqrt	5	100
3	10	sqrt	5	50
2	10	sqrt	2	150
1	NaN	sqrt	5	150

Финалното обучување е направено согласно хиперпараметрите од четвртиот сет од петкратната надворешна валидација, со вредности: 'max_depth': 10, 'max_features': sqrt, 'min_samples_split': 2, 'n_estimators': 150. Процесот на обучување траеше 45s и 371ms. Резултатите добиени за подесениот модел на Random Forest укажуваат на највисока предиктивна способност меѓу анализираните модели, со најниска вредност на RMSE (3.90) и највисока вредност на коефициентот на детерминација $R^2 = 0.589$. Овие резултати покажуваат дека моделот успешно ја фаќа нелинеарната зависност помеѓу влезните параметри и нивото на водата во акумулацијата, обезбедувајќи стабилни и релативно точни предвидувања. Примената на вгнездената вкрстена валидација дополнително ја потврдува робусноста на моделот, бидејќи овозможува оптимизација на хиперпараметрите без преценување на перформансите. Ова го прави Random Forest најсоодветен модел за понатамошна анализа и примена во рамките на овој случај.

6.3.3.4. Модел Одлучувачко дрво (Decision Tree)

Четвртиот обучувачки модел е моделот на Decision Tree Regressor. Во овој случај, моделот е дефиниран со основните параметри (со `random_state = 42` за репродуктивност), без дополнителна регуларизација, што значи дека дрвото може да расте до максимална сложеност и да ги следи податоците детално.

За евалуација на хиперпараметрите е применета петкратна вкрстена валидација, при што податоците се делат на пет сетови, а моделот се обучува и тестира пет пати со различни комбинации.

Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста на моделот. Бидејќи Decision Tree не бара скалирање на влезните податоци, моделот се применува директно врз оригиналните вредности на секоја варијабла. Вредностите за метриците се прикажани во Табела 23.

Важно е да се напомене дека, иако Decision Tree често постигнува ниска просечна апсолутна грешка (MAE), тој може да биде склон кон пренагласено прилагодување и да прави поголеми грешки во поединечни случаи, што се рефлектира во повисоки вредности на MSE и RMSE.

Табела 23. Вредности на метричките показатели од процесот на обучување на DT моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	MSE	RMSE	MAE	R2
1	21.615	4.649	2.797	0.397
2	29.452	5.427	3.387	0.265
3	26.471	5.145	3.165	0.264
4	24.978	4.998	3.084	0.331
5	26.947	5.191	3.245	0.260

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 25.892, Avg RMSE: 5.081, Avg MAE: 3.135, Avg R^2 : 0.303. Процесот на обучување траеше 44ms.

6.3.3.5. Модел К – најблизок сосед (K – nearest neighbor)

Како петти трениран модел е моделот со K-Nearest Neighbors (KNN) регресија. Моделот е дефиниран со параметар `n_neighbors = 5`, што значи дека за секое предвидување се земаат предвид 5 најблиски примероци од податоците за обучување.

За проценка на перформансите е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста на моделот. Конкретните вредности за моделот се дадени во Табела 24. Треба да се напомене дека KNN не гради експлицитен модел (lazy learning), туку ги користи сите тренинг податоци при предвидување, што може да влијае на пресметковната ефикасност кај поголеми податочни множества.

Табела 24. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	MSE	RMSE	MAE	R2
1	21.2968	4.61485	3.34302	0.40579
2	20.5645	4.53481	3.33657	0.48711
3	22.3916	4.73197	3.44831	0.37728
4	19.8642	4.45693	3.19461	0.46766
5	19.9072	4.46174	3.34656	0.45342

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 20.804, Avg RMSE: 4.560, Avg MAE: 3.333, Avg R^2 : 0.438. Процесот на обучување траеше 51ms. Резултатите добиени за моделот на K-Nearest Neighbors укажуваат на умерено добра предиктивна способност, со вредност на коефициентот на детерминација $R^2=0.438$. Вредностите на RMSE (4.56) и MAE (3.33) покажуваат

дека моделот обезбедува релативно стабилни и конзистентни предвидувања, со ниво на точност слично на SVR моделот. Како метод базиран на локална сличност, KNN успешно ги фаќа локалните структури во податоците, но неговата точност може да биде ограничена во случаи со поголема комплексност и варијабилност на системот. Сепак, добиените резултати потврдуваат дека KNN претставува валиден и робустен пристап за моделирање на односот помеѓу влезните параметри и нивото на водата во акумулацијата.

6.3.3.6. Модел со XGBoost (XGBoost Regressor)

Шестиот трениран модел е моделот на XGBoost Regressor. Моделот е дефиниран со следните клучни параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но може да ја зголеми сложеноста на моделот.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое ново дрво. Помали вредности водат до постабилно учење, но бараат повеќе дрвја.

3) `objective = "reg:squarederror"` - Ја дефинира функцијата на загуба, при што се минимизира средноквадратната грешка (MSE).

4) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи XGBoost е базиран на дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета стандардизација на податоците.

За евалуација на моделот применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста и објаснувачката моќ на моделот. Пресметаните вредности на метриките од моделот се прикажани во Табела 25.

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 16.612, Avg RMSE: 4.074, Avg MAE: 2.984, Avg R²: 0.551. Процесот на обучување траеше 443ms.

Табела 25. Вредности на метричките показатели од процесот на обучување на XGBoost моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	MSE	RMSE	MAE	R2
1	16.026	4.003	2.925	0.553
2	16.808	4.100	3.025	0.581
3	16.389	4.048	2.899	0.544
4	15.714	3.964	2.949	0.579
5	18.124	4.257	3.127	0.502

6.3.3.7. Модел со Гаусов Процесен Регресор (GPR)

Седмиот модел за обучување се базира на алгоритмот Гаусов Процесен Регресор (анг. „Gaussian Process Regression“).

Во моделот е користен комбиниран кернел – т.н. „ConstantKernel × RBF (Radial Basis Function)“. Овој кернел овозможува моделирање на мазни нелинеарни односи помеѓу променливите, при што RBF ја контролира „должината на влијание“ (анг. „length scale“), а „ConstantKernel“ ја скалира вредноста на податоците.

Клучни параметри на моделот се:

1) `normalize_y = True` - Ја нормализира целната варијабла, што значително ја подобрува стабилноста на моделот.

2) `n_restarts_optimizer = 5` - Овозможува повеќекратно оптимизирање на параметрите на кернел функцијата, со цел да се избегне локален минимум и да се добие подобро вклопување.

Поради тоа што GPR е чувствителен на растојанија, применета е StandardScaler трансформација

во рамки на методологијата, со што сите податоци од сетот за обучување се стандардизираат во секој сет од петкратната вкрстена валидација. Перформансите се оценуваат преку MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста. Иако GPR има теоретски висока флексибилност и способност за моделирање на сложени соодноси помеѓу различните варијабли, во практични услови може да биде чувствителен на шум и ограничен при работа со поголеми и покомплексни податочни множества.

Вредностите за метричките показатели од процесот на петкратна вкрстена валидација се прикажани во Табела 26.

Табела 26. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	MSE	RMSE	MAE	R2
1	385320.193	620.742	63.024	-10749.967
2	1315.027	36.263	8.211	-31.798
3	145.560	12.065	5.099	-3.048
4	225.441	15.015	5.402	-5.042
5	612149.621	782.400	50.064	-16806.369

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 199831.168, Avg RMSE: 293.296, Avg MAE: 26.359, Avg R^2 : -5519.244. Процесот на обучување траеше 11m и 31s. Резултатите добиени за моделот на Gaussian Process Regression укажуваат на многу слаба и нестабилна предиктивна способност. Се забележуваат екстремно високи вредности на грешките кај одделни сетови (особено сет 1 и сет 5), како и силно негативни вредности на коефициентот на детерминација R^2 , кои достигнуваат и до -16806. Ова укажува дека моделот не само што не ја фаќа зависноста помеѓу влезните и излезната променлива, туку дава резултати значително полоши од едноставна апроксимација со средна вредност. Големата варијабилност помеѓу fold-овите дополнително укажува на нестабилност на моделот, што најверојатно е резултат на чувствителноста на GPR на скалата на податоците, изборот на kernel функцијата и ограничениот број на податоци. Во вакви услови, примената на GPR не се покажува како соодветна за конкретниот проблем.

6.3.3.8. Модел со Поткрени Регресиони Дрва (BT)

Осмиот модел за обучување се базира на алгоритмот на BT, односно Gradient Boosting Regressor, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја. Моделот е дефиниран со следните параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на моделот е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 27. Gradient Boosting, како бустинг метод, е особено ефикасен во моделирање на сложени нелинеарни односи и често постигнува висока предиктивна точност.

Табела 27. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	MSE	RMSE	MAE	R2
1	16.515	4.064	3.222	0.539

Fold	MSE	RMSE	MAE	R2
2	17.939	4.235	3.256	0.553
3	16.763	4.094	3.195	0.534
4	16.686	4.085	3.142	0.553
5	17.540	4.188	3.267	0.518

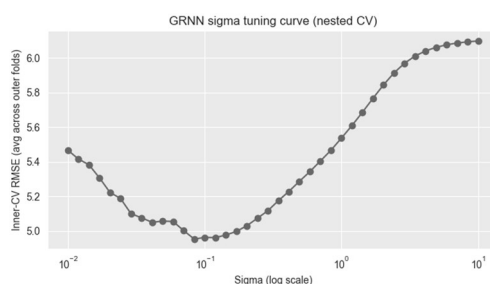
Просечните вредности на метричките показатели, се: Avg MSE: 17.088, Avg RMSE: 4.133, Avg MAE: 3.216, Avg R²: 0.539. Потребното време за обучување беше 274ms. Резултатите добиени за моделот на Boosted Trees укажуваат на стабилна и релативно висока предиктивна способност. Вредностите на RMSE се конзистентни во сите сетови на валидацијата (околу 4.06–4.24), што укажува на ниска варијабилност и добра робусност на моделот. Дополнително, позитивните вредности на коефициентот на детерминација R², кои се движат помеѓу 0.518 и 0.553, покажуваат дека моделот успешно ја објаснува значаен дел од варијабилноста на нивото на водата во акумулацијата. Ваквата конзистентност низ сите сетови укажува дека Boosted Trees претставува сигурен и стабилен пристап за моделирање на нелинеарните поврзаности на параметрите од анализираната студија на случај.

6.3.3.9. Модел со Генерализирана Регресиона Невронска Мрежа (GRNN)

Последниот модел е обучуван на алгоритмот на Генерализирана Регресиона Невронска Мрежа (анг. „General Regression Neural Network“). Во кодот е применета вгнездена трикратна валидација за избор на најдобра вредност на параметарот σ – сигма. Се тестираат вредности во рангот $\sigma = (0.01 - 10)$. Критериумот е да се избере вредност за σ што ќе резултира со минимален RMSE, а во конкретниот случај графикот на варијација на RMSE во зависност од параметарот σ по трикратната вкрстена валидација е прикажан во Прилог 51. Понатаму, со изборот на соодветна вредност за параметарот σ и дополнително се прави петкратна надворешна вкрстена валидација на податоците.

Табела 28. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Fold	Best Sigma	Fold RMSE
1	0.035	4.662
2	0.084	5.037
3	0.084	4.667
4	0.143	4.690
5	0.084	4.764



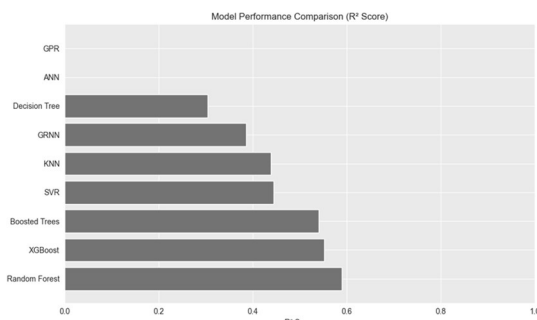
Прилог 51. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Од графикот во Прилог 51 може да се воочи дека минимумот на RMSE = 4.662 се добива за вредност на параметарот $\sigma \approx 0.03$. Резултатите добиени за моделот на General Regression Neural Network укажуваат на умерено добра и релативно стабилна предиктивна способност. Вредностите на RMSE се движат во тесен опсег (приближно 4.66–5.04), што укажува на конзистентност на моделот низ различните сетови. Варијацијата на параметарот σ помеѓу 0.035 и 0.143 покажува дека оптималното ниво на „измазнување“ се прилагодува во зависност од податочниот сет, но без значително влијание врз стабилноста на резултатите. Ова укажува дека GRNN успешно ја фаќа основната структура на податоците, иако со нешто пониска точност во споредба со најдобрите ансамбл модели.

6.3.4. Резултати од обучувањето на моделите

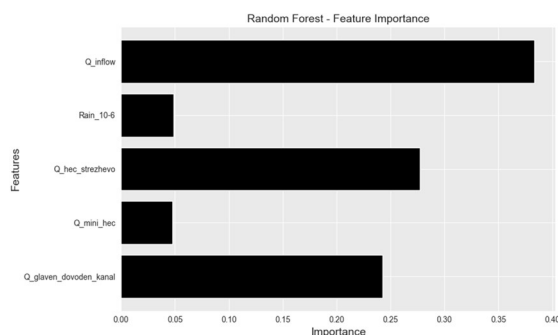
По обучување на девет различни модели базирани на девет различни алгоритми, како што следи презентирани се сумарни резултати од сите модели за претходната фаза во форма на табели и графици.

Резултатите од примената на девет различни модели на машинско учење за предвидување на нивото на водата во акумулацијата „Zwl“ покажуваат јасни разлики во нивната предиктивна способност. Најдобри перформанси се постигнати со ансамбл моделите, при што Random Forest покажува најниска вредност на RMSE и највисока вредност на R^2 , што укажува на највисока точност и робусност. Слично, XGBoost и Boosted Trees покажуваат стабилни и конзистентни резултати, со релативно високи вредности на R^2 и ниска варијабилност помеѓу сетовите. Моделите како Support Vector Regression, K-Nearest Neighbors и General Regression Neural Network покажуваат умерено добра предиктивна способност, при што успешно ја апроксимираат основната зависност, но со нешто пониска точност. Од друга страна, Decision Tree моделот има ограничена способност за генерализација, што се должи на неговата поедноставена структура. Моделот на Artificial Neural Networks покажува нестабилни резултати со негативен R^2 , што укажува на недоволна калибрација и чувствителност на обемот на податоци. Особено слаби резултати се добиени кај Gaussian Process Regression, каде се забележува екстремна варијабилност и многу високи грешки, што го прави моделот несоодветен за овој проблем.



Прилог 52. Приказ на R^2 за секој обучен модел за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

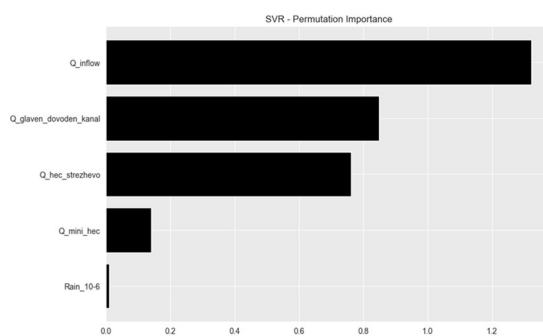
За моделот обучуван со RF, DamResML кодот е програмиран да прикаже график на важност на варијаблите во моделот. Важноста на варијаблите во моделот покажува колкаво е влијанието на предиктор варијаблите во предвидувањето на целната варијабла. За моделот обучуван со RF, овој график е прикажан во Прилог 53. Се согледува дека најголема важност има варијаблата „Q_inflow“ – дотекот во акумулацијата, потоа „Q_hec_strezhevo“ – енергетски обработениот проток во ХЕЦ „Стрежево“ како истекување низводно од браната, и „Q_glaven_dovoden_organ“ како истекување кон каналот за ХМС „Стрежево“ низводно од браната. Предиктор варијаблите „Rain_10-6“ – паднат дожд на повшината на акумулацијата, и „Q_hec_mini“ – енергетски обработен проток во мини хидроцентралата низводно од браната, имаат значително помала важност споредена со претходните три предиктор варијабли.



Прилог 53. График на важност на варијаблите (анг. „Feature importance“) во моделот со RF за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

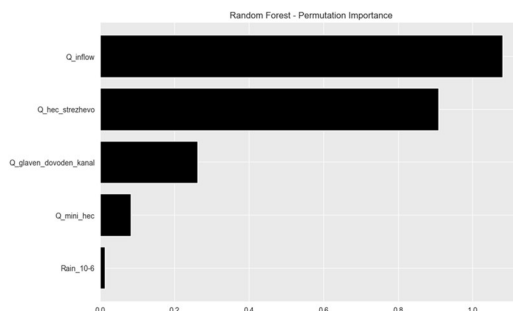
Дополнително, во DamResML кодот програмирано е извлекувањето на пермутациската важност на моделите обучувани со SVR, RF и DT. Пермутациската важност претставува метод за оценување на значењето на предиктор варијаблите врз основа на нивното влијание врз перформансите на моделот. Основната идеја се состои во тоа што вредностите на една променлива случајно се мешаат (пермутираат), со што се нарушува нејзината врска со целната варијабла. Потоа се мери промената во точноста на моделот. Доколку пермутацијата на одредена променлива доведе до значително зголемување на грешката (на пример RMSE), тоа укажува дека моделот во голема мера се потпира на таа променлива. Напротив, ако промената е мала, тогаш нејзиното влијание е ограничено.

На графициите во Прилог 54, Прилог 55 и Прилог 56 прикажани се графици со пермутациска важност за моделите обучени со SVR, RF и DT алгоритмите.

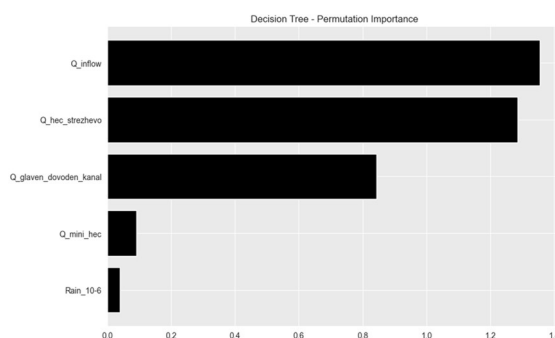


Прилог 54. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со SVR за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Сличен тренд на предиктор варијаблите се провлекува во сите три графици на пермутациска важност – најголема важност има предиктор варијаблата “Q_inflow“, потоа “Q_hec_strezhevo“ и “Q_gl_dovoden_organ“ – т.е. дотекувањата и истекувањата од акумулацијата. Значително помала важност имаат параметрите “Q_mini_hec“ и “Rain_10-6“ што значи дека протокот кој се испушта во мини хидроцентралата како и паднатиот дожд на површината на самата акумулација имаат многу мала важност во формирањето на котата на водното огледало во акумулацијата „Стрежево“.



Прилог 55. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со RF за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.



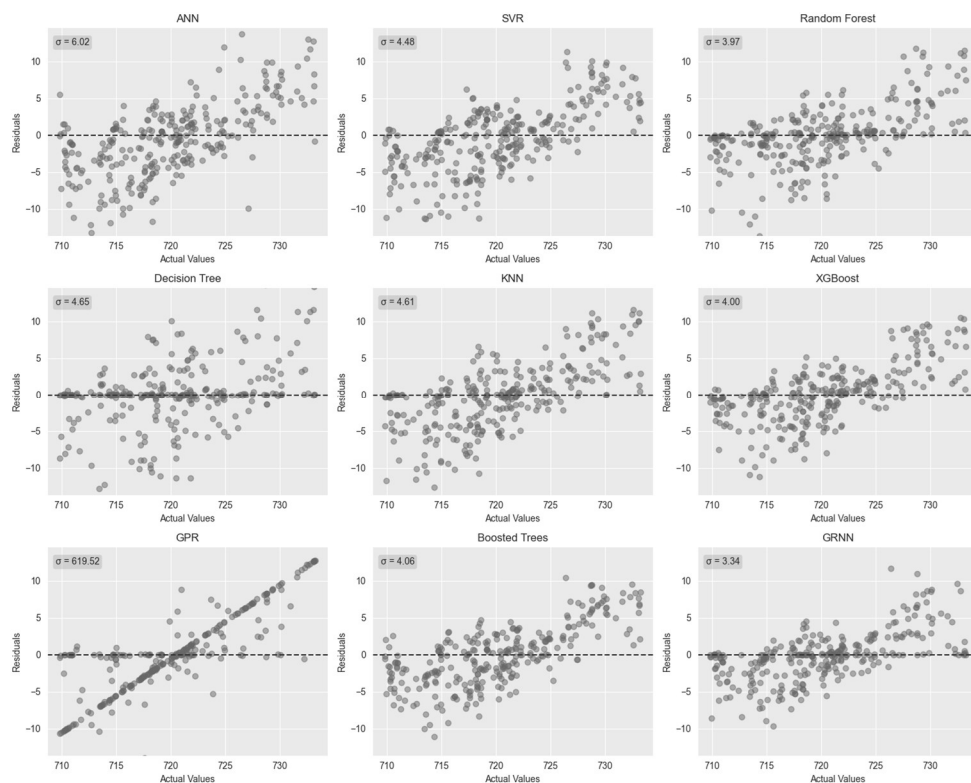
Прилог 56. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со DT за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

За секој обучуван модел, во DamResML кодот програмирано е да се прикаже график на резидуали (остатоци) – грешки во предвидувањата од секој модел. Сумарен график кој ги опфаќа сите 9 модели е прикажан во Прилог 57. Резидуалите се пресметуваат врз основа на кратко предвидување со обучените модели, спроведено на т.н. задржан сет на податоци (анг. „holdout“) кој не бил претходно вклучен во процесот на обучување на моделите, па така истиот е непознат за моделот. Ова издвојување на задржан сет настрана од сетот за обучување кодот го прави уште во првата ќелија од моделот кодот.

Прикажаните графици ја претставуваат резидуалната анализа за девет различни модели на машинско учење. Резидуалите ја претставуваат разликата помеѓу реалните и предвидените вредности, односно грешката на моделот. Во идеален случај, точките треба да бидат случајно распоредени околу хоризонталната нулта линија, без јасен тренд, со мала вертикална дисперзија. Таквата појава укажува дека моделот е стабилен, непристрасен и успешно ја опишува зависноста во податоците.

Анализата на резидуалите за применетите модели на машинско учење дава дополнителен увид во нивната точност и стабилност при предвидувањето на нивото на водата во акумулацијата (Z_{wl}). Кај поголемиот дел од моделите се забележува дека резидуалите се распоредени околу нултата линија, што укажува на отсуство на изразена систематска пристрасност, односно моделите во просек не го преценуваат ниту потценуваат предвидувањето. Сепак, се забележува одреден тренд на зголемување на резидуалите со пораст на реалните вредности, што укажува на постоење на нелинеарност и потенцијално недоволно сфаќање на екстремните вредности од страна на дел од моделите.

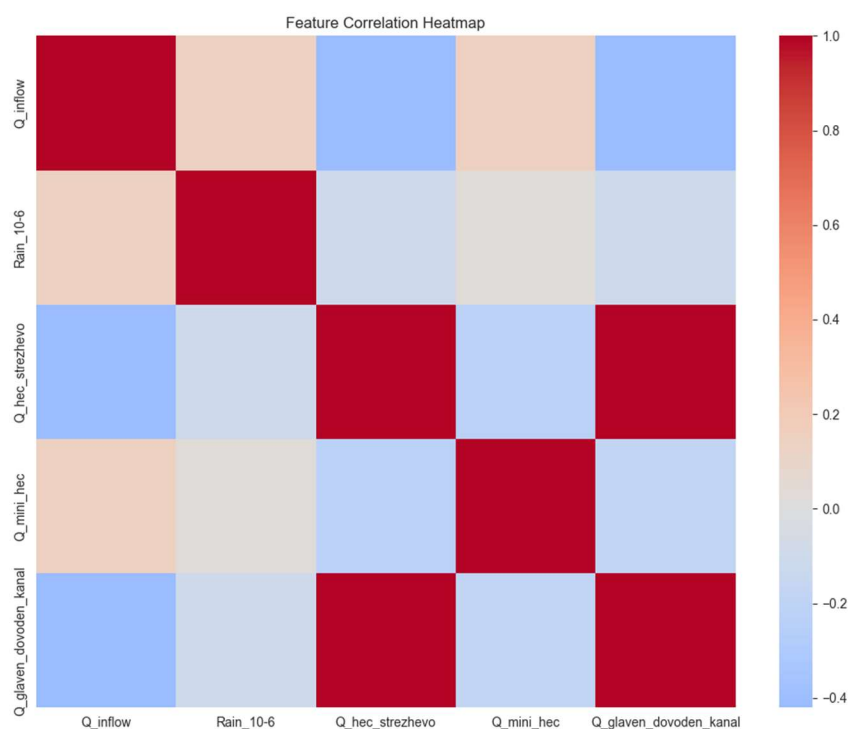
Особено добри резултати покажуваат ансамбл моделите, како Random Forest, XGBoost и Boosted Trees, кај кои резидуалите се концентрирани во релативно тесен опсег околу нулата и покажуваат мала дисперзија. Ова укажува на висока стабилност и способност на овие модели да ја фатат комплексната нелинеарна зависност во системот. Слично, Support Vector Regression и K-Nearest Neighbors покажуваат умерено добра распределба на резидуалите, со нешто поголема варијабилност, но сепак задржуваат задоволителна точност.



Прилог 57. Графици со резидуали за секој обучуван модел за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Од друга страна, Artificial Neural Networks и Decision Tree покажуваат поголема дисперзија на резидуалите, што укажува на помала стабилност и поголема чувствителност на податоците. Најизразени отстапувања се забележуваат кај Gaussian Process Regression, каде се јавуваат екстремни вредности на резидуалите и многу голема дисперзија, што укажува на нестабилност и несоодветност на овој модел за конкретниот проблем. Од друга страна, моделот General Regression Neural Network покажува најмала дисперзија на резидуалите, што укажува на висока конзистентност, иако тоа не секогаш значи и највисока точност во апсолутна смисла.

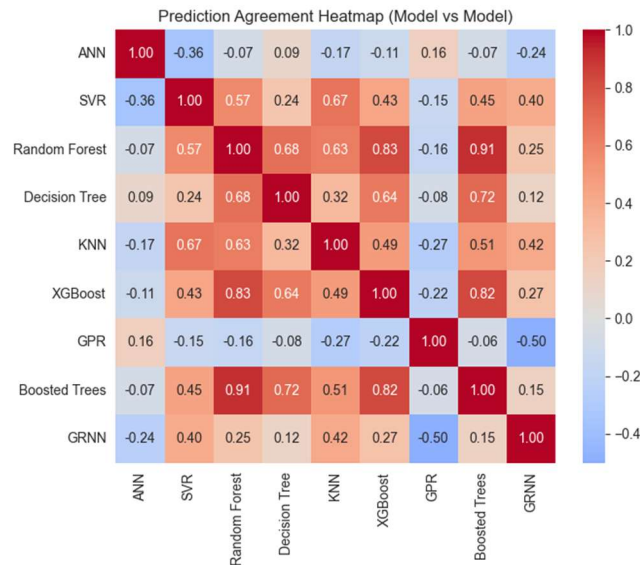
Во оваа анализа, стандардната девијација на резидуалите σ се користи како клучен показател за оценка на стабилноста на моделите. За разлика од RMSE, кој ја мери просечната големина на грешката, σ ја мери дисперзијата на грешките околу нивната средна вредност, односно колку се варијабилни предвидувањата на моделот. Помала вредност на σ укажува на поконзистентни и стабилни предвидувања, додека поголема вредност укажува на присуство на нестабилност и екстремни отстапувања. Ова е особено важно во инженерски апликации, каде стабилноста на моделот е од суштинско значење за доверливост на резултатите.



Прилог 58. Топлотна мапа со приказ на поврзанооста на предиктор варијаблите за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Анализата на корелационата матрица помеѓу предиктор параметрите прикажани во Прилог 58 укажува на постоење на различни степени на зависност помеѓу разгледуваните променливи. Најизразена корелација се забележува помеѓу протоците „Q_hec_strezhevo“ и „Q_glaven_dovoden_kanal“, при што коефициентот на корелација е приближно еднаков на 1, што укажува на речиси целосна линеарна зависност и потенцијална редундантност на овие параметри. Ова укажува дека двете променливи носат речиси идентична информација и нивното истовремено користење во моделот може да не придонесе значително за подобрување на предиктивната способност.

Помеѓу останатите променливи се забележуваат слаби до умерени корелации, што укажува на комплексна и нелинеарна зависност во системот. На пример, умерената позитивна корелација помеѓу врнежите и дотокот е во согласност со физичката природа на процесот, додека негативната корелација помеѓу дотокот и протоците низ хидроцентралите укажува на постоење на управувачки механизми, при кои зголемувањето на дотокот не мора веднаш да резултира со зголемено испуштање на вода.



Прилог 59. Топлотна мапа со приказ на сличноста во предвидувањата од различните модели тренирани на различни алгоритми за анализа на случај 1.3 – $Z_{wl} = f(Q_{inflow}, Rain, Q_{hec_strezhevo}, Q_{mini_hec}, Q_{gl_dovoden_organ})$.

Анализата на сличноста во предвидувањата на различните модели прикажана во топлотната мапа на Прилог 59 покажува постоење на групи на алгоритми со слично однесување. Особено висока корелација се забележува помеѓу ансамбл моделите како Random Forest, XGBoost и Boosted Trees, што укажува дека овие модели ја фаќаат истата основна структура на податоците и даваат речиси идентични предвидувања.

Од друга страна, моделите како Support Vector Regression и K-Nearest Neighbors покажуваат умерена корелација со ансамбл моделите, што укажува дека тие носат дополнителна информација и претставуваат добар избор за комбинирање во ансамбл. Особено значајно е што одредени модели, како Gaussian Process Regression, покажуваат слаба или негативна корелација со останатите, што укажува на нестабилност и различно однесување, па нивната употреба треба внимателно да се разгледа.

6.3.5. Предвидување на целната варијабла

Целна варијабла во следната студија на случај е “ Z_{wl} ” - нивото во акумулацијата „Стрежево“. Со помош на обучени модели од машинско учење, се бара предвидување на целната варијабла со познавање на следните предиктор варијабли:

- 1) „ Q_{inflow} “ – регистриран дотек во акумулацијата во $[m^3/s]$ прикажан во Прилог 47,
- 2) „ $Q_{hec_strezhevo}$ “ – регистриран енергетски обработен проток низ ХЕЦ „Стрежево“ во $[m^3/s]$ прикажан во Прилог 47,
- 3) „ Q_{hec_mini} “ – регистриран енергетски обработен проток низ мини ХЕЦ за биолошки минимум, во $[m^3/s]$ прикажан во Прилог 47,
- 4) „ $Q_{gl_dovoden_organ}$ “ – регистрирано испуштање на вода низ главниот канал за ХМС „Стрежево“ во $[m^3/s]$ прикажан во Прилог 47, и
- 5) „Rain“ – паднат дожд на површината на акумулацијата регистриран во $[10^6 m^3]$ прикажан во Прилог 48.

Обучени се вкупно девет модели со машинско учење базирани на следните алгоритми: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Процесот на предвидување се извршува за 1091 познати вредности на предиктор варијаблите, кои временски се однесуваат на периодот на прибирање на податоците од 1ви јануари 2010 година до 31 декември 2012 година. Овде е интересно да се напомене дека за периодот на предвидување позната е точната вредност на нивото во акумулацијата, со што финално ќе се направи споредба на предвидувањата заедно со она што реално е регистрирано како ниво на вода во акумулацијата „Стрежево“.

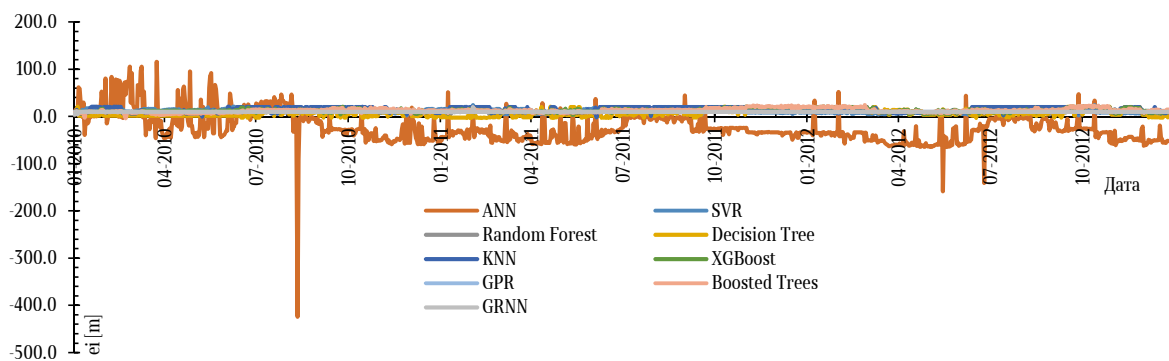
Предиктор варијаблите се внесуваат во DamResML кодот преку датотеката „predictor_variable_for_prediction.xls“ и истата се повикува во ќелијата „Prediction“ по завршување на процесот на обучување на деветте модели.

Резултатите од предвидувањата се прикажани во прилозите кои следат.

За да стане појасно колкави се разликите помеѓу предвидувањата y_{pred} од секој модел и реалните вредности на целата варијабла y , пресметани се грешки e_i – резидуали како разлика помеѓу овие две вредности:

$$e_i = y_i - y_{pred} \quad (6.3.1)$$

Ваквата анализа е особено важна бидејќи не ја прикажува само просечната точност на моделот, туку овозможува увид во неговото однесување низ целиот анализиран период. Доколку резидуалите се блиски до нула и се рамномерно распоредени околу нултата линија, моделот може да се смета за стабилен и непристрасен. Наспроти тоа, доколку резидуалите се претежно позитивни или претежно негативни, тоа укажува на систематско потценување или преценување на реалните вредности.



Прилог 60. Графички приказ на пресметаните разлики - e_i помеѓу предвидените и реалните вредности на целната варијабла, за секој трениран модел.

Во конкретната анализа, графичкиот приказ на разликите помеѓу реалните и предвидените вредности овозможува прикажан во Прилог 60 дава јасна споредба на деветте применети модели. Позитивните резидуали укажуваат дека моделот ја потценил реалната вредност, односно дека предвидената вредност е помала од измерената. Негативните резидуали укажуваат на спротивното, односно дека моделот ја преценил реалната вредност. Затоа, за квалитетен модел не е доволно само грешките да бидат мали по големина, туку тие треба да бидат и без јасно изразена систематска насока. Најдобар модел е оној кај кој отстапувањата се мали, стабилни и приближно центрирани околу нулата.

Од споредбата на моделите може да се заклучи дека Random Forest покажува најповолно однесување. Разликите помеѓу реалните и предвидените вредности кај овој модел се релативно мали и стабилни, без изразени екстремни осцилации. Ова укажува дека моделот успешно ја фаќа основната зависност помеѓу влезните параметри и целната променлива, притоа задржувајќи добра способност за генерализација. Ваквото однесување е во согласност и со претходно добиените метрички показатели, каде Random Forest имаше најниска вредност на RMSE и највисока вредност на R^2 . Дополнително, стабилноста на резидуалите покажува дека моделот не е чувствителен на поединечни вредности и не покажува изразена пристрасност.

Моделот SVR исто така покажува релативно добро однесување, но со нешто поголеми

отстапувања во споредба со Random Forest. Неговите резидуали се во поголем опсег, што укажува дека моделот успева да го следи општиот тренд, но со помала точност при одредени состојби на системот. Ова е очекувано, бидејќи SVR добро моделира нелинеарни зависности, но неговата точност зависи од изборот на параметри, скалирањето на податоците и структурата на влезниот простор.

Кај XGBoost и Boosted Trees се забележува солидна предиктивна способност, но нивните разлики имаат поголема варијабилност. Овие модели успешно ја фаќаат нелинеарната структура на податоците, но кај одредени делови од анализираниот период покажуваат поголеми отстапувања. Тоа укажува дека иако ансамбл методите базирани на бустинг се моќни, во конкретниот случај Random Forest обезбедува постабилна проценка.

GRNN покажува интересно однесување бидејќи разликите се релативно стабилни, но често се групираат околу слична вредност. Ова може да укажува на присуство на систематска пристрасност, односно моделот конзистентно потценува или преценува дел од реалните вредности. Таков модел може да изгледа стабилен, но стабилноста сама по себе не е доволна доколку постои систематска грешка. Поради тоа, GRNN може да се смета за робустен, но не и за најпрецизен модел.

Од друга страна, ANN и GPR покажуваат значително понестабилно однесување. Кај ANN се јавуваат големи позитивни и негативни отстапувања, што укажува на нестабилност и недоволна способност за генерализација. Ова најверојатно е последица на недоволен број на податоци за сигурно обучување на невронска мрежа или несоодветна архитектура на моделот. Кај GPR се забележува речиси константна грешка или нетипично однесување, што покажува дека моделот не успеал соодветно да ја научи зависноста помеѓу променливите. Decision Tree и KNN исто така покажуваат ограничувања, при што Decision Tree може да биде премногу локално прилагоден на тренинг податоците, додека KNN покажува поголема чувствителност на распределбата на податоците.

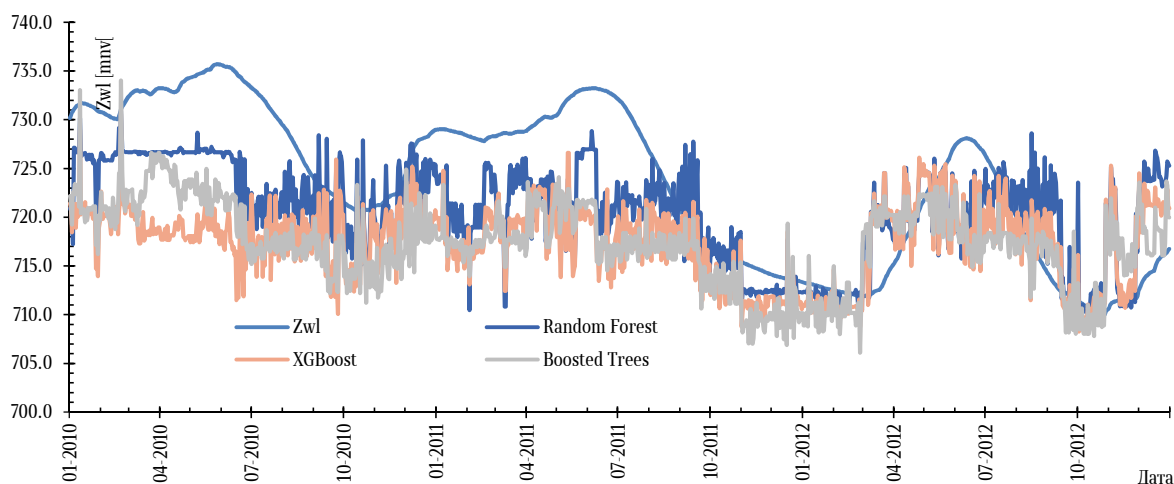
Табела 29. Нумерички вредности на максималните, минималните и просечните разлики помеѓу реалните вредности на целната варијабла и пресметаните вредности со тренираните модели.

	ANN	SVR	Random Forest	Decision Tree	KNN	XGBoost	GPR	Boosted Trees	GRNN
max	115.71	23.93	19.92	20.36	20.15	21.91	9.69	24.08	19.93
min	-424.03	-0.51	-0.67	-3.06	-2.11	-0.44	9.68	-3.85	4.02
average	-24.23	11.04	9.75	7.43	14.78	12.80	9.69	12.92	9.77

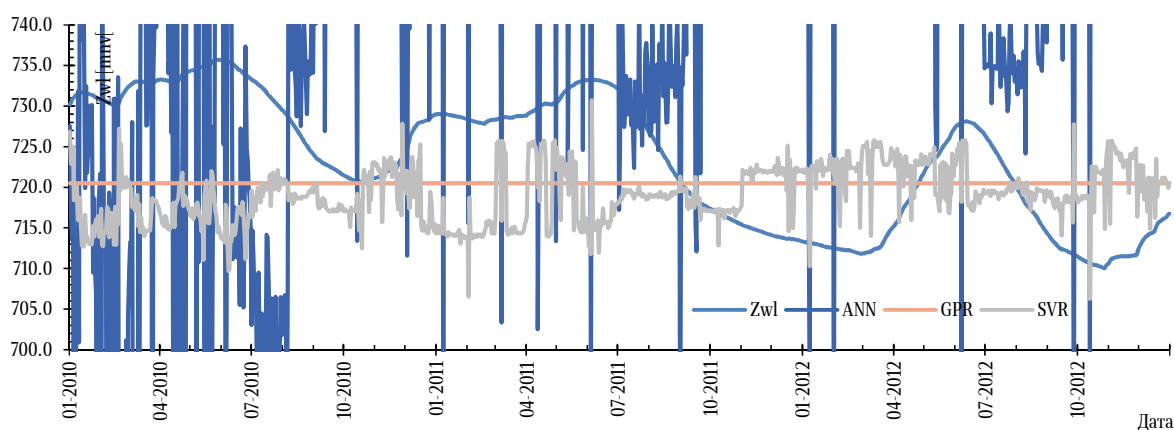
Анализата на разликите помеѓу реалните и предвидените вредности потврдува дека Random Forest е најсоодветен модел за понатамошна примена. Неговата предност не се состои само во најдобрите статистички показатели, туку и во стабилното однесување на резидуалите. Ова е особено важно бидејќи моделот треба да обезбеди доверливи и конзистентни предвидувања во различни работни услови. Затоа, Random Forest може да се усвои како основен модел за понатамошни анализи, вклучително и за примена во хаос оптимизација.

Дополнително, за приказ на предвидените вредности со секој трениран модел, изработени се три прилози со графици каде поединечно може да се согледаат предвидувањата споредени со реалните вредности на нивото на водата во акумулацијата. Графиците се прикажани во Прилог 61, Прилог 62 и Прилог 63. На трите графици прикажани во Прилог 61 со предвидените вредности од моделите XGBoost, RF и BT може да се воочи дека предвидувањата го следат трендот на реалните вредности на целната променлива, но се јавуваат значителни отстапувања во периодите кога е полна акумулацијата, т.е. моделите не можат во целост да ги објаснат максималните нивоа во акумулацијата со познавање на внесените предиктор варијабли. Од друга страна, со анализа на трите графици од предвидувањата пресметани со моделите ANN, GPR и SVR прикажани во Прилог 62, предвидувањата воопшто не го следат трендот на реалните вредности нити пак предвидуваат приближно реални вредности, особено моделот обучуван со ANN. Предвидувањата од моделот со GPR се речиси константна вредност, додека предвидувањата од моделот со SVR значително отстапуваат од реалните вредности на целната варијабла. Резултатите од последните три модели DT, GRNN и KNN се прикажани на графици во Прилог 63. Предвидувањата од DT моделот

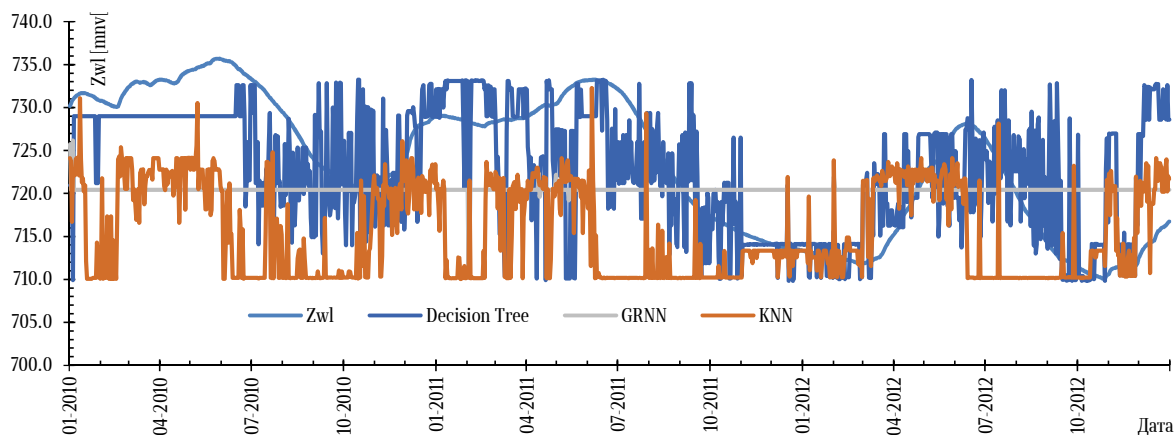
варираат доста често, иако покажуваат следење на трендот на реалните вредности. Честата промена на вредноста која делува како графикот да има постојани скокови, е резултат на нецелосното сфаќање на целната варијабла во однос на предиктор варијаблите. Предвидувањата со KNN моделот делимично го опфаќаат трендот на реалните вредности на целната варијабла, но со значително пониски вредности. Предвидувањата од GRNN моделот се речиси права линија, што претставува целосно потфрлање во предвидувањата.



Прилог 61. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со XGBoost, RF и BT, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.



Прилог 62. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со ANN, GPR и SVR, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.



Прилог 63. Графички приказ на пресметаните предвидувања на целната варијабла со моделите обучени со DT, GRNN и KNN, споредени со реалните одржани коти во акумулацијата за разгледуваниот период прикажани со графикот „Zwl“.

6.3.6. Хаос оптимизација на целната варијабла

Како последен чекор од процесот на подучување на моделите и предвидување на целната варијабла, применета е хаос оптимизација со цел оптимизирање на целната варијабла за конкретни ограничувања на целната варијабла и предиктор варијаблите.

Во конкретниот случај, целната варијабла е водното ниво во акумулацијата „Стрежево“ – „Zwl“, додека предиктор варијаблите се „Q_inflow“ – дотек во акумулацијата, „Rain_10-6“ – паднат дожд на површината на акумулацијата, „Q_hec_mini“ – енергетски обработен проток низ малата хидроцентрала, „Q_hec_strezhevo“ – енергетски обработен проток низ ХЕЦ „Стрежево“ и „Q_gl_dovoden_organ“ – проток испуштен во каналот за мелиорации за ХМС „Стрежево“.

Во оваа фаза, со хаос оптимизација решенијата ќе се бараат согласно множеството предвидувања на целната варијабла добиени со најдобриот модел од претходната евалуација – моделот базиран на Random Forest алгоритмот. Станува збор за вкупно 1091 податок добиен исклучиво со предвидување на целната варијабла.

Во конкретната оптимизациска задача, со примена на хаос оптимизација ќе се пребарува максимум на следната функција:

$$\text{MAX } F(\text{Zwl}): \text{Zwl} = f(\text{Q_inflow}, \text{Rain_10_6}, \text{Q_hec_mini}, \text{Q_hec_strezhevo}, \text{Q_gl_dovoden_kanal})$$

, со земање предвид на следните ограничувања на предиктор варијаблите:

- 1) Дотек во акумулацијата, $\text{Q_inflow} = (0.0, 5.0) \text{ m}^3/\text{s}$,
- 2) Паднат дожд на површината на акумулацијата, $\text{Rain_10-6} = (0.0, 0.10) 10^6 \text{ m}^3$,
- 3) Енергетски обработен проток низ ХЕЦ „Стрежево“, $\text{Q_hec_strezhevo} = (2.0, 9.0) \text{ m}^3/\text{s}$,
- 4) Енергетски обработен проток низ мини ХЕЦ, $\text{Q_hec_mini} = (0.1, 0.3) \text{ m}^3/\text{s}$, и
- 5) Проток кој се испушта кон главен доведен канал за ХМС „Стрежево“, $\text{Q_gl_dovoden_kanal} = (6.0, 9.0) \text{ m}^3/\text{s}$.

Станува збор за оптимизациска задача каде ќе се бара максимално ниво во акумулацијата за периоди на зголемени побарувања за производство на енергија како и зголемени побарувања на вода за наводнување, во услови на намалени дотекувања во акумулацијата.

Зголемените побарувања и намалениот доток се внесени во ограничувањата на предиктор варијаблите, но ограничувања се поставени и на целната варијабла, и тоа оптимумот се бара во следните високи граници:

$$\text{Zwl} = (720, 735) \text{ mnv}$$

По извршеното пребарување, со примена на хаос оптимизацијата пронајден е следниот максимум на функцијата $\text{MAX } (F(\text{Zwl}))$:

$$\text{MAX } F(\text{Zwl}): \text{Zwl} = 729.682 \text{ mnv}$$

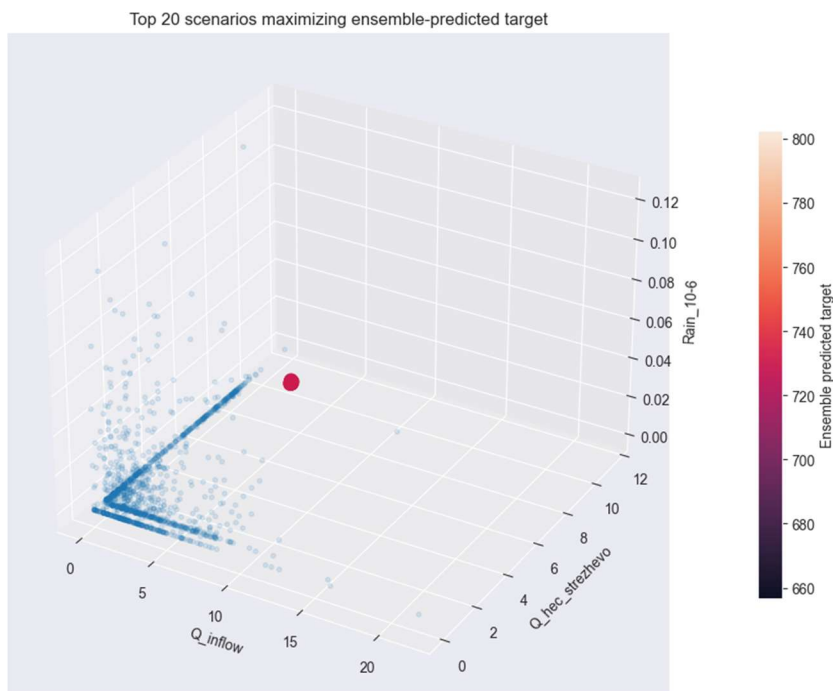
, за следната комбинација на предиктор варијабли:

- 1) Дотек во акумулацијата, $\text{Q_inflow} = 3.747 \text{ m}^3/\text{s}$,
- 2) Паднат дожд на површината на акумулацијата, $\text{Rain_10-6} = 0.0104 \cdot 10^6 \text{ m}^3$,
- 3) Енергетски обработен проток низ ХЕЦ „Стрежево“, $\text{Q_hec_strezhevo} = 8.606 \text{ m}^3/\text{s}$,
- 4) Енергетски обработен проток низ мини ХЕЦ, $\text{Q_hec_mini} = 0.144 \text{ m}^3/\text{s}$, и
- 5) Проток кој се испушта кон главен доведен канал за ХМС „Стрежево“, $\text{Q_gl_dovoden_kanal} = 8.382 \text{ m}^3/\text{s}$.

Бидејќи оптимизационите задачи треба да ни дадат одговор на инженерските оптимизациски проблеми како конкретната функција на максимизација на водното ниво во акумулацијата, потребно е да се елаборира детална анализа на добиените вредности на предиктор и целната варијабла. Најпрво се разгледува целната варијабла со пресметаниот оптимум кој изнесува $\text{Zwl} =$

729.682 mmv. Имајќи предвид дека нормалното ниво во акумулацијата е $Z_{nn} = 735.50$ mmv, а минималното работно ниво во акумулацијата е $Z_{min} = 694$ mmv, пресметаната вредност на целната функција во оптимизациската задача е навистина висока – само 5.818 m под нормалното работно ниво во акумулацијата. Истовремено, за одржување на оваа кота, пресметаните оптимални вредности на предиктор варијаблите изнесуваат $Q_{inflow} = 3.747 \text{ m}^3/\text{s}$ – блиску до средната вредност на дотекот која е $Q_{sr_inflow} = 2.894 \text{ m}^3/\text{s}$, $Rain_{10-6} = 0.0104 \cdot 10^6 \text{ m}^3$ – вредност под 10% од максимумот на регистрираниот дожд паднат на акумулацијата, $Q_{hec_strezhevo} = 8.606 \text{ m}^3/\text{s}$ – повисока вредност и од максималната внесена вредност за обработен енергетски проток низ ХЕЦ „Стрежево“ $Q_{max_hec_strezhevo} = 8.438 \text{ m}^3/\text{s}$, $Q_{mini_hec} = 0.144 \text{ m}^3/\text{s}$ – вредност еднаква на речиси 50% од просечниот енергетски обработен проток низ мини ХЕЦ кој изнесува $Q_{mini_hec} = 0.212 \text{ m}^3/\text{s}$ и $Q_{glaven_dovoden_organ} = 8.382 \text{ m}^3/\text{s}$ – вредност на испуштената вода за ХМС „Стрежево“ речиси еднаква на регистрираниот максимум од побарувачката кој изнесува $Q_{max_glaven_dovoden_organ} = 8.519 \text{ m}^3/\text{s}$. Со оптимизациската задача може да се заклучи дека пронајдени се решенија за предиктор варијаблите кои се во рамки на регистрираните оперативни вредности на хидросистемот, а со тоа истите претставуваат реални состојби кои може да се реплицираат во управувањето со хидросистемот.

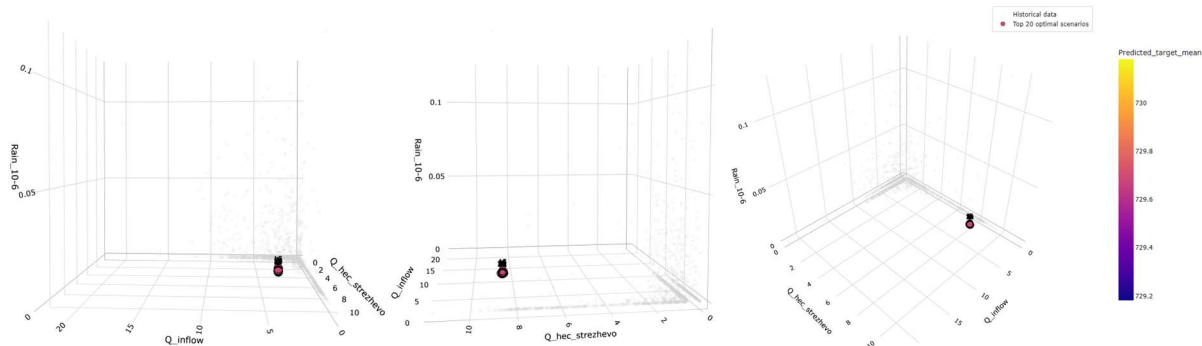
Тридимензионалниот приказ на оптималните сценарија добиени со примена на хаос оптимизацискиот алгоритам (CGO) прикажани во Прилог 64 Прилог 65 овозможува визуелна анализа на односот помеѓу предиктор варијаблите и предвидените вредности на нивото на водата во акумулацијата. На графикот со светло сина боја се прикажани сите вредности од сетот на податоци низ кои се пребарува максимумот на целната функција во 3Д простор на предиктор варијаблите „ Q_{inflow} “, „ $Q_{hec_strezhevo}$ “ и „ $Rain_{10-6}$ “, додека со потемна/црвена точка прикажани се издвоените 20 најдобри пронајдени решенија.



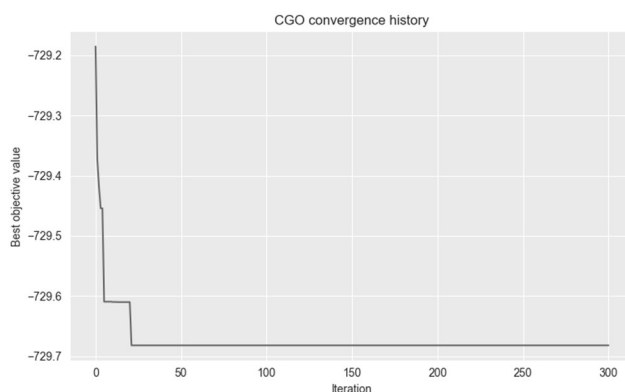
Прилог 64. 3Д визуелизација на пронајдените оптимални максимуми на целната функција $MAX F(Zwl)$; $Zwl = f(Q_{inflow}, Rain_{10-6}, Q_{hec_mini}, Q_{hec_strezhevo}, Q_{gl_dovoden_kanal})$, со обојување на издвоени резултати.

Од графикот може да се забележи дека оптималното сценарио е концентрирано во област со релативно повисоки вредности на дотокот „ Q_{inflow} “ и умерени вредности на протокот низ хидроелектраната „ $Q_{hec_strezhevo}$ “, додека врнежите имаат помала директна варијација во анализираниот простор. Ова укажува дека дотокот претставува доминантен фактор за зголемување на нивото на водата во акумулацијата, што е и физички очекувано од аспект на функционирањето на водостопанскиот систем.

Шарената скала на десната страна ја претставува предвидената вредност на целната варијабла, при што потоплите бои означуваат повисоки предвидени нивоа на водата. На овој начин, графикот овозможува истовремена визуелизација на повеќе аспекти од оптимизацијата: распределбата на сетот на податоци кој се разгледува, положбата на оптималните сценарија и нивната поврзаност со предвидените вредности на целната функција.



Прилог 65. Локација на издвоените 20 максимуми од оптимизацијата на целната варијабла Z_{wl} во 3Д просторот на предиктор варијаблите „Q_inflow“, „Q_hec_strezhevo“ и „Rain_10-6“.



Прилог 66. Конвергенција на оптимизацијата кон пронаоѓање на решение на функцијата $MAX F(Z_{wl})$: $Z_{wl} = f(Q_{inflow}, Rain_{10-6}, Q_{hec_mini}, Q_{hec_strezhevo}, Q_{gl_dovoden_kanal})$.

Покрај пронаоѓањето на оптималните вредности на целната варијабла, хаос оптимизацијата генерира и график на конвергенција на оптимизацијата. Графикот на конвергенција на хаос оптимизациониот алгоритам (CGO) прикажан во Прилог 66 го прикажува историскиот развој на пребарување на најдобрата вредност на целната функција низ оптимизационите итерации. Од графичкиот приказ може да се забележи дека алгоритмот покажува многу брза конвергенција во почетната фаза на оптимизацијата, при што вредноста на целната функција нагло се подобрува во првите неколку итерации. Ова укажува дека алгоритмот успешно го истражува просторот на можни решенија и релативно брзо идентификува области со подобри комбинации на влезните параметри.

По приближно дваесетина итерации, кривата достигнува речиси константна вредност и понатаму останува стабилна до крајот на процесот. Ваквото однесување укажува на стабилизација на алгоритмот и достигнување на оптимално или квази-оптимално решение. Отсуството на значителни осцилации во понатамошните итерации покажува дека алгоритмот не влегува во нестабилни состојби и не покажува тенденција на дивергенција.

Дополнително, рамниот дел од кривата укажува дека понатамошното зголемување на бројот на итерации најверојатно не би довело до значително подобрување на решението, што претставува индикатор за добра ефикасност на оптимизациониот процес. Ова е особено важно од аспект на пресметковната ефикасност, бидејќи покажува дека оптималното решение е достигнато со релативно мал број итерации.

6.4. Сублимат од спроведената студија на случај 1

6.4.1. Согледувања од добиените метрички показатели од студија на случај 1

Со цел појасен приказ на метричките показатели на моделите кои се применуваат во предвидувањата и оптимизацијата, а беа дел од студија на случај 1, изработен е табеларен приказ со акроним на секоја анализа на случај согласно потпоглавјата, и најдобрите метрички показатели на најдобрите обучени модели.

Табела 30. Табеларен приказ на метричките показатели на најдобрите модели во анализите на случај дел од студија на случај 1.

Студија на случај	Конфигурација	Брана	ML	MSE	RMSE	MAE	R ²
1.1	$Q_{\text{discharge}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}})$	Кнежево	RF	0.03	0.16	0.08	0.690
1.2	$P_{\text{kW}} = f(Z_{\text{wl}}, \text{ResVol}, Q_{\text{inflow}}, Q_{\text{energy}})$	Кнежево	BT	49.01	5.95	2.04	0.981
1.3	$Z_{\text{wl}} = f(Q_{\text{inflow}}, \text{Rain}, Q_{\text{hec_strezhevo}}, Q_{\text{mini_hec}}, Q_{\text{gl_dovoden_organ}})$	Стрежево	RF	15.22	3.90	2.84	0.589

Табела 30 ги сумира најдобрите постигнати резултати од моделите на машинско учење за трите анализирани студии на случај во рамките на првата група истражувања, при што се прикажани најуспешните алгоритми и соодветните вредности на метриките MSE, RMSE, MAE и коефициентот на детерминација R².

Кај студијата на случај 1.1, која се однесува на предвидување на испустот $Q_{\text{discharge}}$ најдобри резултати постигнува моделот RF, со R² = 0.690. Ова укажува на добра способност на моделот да ја опише зависноста помеѓу влезните параметри и целната величина, при што малите вредности на MSE (0,03), RMSE (0,16) и MAE (0,08) укажуваат на висока точност на предвидувањата. Добиените резултати потврдуваат дека разгледуваните предиктори содржат доволно информации за успешно моделирање на испустот од акумулацијата.

Анализата покажува дека најдобри резултати се постигнати кај студијата на случај 1.2, во која целната варијабла е генерираната моќност P_kW. Во овој случај, моделот обучен на алгоритмот BT постигнува исклучително висока вредност на коефициентот на детерминација (R² = 0.981), што укажува дека речиси целокупната варијабилност на целната променлива е успешно објаснета со избраните предиктор варијабли. Истовремено, вредностите на RMSE (5,95) и MAE (2,04) остануваат ниски во однос на опсегот на анализираната величина, што дополнително ја потврдува високата точност и стабилност на моделот.

Најниска предиктивна способност е забележана кај студијата на случај 1.3, каде што се предвидува нивото на водата во акумулацијата Стрежево. Иако моделот „Случајна шума“ повторно се покажува како најуспешен алгоритам, постигнатата вредност R² = 0,589 укажува на умерена точност во споредба со останатите анализирани случаи. Овој резултат може да се објасни со покомплексната зависност на нивото на акумулацијата од повеќе хидролошки и оперативни фактори, како и со можни неизмерени влијанија кои не се опфатени во моделот. Сепак, сите три студии покажуваат позитивни вредности на коефициентот R² и релативно ниски грешки, што потврдува дека применетите модели на машинско учење успешно ја опишуваат динамиката на разгледуваните процеси и претставуваат соодветна алатка за поддршка при управувањето со водостопанските системи.

6.4.2. Коментар на добиените резултати од студија на случај 1

Во рамки на Поглавје 6 беа разгледани повеќе анализи на случај поврзани со примена на модели со машинско учење во управување и анализа на работата на хидросистемите „Злетовица“ и „Стрежево“. Истражувањата беа насочени кон предвидување на клучни хидролошки и експлоатациони параметри, како што се истекувањето од акумулацијата, произведената моќност

во хидроцентрала и нивото на вода во акумулацијата, со цел подобро разбирање на влијанието на операторските одлуки врз функционирањето на хидросистемите.

Добиените резултати потврдија дека моделите со машинско учење имаат значителен потенцијал за примена во анализата и оптимизацијата на работата на комплексни водостопански системи. Анализите покажаа дека со примена на соодветни алгоритми може успешно да се моделираат нелинеарните врски помеѓу хидролошките параметри, работните состојби на акумулациите и операторските интервенции.

Во првата анализа на случај, поврзана со предвидување на истекувањето од акумулацијата „Кнежево“, беше утврдено дека ансамбл-моделите базирани на одлучувачки дрвја обезбедуваат најдобри резултати. Random Forest, XGBoost и Boosted Trees покажаа најниски вредности на MSE, RMSE и MAE, како и највисоки вредности на коефициентот R^2 , што укажува на нивната способност успешно да ги моделираат сложените нелинеарни процеси во хидросистемот. Особено значаен резултат претставува фактот што моделот Random Forest успеа да објасни приближно 69% од варијансата на целната променлива, што претставува добар резултат имајќи ја предвид сложеноста на реалниот систем и ограничувањата на достапните податоци.

Анализата на важноста на варијаблите покажа дека дотекот во акумулацијата “Q_inflow” има доминантно влијание врз предвидувањето на истекувањето “Q_discharge”, што е целосно оправдано од инженерски аспект и директно поврзано со начинот на функционирање на хидросистемот „Злетовица“. Дополнително, анализата на пермутациската важност потврди дека различните алгоритми воспоставуваат различни математички интерпретации на системот, при што секој модел различно ги вреднува предиктор варијаблите. Ова укажува дека машинското учење не претставува единствен универзален пристап, туку дека изборот на алгоритам значително влијае врз начинот на кој системот се интерпретира и моделира.

Резидуалните анализи покажаа дека кај повеќето модели постои систематски тренд на потценување на повисоките вредности и преценување на пониските вредности на целната варијабла. Ова укажува дека и покрај добрите метрички показатели, значителен дел од варијабилноста на системот останува необјаснет. Причините за тоа најверојатно се поврзани со ограничениот број предиктор варијабли, непостоењето на податоци за испарување и филтрација, како и комплексната природа на операторските одлуки кои не можат целосно да се опишат само преку хидролошките параметри.

Особено значаен придонес во ова поглавје претставува примената на хаос оптимизација за оптимизирање на испуштањата од акумулацијата. Со примена на алгоритам на хаос оптимизација беа анализирани десетици илјади сценарија со различни комбинации на предиктор варијабли, со цел пронаоѓање на услови кои водат кон максимизација на целната варијабла. Ваквиот пристап покажа дека комбинацијата од модели со машинско учење и оптимизациски алгоритми претставува моќна алатка за анализа на хидросистемите и поддршка при донесување на операторски одлуки.

Во анализата поврзана со предвидување на генерирана моќност во хидроцентралата „ХЕЦ Б“ беше потврдено дека моделите со машинско учење успешно можат да се применат и во енергетски анализи поврзани со хидросистемите. Со предвидување на моќноста врз основа на хидролошките и експлоатационите услови, се создава можност за подобро планирање на производството на електрична енергија и оптимизација на режимот на работа на хидроцентралите.

Дополнително, анализата на акумулацијата „Стрежево“ покажа дека моделите со машинско учење успешно можат да го моделираат нивото на вода во акумулацијата врз основа на познатите дотекувања и истекувања. Ова е особено значајно бидејќи акумулацијата „Стрежево“ претставува комплексен мултифункционален систем со значајна улога во водоснабдувањето, наводнувањето и енергетскиот сектор. Добиените резултати покажаа дека моделите можат да се применат како алатка за анализа на различни сценарија на работа, вклучително и услови на екстремна побарувачка на вода или намалени дотекувања.

Генерално, резултатите од сите анализи потврдуваат дека ансамбл-моделите базирани на

одлучувачки дрвја, особено Random Forest и XGBoost, покажуваат најголема стабилност, најдобра генерализација и највисока предиктивна способност кај анализираните проблеми. Од друга страна, алгоритмите како Gaussian Process Regression и одредени конфигурации на Artificial Neural Networks покажаа значително послаби резултати, особено кај покомлексни и шумни податоци, што укажува дека нивната примена бара внимателна конфигурација и поголеми сетови на податоци.

Наместо сублимат, може да се констатира дека машинското учење претставува исклучително перспективна алатка за современо управување со хидросистеми. Со негова примена е можно успешно моделирање на сложени хидролошки и експлоатациони процеси, предвидување на однесувањето на системите и поддршка при донесување на операторски одлуки. Дополнително, интеграцијата на машинското учење со оптимизациски методи отвора можности за развој на интелигентни системи за управување со акумулации, производство на енергија и распределба на водните ресурси. Ваквиот пристап претставува значаен чекор кон дигитализација и модернизација на водостопанските системи и создава основа за понатамошен развој на интелигентни системи за управување со водни ресурси во услови на климатски промени, зголемена побарувачка на вода и сè покомлексни експлоатациони услови.

7. СТУДИЈА НА СЛУЧАЈ 2

Во студија на случај 2 ќе се разгледуваат седум различни анализи за седум предиктор варијабли кои говорат за однесувањето на разгледуваните брани.

Во првите пет анализи на случај ќе се разгледува состојбата со дренажниот систем за прибирање на филтрациони води под низводното потпорно тело на брана „Кнежево“. Станува збор за три дренажни цевководи – цевковод во лев бок, десен бок и централен цевковод кои имаат задача да ги приберат процедурните води во низводното потпорно тело на браната. Бидејќи станува збор за каменонасипна брана со асфалтбетонско јадро, од исклучителна важност е следењето на овие процедурни води, чија состојба значително влијае на стабилноста на потпорното тело на браната. Зголемување на овие процедурни води потенцијално може да доведе кон дестабилизација на потпорното тело на браната, деформации и слегнувања, кои придонесуваат кон зголемување на ризикот од рушење на браната и загрозување на низводната долина. Поради тоа, како што следи обработени се резултати од пет студии на случај поврзани со прогноза и дефинирање на граници за предупредување на трите дренажни цевководи кои прибираат вода од три различни делови под низводното потпорно тело на браната. Во секоја анализа на случај даден е кус опис на податоците за обучување на моделите со машинско учење, прелиминарна петкратна вкрстена валидација, избор на најдобрите модели, предвидување на целната варијабла за поголем период за анализа и дефинирање на граници за предупредување. Покрај филтрационите текови како целни варијабли, како дополнителни анализи спроведени со цел предвидување на поширок опсег на податоци се предвидувања за водното ниво во пиезометрите под телото на браната кои се наоѓаат во непосредна близина на дренажните колектори. Авторот смета дека, покрај влијанието на хидростатскиот притисок од водното ниво во акумулацијата, може да се земат предвид и водните нивои во пиезометрите кои се во близина на дренажните колектори како вредна информација за состојбата на дренажниот систем.

Во шестата анализа на случај ќе се анализираат вертикалните поместувања во десет ќелии за мерење на вертикалните слегнувања на брана „Кнежево“. Со помош на модели со машинско учење ќе се дополнат неколкугодишните низи на податоци за слегнувањата, со земање предвид на предиктор варијаблите „Zwl“ – ниво на вода во акумулација, и „t“ – време во дневен чекор. Тука се претпоставува дека слегнувањата кај насипните брани имаат долгорочна тенденција да се стаилизираат кон една финална, речиси константна вредност, па со користење на проширените

низи за слегнувања за секоја ќелија ќе се пресметаат временските домени на достигнување на максималните теоретски пресметани слегнувања со примена на екстраполација на сетот од предвидени вредности.

Во седмата анализа на случај се разгледуваат радијалните поместувања во централниот блок на бетонска лачна брана во Франција, во функција од предиктор варијаблите „Zwl“ – ниво на вода во акумулацијата и „T“ – амбиентална температура како највлијателни регистрирани податоци со кои операторот располага. Станува збор за анализа на случај преземена од тематска задача во рамки на 16-тата Интернационална работилница на нумерички анализи за брани (анг. „16th International Benchmark Workshop on Numerical Analysis of Dams“) во организација на Комитетот за компјутерски аспекти на анализи и проектирање на брани (анг. „Committee on Computational Aspects of Analysis and Design of Dams“) на Интернационалната комисија за големи брани (анг. „ICOLD - International Commission on Large Dams“), Тема А. Анализата на случај опфаќа период на обучување на модели со машинско учење за понатамошна анализа од аспект на предвидување на радијалните поместувања за период кога истите не се регистрирани.

Во продолжение се дадени детални појаснувања за седумте анализи на случај, а во завршно потпоглавје сумиран е сублимат од спроведените анализи и применливоста на моделите со машинско учење во анализите на случај.

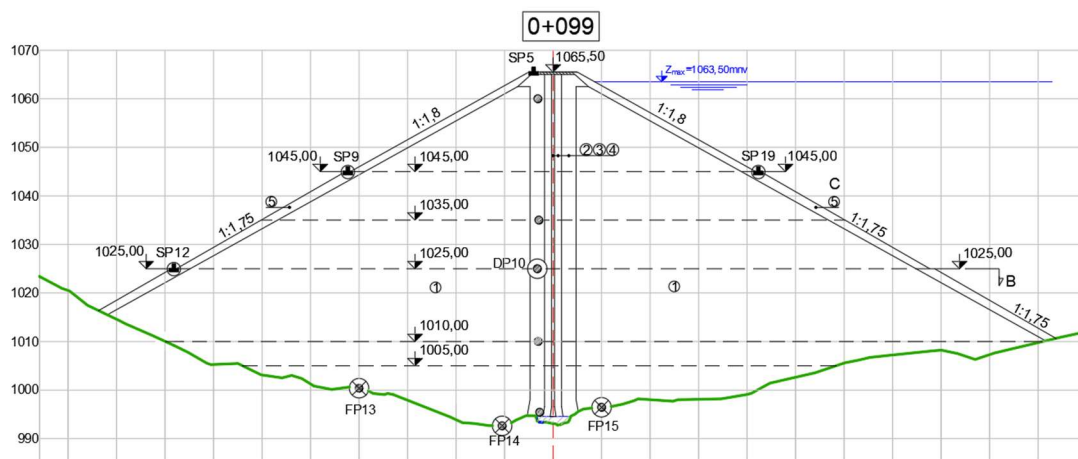
7.1. Анализа на случај 2.1 – $FP_{14}=f(Zwl)$ и $FP_{15}=f(Zwl)$

7.1.1. Подготовка на податоците

За потребите за понатамошни анализи поврзани со прогноза на процедурите води мерени во лев бок под брана „Кнежево“, изработена е предметната анализа на случај каде се извршува предвидување на нивоите во пиезометрите во фундаментот под брана „Кнежево“ именувани како „FP₁₄“ и „FP₁₅“. Станува збор за пиезометри кои се наоѓаат возводно и низводно од асфалтбетонското јадро на браната, во левиот бок од браната. Предметните анализи се изработени со претпоставка дека постои физичка поврзаност помеѓу мерените филтрациони води во лев бок и нивоите во предметните пиезометри. Попречниот пресек со висинската поставеност на пиезометрите е прикажана на Слика 14.

Нивоите во пиезометарот „FP₁₅“ кој се наоѓа на возводната страна од асфалтбетонското јадро се очекува да покажува хидростатски нивои кои се блиски на мерените нивои на вода во акумулацијата. Од друга страна, поради постоењето на инјекциона завеса под асфалтбетонското јадро, кај пиезометарот „FP₁₄“ се очекуваат значително пониски мерени нивои.

Вредноста на нивоите во овие пиезометри не се оскултира редовно па поради тоа во моментот поседуваме низа од 82 мерени податоци за нивоите во двата предметни пиезометри.



Слика 14. Приказ на попречен пресек на брана „Кнежево“ со локација на пиезометрите FP₁₄ и FP₁₅ во фундаментот под браната.

Целта на анализата на случај е да се предвидат нивоите во пиезометрите „FP_14“ и „FP_15“ во однос на еден предиктор – нивото на вода во акумулацијата „Zwl“ за периодот помеѓу 2021 и 2025 година кога имаме голем дел од временскиот домен кога не се регистрираат податоци за нивоите во пиезометрите. Со цел збогатување на опусот на предиктор варијабли, се воведуваат уште две изведени предиктор варијабли – „dZw“ и „p=dZw/t“.

Првата изведена варијабла „dZw“ претставува пресметана вредност како разлика на водните нивои во акумулацијата помеѓу две регистрирани мерења:

$$dZ_w = |Z_{w_i} - Z_{w_{i-1}}| = \dots [m] \quad 7.1.1$$

, каде:

dZw – апсолутна вредност на разлика во водните нивои помеѓу две последователни регистрирани мерења, во [m],

Z_{wi} – регистрирана вредност на водното ниво во акумулацијата во временски чекор i,

Z_{wi-1} – регистрирана вредност на водното ниво во акумулацијата во временски чекор i-1.

Втората изведена предиктор варијабла „p=dZw/t“ претставува степен на промена на водното ниво во однос на разликата во временските интервали „t“, и се пресметува како:

$$p = \frac{dZ_w}{t} = \frac{|Z_{w_i} - Z_{w_{i-1}}|}{t} \quad 7.1.2$$

$$t = t_{Zw_i} - t_{Zw_{i-1}} \quad 7.1.3$$

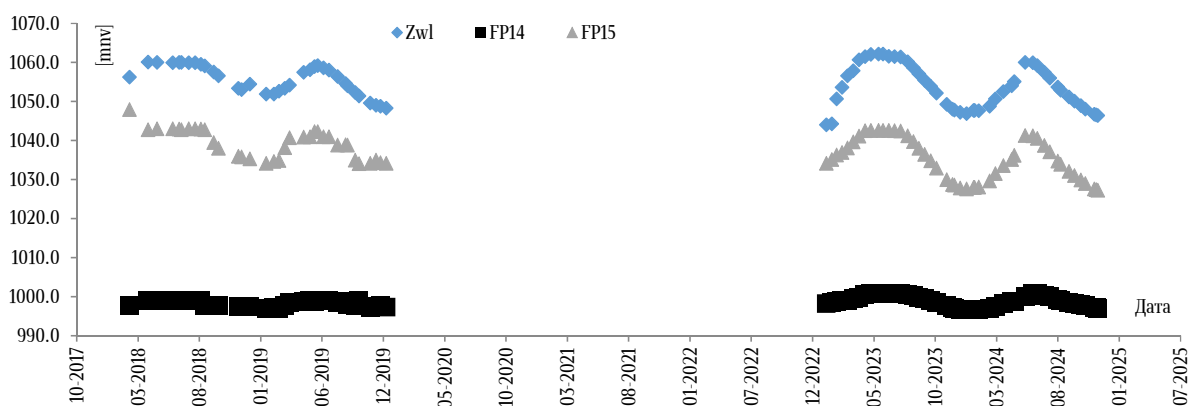
, каде:

t – разлика во денови,

t_{Zw_i} – ден на кој е регистрирано водното ниво во акумулацијата Z_{wi},

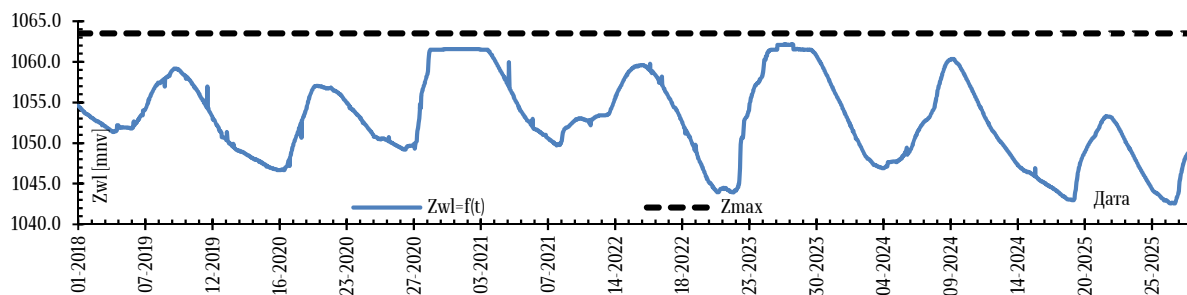
t_{Zw_i-1} – ден на кој е регистрирано водното ниво во акумулацијата Z_{wi-1}.

Иако се зборува за една анализа на случај, изработени се два засебни модели каде посебно се тренира пресметката на нивото во пиезометарот „FP_14“ во однос на нивото во акумулацијата „Zwl“ и изведените варијабли „dZw“ и „p“; и посебно се тренира пресметката на нивото во пиезометарот „FP_15“ во однос на нивото во акумулацијата „Zwl“ и изведените варијабли „dZw“ и „p“.



Прилог 67. Приказ на вредностите на целните варијабли „FP_14“ и „FP_15“, како и предиктор варијаблата „Zwl“, користени во фазата на обучување на моделите.

Низата на податоци со кои ќе се извршува процесот на обучување на моделите е прикажан на Прилог 67, додека низата на податоци на предиктор варијаблата „Zwl“ со која ќе се извршува предвидување на целните варијабли е прикажана на Прилог 68. Во процесот на предвидување ќе се извршува прогноза на целните варијабли за вкупно 2606 податоци за познатата предиктор варијабла „Zwl“ и пресметаните предиктор варијабли „dZw“ и „p“.



Прилог 68. Приказ на вредностите на предиктор варијаблата „Zwl“ за временскиот период за кој ќе се извршува предвидување на целните варијабли - нивоите во пиезометрите „FP_14“ и „FP_15“.

7.1.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Двете целни варијабли се користени во оригинална форма, без примена на нормализација или стандардизација на податоците. Од друга страна, предиктор варијаблата беше стандардизирана во рамки на секој сет од петкратната валидација, но само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација за предиктор варијаблата „FP_14“ се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 31.

Табела 31. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_14“.

5-fold CV					
No.	Model	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	ANN	8111.388	58.797	20.548	-6708.421
2	GPR	1.483	1.199	1.032	-0.017
3	Decision Tree	0.665	0.784	0.596	0.499
4	GRNN	0.625	0.788	0.626	0.518
5	XGBoost	0.539	0.711	0.553	0.579
6	SVR	0.502	0.687	0.503	0.585
7	Boosted Trees	0.473	0.659	0.494	0.629
8	KNN	0.462	0.669	0.551	0.653
9	Random Forest	0.444	0.649	0.515	0.658
max value:		8111.388	58.797	20.548	0.658
min value:		0.444	0.649	0.494	-6708.421

Прелиминарната петкратна вкрстена валидација на моделите за предвидување на таргет варијаблата „FP_14“, со примена на предиктор варијаблите „Zwl“, „dZwl“ и „ $p = dZwl/t$ “, покажа дека воведувањето на изведените варијабли придонесува за подобрување на точноста кај поголемиот дел од применетите алгоритми. Метричките показатели од петкратната вкрстена валидација се прикажани во Табела 31. Најдобри резултати постигна моделот Случајна шума (Random Forest), со просечни вредности MSE = 0.444, RMSE = 0.649, MAE = 0.515 и $R^2 = 0.658$,

непосредно следни се моделите K-најблиски соседи (KNN), Поткренати регресиони дрва (Boosted Trees) и Регресија со потпомогнати вектори (SVR), кои исто така покажаа високи вредности на коефициентот на детерминација ($R^2 > 0.58$). XGBoost, Генерализираната регресиона невронска мрежа (GRNN) и Одлучувачкото дрво (Decision Tree) постигнаа задоволителни резултати, додека Гаусовиот процесен регресор (GPR) повторно покажа ограничена способност за генерализација со речиси нулта вредност на коефициентот R^2 . Како и во претходните анализи, моделот на вештачки невронски мрежи (ANN) оствари значително неповолни метрички показатели, што укажува на неговата несоодветност за разгледуваниот податочен сет.

Табела 32. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_15“.

5-fold CV					
No.	Model	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	Boosted Trees	3.872	1.925	1.351	0.817
2	Support Vector Regressor	3.972	1.931	1.193	0.814
3	Random Forest	4.195	2.014	1.392	0.804
4	GRNN	4.735	2.163	1.571	0.787
5	K-Nearest Neighbors	5.013	2.218	1.621	0.777
6	Decision Tree	5.636	2.258	1.348	0.727
7	XGBoost	5.239	2.276	1.639	0.776
8	Gaussian Process	5.507	2.316	1.537	0.750
9	Artificial Neural Networks	68704.771	194.567	115.485	-3314.162
max value:		68704.771	194.567	115.485	0.817
min value:		3.872	1.925	1.193	-3314.162

Прелиминарната петкратна вкрстена валидација на моделите за предвидување на таргет варијаблата „FP_15“ покажа дека поголемиот дел од применетите алгоритми обезбедуваат висока точност и добра способност за генерализација (Табела 32). Најдобри резултати постигна моделот Поткренати регресиони дрва (Boosted Trees), со просечни вредности $MSE = 3.872$, $RMSE = 1.925$, $MAE = 1.351$ и $R^2 = 0.817$. Многу блиски резултати остварија и Регресијата со потпомогнати вектори (Support Vector Regressor) и Случајната шума (Random Forest), со вредности на коефициентот на детерминација од 0.814 и 0.804, соодветно. Генерализираната регресиона невронска мрежа (GRNN), K-најблиски соседи (KNN), XGBoost, Гаусовиот процесен регресор (GPR) и Одлучувачкото дрво (Decision Tree) исто така покажаа задоволителни резултати, при што сите модели, со исклучок на Вештачките невронски мрежи (ANN), остварија вредности на коефициентот R^2 поголеми од 0.72. За разлика од останатите алгоритми, моделот ANN повторно покажа исклучително неповолни метрички показатели, што укажува на неговата несоодветност за анализираниот податочен сет. Генерално, резултатите од прелиминарната вкрстена валидација укажуваат дека повеќето применети алгоритми успешно ја моделираат зависноста на таргет варијаблата „FP_15“, при што ансамбл-моделите и алгоритмите базирани на потпомогнати вектори покажуваат најдобри перформанси.

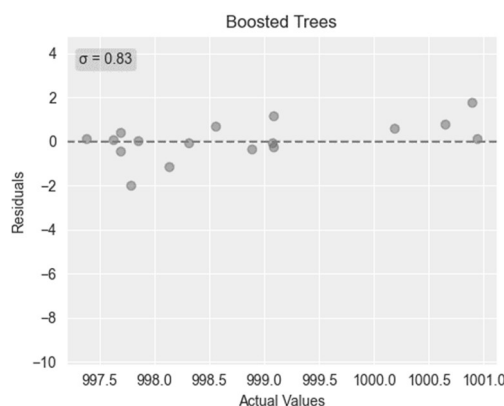
7.1.3. Обучување на моделите

Како алгоритми одбрани за понатамошно разгледување и примена во фаза на предвидување на целните варијабли, применети се оние со најдобри метрички показатели од претходниот чекор во моделирањето – за целната варијабла „FP_14“ одбран е алгоритмот Boosted Trees за понатамошно обучување, додека за целната варијабла „FP_15“ одбран е алгоритмот Random Forest.

Во процесот на обучување на моделите со напоменатите алгоритми, DamResML пресметува и одредени графици кои зборуваат за важноста на воведените предиктор варијабли како и точноста на обучените модели, сето тоа врз основа на задржан сет од 20% на податочното множество претставено во фазата на обучување на моделите. Како што следи, даден е приказ на дел од добиените резултати од обучувањето на моделите.

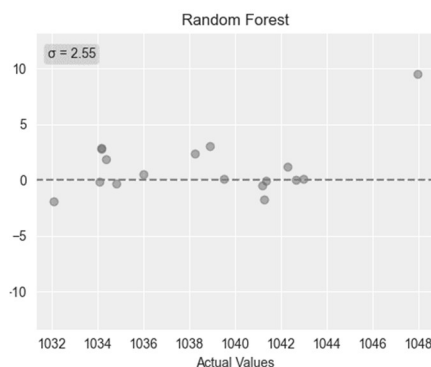
Резидуалите добиени при предвидувањето на таргет варијаблата „FP_14“ со моделот Поткренати регресиони дрва (Boosted Trees), користејќи го задржаниот 20% сет на податоци за

валидација, покажуваат рамномерна распределба околу нултата вредност, без изразена систематска пристрасност или забележлив тренд. Најголемиот дел од резидуалите се концентрирани во близина на нулата, што укажува на добра согласност помеѓу предвидените и измерените вредности. Дополнително, релативно малата стандардна девијација на резидуалите ($\sigma = 0.83$) укажува на стабилност на моделот и задоволителна точност при предвидувањето на варијаблата „FP_14“. Графикот со резидуали е прикажан во Прилог 69.



Прилог 69. Пресметани резидуали по обучување на моделот за целната варијабла „FP_14“ со примена на 20% задржан сет на податоци за валидација на моделот.

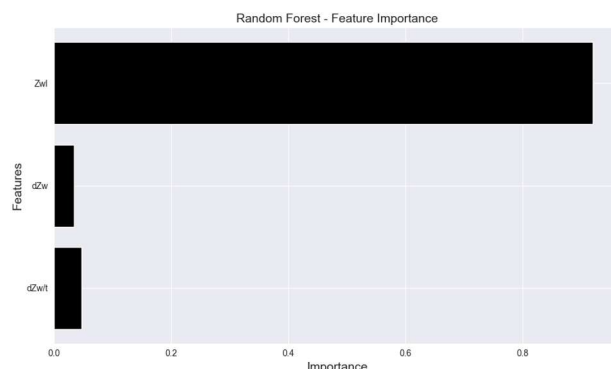
Резидуалите добиени при предвидувањето на таргет варијаблата „FP_15“ со моделот Случајна шума (анг. „Random Forest“), користејќи го задржаниот 20% сет на податоци за валидација, се распределени околу нултата вредност без изразен систематски тренд, што укажува на добра согласност помеѓу предвидените и измерените вредности (Прилог 70). Најголемиот дел од резидуалите имаат релативно мали апсолутни вредности, додека е забележано едно поголемо позитивно отстапување кое придонесува за зголемување на стандардната девијација на резидуалите ($\sigma = 2.55$). И покрај ова поединечно отстапување, распределбата на резидуалите укажува дека моделот успешно ја опишува зависноста на таргет варијаблата „FP_15“ и обезбедува задоволителна точност при предвидувањето.



Прилог 70. Пресметани резидуали по обучување на моделот за целната варијабла „FP_15“ со примена на 20% задржан сет на податоци за валидација на моделот.

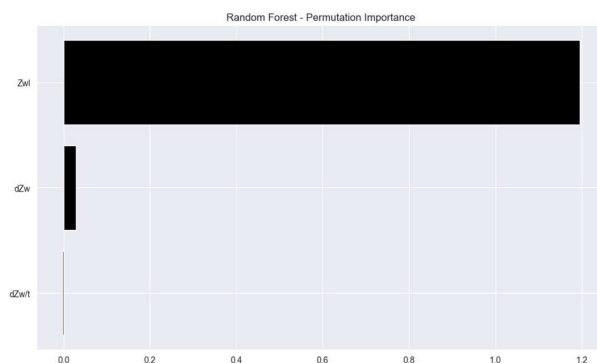
Кај вториот модел каде целната варијабла е водното ниво во пиезометарот „FP_15“ обучуван со помош на алгоритмот RF, анализата на важноста на предиктор варијаблите со примена на моделот Случајна шума (анг. „Random Forest“) покажува дека нивото на водата во акумулацијата „Zwl“ претставува убедливо најзначајниот параметар за предвидување на таргет варијаблата „FP_15“ (Прилог 71).

Анализата на важноста на варијаблите (анг. „Feature Importance“) и пермутационската анализа (анг. „Permutation Importance“) прикажани во графиконите во Прилог 71 и Прилог 72 покажуваат изразена доминација на варијаблата Zwl, додека останатите предиктор варијабли имаат значително помало влијание врз предвидувањата на моделот.



Прилог 71. Важност на варијаблите пресметана за моделот RF за анализа на случај $FP_{14}=f(Zw1, dZw, p)$.

Ваквата согласност помеѓу двата пристапи за оценување на важноста на предикторите укажува на стабилност и доверливост на добиените резултати и потврдува дека промените на нивото на водата во акумулацијата претставуваат примарен фактор кој влијае врз флуктуацијата на водното ниво во пиезометарот „FP_15“.



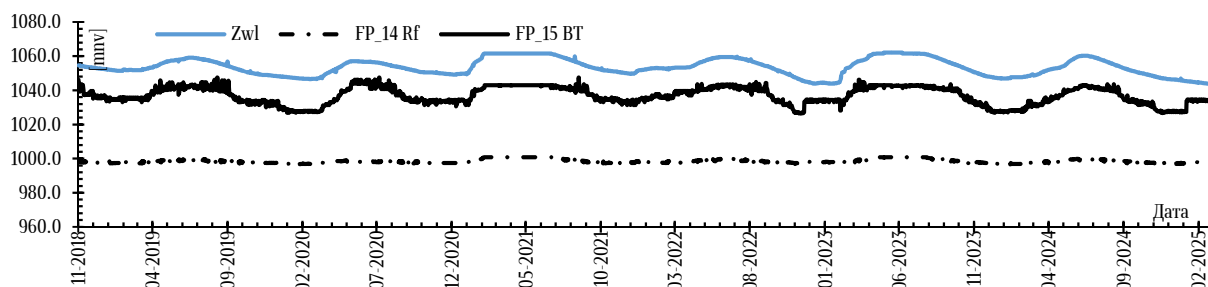
Прилог 72. Пермутациона важност пресметана за моделот RF за анализа на случај $FP_{14}=f(Zw1, dZw, p)$.

7.1.4. Предвидување на целната варијабла

По извршеното обучување на моделите со алгоритмот RF за целната варијабла „FP_14“ и BT за целната варијабла „FP_15“, извршено е предвидување на целните варијабли на непознат сет на податоци на предиктор варијаблата „Zw1“ – ниво на вода во акумулацијата за периодот од 2018 до 2025 година и пресметаните предиктор варијабли „dZw1“ и „ $p = dZw1/t$ “.

Предвидувањата за целните варијабли се прикажани Прилог 73.

Вредностите на нивоите се во рамки на очекуваните, т.е. нивото во пиезометарот „FP_14“ кој се наоѓа на низводната страна од асфалтбетонското јадро и инјекционата завеса покажува речиси константна вредност од 1000mmv со мали осцилации кои кореспондираат на промената на водното ниво во акумулацијата. Од друга страна, осцилациите на нивото на вода во пиезометарот „FP_15“ се во поголеми варијанси и поблиски до нивото на вода во акумулацијата – повторно, очекувано, поради локациската поставеност на пиезометарот на возводната страна од асфалтбетонското јадро и инјекционата завеса под браната.



Прилог 73. График со приказ на предвидените вредности на целните варијабли „FP_14“ и „FP_15“ со моделите тренирани со алгоритмите RF и BT, соодветно.

Податоците од предвидувањата се применуваат во наредната анализа на случај каде се проценуваат процедурните филтрациони води од левиот бок под брана „Кнежево“, т.е. се спроведува анализа за зависноста на процедурните води со водните нивои во пиезометрите „FP_14“ и „FP_15“ кои се наоѓаат во близина на дренажниот колектор за процедурни води.

7.2. Анализа на случај 2.2 – $V_{\text{notch_left}} = f(Z_{\text{wl}}, Q_{\text{inflow}}, \text{FP}_{14}, \text{FP}_{15})$

7.2.1. Подготовка на податоците

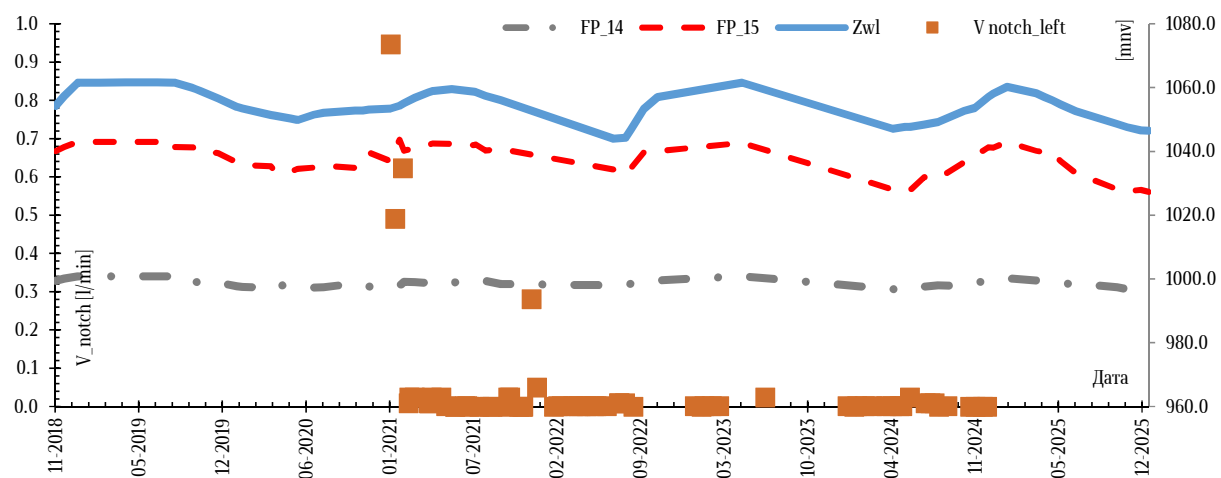
За анализа на процедурните води прибрани со дренажен колектор во низводното потпорно тело во лев бок на брана „Кнежево“, се извршува анализа на случај 2.2 каде целна варијабла е пресметката на процедурните води „ $V_{\text{notch_left}}$ “ во [l/min] со земање предвид на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „ Z_{wl} “ во [mnv] прикажано во Прилог 74,
- 2) Дотек во акумулацијата, „ Q_{inflow} “ во [m³/s] прикажано во Прилог 75,
- 3) Ниво на вода во пиезометарот „FP_14“ лоциран во фундаметот на браната, возводно од асфалтбетонското јадро и инјекционата завеса, „FP_14“ во [mnv] прикажано во Прилог 74, и
- 4) Ниво на вода во пиезометарот „FP_15“ лоциран во фундаметот на браната, низводно од асфалтбетонското јадро и инјекционата завеса, „FP_15“ во [mnv] прикажано во Прилог 74.

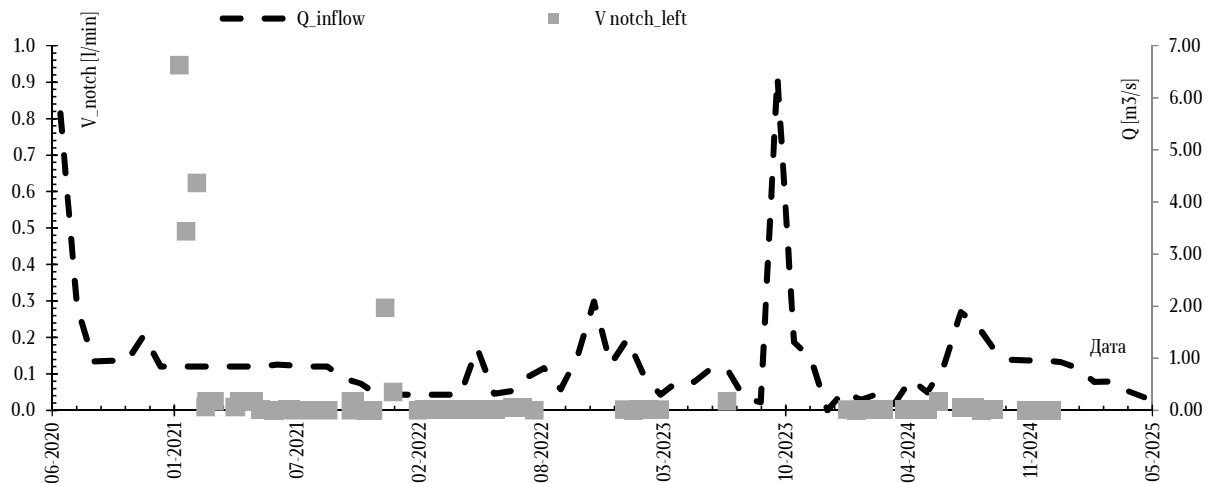
Целната варијабла – процедурни води од лев бок под тело на браната „ $V_{\text{notch_left}}$ “ се регистрира низводно од брана „Кнежево“ со периодични мерења во [l/min]. Со цел проценка на овие води за целосниот период од 2018 до 2025 година, се изработува предметната анализа со модели базирани на машинско учење кои треба да извршат предвидување на целната варијабла во однос на посочените предиктор варијабли.

Процесот на обучување се извршува за 66 податоци на целната варијабла и предиктор варијаблите (Прилог 74 и Прилог 75), додека предвидувањето се извршува за 2606 податоци на предиктор варијаблите. Во процесот на предвидување, вредноста на целната варијабла е непозната. Вредностите на предиктор варијаблите користени во процесот на предвидување се прикажани во Прилог 76.

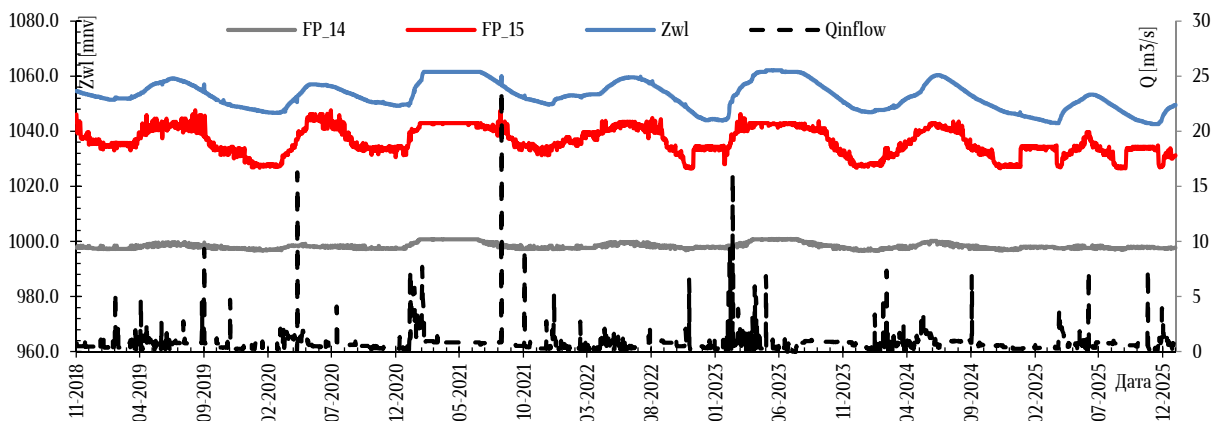
Вредностите на предиктор варијаблите „ Z_{wl} “ и „ Q_{inflow} “ се преземени од теренски мерења, додека вредностите на предиктор варијаблите „FP_14“ и „FP_15“ се преземени од студијата на случај 2.1. каде нивоите во пиезометрите се пресметуваат со помош на модели со машинско учење.



Прилог 74. График со приказ на вредностите на целната варијабла „ $V_{\text{notch_left}}$ “ и предиктор варијаблите „ Z_{wl} “, „FP_14“ и „FP_15“ кои се користат во процесот на обучување на моделите.



Прилог 75. График со приказ на вредностите на целната варијабла „V_notch_left“ и предиктор варијаблата „Q_inflow“ кои се користат во процесот на обучување на моделите.



Прилог 76. График со приказ на предиктор варијаблите „Zwl“, „FP_14“, „FP_15“ и „Q_inflow“ за периодот на предвидување на целната варијабла „V_notch_left“.

7.2.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Перцепсија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Целната варијабла е користена во нејзината оригинална форма, без примена на нормализација или стандардизација. Од друга страна, предиктор варијаблите беа стандардизирани во рамки на секој сет од петкратната валидација, и тоа само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 33.

Табела 33. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за моделот $V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15)$.

5-fold CV					
No.	Algorithm	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	K-Nearest Neighbors	0.0229	0.1215	0.0535	-4.6695
2	Support Vector Regressor	0.0278	0.1492	0.1033	-40.9943
3	GRNN	0.0389	0.1537	0.0730	-8.1779
4	Artificial Neural Networks	0.0392	0.1587	0.0897	-21.7265
5	Gaussian Process	0.0351	0.1606	0.0813	-12.3135
6	Random Forest	0.0403	0.1814	0.0882	-30.2408
7	XGBoost	0.0493	0.2043	0.1193	-61.3527
8	Decision Tree	0.0680	0.2506	0.1090	-166.6811
9	Boosted Trees	0.0673	0.2509	0.1069	-155.7118
max value:		0.0680	0.2509	0.1193	-4.6695
min value:		0.0229	0.1215	0.0535	-166.6811

Целната променлива „V_notch_left“ го претставува количеството на процедни води измерени на левиот мерен прелив на браната Кнежево, што претставува важен индикатор за следење на филтрациониот режим и филтрационата безбедност на браната.

Во анализата беа користени 66 податочни записи и четири влезни параметри, при што е забележано дека самата целна променлива покажува релативно мала варијабилност, односно регистрираните вредности на целната варијабла кои се користат во процесот на подучување немаат значителни осцилации.

За оценка на предиктивната способност на моделите беше применета петкратна вкрстена валидација, при што беа споредени девет различни алгоритми на машинско учење преку метриките MSE, RMSE, MAE и R^2 . Врз основа на добиените резултати може да се констатира дека моделот K-Nearest Neighbors (KNN) постигна најдобри вкупни перформанси и се издвојува како најсоодветен модел за предвидување на оваа целна променлива. Имено, KNN ја постигна најниската вредност на средната квадратна грешка ($MSE = 0.0229$), што укажува дека овој модел има најмали квадратни отстапувања помеѓу измерените и предвидените вредности. Дополнително, истиот модел ја постигна и најниската вредност на $RMSE = 0.1215$, што покажува дека просечното отстапување во реални мерни единици е најмало токму кај овој алгоритам. Исто така, KNN ја даде и најниската вредност на $MAE = 0.0534$, што значи дека просечната апсолутна грешка при предвидувањето е најмала во споредба со останатите модели. Овие резултати јасно укажуваат дека KNN најуспешно ги следи малите промени и суپтилните варијации на филтрационата вода.

Покрај него, солидни резултати покажаа и моделите GRNN, Gaussian Process и Support Vector Regressor (SVR), кои исто така имаат релативно ниски вредности на грешките и можат да се сметаат за соодветни алтернативи. GRNN постигна добар баланс помеѓу стабилност и точност, додека Gaussian Process покажа добра способност за моделирање на нелинеарни зависности. SVR, иако со нешто повисоки вредности на MAE, сепак покажа конкурентни резултати кај MSE и RMSE. Од друга страна, Random Forest, XGBoost, Boosted Trees и особено Decision Tree покажаа послаби резултати во споредба со претходно наведените модели, што укажува дека алгоритмите базирани на дрва не се најсоодветни за овој тип релативно мал и стабилен сет на податоци.

Кај сите анализирани модели беа добиени негативни вредности на коефициентот на детерминација R^2 , што на прв поглед може да укажува на слаба предиктивна способност. Сепак, во конкретниот случај ваквите резултати треба внимателно да се интерпретираат, бидејќи негативните R^2 вредности најверојатно се последица на ограничениот број записи, малиот број тест примероци во секој сет, како и ниската варијабилност на целната променлива. Кога целната променлива е речиси константна или се движи во тесен опсег, R^2 станува чувствителен и често не ја отсликува реалната точност на моделот. Поради тоа, кај ова истражување значително порелевантни показатели за успешност се MSE, RMSE и MAE.

Од инженерски аспект, добиените резултати се логични, бидејќи филтрационата вода претставува параметар кој во нормални експлоатациони услови обично се менува постепено и без нагли осцилации, што ја прави задачата за предвидување специфична и чувствителна на мали грешки.

Врз основа на целокупната анализа може да се заклучи дека методите на машинско учење можат успешно да се применат за предвидување на количеството филтрациона вода на браната Кнежево, при што моделот K-Nearest Neighbors се покажува како најсоодветно решение за конкретниот сет на податоци. Овие резултати потврдуваат дека и кај релативно мал број мерења, со правилен избор на алгоритам, можно е да се добијат корисни модели за поддршка на системите за мониторинг и рано детектирање на потенцијални промени во филтрациониот режим на браната.

7.2.3. Обучување на моделот со GRNN

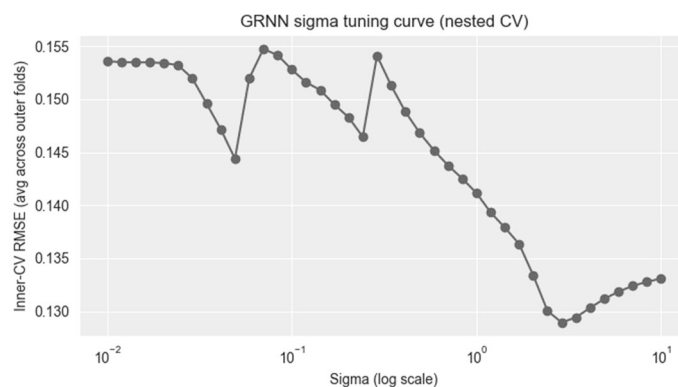
По разгледување на добиените податоци од прелиминарната петкратна вкрстена валидација, избран е алгоритмот GRNN како алгоритам кој со влезниот податочен сет за обука обезбедува највисока вредност на коефициентот на детерминација $R^2 = -9.74$ и една од најниските средни апсолутни грешки $MAE = 0.064$ l/min.

Алгоритмот GRNN не гради експлицитна параметарска функција, туку предвидувањата ги врши преку тежинско осреднување на сите обучувани примероци, а тежинските коефициенти зависат помеѓу растојанието помеѓу примероците. Основен параметар на моделот е σ – сигма. Овој параметар ја контролира ширината на Гаус функцијата, односно степенот на „влијание“ на соседните точки. Помали вредности водат до локално чувствителен модел (ризик од претерување во предвидувањата), додека поголеми вредности создаваат помазни предвидувања (ризик од потценување на предвидувањата).

Во кодот е применета вгнездена трикратна валидација за избор на најдобра вредност на параметарот σ – сигма. Се тестираат вредности во рангот $\sigma = (0.01 - 10)$. Критериумот е да се избере вредност за σ што ќе резултира со минимален RMSE, а во конкретниот случај графикот на варијација на RMSE во зависност од параметарот σ по трикратната вкрстена валидација е прикажан во Прилог 77. Понатаму, со изборот на соодветна вредност за параметарот σ и дополнително се прави петкратна надворешна вкрстена валидација на податоците.

Табела 34. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација анализа на случај V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15).

Fold	Best Sigma	Fold RMSE
1	0.084	0.170
2	2.424	0.161
3	10.000	0.288
4	1.701	0.038
5	0.084	0.170



Прилог 77. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација анализа на случај V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15).

Прикажаната крива во Прилог 77 ја претставува постапката за определување на оптималната вредност на параметарот σ кај моделот Генерализирана регресиона невронска мрежа (GRNN) со примена на вгнездена вкрстена валидација. За секоја анализирана вредност на параметарот σ е пресметана просечната вредност на RMSE преку надворешните вкрстени поделби, при што е избрана вредноста која обезбедува најмала грешка при предвидувањето. Резултатите покажуваат дека со зголемување на параметарот σ првично се намалува просечната вредност на RMSE, при што најдобри перформанси се постигнуваат при $\sigma \approx 3$. По оваа вредност се забележува благо зголемување на грешката, што укажува дека дополнителното измазнување на моделот не придонесува за подобрување на неговата способност за генерализација. Ваквата анализа овозможува објективен избор на параметарот σ и придонесува кон формирање на постабилен и попрецизен GRNN модел.

7.2.4. Предвидување на целната варијабла

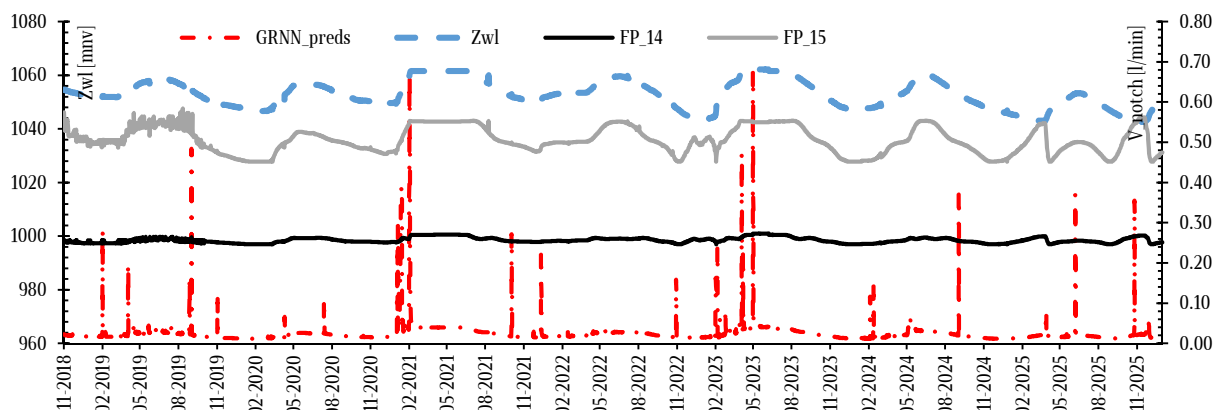
Целна варијабла во предметната студија на случај е количината на филтрациони води „V_notch_left“ измерени во прелив од лев бок на браната „Кнежево“. Со помош на обучениот модел од машинско учење, се бара предвидување на целната варијабла со познавање на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „Zwl“ во [mnv] прикажано во Прилог 74,
- 2) Дотек во акумулацијата, „Q_inflow“ во [m³/s] прикажано во Прилог 75,
- 3) Ниво на вода во пиезометарот FP_14 лоциран во фундаметот на браната, возводно од асфалтбетонското јадро и инјекционата завеса, „FP_14“ во [mnv] прикажано во Прилог 74, и
- 4) Ниво на вода во пиезометарот „FP_15“ лоциран во фундаметот на браната, низводно од асфалтбетонското јадро и инјекционата завеса, „FP_15“ во [mnv] прикажано во Прилог 74.

Обучен е еден модел со машинско учење, и тоа модел базиран на алгоритмот GRNN.

Процесот на предвидување се извршува за 2606 познати вредности на предиктор варијаблите, кои временски се однесуваат на периодот на прибирање на податоците од 1^{ви} ноември 2018 година до 31^{ви} декември 2025 година. За анализираниот период нема континуирани вредности на целната варијабла бидејќи истата се регистрира од страна на операторот во нееднакви временски интервали. Предиктор варијаблите се внесуваат во DamResML кодот преку датотеката „predictor_variable_for_prediction.xls“ и истата се повикува во ќелијата „Prediction“ по завршување на процесот на обучување на моделот.

Резултатите од предвидувањата се прикажани во Прилог 78 и Прилог 79.

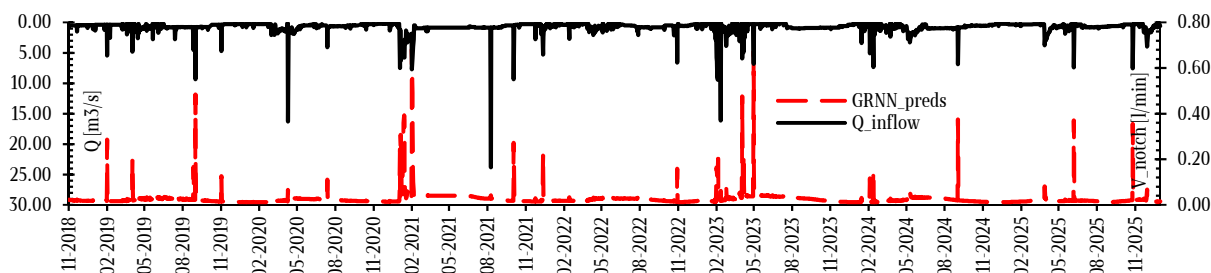


Прилог 78. График со приказ на предвидените вредности на целната варијабла 'V_notch_left' (график GRNN_preds), предиктор варијаблите „Zwl“, „FP_14“ и „FP_15“.

Во Прилог 78 даден е графички приказ на целната варијабла предвидена со моделот трениран со KNN, споредено со вредностите на предиктор варијаблите „Zwl“, „FP_14“ и „FP_15“. Од графикот може да се согледа дека предвидените вредности на филтрационите води го следат трендот на

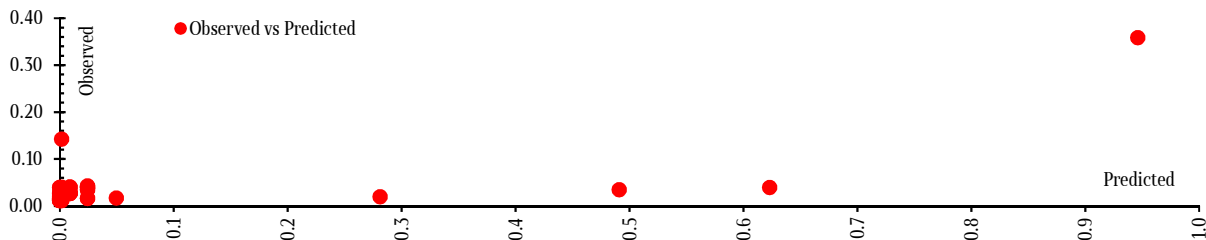
промената на нивото на вода во акумулацијата „Zwl“. Во Прилог 79 каде е даден приказ на целната варијабла „V_notch_left“ и предиктор варијаблата „Q_inflow“ може повторно да се согледа дека нецелосно предвидувањата го следат трендот на дотекувања.

Минималната вредност на целната варијабла предвидена со моделот изнесува $V_notch_left_{min} = 3.91 \cdot 10^{-5}$ [l/min] додека максималната вредност изнесува $V_notch_left_{max} = 0.941 \cdot 10^{-5}$ [l/min]. Овде е интересно да се посочи дека во процесот на обучување се јавуваат нулти мерења регистрирани од страна на операторот, но моделот во предвидувањата никогаш не предвидува нулти истекувања. Максималната предвидена вредност е пониска од максималната вредност во периодот на обучување која изнесува 0.946 [l/min].



Прилог 79. График со приказ на предвидените вредности на целната варијабла „V_notch_left“ (график GRNN_preds), и предиктор варијаблата „Q_inflow“.

На графичкиот приказ во Прилог 80 може да се забележи дека не постои јасна согласност помеѓу регистрираните и пресметаните вредности на филтрационите истекувања. Најголемиот дел од измерените вредности се групирани во многу тесен опсег со ниски вредности, додека моделот GRNN за одредени случаи предвидува значително повисоки вредности, што укажува на тенденција за преценување на филтрационите истекувања. Ваквата распределба на резултатите потврдува дека моделот не успеал соодветно да ја опише зависноста помеѓу влезните и целната променлива, што е во согласност со добиените негативни вредности на коефициентот на детерминација (R^2) и може да се објасни со ограничениот број достапни податоци и нивната мала варијабилност.



Прилог 80. График со приказ на регистрираните филтрациони протекувања „V_notch_left (observed)“ и пресметаните филтрациони истекувања (predicted) согласно предвидувањата од моделот обучен со GRNN.

7.2.5. Дефинирање на граници за предупредување (thresholds)

Филтрационите води претставуваат еден од најзначајните индикатори за безбедносната состојба на браните, бидејќи нивното однесување директно ја одразува состојбата на телото на браната, фундаментот под браната и состојбата на дренажниот систем. Самото постоење на филтрација не претставува опасност, бидејќи одредено количество процедни води е очекувана појава кај секоја брана. Сепак, ненадејно зголемување на филтрационите количини, промена на зависноста помеѓу котата на акумулацијата и процедените води, долгорочен растечки тренд или појава на заматување на филтрационите води може да претставуваат индикација за нови филтрациони патишта, внатрешна ерозија, деградација на фундаментот или намалена ефикасност на дренажниот систем.

Со примена на модели на машинско учење може да се дефинира очекуваното однесување на филтрацијата во зависност од работните услови, а секое значајно отстапување помеѓу измерените и предвидените вредности може да се користи како систем за рано предупредување и поддршка

при донесувањето на одлуки за санација на дренажниот систем, итно снижување на нивото на вода во акумулација и слично.

Во предметната анализа на случај, резултатите од предвидените вредности за филтрациониот тек мерен во левиот дренажен колектор под брана „Кнежево“ именувани како „V_notch_left“ се применуваат за дефинирање на граници за предупредување (анг. „thresholds“). Понатаму, со дефинираните граници на предвидување се врши споредба на реалните измерени податоци за филтрацијата „V_notch_left“ и се донесуваат заклучоци за тоа дали мерените вредности се во граници на очекуваните, или излегуваат од нив.

Границите на предупредување се дефинираат во однос на пресметана σ - стандардна девијација од предвидените 2606 податоци на целната варијабла. Предложената методологија за примена на граници за предупредување пресметани во однос на стандардната девијација на сетот на податоци е применета во студија на случај каде се истражуваат границите за предупредување од претерана филтрација кај неколку насипни брани во Кина изработена од Чен и соработниците (Chen, et al., 2017).

Стандардната девијација во конкретниот код се пресметува во ќелијата „Statistics of predictions“ во DamResML кодот, веднаш по пресметка на предвидувањата. Стандардната девијација математички се пресметува како:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - x_{ave})^2} \quad (7.2.1)$$

, каде што:

σ - стандардна девијација на анализираниот сет на податоци,

n - број на податоци во сетот,

x_{ave} - средна вредност на податоците во сетот,

x_i - поединечна вредност на секој податок во сетот.

Пресметаната стандардна девијација од сетот на предвидени вредности на „V_notch_left“ изнесува $\sigma = 0.0309$. Средната вредност на сетот на податоци изнесува $x_{ave} = 0.01699$.

Првата, пониска граница на предупредување се дефинира како:

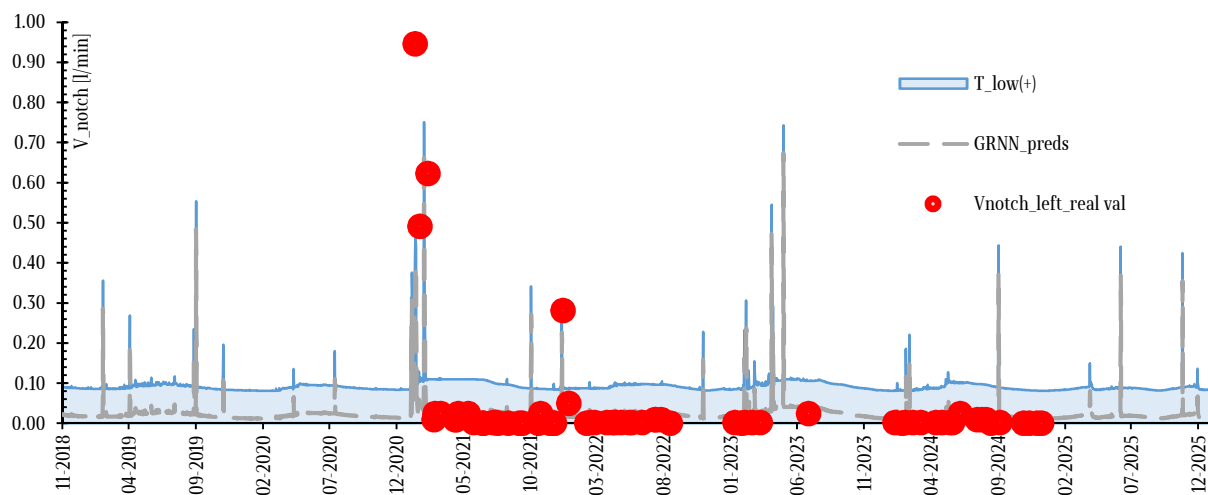
$$T_{low}^+ = x_i + 2 \cdot \sigma; \quad T_{low}^- = x_i - 2 \cdot \sigma; \quad (7.2.2)$$

Втората, повисока граница на предупредување се дефинира како:

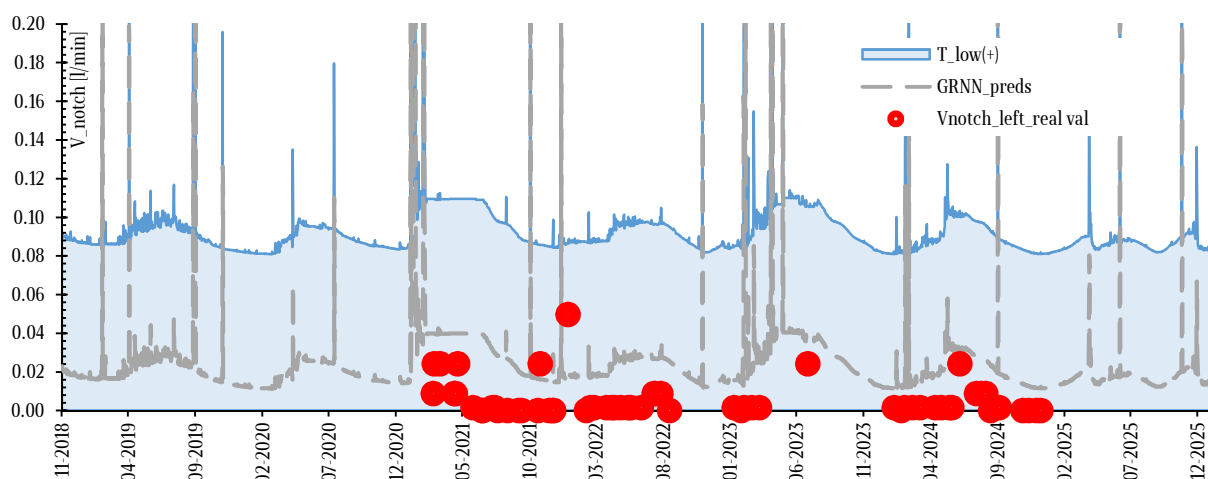
$$T_{high}^+ = x_i + 3 \cdot \sigma; \quad T_{low}^- = x_i - 3 \cdot \sigma; \quad (7.2.3)$$

Двете граници на предупредување варираат во однос на варијацијата на предвидените вредности на целната варијабла x_i . Границите претставуваат оддалечени вредности од предвидените вредности на целната варијабла, согласно математичката релација во равенките 7.2.2 и 7.2.3.

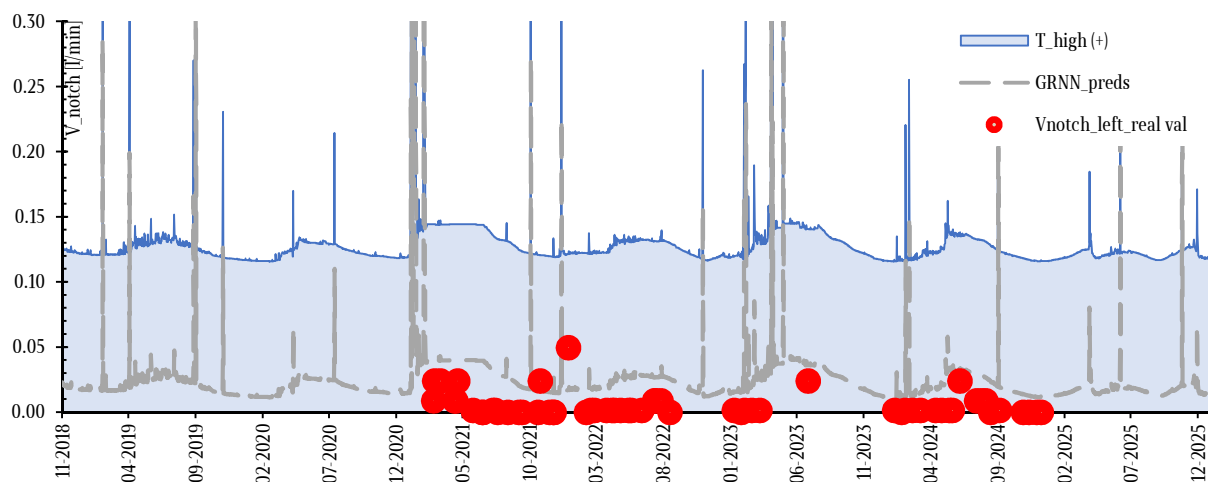
Доколку овие граници се вратат на графички приказ со сетот на предвидени податоци и сетот на мерени податоци на целната варијабла „V_notch_left“ (Прилог 81), може да се согледаат четири инстанци каде мерената филтрација излегува надвор од пресметаните граници. Бидејќи само 4 вредности отстапуваат од познатите 66 мерени податоци (околу 6% од податоците), се претпоставува дека овие вредности се одредна грешка во мерењата. Поради тоа нивниот приказ е изоставен од графици во Прилог 82 и Прилог 83, по што се согледува дека сетот на мерени податоци во целост влегува во границите на предупредување и нема отстапувања. Но, за целосна анализа за состојбата со филтрационите текови под брана „Кнежево“ дополнително се изработени модели за проценка на процедурите води од десен бок (десна дренажа) и од средината на низводното потпорно тело (средна дренажа), по што ќе се донесат соодветни заклучоци за тоа дали отстапувањата над горната граница на филтрацијата од лев бок е аларматна или е сепак појава која е во граници на очекувани вредности.



Прилог 81. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“ без изостанување на екстремните измерени вредности кои значително отстапуваат од границите.



Прилог 82. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“.



Прилог 83. График со приказ на повисоките граници на предупредување „T_high“ на предвидените податоци за целната варијабла „V_notch_left“, со внесени мерени податоци за целната варијабла за периодот на набљудување „V_notch_left_real_val“.

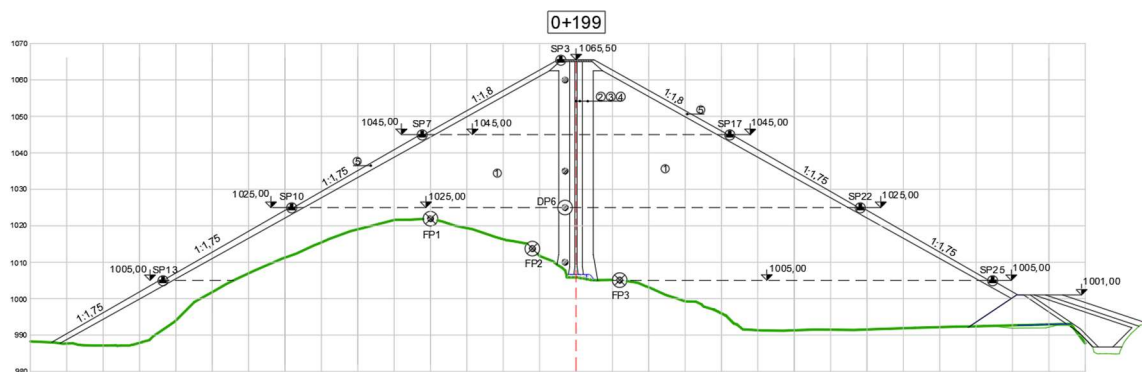
7.3. Анализа на случај 2.3 – $FP_2=f(Zwl)$ и $FP_3=f(Zwl)$

7.3.1. Подготовка на податоците

За потребите за понатамошни анализи поврзани со прогноза на процедните води мерени во десен бок под брана „Кнежево“, изработена е предметната анализа на случај каде се извршува предвидување на нивоите во пиезометрите во фундаментот под брана „Кнежево“ именувани како „FP_2“ и „FP_3“. Станува збор за пиезометри кои се наоѓаат возводно и низводно од асфалтбетонското јадро на браната, во десниот бок од браната. Спроведените анализи се изработени со претпоставка дека постои поврзаност помеѓу измерените филтрациони води во десен бок со нивоите во пизометрите кои се предмет на разгледување. Попречниот пресек со висинската поставеност на пиезометрите е прикажана на Слика 15.

Нивоите во пиезометарот „FP_3“ кој се наоѓа на возводната страна од асфалтбетонското јадро се очекува да покажува хидростатски нивои кои се блиски на мерените нивои на вода во акумулацијата. Од друга страна, поради постоењето на инјекциона завеса под асфалтбетонското јадро, кај пиезометарот „FP_2“ се очекуваат значително пониски мерени нивои.

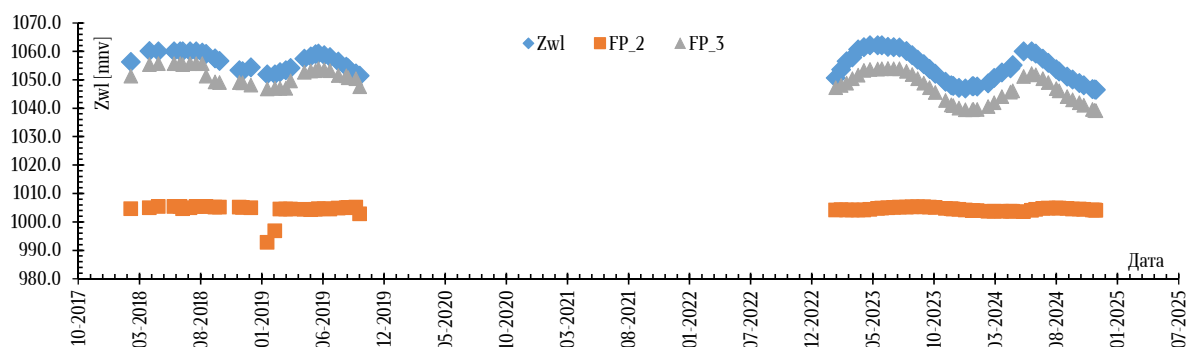
Вредноста на нивоите во овие пиезометри не се оскултира редовно, па поради тоа во моментот поседуваме низа од 76 мерени податоци за нивоите во двата предметни пиезометри.



Слика 15. Приказ на попречен пресек на брана „Кнежево“ со локација на пиезометрите „FP_2“ и „FP_3“ во фундаментот под браната.

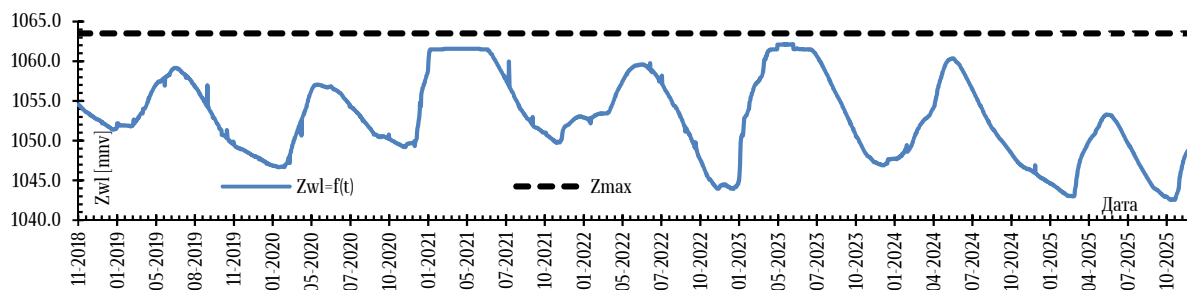
Целта на анализата на случај е да се предвидат нивоите во пиезометрите „FP_2“ и „FP_3“ во однос на еден и единствен предиктор – нивото на вода во акумулацијата за периодот помеѓу 2021 и 2025 година кога имаме голем дел од временскиот домен кога не се регистрираат податоци за нивоите во пиезометрите.

Иако се зборува за една анализа на случај, изработени се два засебни модели каде посебно се тренира пресметката на нивото во пиезометарот „FP_2“ во однос на нивото во акумулацијата „Zwl“, и посебно се тренира пресметката на нивото во пиезометарот „FP_3“ во однос на нивото во акумулацијата Zwl.



Прилог 84. Приказ на вредностите на целните варијабли „FP_2“ и „FP_3“, како и предиктор варијаблата „Zwl“.

Низата на податоци со кои ќе се извршува процесот на обучување на моделите е прикажан на Прилог 84, додека низата на податоци на предиктор варијаблата „Zwl“ со која ќе се извршува предвидување на целните варијабли е прикажана на Прилог 85. Во процесот на предвидување ќе се извршува прогноза на целните варијабли за вкупно 2606 податоци за позната предиктор варијабла.



Прилог 85. Приказ на вредностите на предиктор варијаблата „Zwl“ за временскиот период за кој ќе се извршува предвидување на целните варијабли - нивоите во пиезометрите „FP_2“ и „FP_2“.

7.3.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Перспесија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Двете целни варијабли се користени во оригинална форма, без примена на нормализација или стандардизација на податоците. Од друга страна, предиктор варијаблата беше стандардизирана во рамки на секој сет од петкратната валидација, и тоа само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација за предиктор варијаблата „FP_2“ се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 35.

Табела 35. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_2“.

5-fold CV					
No.	Model	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	GPR	554.814	15.500	4.506	-1715.177
2	ANN	75.819	8.657	6.991	-209.803
3	KNN	2.518	1.407	0.770	-1.749
4	GRNN	2.625	1.428	0.746	-1.588
5	XGBoost	2.817	1.297	0.773	-0.333
6	Decision Tree	0.927	0.907	0.630	-0.245
7	SVR	2.837	1.331	0.669	-0.107
8	Random Forest	1.320	0.940	0.586	-0.078
9	Boosted Trees	0.847	0.860	0.592	-0.047
max value:		554.814	15.500	6.991	-0.047
min value:		0.847	0.860	0.586	-1715.177

Врз основа на добиените резултати, како најуспешен модел се издвојува Boosted Trees, кој

постигна најниски вредности на MSE (0.846) и RMSE (0.860), како и највисока вредност на коефициентот R^2 (-0.047), што претставува најдобар вкупен баланс помеѓу точност и стабилност на анализираниите модели. Многу блиску зад него се наоѓа моделот Random Forest, кој ја постигна најниската вредност на MAE (0.586), што укажува на најмало просечно апсолутно отстапување. Земајќи ги предвид резултатите од прелиминарната петкратна вкрстена валидација, како алгоритам кој се применува во наредниот чекор за обучување на модел за предикција се применува ВТ алгоритмот. Останатите алгоритми, поради неповолните метрички показатели, се исклучуваат од процесот на обучување.

Резултатите од петкратната вкрстена валидација за предиктор варијаблата „FP_3“ се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 36. Во табелата може да се воочат многу подобри метрички показатели од првичната петкратна вкрстена валидација споредено со резултатите на пиезометарот „FP_2“. Очекувано, поради локацијата на пиезометарот „FP_3“, алгоритмите прават поточни и постабилни модели кои ја опишуваат поврзаноста на целната варијабла „FP_3“ и предиктор варијаблата „Zwl“. Имено, согледувајќи ги влезните податоци во Прилог 84, водното ниво во „FP_3“ неприкосновено го следи нивото на водата во акумулацијата „Zwl“, па оттука, алгоритмите со машинско учење успеваат да воспостават стабилни поврзувања помеѓу предиктор и целната варијабла, а со тоа – и подобри метрички показатели на самите модели што директно се согледува од резултатите во Табела 36.

Табела 36. Преглед на вредности на метричките показатели по изработена петкратна валидација на сетот на податоци за обучување на моделите за деветте алгоритми од машинско учење за предиктор варијаблата „FP_3“.

5-fold CV					
No.	Model	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	GPR	5145.093	46.945	14.369	-182.639
2	ANN	85.869	9.194	7.632	-2.515
3	Decision Tree	5.272	2.281	1.685	0.774
4	Boosted Trees	4.366	2.078	1.573	0.813
5	XGBoost	3.602	1.885	1.353	0.847
6	Random Forest	3.512	1.865	1.473	0.853
7	SVR	3.028	1.716	1.402	0.870
8	KNN	2.924	1.697	1.355	0.875
9	GRNN	2.397	1.531	1.262	0.890
max value:		5145.093	46.945	14.369	0.890
min value:		2.397	1.531	1.262	-182.639

Резултатите од петкратната вкрстена валидација за предвидување на нивото на вода во пиезометарот „FP_3“ покажуваат разлики во перформансите помеѓу применетите модели. Најдобри резултати постигнува моделот GRNN, со највисока вредност на коефициентот на детерминација ($R^2 = 0.890$) и најниски вредности на грешките (MSE = 2.397, RMSE = 1.531, MAE = 1.262), што укажува на неговата способност прецизно да ги апроксимира нелинеарните односи во пиезометриските податоци. Многу блиски перформанси покажуваат и K-Nearest Neighbors ($R^2 = 0.875$) и Support Vector Regressor ($R^2 = 0.870$), што дополнително потврдува дека методите базирани на локална сличност и кернел функции се особено погодни за овој тип на проблем. Следуваат Random Forest ($R^2 = 0.853$) и XGBoost ($R^2 = 0.847$), кои иако се моќни ансамбл методи, во овој случај покажуваат нешто послаби резултати, веројатно поради ограничениот обем на податоци и специфичната структура на зависностите. Boosted Trees ($R^2 = 0.813$) и Decision Tree ($R^2 = 0.774$) имаат најниски вредности на R^2 меѓу моделите со стабилно однесување, што е очекувано поради нивната ограничена способност за генерализација и склоност кон претерувања во предвидувањата. Од друга страна, моделите Gaussian Process Regression и Artificial Neural Networks покажуваат исклучително лоши резултати, со негативни вредности на R^2 (-182.639 и -2.515 соодветно) и значително високи грешки, особено кај GPR (MSE = 5145.093), што укажува на целосен неуспех во моделирањето и дека нивните предвидувања се полоши од едноставна апроксимација со средна вредност. Ваквите резултати најверојатно се должат на несоодветна конфигурација, нестабилност

при тренинг или несоодветност на моделот за конкретниот сет на податоци. Генерално, може да се заклучи дека за предвидување на нивото во пиезометарот „FP_3“ најсоодветни се GRNN, KNN и SVR, кои обезбедуваат најдобар баланс помеѓу точност и стабилност и претставуваат најперспективен избор за понатамошна примена во анализа и мониторинг на однесувањето на браната.

Како заклучок од анализата на метричките показатели од прелиминарната петкратна вкрстена валидација, за понатамошна анализа на целната варијабла „FP_2“ разгледувани се резултатите од BT обучениот модел, додека за целната варијабла „FP_3“ разгледувани се резултатите од GRNN обучениот модел.

7.3.3. Обучување на моделите

7.3.3.1. Модел со Поткрени Регресиони Дрва (BT) за предвидување на $FP_2=f(Zwl)$

Единствениот усвоен модел за обучување на целната варијабла „FP_2“ се базира на алгоритмот на Boosted Regression Trees, односно Gradient Boosting Regressor, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја.

Моделот е дефиниран со следните параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на перформансите на моделот применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 37. Gradient Boosting, како бустинг метод, е особено ефикасен во моделирање на сложени нелинеарни односи и често постигнува висока предиктивна точност.

Табела 37. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $FP_2=f(Zwl)$.

Fold	MSE	RMSE	MAE	R2
1	0.473	0.688	0.595	-1.080
2	1.428	1.195	0.759	0.651
3	0.357	0.597	0.459	-0.390
4	1.716	1.310	0.763	0.809
5	0.260	0.510	0.381	-0.225

Просечните вредности на метричките показатели, се: Avg MSE: 0.846, Avg RMSE: 0.860, Avg MAE: 0.591, Avg R^2 : -0.047. Потребното време за обучување беше 96ms.

7.3.3.2. Модел со Генерализирана Регресиона Невронска Мрежа (GRNN) за предвидување на $FP_3=f(Zwl)$

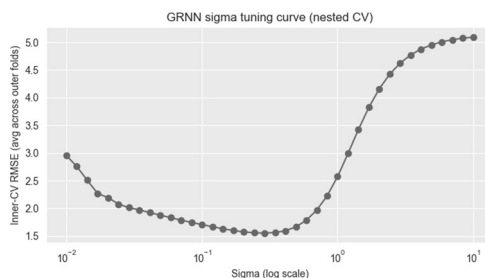
Единствениот обучен модел за целната варијабла „FP_3“ е изработен со примена на алгоритмот на Генерализирана Регресиона Невронска Мрежа (анг. „General Regression Neural Network“).

Во кодот е применета вгнездена трикратна валидација за избор на најдобра вредност на параметарот σ – сигма. Се тестираат вредности во рангот $\sigma = (0.01 - 10)$. Критериумот е да се избере вредност за σ што ќе резултира со минимален RMSE, а во конкретниот случај графикот на варијација на RMSE во зависност од параметарот σ по трикратната вкрстена валидација е прикажан во Прилог 86. Понатаму, со изборот на соодветна вредност за параметарот σ и

дополнително се прави петкратна надворешна вкрстена валидација на податоците.

Табела 38. Вредност на параметарот σ и RMSE за секој сет од петкратната вкрстена валидација за моделот $FP_3=f(Zwl)$.

Fold	Best Sigma	Fold RMSE
1	0.289	1.524
2	0.242	1.723
3	0.289	1.097
4	0.346	1.623
5	0.289	1.692



Прилог 86. Вредност на параметарот σ добиен со вгнездената трикратна вкрстена валидација за моделот $FP_3=f(Zwl)$.

Од графикот во Прилог 86 може да се воочи дека минимумот на $RMSE = 1.097$ се добива за вредност на параметарот $\sigma \approx 0.289$.

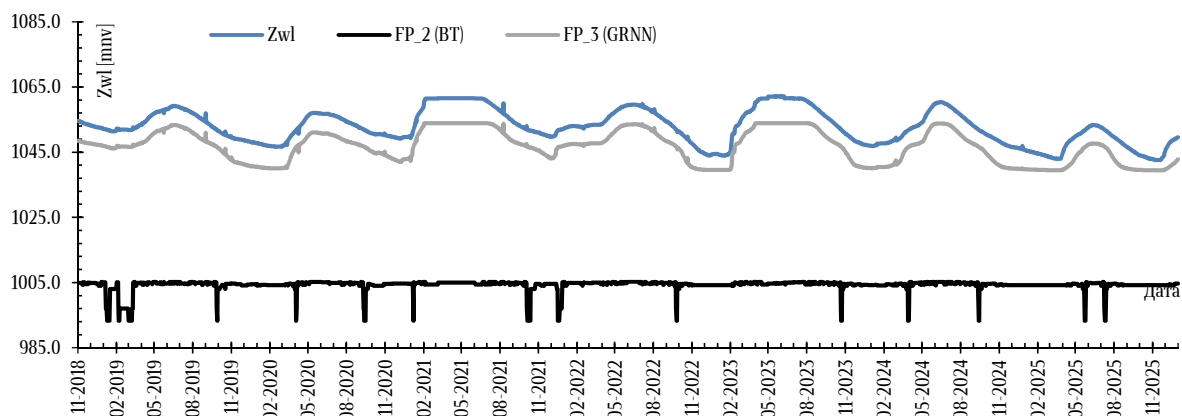
7.3.4. Предвидување на целните варијабли $FP_2=f(Zwl)$ и $FP_3=f(Zwl)$

По извршеното обучување на моделите BT – за предвидување на целната варијабла „FP_2“, и GRNN – за предвидување на целната варијабла „FP_3“, извршени се предвидувања на целните варијабли врз основа на позната предиктор варијабла „Zwl“ – ниво на вода во акумулацијата за периодот од 2018 до 2025 година. Станува збор за вкупно 2606 податоци на предиктор варијаблата за период кога постојат местимични мерења на целните варијабли, прикажани на графиците во Прилог 84 и Прилог 85.

Предвидувањата на целните варијабли се прикажани во графикот во Прилог 87. Од резултатите може да се воочи дека нивото во пиезометарот „FP_3“ во целост го следи нивото на вода во акумулацијата, додека водното ниво во „FP_2“ осцилира во согласност со флукуацијата на водното ниво во акумулацијата, но на значително пониско ниво (под 1005 mnnv) и во потесен коридор, со голем дел од периодот на предвидување каде речиси нивото е константно.

Минималното ниво во „FP_2“, согласно резултатите, е на кота 993.226 mnnv, додека максималното ниво во „FP_2“ е 1005.208 mnnv.

Кај пиезометарот „FP_3“, минималното ниво од резултатите со предвидувањата на вредноста изнесува 1039.410 mnnv, додека максималното ниво изнесува 1053.914 mnnv.



Прилог 87. График со приказ на предвидувањата на целните варијабли FP_2 и FP_3 во однос на предиктор варијаблата Zwl за разгледуваниот временски период од 2018 до 2025 година.

7.4. Анализа на случај 2.4 – $V_{\text{notch_right}} = f(Z_{wl}, Q_{\text{inflow}}, FP_2, FP_3)$

7.4.1. Подготовка на податоците

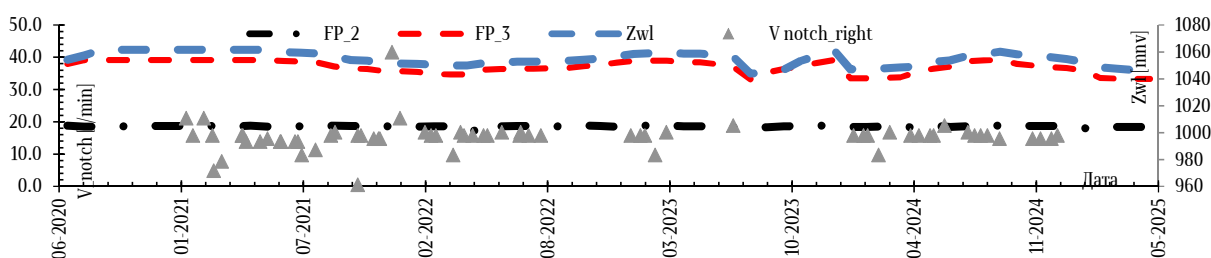
За анализа на процедните води прибрани со дренажен колектор во низводното потпорно тело во десен бок на брана „Кнежево“, се извршува анализа на случај 2.4 каде целна варијабла е пресметката на процедните води „ $V_{\text{notch_right}}$ “ во [l/min] со земање предвид на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „ Z_{wl} “ во [mnv] прикажано во Прилог 88,
- 2) Дотек во акумулацијата, „ Q_{inflow} “ во [m³/s] прикажано во Прилог 89,
- 3) Ниво на вода во пиезометарот „ FP_3 “ лоциран во фундаметот на браната, возводно од асфалтбетонското јадро и инјекционата завеса, „ FP_3 “ во [mnv] прикажано во Прилог 88, и
- 4) Ниво на вода во пиезометарот „ FP_2 “ лоциран во фундаметот на браната, низводно од асфалтбетонското јадро и инјекционата завеса, „ FP_2 “ во [mnv] прикажано во Прилог 89.

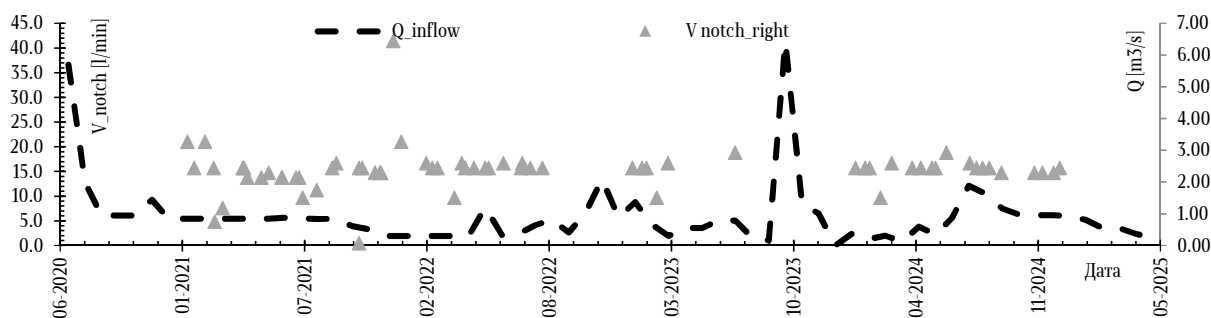
Целната варијабла – процедни води од десен бок под тело на браната „ $V_{\text{notch_right}}$ “ се регистрира низводно од брана „Кнежево“ со периодични мерења во [l/min]. Со цел проценка на овие води за целосниот период од 2018 до 2025 година, се изработува предметната анализа со модели базирани на машинско учење кои треба да извршат предвидување на целната варијабла во однос на посочените предиктор варијабли.

Процесот на обучување се извршува за 66 податоци на целната варијабла и предиктор варијаблите (Прилог 88 и Прилог 89), додека предвидувањето се извршува за 2606 податоци на предиктор варијаблите. Во процесот на предвидување, вредноста на целната варијабла е непозната. Вредностите на предиктор варијаблите користени во процесот на предвидување се прикажани во Прилог 90.

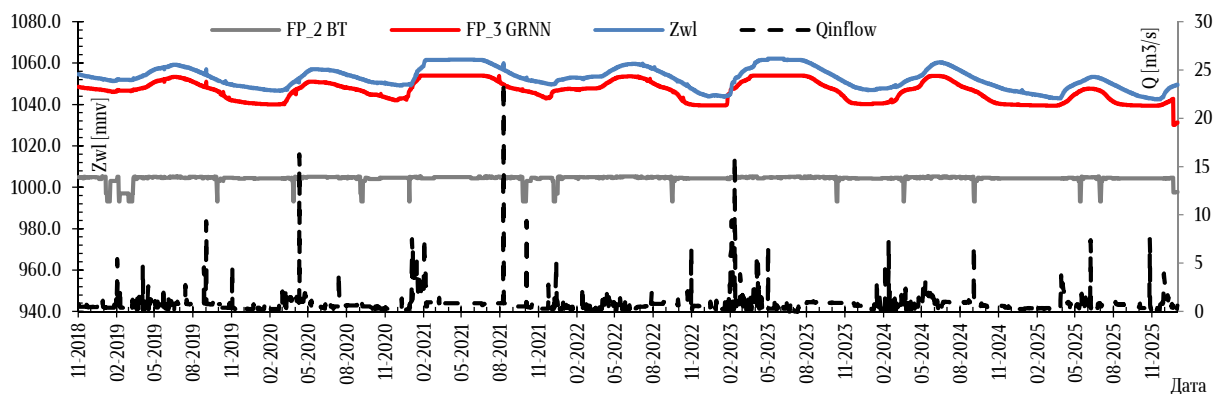
Вредностите на предиктор варијаблите „ Z_{wl} “ и „ Q_{inflow} “ се преземени од теренски мерења, додека вредностите на предиктор варијаблите „ FP_2 “ и „ FP_3 “ се преземени од студијата на случај 2.3. каде нивоите во пиезометрите се пресметуваат со помош на модели со машинско учење.



Прилог 88. График со приказ на вредностите на целната варијабла „ $V_{\text{notch_right}}$ “ и предиктор варијаблите „ FP_2 “, „ FP_3 “ и „ Z_{wl} “ применети во процесот на обучување на моделите.



Прилог 89. График со приказ на вредностите на целната варијабла „ $V_{\text{notch_right}}$ “ и предиктор варијаблата „ Q_{inflow} “ кои се применети во процесот на обучување на моделите.



Прилог 90. График со приказ на вредностите на предиктор варијаблите „FP_2“, „FP_3“, „Zwl“ и „Q_inflow“ применети во процесот на предвидување на целната варијабла „V_notch_right“.

7.4.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

За проценка на моделите и нивната предикциона способност, за анализираниот систем $V_{\text{notch_right}} = f(Zwl, Q_{\text{inflow}}, FP_2, FP_3)$ изработена е прелиминарна петкратна вкрстена валидација со деветте алгоритми за машинско учење. За оцена на креираните модели, прикажани се резултати од метричките параметри во Табела 39.

Табела 39. Резултати од метричките показатели од прелиминарната петкратна вкрстена валидација на моделите за $V_{\text{notch_right}} = f(Zwl, Q_{\text{inflow}}, FP_2, FP_3)$.

5-fold CV					
No.	Algorithm	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	XGBoost	38.43	5.75	3.40	-10.54
2	ANN	24.24	4.81	3.63	-8.53
3	Decision Tree	41.36	5.82	3.75	-8.13
4	Boosted Trees	34.36	5.28	3.45	-6.49
5	GPR	25.27	4.40	2.54	-5.29
6	KNN	24.56	4.26	2.88	-1.54
7	Random Forest	24.09	4.18	2.64	-1.33
8	SVR	23.18	4.05	2.53	-0.79
9	GRNN	25.98	4.73	2.68	-0.56
max value:		41.36	5.82	3.75	-0.56
min value:		23.18	4.05	2.53	-10.54

Резултатите добиени од примената на петкратната вкрстена валидација за предвидување на променливата „V_notch_right“ врз основа на влезните параметри „Zwl“, „FP_2“, „FP_3“ и „Q_inflow“ укажуваат на генерално незадоволителни перформанси кај сите разгледувани модели, што се потврдува преку негативните вредности на коефициентот на детерминација R^2 . Ова значи дека ниту еден модел не успева да ја објасни варијабилноста на целната променлива подобро од едноставна апроксимација со нејзината средна вредност. Сепак, дополнителната анализа на опсегот на податоците, при што „V_notch_right“ се движи од 0.49 до 41.516 l/min, овозможува подлабоко разбирање на причините за ваквото однесување. Имено, станува збор за променлива со многу широк динамички опсег, што укажува на висока варијабилност на процесот на филтрација. Во такви услови, иако апсолутните грешки ($RMSE \approx 4-6$ l/min) на прв поглед може да изгледаат умерени, тие претставуваат значителни релативни отстапувања кај пониските вредности на протекување, додека кај повисоките вредности имаат помало влијание. Оваа нерамномерна распределба на грешката долж опсегот на податоците доведува до неможност моделите да ја репродуцираат вкупната варијанса, што резултира со негативни вредности на R^2 .

Ваквите резултати укажуваат дека однесувањето на системот не е едноставно и не може да се опише со единствена глобална функција, туку најверојатно е резултат на сложени, нелинеарни и

режимски зависни процеси. Можно е при ниски вредности на дотекувањата во акумулацијата, врз филтрацијата да доминираат локални ефекти како геомеханичките карактеристики на материјалите, додека при повисоки вредности на дотекувањата, количеството на вода да е контролирана од различни управувачки механизми како зголемено испуштање на протек низ темелниот испуст. Дополнително, треба да се нагласи дека филтрациските процеси кај браните имаат изразен временски карактер, односно одложен одговор на промените во нивото на акумулацијата и протокот, што не е опфатено со анализираниот статички модел. Во комбинација со релативно малиот број на податоци (66), ова значително ја ограничува способноста на моделите да идентификуваат стабилни и генерализирачки зависности.

Оттука, може да се заклучи дека слабите перформанси на моделите не се должат исклучиво на изборот на алгоритмите за подучување на моделите, туку пред сè на комплексноста на физичкиот процес и на недоволната репрезентативност на влезните параметри. Овие резултати укажуваат дека за адекватно моделирање на „V_notch_right“ е неопходен понапреден пристап кој ќе ја земе предвид временската динамика на системот, можната нелинеарност и различните режими на однесување, како и трансформација на податоците со цел намалување на варијабилноста. Иако на прв поглед резултатите се незадоволителни, тие претставуваат значаен резултат, бидејќи јасно укажуваат дека процесот на филтрација кај разгледуваниот систем не може да се опише со едноставни статички релации, што е важна основа за понатамошно унапредување на моделите и подобро разбирање на однесувањето на браната.

7.4.3. Обучување на моделите

По извршената прелиминарна петкратна вкрстена валидација, обучени се три модели – RF, KNN и SVR - за понатамошна предикција на целната варијабла „V_notch_right“ врз основа на познати вредности на предиктор варијабли „Zwl“, „FP_2“, „FP_3“ и „Q_inflow“ и самата целна варијабла „V_notch_right“. Станува збор за краток сет на податоци од само 66 вредности регистрирани во рамки на редовната оскултација на брана „Кнежево“. Како што следи дадени се кратки информации за обучените модели.

7.4.3.1. Модел на Случајна шума (RF)

Првиот обучен модел е моделот на Random Forest перцепсија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја.

За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators` (50, 100, 150) - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth` (None, 10, 20) - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split` (2, 5, 10) - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features` ("sqrt", "log2", None) - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијабли.

За добивање на објективна проценка на перформансите применета вгнездена n-кратна валидација, и тоа:

- Внатрешна трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку т.н. GridSearchCV.
- Надворешна петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Овој пристап обезбедува непристрасна проценка на перформансите и ја намалува веројатноста за пренагласено прилагодување на моделот кон податоците. Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување.

Табела 40. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$.

Fold	max_depth	max_features	min_samples_split	n_estimators
1	None	sqrt	10	150
2	None	sqrt	10	150
3	None	sqrt	10	150
4	None	None	10	50
5	None	sqrt	10	100

Финалното обучување е направено согласно хиперпараметрите од петтиот сет од петкратната надворешна валидација, со вредности: 'max_depth': None, 'max_features': sqrt, 'min_samples_split': 10, 'n_estimators': 100. Процесот на обучување траеше 39s и 829ms.

7.4.3.2. Модел на К – најблиски соседи (KNN)

Вториот алгоритам земен предвид во предметната студија на случај е алгоритмот К-најблизок сосед (анг. „K-nearest neighbor“) или накратко KNN за обучување со сетот на влезни податоци како единствен модел кој покажува задоволителни резултати на метричките показатели.

Моделот е дефиниран со параметар $n_neighbors = 5$, што значи дека за секое предвидување се земаат предвид 5 најблиски примероци од податоците за обучување.

Бидејќи KNN се базира на растојанија, примената на т.н. StandardScaler во рамки на моделирањето е од клучно значење. На овој начин, сите предиктор варијабли се стандардизираат (средна вредност 0 и стандардна девијација 1), со што се избегнува доминација на променливи со поголем размер. Скалирањето се извршува внатре во секој сет од петкратна вкрстена валидација, што обезбедува правилна и непристрасна евалуација.

За проценка на перформансите е применета петкратна вкрстена валидација, при што моделот се обучува и тестира на различни подсетови од податоците.

Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста на моделот. Конкретните вредности за моделот се дадени во Табела 41.

Табела 41. Вредности на метричките показатели од процесот на обучување на KNN моделот по спроведена петкратна валидација за студија на случај $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$.

Fold	MSE	RMSE	MAE	R2
1	13.286	3.645	2.753	-0.373
2	84.213	9.177	5.248	-0.063
3	13.669	3.697	2.617	-0.385
4	7.261	2.695	2.006	-1.699
5	4.352	2.086	1.778	-5.203

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 24.55, Avg RMSE: 4.2599, Avg MAE: 2.88, Avg R^2 : -1.544. Процесот на обучување траеше 46ms.

7.4.3.3. Модел на Регресија со потпомогнати вектори (SVR)

Третиот обучен модел е моделот на Регресија со потпомогнати вектори или накратко SVR (анг. „Support Vector Regressor“).

Моделот е дефиниран со следните параметри:

1) kernel = "rbf" - Се користи радијална базисна функција (Radial Basis Function), која овозможува

моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) $C = 100$ - Регуларизациски параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на C значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување.

3) $\epsilon = 0.1$ - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

Дополнително, моделот е вклучен во кодот заедно со `StandardScaler`, со што преиктор варијаблите се стандардизираат (средна вредност 0 и стандардна девијација 1) во рамки на секој сет од петкратна вкрстена валидација. Ова е особено важно за SVR, бидејќи моделот е чувствителен на размерот на влезните податоци.

За евалуација на моделот е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Средната вредност на метричките показатели по обучувањето на моделот изнесуваат: Avg MSE = 23.183, Avg RMSE = 4.050, Avg MAE = 0.2531, Avg R^2 = -0.793. Процесот на обучување траеше 51ms.

7.4.4. Концепт на тежинско осреднување (анг. „weighted averaging“)

Во ситуации кога поединечните модели на машинско учење не покажуваат задоволителни перформанси, особено кога вредностите на коефициентот на детерминација R^2 се негативни како што е случајот кај анализираниот проблем на предвидување на филтрацијата „V_notch_right“, оправдан пристап претставува примена на ансамбл техники, односно комбинирање на резултатите од предвидувањето на повеќе модели со цел добивање на постабилна прогноза. Една од наједноставните, но ефективни техники во рамките на ансамбл пристапите е т.н. тежинско осреднување (анг. „weighted averaging“), при што конечната предвидена вредност се добива како тежински просек од предвидувањата на повеќе поединечни модели. Како што следи даден е краток опис на тежинското среднување, кое во наредна фаза ќе се примени по извршеното предвидување на целната варијабла.

Доколку се разгледаат n различни модели, при што секој модел дава предвидување y_{pred} , конечната прогноза се дефинира како $y_{pred} = \sum_{i=1}^n w_i y_i$, при што w_i претставуваат тежините доделени на секој модел, со услов $\sum_{i=1}^n w_i = 1$. Клучниот аспект во оваа метода е определување на тежините, кои во практиката најчесто се дефинираат врз основа на точноста на моделите, односно обратно пропорционално на нивната грешка. Во овој труд, тежините се пресметуваат преку метриката RMSE (анг. „Root Mean Square Error“), според изразот $w_i = (1/RMSE_i) / (\sum_{j=1}^n 1/RMSE_j)$, со што моделите со помала грешка добиваат поголемо влијание врз конечната прогноза, додека моделите со поголема грешка имаат ограничен придонес. Ваквиот пристап е особено соодветен во услови кога податоците се ограничени по обем (на пример 66 мерења), содржат значителен шум и кога зависноста помеѓу влезните и излезната променлива е слаба или комплексна, како што е често случај кај процесите на филтрација и проценката на проток низ тело на брана или нејзината основа. Со комбинирањето на повеќе модели се постигнува намалување на варијансата на грешката, бидејќи различните модели имаат различни слабости и прават различни грешки, кои со осреднувањето делумно се поништуваат, што резултира со постабилни и физички прифатливи предвидувања. Дополнително, ансамбл моделите покажуваат поголема робустност кон шум во податоците и намалена зависност од специфичните карактеристики на поединечен алгоритам.

Практичната имплементација на овој пристап подразбира избор на неколку најдобри модели (вообичаено 2 до 4) врз основа на нивните перформанси (најнизок RMSE), пресметка на нивните тежини, и потоа комбинирање на нивните предвидувања за секој нов влезен пример. Сепак, важно е да се нагласи дека иако тежинското усреднување може да доведе до подобрување на резултатите,

тоа не го елиминира фундаменталниот проблем поврзан со квалитетот и количината на податоците, ниту пак создава нова информација, туку претставува оптимално искористување на постоечките модели. Оттука, добиените резултати треба да се интерпретираат како најдобра можна апроксимација во рамките на достапните податоци, а не како високо прецизна прогноза.

Враќајќи се еден чекор назад, со анализа на резултатите добиени преку петкратната вкрстена валидација согледуваме дека ниту еден од применетите модели не постигнува задоволителна предиктивна способност, што е евидентно од негативните вредности на коефициентот на детерминација R^2 кај сите алгоритми. Ова укажува дека моделите не успеваат адекватно да ја опишат зависноста помеѓу влезните променливи и целната величина, при што нивните предвидувања се положи во однос на едноставен модел базиран на средна вредност. Сепак, според метриката RMSE, која во овој контекст претставува најрелевантен индикатор за точноста, може да се изврши релативна споредба на перформансите. Најдобри резултати се постигнати со Support Vector Regression (RMSE = 4.05), следен од Random Forest (RMSE = 4.18) и K-Nearest Neighbors (RMSE = 4.26). Овие модели покажуваат најмала грешка и релативно подобра стабилност во споредба со останатите алгоритми, иако нивните R^2 вредности остануваат негативни. Од друга страна, моделите базирани на градиентно бустирање и поединечни дрва (како што се XGBoost, Decision Tree и Boosted Trees) покажуваат значително полоши резултати, што укажува на нивна несоодветност за конкретниот проблем и достапниот сет на податоци. Поради тоа, наместо избор на еден модел, се препорачува примена на ансамбл пристап базиран на тежинско осреднување, при што во финалната прогноза се комбинираат предвидувањата од најдобрите модели (SVR, Random Forest и KNN), со тежини дефинирани обратно пропорционално на нивната грешка.

Пресметаните тежини на RMSE од петкратната вкрстена валидација на избраните три модели со најдобри вредности на RMSE, се дадени како што следи.

Најпрво, општата равенка за пресметка на тежинските коефициенти за осреднување, гласи:

$$w_i = \frac{\left(\frac{1}{RMSE_i}\right)}{\left(\sum_{j=1}^n \frac{1}{RMSE_j}\right)} \quad (7.4.1)$$

Инверзните RMSE вредности на RMSE за секој модел се пресметуваат како:

$$\frac{1}{RMSE_{SVR}} = \frac{1}{4.05} = 0.2469; \frac{1}{RMSE_{RF}} = \frac{1}{4.18} = 0.2392; \frac{1}{RMSE_{KNN}} = \frac{1}{4.26} = 0.2347 \quad (7.4.2)$$

Сумата на инверзните вредности на RMSE изнесува:

$$SUM_{RMSE} = 0.7208 \quad (7.4.3)$$

Оттука, тежините се пресметуваат како:

$$w_{SVR} = \frac{0.2469}{0.7208} = 0.3425; w_{RF} = \frac{0.2392}{0.7208} = 0.3318; w_{KNN} = \frac{0.2347}{0.7208} = 0.3257 \quad (7.4.4)$$

Тежинските коефициенти ќе се применуваат во наредниот чекор, по пресметка на предвидувањата на целната варијабла „V_notch_right“. Имајќи ги предвид тежинските коефициенти, крајните вредности на целната варијабла кои ќе се земаат предвид како релевантни, ќе се пресметаат како:

$$y_{pred} = w_{SVR} \cdot y_{pred_{SVR}} + w_{RF} \cdot y_{pred_{RF}} + w_{KNN} \cdot y_{pred_{KNN}} \quad (7.4.5)$$

7.4.5. Предвидување на целната варијабла

Целна варијабла во предметната студија на случај е количината на филтрациони води „V_notch_right“ измерени во прелив од десен бок на браната „Кнежево“. Со помош на обучениот модел од машинско учење, се бара предвидување на целната варијабла со познавање на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „Zwl“ во [mnv] прикажано во Прилог 88,
- 2) Дотек во акумулацијата, „Q_inflow“ во [m³/s] прикажано во Прилог 89,

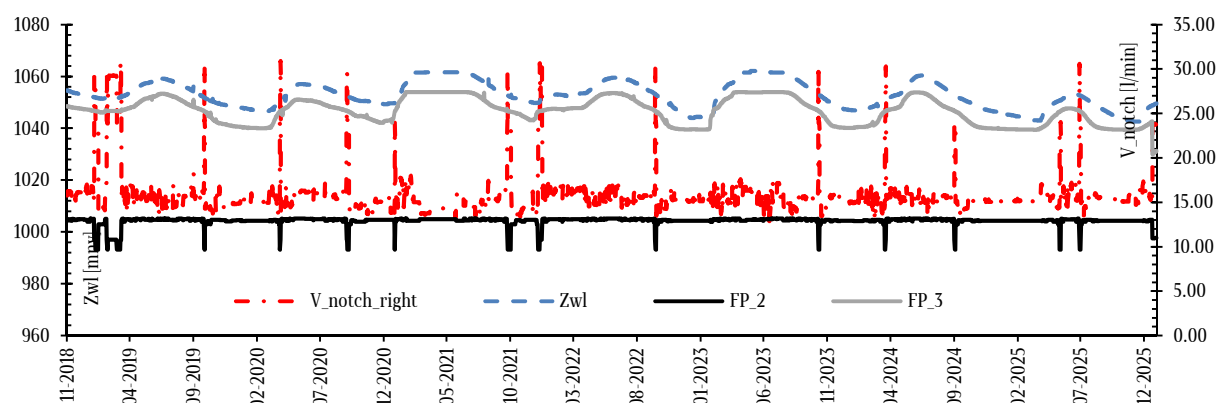
3) Ниво на вода во пиезометарот „FP_3“ лоциран во фундаметот на браната, возводно од асфалтбетонското јадро и инјекционата завеса, „FP_3“ во [mnv] прикажано во Прилог 88, и

4) Ниво на вода во пиезометарот „FP_2“ лоциран во фундаметот на браната, низводно од асфалтбетонското јадро и инјекционата завеса, „FP_2“ во [mnv] прикажано во Прилог 89.

Обучени се вкупно три модели со машинско учење со три различни алгоритми кои се проценети како најдобри од претходната фаза на прелиминарна петкратна вкрстена валидација. Станува збор за моделите SVR, RF и KNN.

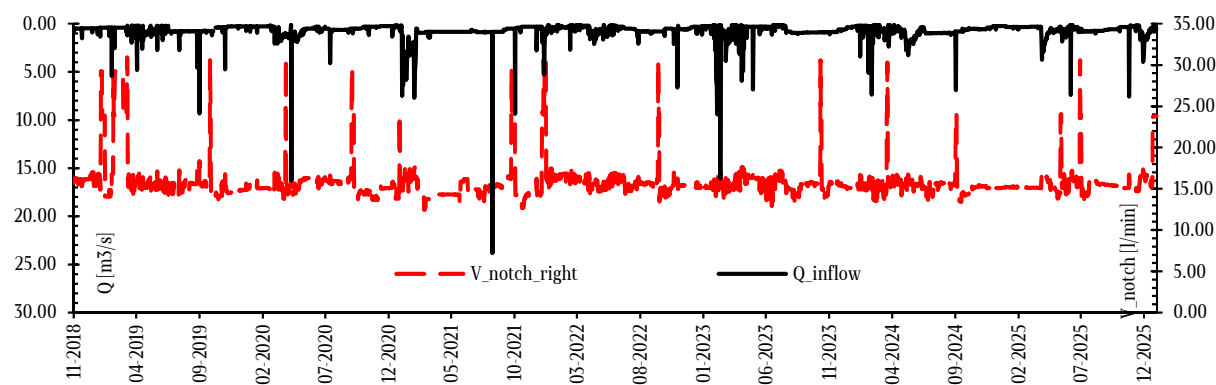
Процесот на предвидување се извршува за 2606 познати вредности на предиктор варијаблите, кои временски се однесуваат на периодот на прибирање на податоците од 1^{ви} ноември 2018 година до 31 декември 2025 година. За анализираниот период нема континуирани вредности на целната варијабла бидејќи истата се регистрира од страна на операторот во нееднакви временски интервали.

Пресметаните предвидувања, согласно согледувањата од метричките показатели на моделите, се осреднети со тежински коефициенти како што е во детали појаснето во Поглавје 7.4.3. Резултатите од осреднетите предвидувања се прикажани во Прилог 91 и Прилог 92. Прикажаните резултати од предвидувањата се тежински осреднети вредности на целната варијабла, со земање предвид на добиените предвидувања од моделите тренирани со алгоритмите SVR, RF и KNN.



Прилог 91. График со приказ на предвидените вредности на целната варијабла „V_notch_right“, предиктор варијаблите „Zwl“, „FP_2“ и „FP_3“.

Во Прилог 91 прикажани се резултати од предвидувањето на целната варијабла „V_notch_right“ како и предиктор варијаблите „Zwl“, „FP_2“ и „FP_3“, додека во Прилог 92 е прикажан график на целната варијабла и предиктор варијаблата „Q_inflow“. Од графичкиот приказ може да се воочи дека филтрационите води не следат никаков особен тренд од предиктор варијаблите. Предвидувањата на целната варијабла се движат во границите од 12.436 [l/min] до 30.926 [l/min] или средна вредност од 15.785 [l/min].



Прилог 92. График со приказ на предвидените вредности на целната варијабла „V_notch_right“, и предиктор варијаблата „Q_inflow“.

7.4.6. Дефинирање на граници за предупредување (thresholds)

Слично на анализата за дефинирање на граници за предупредување кај „V_notch_left“ на предвидените вредности на филтрационите води од лев бок, изработена е пресметка на границите за предупредување за целната варијабла која е предмет на анализа на студија на случај 2.4 – „V_notch_right“.

Границите на предупредување се дефинираат во однос на пресметана σ - стандардна девијација од осреднетите предвидени вредности податоци на целната варијабла, вкупно 2606 на број.

По дефинирање на границите на предупредување како оддалечени вредности од предвидените, во пресметаните коридори на границите се вметнуваат мерените податоци од филтрационото количество низ десниот дренажен колектор и се прави споредба на мерењата согласно математички дефинираните граници. Целта е да се воочат отстапувања на мерењата надвор од границите на предупредување, што би резултирало со алармирање на операторот за невообичаени истекувања низ дренажниот колектор.

Постапката за пресметка на границите за предупредување е слична на пресметката изработена во Поглавје 7.2.5.

Стандардната девијација во конкретниот код се пресметува во ќелијата ‘Statistics of predictions’ во Пајтон кодот, веднаш по пресметка на предвидувањата. Математичката релација за пресметка на стандардната девијација од податочното множество, гласи:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - x_{ave})^2} \quad (7.4.6)$$

, каде што:

σ – стандардна девијација на анализираниот сет на податоци,

n – број на податоци во сетот,

x_{ave} – средна вредност на податоците во сетот,

x_i – поединечна вредност на секој податок во сетот.

Пресметаната стандардна девијација од сетот на предвидени вредности на ‘V_notch_right’ изнесува $\sigma = 2.3142$. Средната вредност на сетот на податоци изнесува $x_{ave} = 15.7855$ [l/min].

Првата, пониска граница на предупредување се дефинира како:

$$T_{low}^+ = x_i + 2 \cdot \sigma; T_{low}^- = x_i - 2 \cdot \sigma; \quad (7.4.7)$$

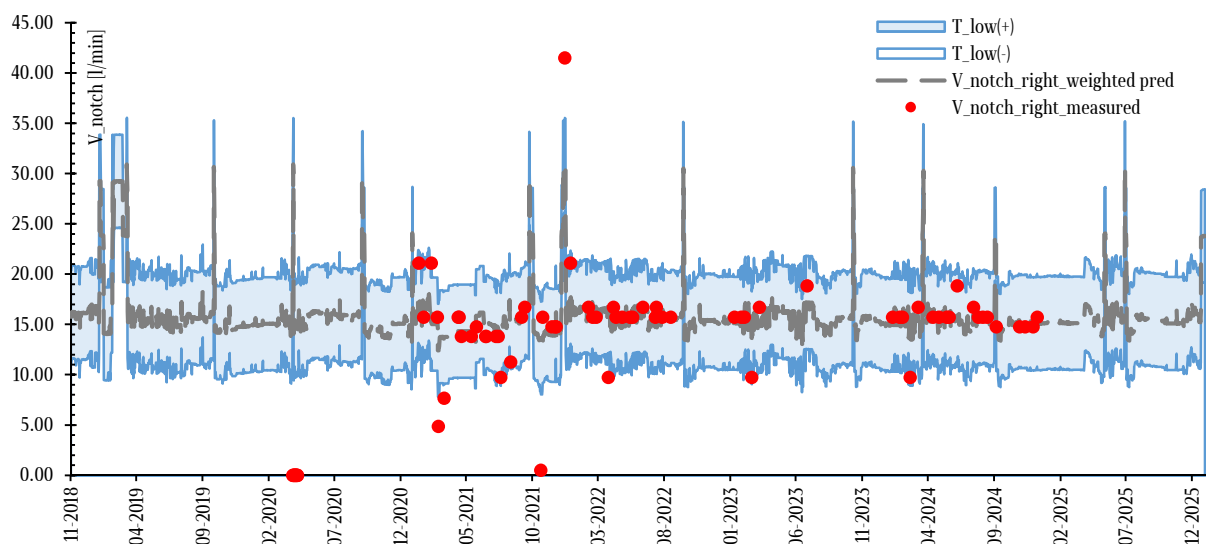
Втората, повисока граница на предупредување се дефинира како:

$$T_{high}^+ = x_i + 3 \cdot \sigma; T_{low}^- = x_i - 3 \cdot \sigma; \quad (7.4.8)$$

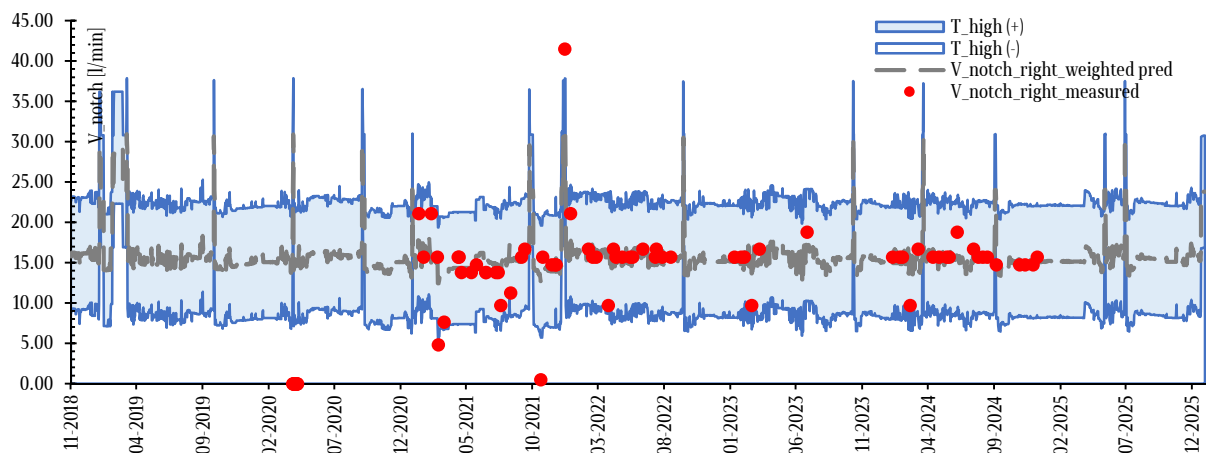
Двете граници на предупредување варираат во однос на осцилацијата на предвидените вредности на целната варијабла x_i . Границите претставуваат оддалечени вредности од предвидените вредности на целната варијабла, согласно математичката релација во равенките 7.4.7 и 7.4.8. Пресметаните коридори на дозволените граници на филтрационото количество низ десниот дренажен колектор се прикажани во Прилог 93 и Прилог 94.

Во Прилог 93 е даден приказ на коридорот дефиниран со пониските граници на предупредување T_{low} , додека во Прилог 94 е даден приказ на коридорот дефиниран со повисоките граници на предупредување T_{high} . Од двата прилози може да се заклучи дека мерените податоци за филтрацијата низ десниот дренажен колектор се во граници на очекуваните вредности. Најголем дел од мерените вредности; 60 од вкупно 66 се во граници на дефинираните коридори, а единствено 6 вредности отстапуваат од пресметаните граници.

Со добиените резултати може да се заклучи дека филтрацијата е во граници на очекуваните вредности.



Прилог 93. График со приказ на пониските граници на предупредување „T_low“ на предвидените податоци за целната варијабла „V_notch_right“, со приказ на измерените вредности за филтрацијата во разгледуваниот период „V_notch_right_measured“.



Прилог 94. График со приказ на повисоките граници на предупредување „T_high“ на предвидените податоци за целната варијабла „V_notch_right“, со приказ на измерените вредности за филтрацијата во разгледуваниот период „V_notch_right_measured“.

7.5. Анализа на случај 2.5 – $V_{notch_middle}=f(Zwl, Q_{inflow})$

7.5.1. Подготовка на податоците

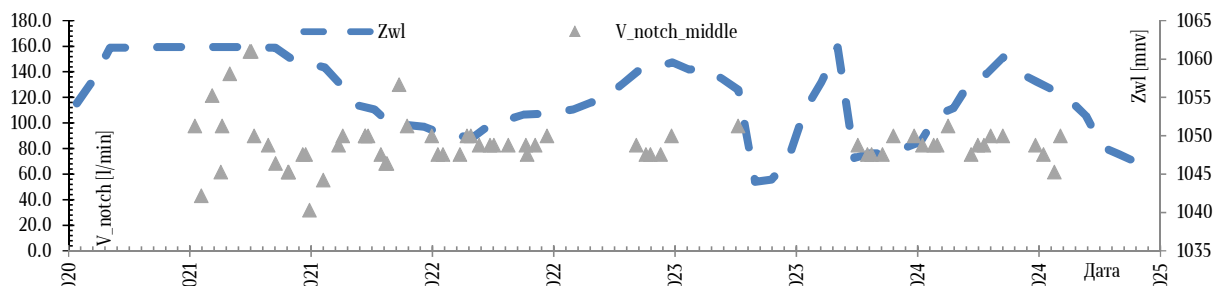
За анализа на процедурните води прибрани со дренажен колектор кој минува централно низ потпорното низводно тело на брана „Кнежево“, се извршува анализа на случај 2.5 каде целна варијабла е пресметката на процедурните води „V_notch_middle“ во [l/min] со земање предвид на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „Zwl“ во [mnv] прикажано во Прилог 88,
- 2) Дотек во акумулацијата, „Q_inflow“ во [m³/s] прикажано во Прилог 89,

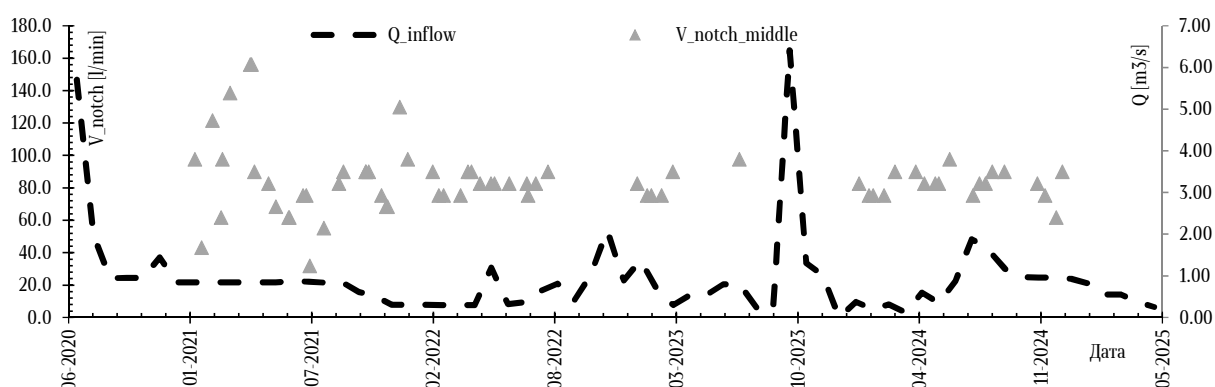
Целната варијабла – процедурни води од централен појас под тело на браната „V_notch_middle“ се регистрира низводно од брана „Кнежево“ со периодични мерења во [l/min]. Со цел проценка на овие води за целосниот период од 2018 до 2025 година, се изработува предметната анализа со модели базирани на машинско учење кои треба да извршат предвидување на целната варијабла во однос на посочените предиктор варијабли.

Процесот на обучување се извршува за 66 податоци на целната варијабла и предиктор варијаблите (Прилог 95 и Прилог 96), додека предвидувањето се извршува за 2606 податоци на предиктор варијаблите. Во процесот на предвидување, вредноста на целната варијабла е непозната. Вредностите на предиктор варијаблите користени во процесот на предвидување се прикажани во Прилог 97.

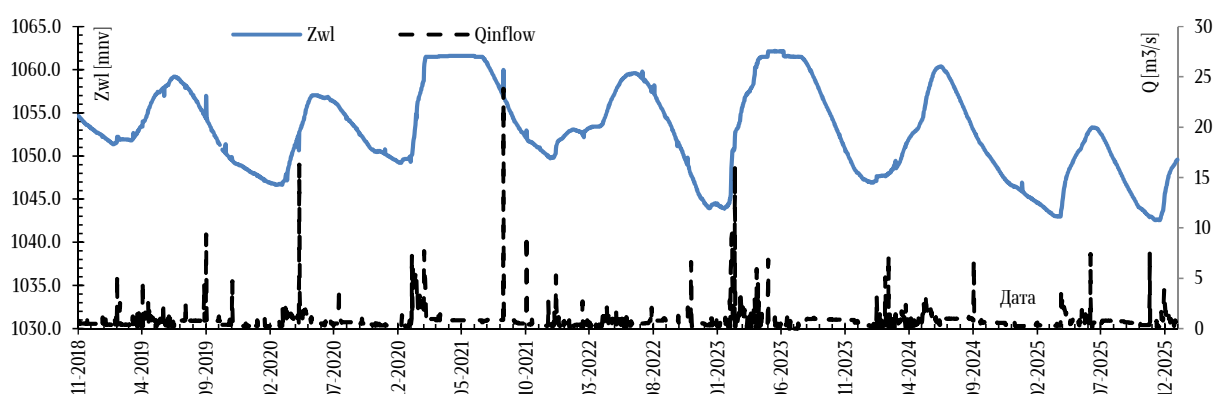
Вредностите на предиктор варијаблите „Zwl“ и „Q_inflow“ се преземени или пресметани од теренски мерења регистрирани во невоедначени временски интервали од страна на операторот на хидросистемот.



Прилог 95. График со приказ на вредностите на целната варијабла „V_notch_middle“ и предиктор варијаблата „Zwl“ применети во процесот на обучување на моделите.



Прилог 96. График со приказ на вредностите на целната варијабла „V_notch_middle“ и предиктор варијаблата „Q_inflow“ кои се применети во процесот на обучување на моделите.



Прилог 97. График со приказ на вредностите на предиктор варијаблите „Zwl“ и „Q_inflow“ применети во процесот на предвидување на целната варијабла „V_notch_middle“.

7.5.2. Прелиминарна петкратна вкрстена валидација, пред обучување на моделите

За проценка на моделите и нивната предикциона способност, за анализираниот систем $V_notch_middle = f(Zwl, Q_inflow)$ изработена е прелиминарна петкратна вкрстена валидација со

деветте алгоритми за машинско учење. За оцена на креираните модели, прикажани се резултати од метричките параметри во Табела 42.

Табела 42. Табеларен приказ на метричките показатели од изработената петкратна вкрстена валидација на студија на случај V_notch_middle=f(Zwl,Qinf).

5-fold CV					
No.	Algorithm	Avg MSE	Avg RMSE	Avg MAE	Avg R2
1	ANN	731.26	26.48	21.14	-1.08
2	XGBoost	524.12	22.62	16.29	-0.44
3	Decision Tree	483.76	21.25	14.72	-0.31
4	KNN	466.93	21.49	14.03	-0.25
5	Boosted Trees	435.16	20.37	14.28	-0.16
6	GRNN	415.86	19.95	13.08	-0.05
7	SVR	400.58	19.78	12.79	-0.04
8	GPR	389.07	19.53	12.42	-0.03
9	Random Forest	376.18	19.13	13.63	-0.02
	max value:	731.26	26.48	21.14	-0.02
	min value:	376.18	19.13	12.42	-1.08

Резултатите добиени од петкратната вкрстена валидација за предвидување на филтрациониот протек на локацијата „V_notch_middle“ посочуваат на ограничена предиктивна способност на применетите модели на машинско учење. Кај сите разгледувани алгоритми се добиени негативни вредности на коефициентот на детерминација R^2 , кои се движат од -1.08 кај моделот на вештачки невронски мрежи до -0.02 кај моделот Random Forest. Негативните вредности на R^2 значат дека моделите не успеваат да ја надминат едноставната апроксимација базирана на средната вредност на целната променлива, кој е знак на слаба зависност помеѓу избраните влезни променливи и филтрацискиот проток. Сепак, и покрај ваквата ограничена апсолутна точност, споредбената анализа врз основа на грешките (RMSE и MAE) овозможува идентификација на релативно подобрите модели.

Најдобри резултати се постигнати со моделот Random Forest, кој има најниска вредност на RMSE од 19.13 и R^2 вредност најблиска до нула (-0.02). Ова укажува дека моделот, иако не обезбедува висока точност, сепак успева делумно да го фати општиот тренд на промената на филтрацискиот проток. Слични перформанси покажуваат и Gaussian Process Regression и Support Vector Regression, со RMSE вредности од 19.53 и 19.78, и соодветни R^2 вредности од -0.03 и -0.04. Малите разлики помеѓу овие модели укажуваат на конзистентно ниво на предвидливост, независно од типот на алгоритмот, што дополнително потврдува дека ограничувањето не произлегува од изборот на модел, туку од карактеристиките на податоците.

Од друга страна, моделите со поголема комплексност, како што се вештачките невронски мрежи и XGBoost, покажуваат значително полоши резултати, со високи вредности на грешките и поизразено негативни вредности на R^2 . Ова може да се објасни со малиот број на податоци (66 мерења), кој не е доволен за ефективно тренирање на сложени модели, што резултира со претерување во предвидувањата или нестабилност на моделот. Во вакви услови, поедноставните и поробусни модели, како Random Forest и SVR, се покажуваат како попогодни.

Од инженерски аспект, добиените резултати се очекувани и во согласност со физичката природа на процесот. Филтрацијата во низводниот дел од браната е сложен процес кој зависи од повеќе фактори, вклучувајќи ја распределбата на пермеабилноста, состојбата на дренажниот систем, хидрауличкиот градиент и распределбата на порниот притисок во подлогата под браната. Во конкретната анализа се користени само две влезни променливи – нивото на водата во акумулацијата „Zwl“ и дотекот во акумулацијата „Qinflow“, кои иако се релевантни, не ја опфаќаат целосно физичката состојба на системот. Како резултат на тоа, моделите можат да го апроксимираат општиот тренд, но не и да ги репродуцираат локалните варијации и динамиката на филтрацијата.

Сепак, фактот дека најдобрите модели имаат R^2 вредности блиску до нула укажува дека постои

одредено ниво на зависност помеѓу влезните и излезната променлива, односно дека достапните параметри содржат ограничена, но релевантна информација. Во насока на подобрување на стабилноста на предвидувањата, оправдано е користење на ансамбл пристап, при што се комбинираат резултатите од најдобрите модели – Random Forest, Gaussian Process Regression и Support Vector Regression. Овој пристап овозможува намалување на варијансата и обезбедува поробусна проценка на филтрацискиот проток во услови на ограничени податоци.

Во заклучок, може да се констатира дека примената на методите на машинско учење во вакви услови има ограничена предиктивна моќ, пред сè поради недоволниот број на податоци и ограничениот број на влезни параметри. За значително подобрување на резултатите, неопходно е вклучување на дополнителни физички релевантни променливи, како што се пиезометриските нивоа. Сепак, и покрај овие ограничувања, анализата дава вреден увид во применливоста и ограничувањата на машинското учење во моделирањето на филтрациски процеси кај насипни брани.

7.5.3. Обучување на моделите

По извршената прелиминарна петкратна вкрстена валидација, обучени се три модели – RF, GPR и SVR - за понатамошна предикција на целната варијабла „V_notch_middle“ врз основа на познати вредности на предиктор варијабли „Zwl“ и „Q_inflow“, и самата целна варијабла. Станува збор за краток сет на податоци од само 66 вредности регистрирани во рамки на редовната оскултација на брана „Кнежево“. Како што следи дадени се кратки информации за обучените модели.

7.5.3.1. Модел на Случајна шума (RF)

Првиот обучен модел е моделот на Random Forest регресија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја..

За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators` (50, 100, 150) - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth` (None, 10, 20) - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split` (2, 5, 10) - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features` ("sqrt", "log2", None) - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијабли.

Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R^2 прикажани во Табела 43, што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување.

Табела 43. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај V_notch_middle=f(Zwl, Q_inflow).

Fold	max_depth	max_features	min_samples_split	n_estimators
1	NaN	None	10	150
2	10	None	2	150
3	NaN	None	10	100
4	NaN	None	2	50
5	10	None	5	150

Финалното обучување е направено согласно хиперпараметрите од петтиот сет од петкратната надворешна валидација, со вредности: 'max_depth': None, 'max_features': none, 'min_samples_split':

10, 'n_estimators': 150. Процесот на обучување траеше 37s и 214ms.

7.5.3.2. Модел со Гаусов Процесен Регресор (GPR)

Вториот модел за обучување се базира на алгоритмот Гаусов Процесен Регресор (анг. „Gaussian Process Regression“).

Клучни параметри на моделот се:

1) `normalize_y = True` - Ја нормализира целната варијабла, што значително ја подобрува стабилноста на моделот.

2) `n_restarts_optimizer = 5` - Овозможува повеќекратно оптимизирање на параметрите на кернел функцијата, со цел да се избегне локален минимум и да се добие подобро вклопување.

Поради тоа што GPR е чувствителен на растојанија, применета е StandardScaler трансформација во рамки на методологијата, со што сите податоци од сетот за обучување се стандардизираат во секој сет од петкратната вкрстена валидација. Перформансите се оценуваат преку MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста. Иако GPR има теоретски висока флексибилност и способност за моделирање на сложени соодноси помеѓу различните варијабли, во практични услови може да биде чувствителен на шум и ограничен при работа со поголеми и покомплексни податочни множества.

Вредностите за метричките показатели од процесот на петкратна вкрстена валидација се прикажани во Табела 44.

Табела 44. Вредности на метричките показатели од процесот на обучување на GPR моделот по спроведена петкратна валидација за студија на случај $V_notch_middle=f(Z_{wl}, Q_inflow)$.

Fold	MSE	RMSE	MAE	R2
1	391.645	19.790	12.701	-0.149
2	615.286	24.805	16.068	-0.091
3	311.052	17.637	10.295	-0.126
4	297.601	17.251	11.740	-0.055
5	329.749	18.159	11.302	0.293

Просечните вредности на метричките показатели изнесуваат: Avg MSE: 389.066, Avg RMSE: 19.528, Avg MAE: 12.420, Avg R^2 : -0.025. Процесот на обучување траеше 258ms.

7.5.3.3. Обучување на моделите со SVR

Последниот обучуван модел во анализираната студија на случај е моделот на Support Vector Regression (SVR), кој претставува метод базиран на поддржувачки вектори за решавање на регресиони проблеми.

Моделот е дефиниран со следните параметри:

1) `kernel = "rbf"` - Се користи радијална базна функција (Radial Basis Function), која овозможува моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) `C = 100` - Регулаторен параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на C значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување.

3) `epsilon = 0.1` - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

За евалуација на моделот за V_notch_middle е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриките MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Средната вредност на метричките показатели по обучувањето на моделот за V_notch_middle изнесуваат: Avg MSE = 400.578, Avg RMSE = 19.780, Avg MAE = 12.785, Avg R^2 = -0.039. Процесот на

обучување траеше 43ms.

7.5.4. Предвидување на целната варијабла со тежинско осреднување

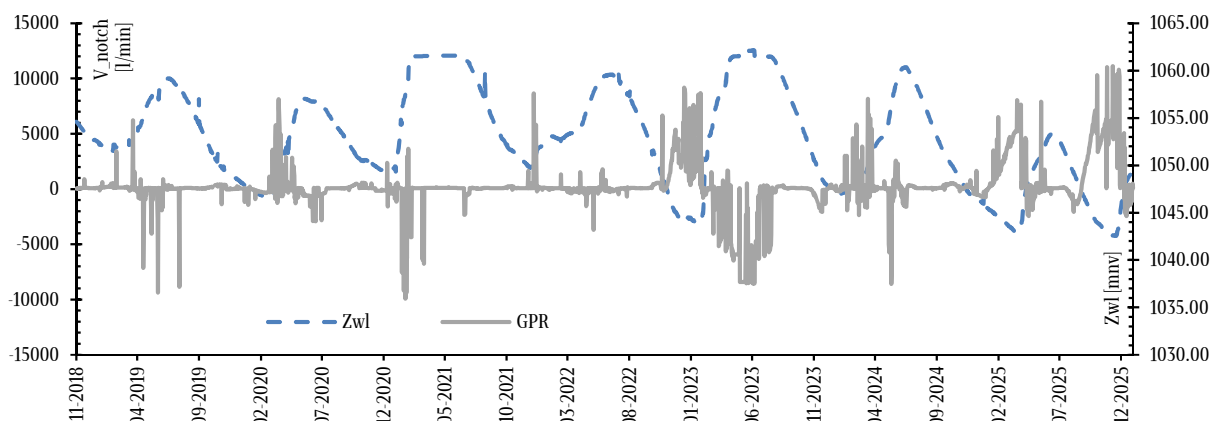
Целна варијабла во предметната студија на случај е количината на филтрациони води „V_notch_middle“ измерени во прелив од централна дренажа на низводното потпорно тело на браната „Кнежево“. Со помош на обучените модели од машинско учење, се бара предвидување на целната варијабла со познавање на следните предиктор варијабли:

- 1) Ниво на вода во акумулација, „Zwl“ во [mnv] прикажано во Прилог 97, и
- 2) Дотек во акумулацијата, „Q_inflow“ во [m³/s] прикажано во Прилог 97.

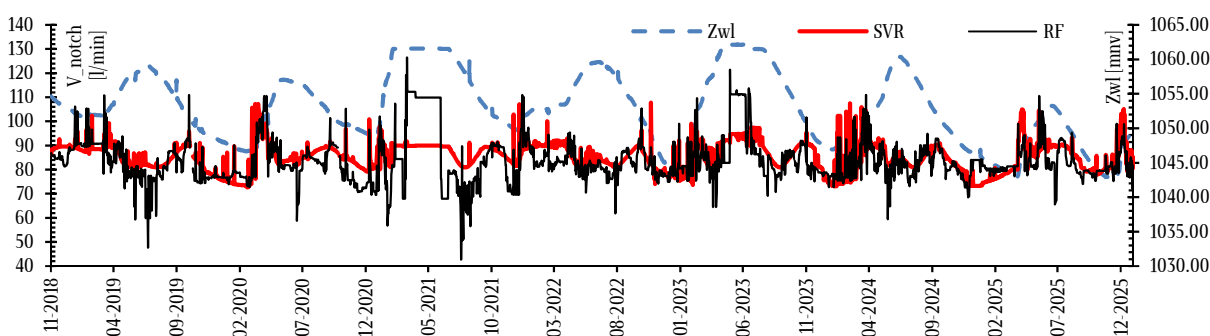
Обучени се вкупно три модели со машинско учење со три различни алгоритми кои се проценети како најдобри од претходната фаза на прелиминарна петкратна вкрстена валидација. Станува збор за моделите RF, GPR и SVR.

Процесот на предвидување се извршува за 2606 познати вредности на предиктор варијаблите, кои временски се однесуваат на периодот на прибирање на податоците од 1^{ви} ноември 2018 година до 31 декември 2025 година. За анализираниот период нема континуирани вредности на целната варијабла бидејќи истата се регистрира од страна на операторот во нееднакви временски интервали. Предиктор варијаблите се внесуваат во DamResML кодот преку датотеката ‘predictor_variable_for_prediction.xls’ и истата се повикува во ќелијата ‘Prediction’ по завршување на процесот на обучување на моделот.

Од резултатите прикажани во Прилог 98 може да се воочи дека иако во процесот на петкратна вкрстена валидација и во процесот на обучување моделот со GPR даваше релативно добри метрички показатели споредено со останатите модели, во фазата на предвидување овој модел предвидува дури и негативни вредности на филтрацијата „V_notch_middle“, и тоа во неколку илјади [l/min] што не е возможно поради тоа што мерените вредности на анализираниот прелив не надминуваат 156 [l/min].



Прилог 98. Графички приказ на пресметаните предвидувања на целната варијабла „V_notch_middle“ со моделот трениран на GPR алгоритмот.



Прилог 99. Графички приказ на пресметаните предвидувања на целната варијабла „V_notch_middle“ со моделите тренирани на RF и SVR алгоритмите.

Бидејќи резултатите од GPR моделот се незадоволителни, како релевантни за дискусија ќе се земат предвид единствено резултатите од обучените модели на RF и SVR. Вкупното предвидување, слично како во претходната студија на случај 2.4 – $V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)$ ќе се пресмета со тежинско осреднување на предвидувањата добиени како предвидувања од изработените модели.

Враќајќи се еден чекор назад, со анализа на резултатите добиени преку петкратната вкрстена валидација, согледуваме дека ниту еден од применетите модели не постигнува задоволителна предиктивна способност, што е евидентно од негативните вредности на коефициентот на детерминација R^2 кај сите алгоритми. Ова укажува дека моделите не успеваат адекватно да ја опишат зависноста помеѓу влезните променливи и целната величина, при што нивните предвидувања се полоши во однос на едноставен модел базиран на средна вредност. Поради тоа, наместо избор на еден модел, се препорачува примена на ансамбл пристап базиран на тежинско осреднување, при што во финалната прогноза се комбинираат предвидувањата од најдобрите модели (SVR и Random Forest), со тежини дефинирани обратно пропорционално на RMSE.

Пресметаните тежини на RMSE од петкратната вкрстена валидација на избраните два модели со најдобри вредности на RMSE, се дадени како што следи.

Најпрво, општата равенка за пресметка на тежинските коефициенти за осреднување, гласи:

$$w_i = \frac{\left(\frac{1}{RMSE_i}\right)}{\left(\sum_{j=1}^n \frac{1}{RMSE_j}\right)} \quad (7.5.1)$$

Инверзните RMSE вредности на RMSE за секој модел се пресметуваат како:

$$\frac{1}{RMSE_{RF}} = \frac{1}{19.73} = 0.05227; \frac{1}{RMSE_{SVR}} = \frac{1}{19.78} = 0.05056 \quad (7.5.2)$$

Сумата на инверзните вредности на RMSE изнесува:

$$SUM_{RMSE} = 0.10283 \quad (7.5.3)$$

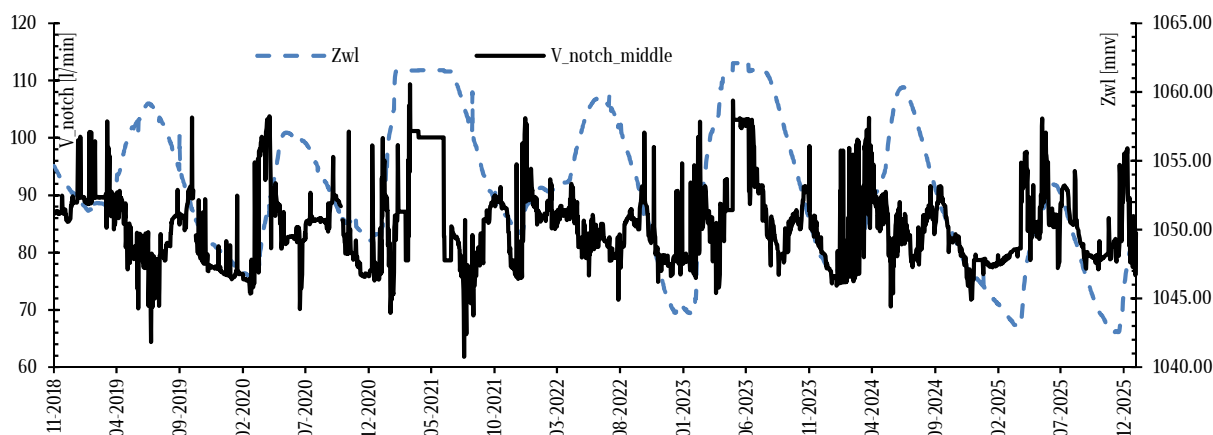
Оттука, тежините се пресметуваат како:

$$w_{RF} = \frac{0.05227}{0.10283} = 0.508; w_{SVR} = \frac{0.05056}{0.10283} = 0.492 \quad (7.5.4)$$

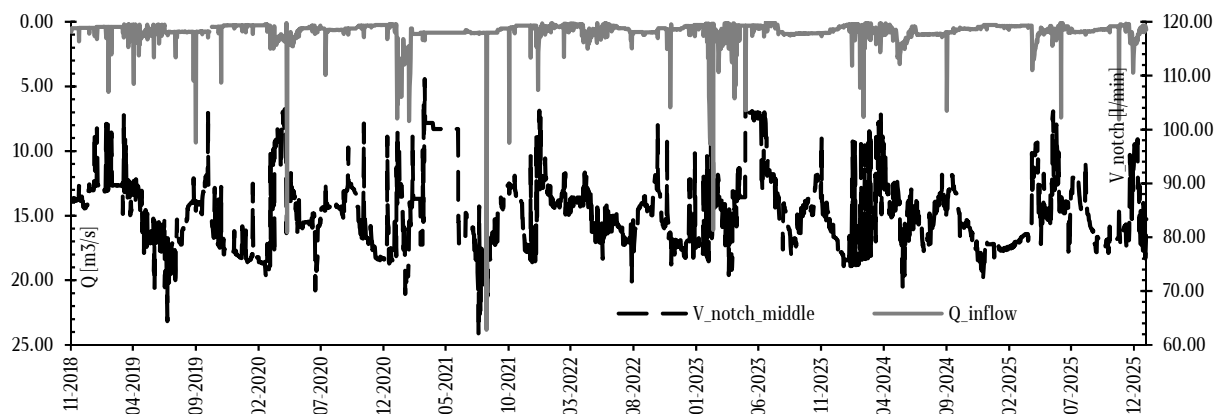
Имајќи ги предвид тежинските коефициенти, крајните вредности на целната варијабла кои ќе се земаат предвид како релевантни, ќе се пресметаат како:

$$y_{pred} = w_{SVR} \cdot y_{pred_{SVR}} + w_{RF} \cdot y_{pred_{RF}} \quad (7.5.5)$$

Резултатите од осреднетите предвидувања се прикажани во графициите на Прилог 100 и Прилог 101.



Прилог 100. Графички приказ на осреднетите вредности на целната варијабла „V_notch_middle“ и предиктор варијаблата „Zwl“.



Прилог 101. Графички приказ на осреднетите вредности на целната варијабла „V_notch_middle“ и предиктор варијаблата „Q_inflow“.

7.5.5. Дефинирање на граници за предупредување (thresholds)

Слично на анализата за дефинирање на граници за предупредување кај целните варијабли „V_notch_left“ и „V_notch_right“ на предвидените вредности на филтрационите води од лев и десен бок, изработена е пресметка на границите за предупредување за тежински осреднетата предвидена целна варијабла која е предмет на анализа на студија на случај 2.5 – „V_notch_middle“.

Границите на предупредување се дефинираат во однос на пресметана σ - стандардна девијација од осреднетите предвидени вредности податоци на целната варијабла, вкупно 2606 на број.

По дефинирање на границите на предупредување како оддалечени вредности од предвидените, во пресметаните коридори на границите се вметнуваат мерените податоци од филтрационото количество низ средниот дренажен колектор и се прави споредба на мерењата согласно математички дефинираните граници. Целта е да се воочат отстапувања на мерењата надвор од границите на предупредување, што би резултирало со алармирање на операторот за невообичаени истекувања низ дренажниот колектор.

Постапката за пресметка на границите за предупредување е слична на пресметката изработена во Поглавје 7.2.5.

Пресметаната стандардна девијација од сетот на предвидени вредности на „V_notch_middle“ изнесува $\sigma = 6.87$. Средната вредност на сетот на податоци изнесува $x_{ave} = 84.538$ [l/min].

Првата, пониска граница на предупредување се дефинира како:

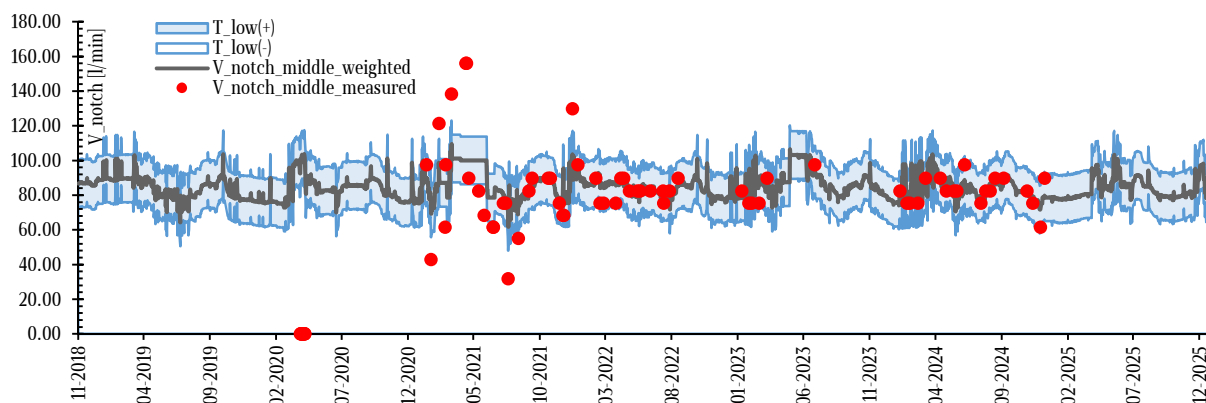
$$T_{low}^{+} = x_i + 2 \cdot \sigma; T_{low}^{-} = x_i - 2 \cdot \sigma; \quad (7.5.6)$$

Втората, повисока граница на предупредување се дефинира како:

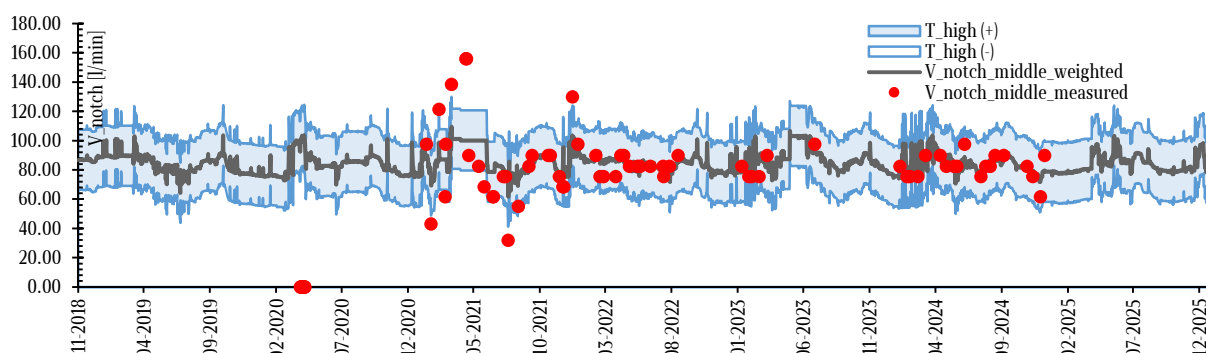
$$T_{high}^{+} = x_i + 3 \cdot \sigma; T_{high}^{-} = x_i - 3 \cdot \sigma; \quad (7.5.7)$$

Двете граници на предупредување варираат во однос на осцилацијата на предвидените вредности на целната варијабла x_i . Границите претставуваат оддалечени вредности од предвидените вредности на целната варијабла, согласно математичката релација во равенките 7.5.6 и 7.5.7. Пресметаните коридори на дозволените граници на филтрационото количество низ средниот дренажен колектор се прикажани во Прилог 102 и Прилог 103.

Во Прилог 102 е даден приказ на коридорот дефиниран со пониските граници на предупредување T_{low} , додека во Прилог 103 е даден приказ на коридорот дефиниран со повисоките граници на предупредување T_{high} . Од двата прилози може да се заклучи дека мерените податоци за филтрацијата низ средниот дренажен колектор се во граници на очекуваните вредности. Најголем дел од мерените вредности; 55 од вкупно 66; се во границите на предупредување, т.е. 16.67% од мерените податоци ја надминуваат пониската и повисоката граница на предупредување.



Прилог 102. График со приказ на коридорот со пониските граници на предупредување „T_{low}“ на предвидените податоци за целната варијабла „V_notch_middle“, со приказ и на измерените вредности на целната варијабла во разгледуваниот период.



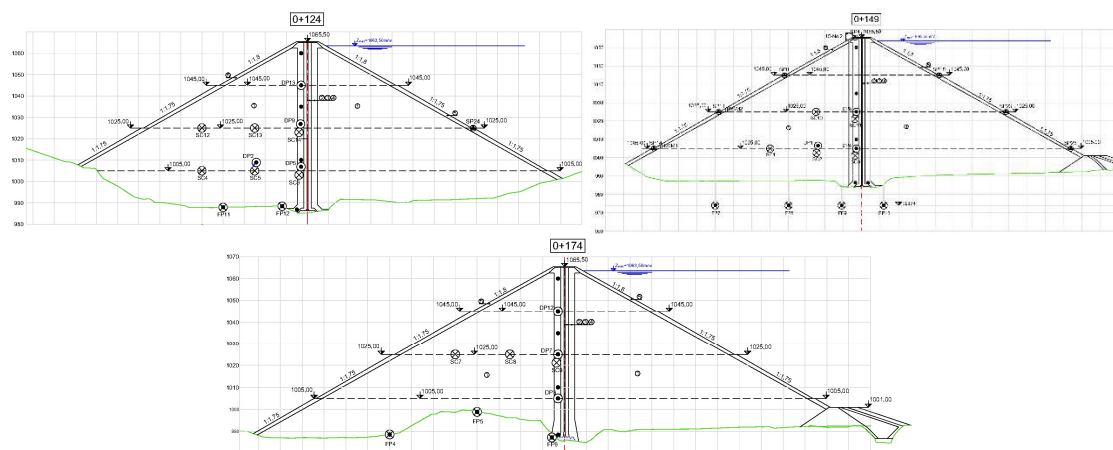
Прилог 103. График со приказ на коридорот со повисоките граници на предупредување „T_{low}“ на предвидените податоци за целната варијабла „V_notch_middle“, со приказ и на измерените вредности на целната варијабла во разгледуваниот период.

7.6. Анализа на случај 2.6 – $SC_i=f(Zwl,t)$

7.6.1. Подготовка на податоците

Во следната анализа на случај 2.6 од серијата студии на случај 2 ќе се анализира зависноста помеѓу вертикалните слегнувања „SC“ како целна варијабла, и предиктор варијаблите „Zwl“ – ниво на вода во акумулацијата и „t“ – временска компонента на слегнувањата во денови. Станува збор за вертикални слегнувања кои се мерени во ќелии за мерење на слегнувањата (анг. „Settlements cells“) во низводното потпорно тело на брана „Кнежево“. Мерењата се извршени од страна на операторот во неправилни временски интервали, во периодот од 2018 година до 2024 година, со вкупно 134 податоци регистрирани во напоменатиот период, за секоја ќелија поединечно. Ќелиите за мерење на вертикалните слегнувања се поставени во телото на браната на повеќе локации и различни висински нивои.

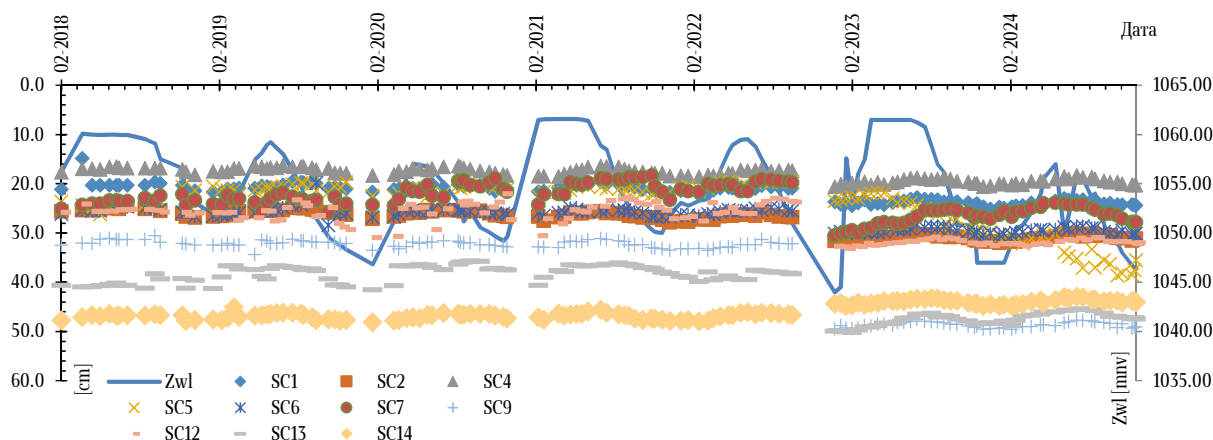
Во низводното потпорно тело на брана „Кнежево“ вградени се поголем број на ќелии, од кои во функција се 10: „SC1“, „SC2“, „SC4“, „SC5“, „SC6“, „SC7“, „SC9“, „SC12“, „SC13“ и „SC14“. Ќелиите „SC12“, „SC13“ и „SC14“ се поставени во низводното потпорно тело на стационожа km 0+124 од круната на браната на висинско ниво од 1025mnv. Ќелиите „SC4“, „SC5“ и „SC6“ се поставени во низводното потпорно тело на стационожа km 0+124 од круната на браната на висинско ниво од 1005mnv. Ќелиите „SC1“ и „SC2“ се поставени во низводното потпорно тело на стационожа km 0+149 од круната на браната на висинско ниво од 1005mnv. Ќелиите „SC7“, „SC8“ и „SC9“ се поставени во низводното потпорно тело на стационожа km 0+174 од круната на браната на висинско ниво од 1025mnv. Попречните пресеци со локација на ќелиите се прикажани во Слика 16.



Слика 16. Приказ на попречните пресеци на стационата km 0+124, 0+149 и 0+174 по оска на круна на браната „Кнежево“ со позиција на функционалните ќелии за мерење на слегнувањата.

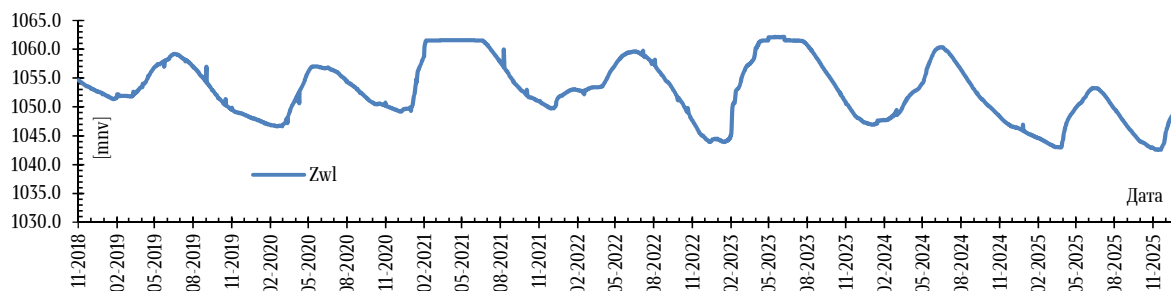
Оваа анализа ќе се спроведува во две фази. Првата фаза ќе опфати дополнување на низата на податоци за слегнувањата во секоја ќелија со примена на обучени модели со машинско учење. Имено, 134-те измерени податоци во периодот од 2018 до 2024 година ќе се користат за обучување на модели со машинско учење за секоја ќелија засебно, а потоа, врз основа на познати податоци за предиктор варијаблите „Zwl“ и „t“ за секој ден од 01.11.2018 до 31.12.2025 година, со предвидување на целната варијабла ќе се пополни во целост низата на слегнувања, повторно – за секоја ќелија. Слегнувањата во секоја ќелија ќе бидат предмет на подучување на различни модели со машинско учење, но во фазата на предвидување ќе се одбираат само моделите со најдобри метрички показатели. Втората фаза ќе опфати екстраполација на слегнувањата со цел определување на датум на достигнување на 90%, 95% и 100% од теоретски пресметаните слегнувања за секоја ќелија.

Во фазата на обучување на секој модел користени се податоци од мерени слегнувања во секоја ќелија. Станува збор за вкупно 134 податоци, како што беше и претходно напоменато. Во овој период на мерења, покрај целната варијабла „SC“ за секоја ќелија, вршени се мерења и за водното ниво во акумулацијата „Zwl“ кое се јавува како предиктор варијабла во анализите. Бидејќи слегнувањата кај насипните брани имаат реолошка карактеристика – со текот на времето истите се доближуваат кон константна вредност, со изразен висок градиент во почетниот период на експлоатација на хидросистемот и намалување на градиентот на слегнувања во експлоатациониот период – покрај водното ниво кое влијае на слегнувањата „Zwl“ во анализите е земено предвид и времето како карактеристика, во дневен интервал „t“. Податоците кои се земени предвид во периодот на обучување на сите десет модели за слегнувања во ќелиите се прикажани графички во Прилог 104.



Прилог 104. Графички приказ на мерените слегнувања „SC“ во сите 10 ќелии кои се предмет на анализа за периодот од 2018 до 2024 година, заедно со мереното ниво во акумулацијата „Zwl“ во истите временски периоди на регистрирање на слегнувањата.

Во периодот на предвидувања, за целосно пополнување на низата на слегнувањата за периодот од 01.11.2018 година па сè до 31.12.2025 година се користат познати податоци за дневните нивоа во акумулацијата „Кнежево“ – „Zwl“ прикажани во Прилог 105. Станува збор за вкупно 2606 податоци за нивото во акумулацијата, мерени секојдневно во напоменатиот временски период.



Прилог 105. Графички приказ на водното ниво во акумулацијата „Zwl“ за секој ден почнувајќи од 01.11.2018 година па се до 31.12.2025 година - податоци кои се користат во сите модели за слегнувања во функција на предиктор варијабли.

Како што следи, даден е приказ од избраните најдобри модели со машинско учење и нивната понатамошна примена во проценка на датумот на достигнување на максималните пресметани теоретски големини на слегнувањата во секоја ќелија. Нивната нумеричка вредност е прикажана во Табела 45.

Табела 45. Максимални теоретски пресметани слегнувања за секоја ќелија која е предмет на анализа, во [cm].

SC	SC1_max	SC2_max	SC4_max	SC5_max	SC6_max	SC7_max	SC12_max	SC9_max	SC13_max	SC14_max
[cm]	36	41.7	35.1	42.6	48	37.4	43.3	60.60	59.7	66.7

7.6.2. Прелиминарна петкратна вкрстена валидација

За проценка на точноста кој секој алгоритам може да ја обезбеди во негова понатамошна примена во фаза на предвидувања на слегнувањата, најпрво е направена петкратна вкрстена валидација на сите девет алгоритми со машинско учење. Процесот започнува со повикување на датотека во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Регресија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен регресор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“). Резултатите за секоја ќелија се прикажани како што следи.

Табела 46. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC1=f(Zwl,t)$.

	Model	MSE	RMSE	MAE	R2
1	ANN	4.860	1.610	1.271	-0.444
2	GPR	2.078	1.424	1.154	0.262
3	XGBoost	0.751	0.737	0.397	0.731
4	Decision Tree	0.707	0.744	0.410	0.734
5	KNN	0.691	0.798	0.563	0.767
6	Boosted Trees	0.596	0.668	0.356	0.782
7	SVR	0.603	0.733	0.502	0.804
8	GRNN	0.584	0.714	0.477	0.804
9	Random Forest	0.460	0.615	0.405	0.856
	min:	0.460	0.615	0.356	-0.444
	max:	4.860	1.610	1.271	0.856

Табела 47. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC2=f(Zw1,t)$.

	Model	MSE	RMSE	MAE	R2
1	ANN	7.619	1.859	1.551	-0.137
2	GPR	3.589	1.888	1.563	0.358
3	SVR	0.564	0.749	0.498	0.899
4	KNN	0.562	0.736	0.512	0.902
5	GRNN	0.505	0.678	0.485	0.915
6	XGBoost	0.284	0.461	0.278	0.952
7	Decision Tree	0.155	0.386	0.290	0.971
8	Random Forest	0.119	0.343	0.270	0.978
9	Boosted Trees	0.099	0.313	0.244	0.982
min:		0.099	0.313	0.244	-0.137
max:		7.619	1.888	1.563	0.982

Табела 48. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC4=f(Zw1,t)$.

	Model	MSE	RMSE	MAE	R2
1	ANN	3.294	1.252	1.025	-0.848
2	GPR	1.028	1.012	0.820	0.291
3	KNN	0.354	0.590	0.450	0.761
4	SVR	0.341	0.581	0.431	0.766
5	GRNN	0.328	0.564	0.437	0.788
6	XGBoost	0.220	0.443	0.310	0.853
7	Decision Tree	0.184	0.428	0.333	0.871
8	Random Forest	0.136	0.363	0.281	0.908
9	Boosted Trees	0.133	0.362	0.282	0.909
min:		0.133	0.362	0.281	-0.848
max:		3.294	1.252	1.025	0.909

Табела 49. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC5=f(Zw1,t)$.

	Model	MSE	RMSE	MAE	R2
1	GPR	14.832	3.790	2.734	0.379
2	ANN	5.356	1.857	1.405	0.815
3	KNN	2.997	1.677	1.206	0.879
4	GRNN	2.063	1.419	1.007	0.919
5	SVR	1.459	1.184	0.849	0.940
6	Decision Tree	1.320	1.137	0.769	0.945
7	XGBoost	1.071	1.015	0.759	0.952
8	Boosted Trees	0.976	0.979	0.716	0.959
9	Random Forest	0.876	0.932	0.663	0.963
min:		0.876	0.932	0.663	0.379
max:		14.832	3.790	2.734	0.963

Табела 50. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC6=f(Zw1,t)$.

	Model	MSE	RMSE	MAE	R2
1	ANN	7.929	2.040	1.591	-0.757
2	GPR	3.007	1.715	1.416	0.283
3	XGBoost	1.002	0.860	0.438	0.742
4	KNN	0.954	0.961	0.589	0.775
5	SVR	0.923	0.921	0.547	0.790

	Model	MSE	RMSE	MAE	R2
6	GRNN	0.733	0.827	0.533	0.835
7	Decision Tree	0.650	0.715	0.407	0.863
8	Random Forest	0.536	0.660	0.390	0.885
9	Boosted Trees	0.317	0.516	0.305	0.929
	min:	0.317	0.516	0.305	-0.757
	max:	7.929	2.040	1.591	0.929

Табела 51. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC7=f(Zwl,t)$.

	Model	MSE	RMSE	MAE	R2
1	GPR	6.365	2.506	1.877	0.242
2	ANN	7.352	2.405	1.897	0.277
3	GRNN	3.002	1.675	1.117	0.638
4	KNN	3.044	1.728	1.231	0.644
5	SVR	2.734	1.633	1.104	0.673
6	XGBoost	1.740	1.184	0.739	0.766
7	Decision Tree	0.996	0.988	0.698	0.870
8	Random Forest	0.758	0.853	0.612	0.900
9	Boosted Trees	0.756	0.854	0.626	0.900
	min:	0.756	0.853	0.612	0.242
	max:	7.352	2.506	1.897	0.900

Табела 52. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC9=f(Zwl,t)$.

	Model	MSE	RMSE	MAE	R2
1	GPR	33.469	5.743	4.627	0.429
2	ANN	7.308	2.689	1.974	0.874
3	KNN	4.372	1.945	0.912	0.929
4	SVR	4.019	1.991	1.140	0.932
5	GRNN	3.712	1.723	0.858	0.942
6	XGBoost	2.498	1.087	0.483	0.962
7	Decision Tree	0.382	0.609	0.390	0.993
8	Boosted Trees	0.284	0.523	0.358	0.995
9	Random Forest	0.245	0.491	0.352	0.996
	min:	0.245	0.491	0.352	0.429
	max:	33.469	5.743	4.627	0.996

Табела 53. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај $SC12=f(Zwl,t)$.

	Model	MSE	RMSE	MAE	R2
1	GPR	8.195	2.858	2.397	0.241
2	ANN	8.667	2.411	1.943	0.304
3	GRNN	3.008	1.729	1.145	0.717
4	KNN	3.055	1.728	1.175	0.719
5	SVR	2.900	1.697	1.085	0.732
6	Decision Tree	2.995	1.688	1.008	0.733
7	XGBoost	2.063	1.421	0.876	0.810
8	Boosted Trees	1.986	1.394	0.870	0.818
9	Random Forest	1.613	1.254	0.788	0.851
	min:	1.613	1.254	0.788	0.241
	max:	8.667	2.858	2.397	0.851

Табела 54. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај SC13= $f(Zwl, t)$.

	Model	MSE	RMSE	MAE	R2
1	GPR	13.441	3.655	2.980	0.333
2	ANN	17.335	3.197	2.612	0.341
3	KNN	3.300	1.757	1.163	0.849
4	GRNN	3.173	1.662	1.119	0.853
5	SVR	2.853	1.673	1.088	0.862
6	XGBoost	1.771	1.198	0.677	0.910
7	Decision Tree	1.226	1.085	0.666	0.941
8	Boosted Trees	0.813	0.898	0.599	0.959
9	Random Forest	0.732	0.838	0.543	0.963
min:		0.732	0.838	0.543	0.333
max:		17.335	3.655	2.980	0.963

Табела 55. Резултати од прелиминарна петкратна вкрстена валидација на студија на случај SC14= $f(Zwl, t)$.

	Model	MSE	RMSE	MAE	R2
1	ANN	45.892	4.638	3.816	-24.367
2	GPR	1.725	1.297	1.082	0.213
3	SVR	0.302	0.546	0.393	0.846
4	KNN	0.298	0.537	0.397	0.846
5	Decision Tree	0.292	0.512	0.351	0.847
6	GRNN	0.271	0.514	0.390	0.878
7	XGBoost	0.237	0.473	0.323	0.884
8	Random Forest	0.197	0.429	0.304	0.898
9	Boosted Trees	0.178	0.407	0.290	0.907
min:		0.178	0.407	0.290	-24.367
max:		45.892	4.638	3.816	0.907

Врз основа на резултатите од петкратната вкрстена валидација за сите анализирани ќелии на слегнување кај браната Кнежево, може да се заклучи дека алгоритмите базирани на ансамбли од одлучувачки дрва, особено алгоритмите Random Forest и Boosted Trees, покажуваат највисока точност и стабилност во процесот на моделирање и предвидување на слегнувањата. Кај најголем дел од анализираниите студии на случај, овие алгоритми обезбедуваат најниски вредности на грешките MSE, RMSE и MAE, како и највисоки вредности на коефициентот на детерминација R^2 , што укажува на нивна висока способност за репрезентација на нелинеарните зависимости помеѓу временскиот фактор, нивото на водата во акумулацијата и слегнувањата на браната.

За студијата на случај SC1= $f(Zwl, t)$, најдобри резултати покажува алгоритмот Random Forest, кој постигнува највисока вредност на $R^2=0.856$, како и најниски вредности на MSE=0.460 и RMSE=0.615. Иако алгоритмите SVR и GRNN покажуваат идентична вредност на $R^2=0.804$, моделот Random Forest обезбедува значително помали грешки и подобра генерализација, поради што е избран како најсоодветен модел за понатамошни предвидувања.

Кај SC2= $f(Zwl, t)$, највисока точност е постигната со алгоритмот Boosted Trees, кој обезбедува највисока вредност на коефициентот $R^2=0.982$ и најниски вредности на грешките MSE=0.099, RMSE=0.313 и MAE=0.244. Веднаш зад него се алгоритмите Random Forest и Decision Tree, но разликите укажуваат дека Boosted Trees најуспешно ја моделира динамиката на слегнувањето за оваа ќелија.

За SC4= $f(Zwl, t)$, алгоритмите Boosted Trees и Random Forest повторно покажуваат најдобри резултати, при што Boosted Trees има највисока вредност на $R^2=0.909$ и минимални вредности на грешките. Иако разликата помеѓу двата модели е мала, резултатите укажуваат на благо подобра способност на Boosted Trees за репродукција на мерените вредности на слегнување.

Кај SC5= $f(Zwl, t)$, најдобри резултати покажува Random Forest со $R^2=0.963$ и најниски вредности

на $MSE=0.876$, $RMSE=0.932$ и $MAE=0.663$. Алгоритмот Boosted Trees исто така покажува многу висока точност, но Random Forest обезбедува најдобра севкупна статистичка изведба.

За $SC6=f(Zwl,t)$, алгоритмот Boosted Trees јасно се издвојува како најуспешен модел со $R^2=0.929$ и најниски вредности на сите анализирани грешки. Алгоритмите Random Forest и Decision Tree исто така покажуваат добри резултати, но со нешто пониска точност во споредба со Boosted Trees.

Кај $SC7=f(Zwl,t)$, алгоритмите Random Forest и Boosted Trees покажуваат практично идентични резултати со $R^2=0.900$. Сепак, моделот Random Forest има минимално пониска вредност на MSE и MAE , поради што може да се смета за најсоодветен модел за оваа студија на случај.

За $SC9=f(Zwl,t)$, моделот Random Forest покажува исклучително висока точност со $R^2=0.996$ и најниски вредности на $MSE=0.245$, $RMSE=0.491$ и $MAE=0.352$. Резултатите укажуваат дека слегнувањата кај оваа ќелија можат да се предвидуваат со многу висока сигурност со примена на ансамбл модели.

Кај $SC12=f(Zwl,t)$, повторно најдобар модел е Random Forest, кој постигнува највисока вредност на $R^2=0.851$ и најниски вредности на сите анализирани показатели за грешка. Иако Boosted Trees и XGBoost покажуваат добри резултати, Random Forest обезбедува најдобар баланс помеѓу стабилноста и точноста на предвидувањето.

За $SC13=f(Zwl,t)$, алгоритмот Random Forest повторно покажува најдобра статистичка изведба со $R^2=0.963$ и најниски вредности на грешките. Алгоритмот Boosted Trees покажува многу блиски резултати, но Random Forest обезбедува нешто подобра генерализација на податоците.

Кај последната анализирана студија на случај $SC14=f(Zwl,t)$, најдобри резултати се добиени со алгоритмот Boosted Trees, кој обезбедува највисока вредност на $R^2=0.907$ и најниски вредности на MSE , $RMSE$ и MAE . Алгоритмот Random Forest покажува многу блиски резултати, но Boosted Trees демонстрира подобра способност за моделирање на зависностите кај оваа ќелија.

Генерално, резултатите од петкратната вкрстена валидација покажуваат дека ансамбл алгоритмите базирани на одлучувачки дрва (Random Forest и Boosted Trees) претставуваат најсоодветни модели за анализа и предвидување на слегнувањата кај браната Кнежево. Овие алгоритми обезбедуваат висока точност, стабилност и способност за моделирање на сложени нелинеарни процеси, што ги прави особено погодни за примена во системи за следење на долгорочното однесување и безбедност на браните. Од друга страна, алгоритмите како ANN и GPR во повеќето случаи покажуваат значително послаби резултати, со пониски вредности на R^2 и повисоки вредности на грешките, што укажува на нивна ограничена способност за соодветно моделирање на расположливите податоци во конкретната студија на случај.

7.6.3. Обучување на моделите

Десетте модели за слегнувањата во разгледуваните ќелии ќе се обучуваат со избраните алгоритми кои се издвојуваат како најдобри во однос на метричките показатели од прелиминарната петкратна вкрстена валидација, појаснети во претходното поглавје. Станува збор за само два алгоритми – алгоритмот на Случајна шума (RF) со кој ќе се обучуваат моделите за слегнувања во ќелиите SC1, SC2, SC5, SC7, SC9, SC12 и SC13, и алгоритмот на Поткрени дрва (BT) со кој ќе се обучуваат моделите за слегнувања во ќелиите SC4, SC6 и SC14. Во продолжение е даден кус опис на моделите и избраните хиперпараметри за финалното обучување пред предвидувања на целните варијабли.

7.6.3.1. Модел на Случајна шума (RF)

Најголем дел од обучените модели кои понатаму се применуваат за предвидување на целната варијабла – слегнувањата кај брана „Кнежево“ во ќелиите SC1, SC2, SC5, SC7, SC9, SC12 и SC13, се обучени со алгоритмот на Random Forest регресија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја.

За оптимизација на моделите применет е пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators` (50, 100, 150) - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth` (None, 10, 20) - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split` (2, 5, 10) - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features` ("sqrt", "log2", None) - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијаблите.

За добивање на објективна проценка на перформансите применета вгнездена n -кратна валидација, и тоа:

- Внатрешна (inner) трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку GridSearchCV.
- Надворешна (outer) петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Овој пристап обезбедува непристрасна проценка на перформансите и ја намалува веројатноста за пренагласено прилагодување на моделот кон податоците. Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување.

Кај секој модел за слегнувањата во различните ќелии SC1, SC2, SC5, SC7, SC9, SC12 и SC13 добиени се различни вредности на хиперпараметрите за избор на финалниот модел кој подоцна се користи во фазата на предвидувања. Бидејќи станува збор за седум различни модели, оптималната комбинација на хиперпараметрите е прикажана во Табела 56.

Табела 56. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за седумте студии на случај $SC1=f(Zw1,t)$, $SC2=f(Zw1,t)$, $SC5=f(Zw1,t)$, $SC7=f(Zw1,t)$, $SC9=f(Zw1,t)$, $SC12=f(Zw1,t)$ и $SC13=f(Zw1,t)$.

SC	Fold	max_depth	max_features	min_samples_split	n_estimators
SC1	3	10	sqrt	2	150
SC2	4	10	None	2	150
SC5	1	None	None	2	150
SC7	4	None	None	2	50
SC9	2	None	None	2	50
SC12	1	10	sqrt	2	150
SC13	2	None	None	2	50

7.6.3.2. Модел со Поткрнати Регресиони Дрва (BT)

Обучувањето на останатите модели за предвидување на слегнувањата во ќелиите SC4, SC6 и SC14 се подготвени со примена на алгоритмот на Boosted Regression Trees, односно Gradient Boosting Regressor, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја.

Секој модел е дефиниран со следните параметри:

1) `n_estimators` = 100 - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate` = 0.1 - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state` = 42 - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на перформансите на секој од моделите применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 57. Gradient Boosting, како бустинг метод, е особено ефикасен во моделирање на сложени нелинеарни односи и често постигнува висока предиктивна точност.

Табела 57. Просечна вредност на метричките показатели од процесот на обучување на BT моделите по спроведена петкратна валидација, за студии на случај $SC4=f(Zwl,t)$, $SC6=f(Zwl,t)$ и $SC14=f(Zwl,t)$.

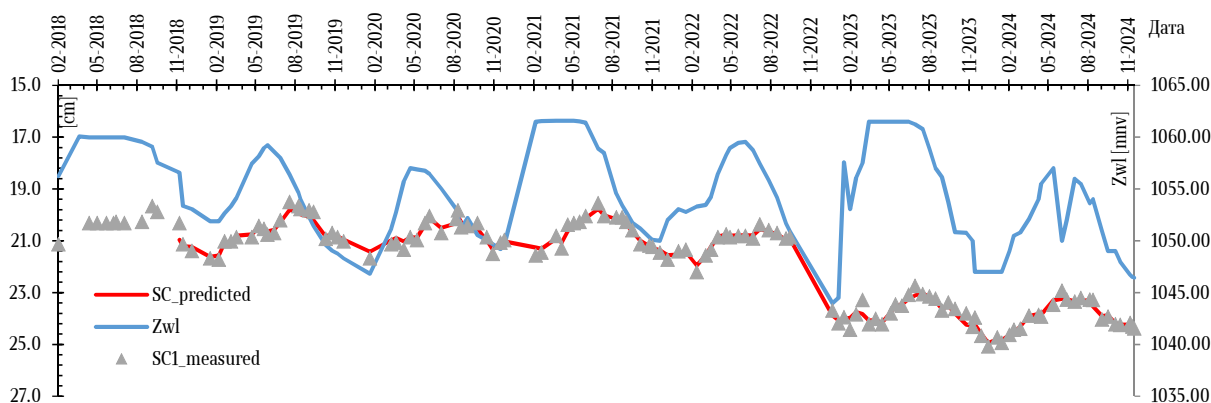
SC	MSE	RMSE	MAE	R2
SC4	0.133	0.362	0.282	0.908
SC6	0.317	0.516	0.304	0.929
SC14	0.178	0.406	0.29	0.906

7.6.4. Предвидување на целните варијабли $SC_i=f(Zwl,t)$

По извршеното обучување на моделите за слегнувањата во секоја анализирана ќелија со примена на најсоодветните алгоритми за машинско учење, следи извршување на предвидувањата на целните варијабли со обучените модели.

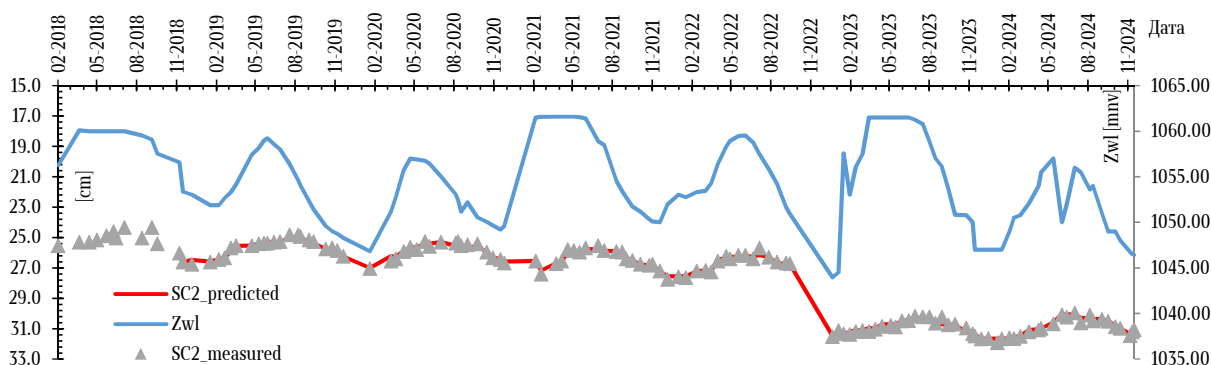
Предвидувањата се извршуваат за познати вредности на предиктор варијаблите „Zwl“ и „t“. Станува збор за низа на податоци од 2606 вредности кои се однесуваат на мерења на нивото на водата во акумулацијата за периодот од 01.11.2018 година до 31.12.2025 година. Целта на предвидувањата е да се пополни временската низа од 01.11.2018 до 31.12.2025 година со вредности на слегнувањата кои операторот на хидросистемот не ги регистрира на дневно ниво, туку ги регистрира на нееднакви временски интервали.

За студијата на случај $SC1=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC1“. Предвидувањата се прикажани графички во Прилог 106. На графикот може да се воочи дека предвидените слегнувања „SC1_predicted“ споредено со регистрираните слегнувања SC1_measured за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 1.9% до 3.1%.



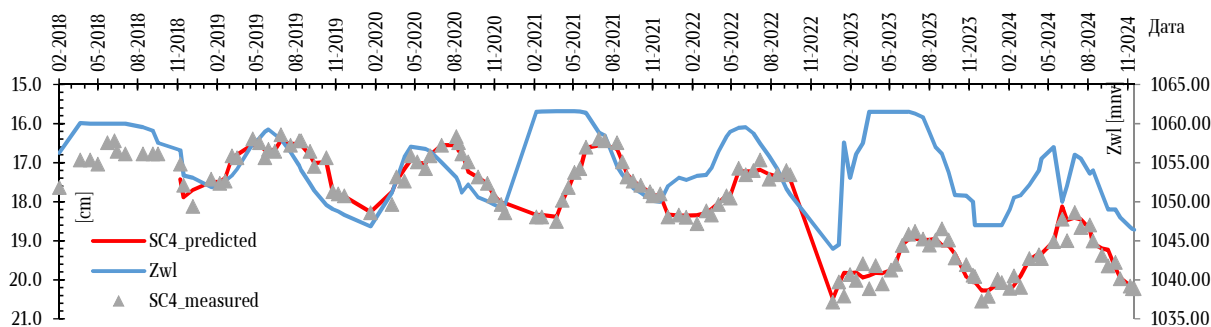
Прилог 106. Графички приказ на пресметаните предвидувања за $SC1=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC1, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

За студијата на случај $SC1=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC2“. Предвидувањата се прикажани графички во Прилог 107. На графикот може да се воочи дека предвидените слегнувања SC2_predicted споредено со регистрираните слегнувања „SC2_measured“ за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 1.3% до 1.7%.



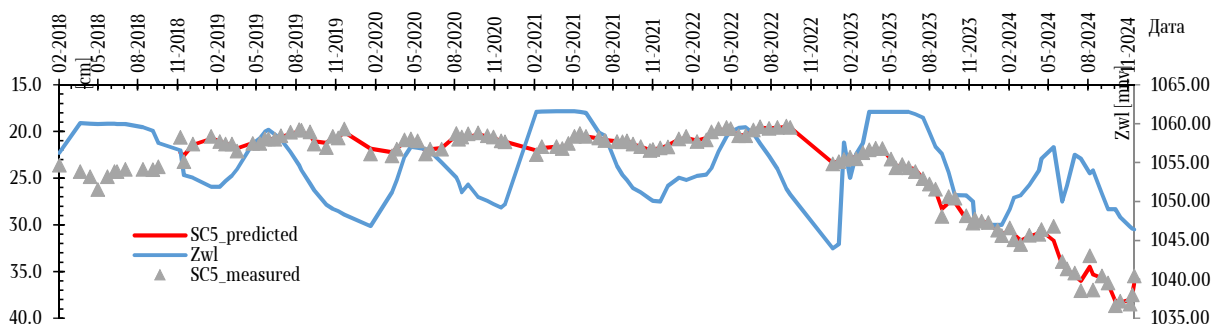
Прилог 107. Графички приказ на пресметаните предвидувања за $SC2=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC2_measured, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

За студијата на случај $SC4=f(Zwl,t)$ обучен е модел со алгоритмот BT, по што пресметани се предвидувања на целната варијабла „SC4“. Предвидувањата се прикажани графички во Прилог 108. На графикот може да се воочи дека предвидените слегнувања SC4_predicted споредено со регистрираните слегнувања SC4_measured за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 2.3% до 2.9%.



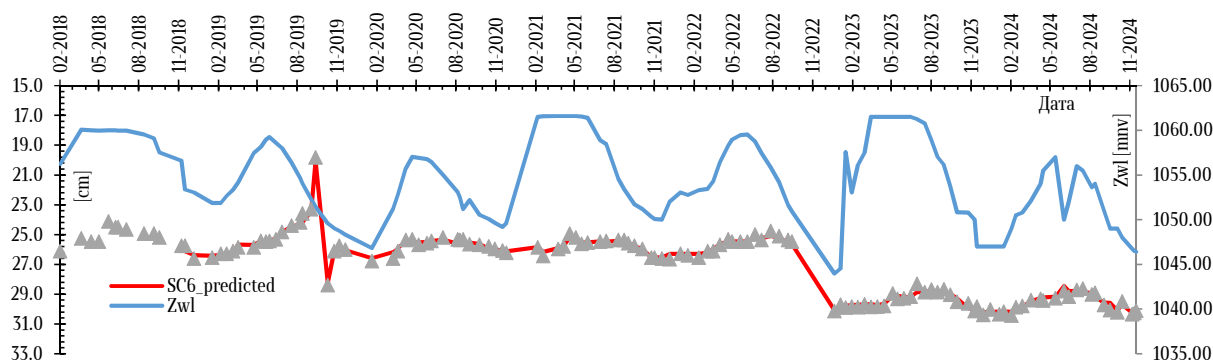
Прилог 108. Графички приказ на пресметаните предвидувања за $SC4=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC4_measured, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

За студијата на случај $SC5=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC5“. Предвидувањата се прикажани графички во Прилог 109. На графикот може да се воочи дека предвидените слегнувања SC5_predicted споредено со регистрираните слегнувања SC5_measured за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 4.6% до 7.8%.



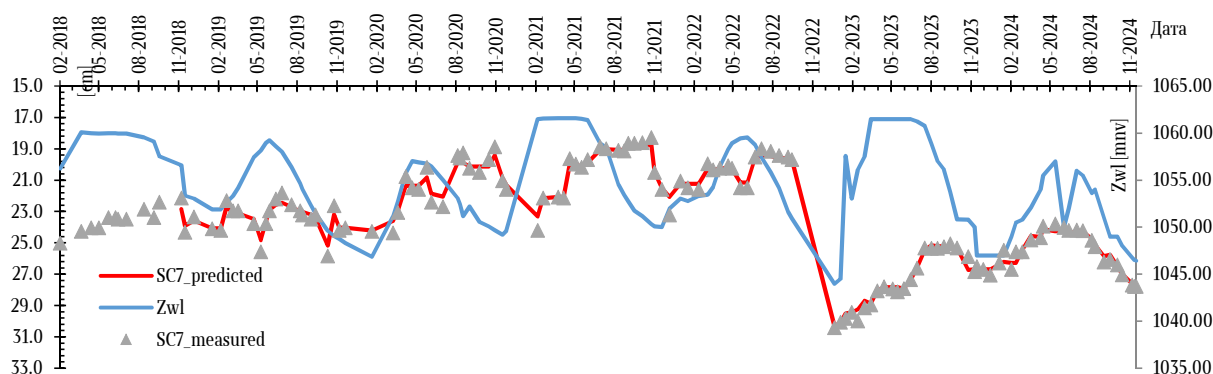
Прилог 109. Графички приказ на пресметаните предвидувања за $SC5=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата SC5_measured, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

За студијата на случај $SC6=f(Zwl,t)$ обучен е модел со алгоритмот BT, по што пресметани се предвидувања на целната варијабла „SC6“. Предвидувањата се прикажани графички во Прилог 110. На графикот може да се воочи дека предвидените слегнувања SC6_predicted споредено со регистрираните слегнувања SC6_measured за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 1.7% до 2.5%.



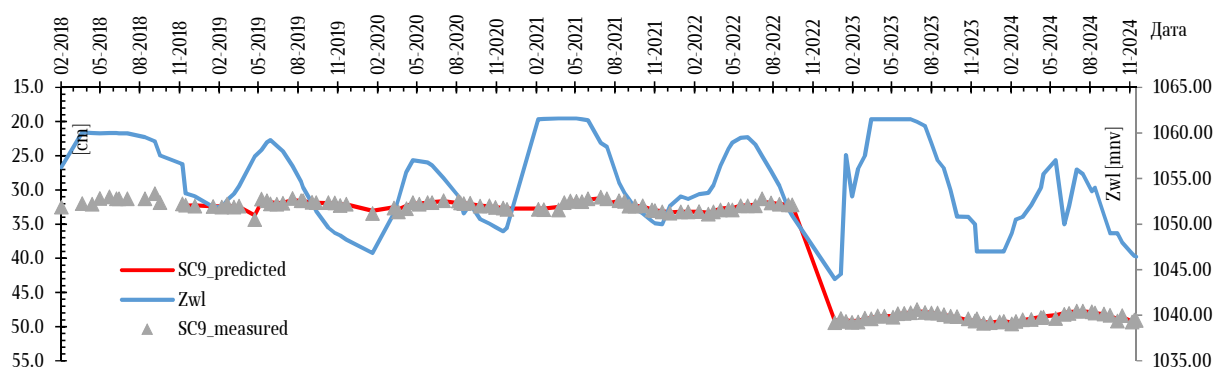
Прилог 110. Графички приказ на пресметаните предвидувања за $SC6=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC6_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

За студијата на случај $SC7=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC7“. Предвидувањата се прикажани графички во Прилог 111. На графикот може да се воочи дека предвидените слегнувања $SC7_predicted$ споредено со регистрираните слегнувања $SC7_measured$ за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 3.3% до 4.9%.



Прилог 111. Графички приказ на пресметаните предвидувања за $SC7=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC7_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

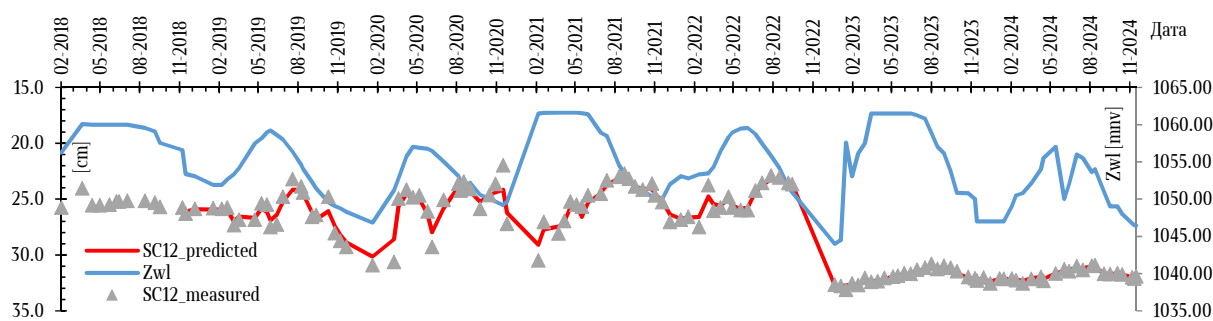
За студијата на случај $SC9=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC9“. Предвидувањата се прикажани графички во Прилог 112. На графикот може да се воочи дека предвидените слегнувања $SC9_predicted$ споредено со регистрираните слегнувања $SC9_measured$ за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 1.8% до 2.3%.



Прилог 112. Графички приказ на пресметаните предвидувања за $SC9=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC9_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „Zwl“.

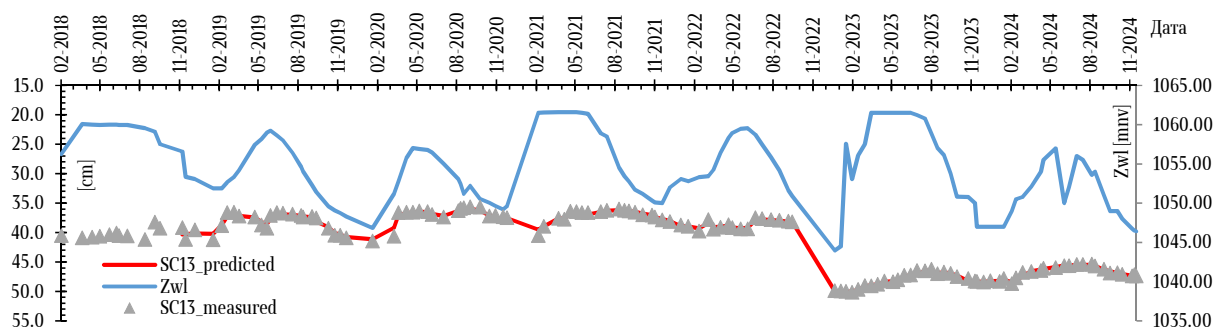
За студијата на случај $SC12=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „SC12“. Предвидувањата се прикажани графички во Прилог 113. На графикот може да се воочи дека предвидените слегнувања $SC12_predicted$ споредено со регистрираните слегнувања $SC12_measured$ за датумите на кои им се преклопуваат вредностите,

се доста блиски, со разлика при споредба на резултатите која се движи од 6.7% до 9.8%.



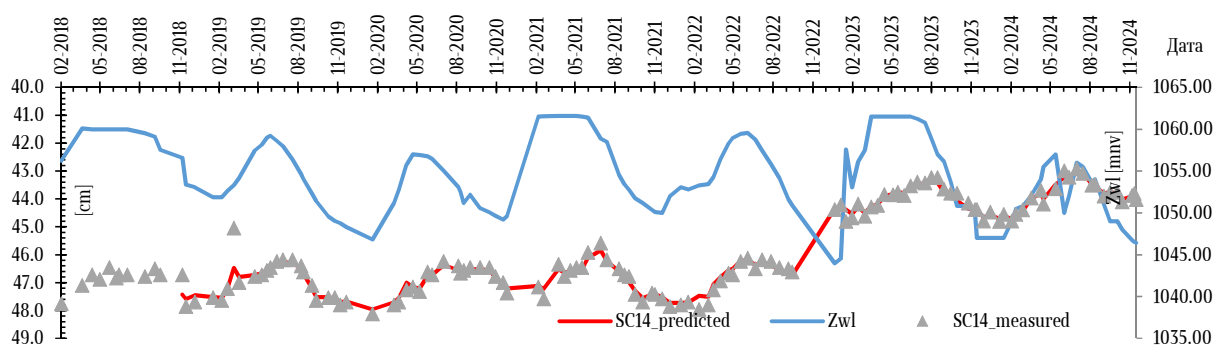
Прилог 113. Графички приказ на пресметаните предвидувања за $SC12=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC12_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „ Zwl “.

За студијата на случај $SC13=f(Zwl,t)$ обучен е модел со алгоритмот RF, по што пресметани се предвидувања на целната варијабла „ $SC13$ “. Предвидувањата се прикажани графички во Прилог 114. На графикот може да се воочи дека предвидените слегнувања $SC13_predicted$ споредено со регистрираните слегнувања $SC13_measured$ за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 2.8% до 3.7%.



Прилог 114. Графички приказ на пресметаните предвидувања за $SC13=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC13_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „ Zwl “.

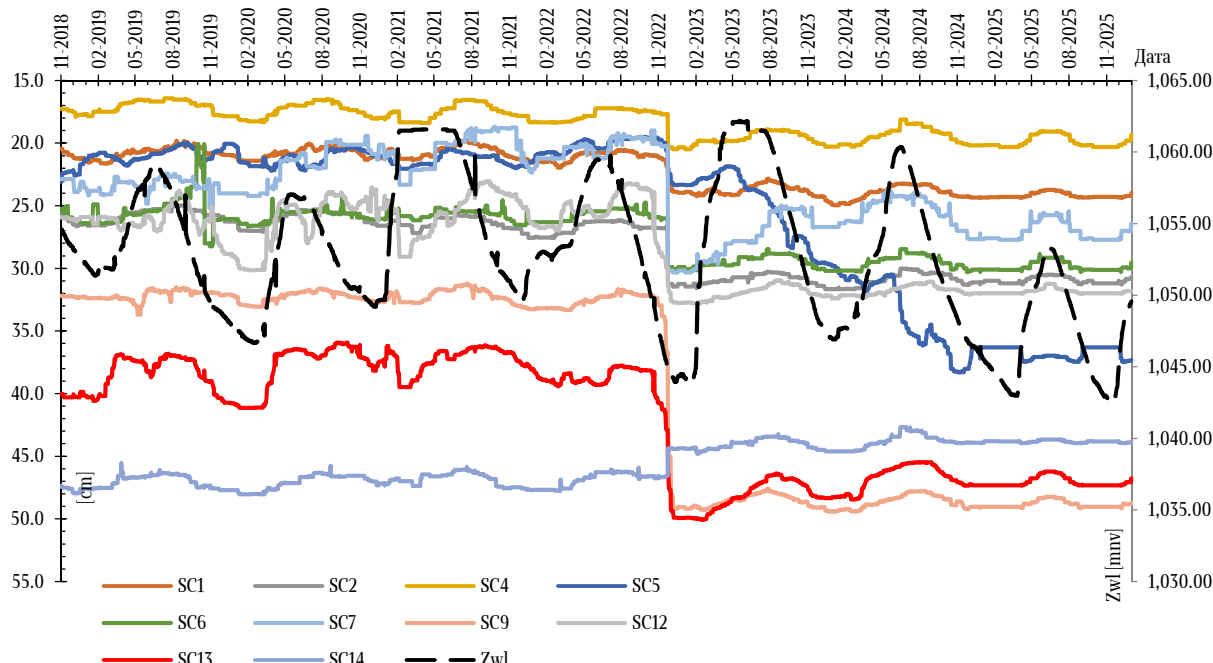
За студијата на случај $SC14=f(Zwl,t)$ обучен е модел со алгоритмот BT, по што пресметани се предвидувања на целната варијабла „ $SC14$ “. Предвидувањата се прикажани графички во Прилог 115. На графикот може да се воочи дека предвидените слегнувања $SC14_predicted$ споредено со регистрираните слегнувања $SC14_measured$ за датумите на кои им се преклопуваат вредностите, се доста блиски, со разлика при споредба на резултатите која се движи од 1.0% до 3.2%.



Прилог 115. Графички приказ на пресметаните предвидувања за $SC14=f(Zwl,t)$ со измерените вредности на слегнувањата во ќелијата $SC14_measured$, и приказ на мереното водно ниво во акумулацијата – предиктор варијаблата „ Zwl “.

Со помош на обучените модели направени се пресметки за предвидување на целните варијабли $SC_i=f(Zwl,t)$ за 2606 дена т.е. 2606 вредности на предиктор варијаблите. Направените пресметки за предвидувањата послужуваат како основа за наредната пресметка – време на достигнување на

врвните вредности на теоретски пресметаните слегнувања во анализираните ќелии, со примена на код во Пајтон. Низата на податоци со кои кодот ќе прави пресметки, која претставува и низа од предвидувањата на слегнувањата во секоја од десетте ќелии, е прикажана во Прилог 116.



Прилог 116. Пресметани предвидувања за слегнувањата во ќелиите “SC1”, “SC2”, “SC4”, “SC5”, “SC6”, “SC7”, “SC9”, “SC12”, “SC13” и “SC14” со дополнет приказ на предиктор варијаблата „Zwl“ – ниво на вода во акумулацијата.

7.6.5. Екстраполација на слегнувањата

7.6.5.1. Методологија на пресметката

Во оваа фаза од анализата извршена е екстраполација на слегнувањата на браната врз основа на претходно добиениот сет на предвидени вредности за поединечните ќелии за мерење на слегнувањата. Методологијата е имплементирана во Python, при што основната цел е да се определи очекуваниот развој на слегнувањата во подолг временски период, односно до крајот на избраниот проектен хоризонт, во конкретниот случај до 2100 година. Кодот започнува со вчитување на влезната Excel-датотека `Input_settlements_calc.xlsx`, во која првата колона претставува датум во денови, додека останатите колони ги содржат вредностите на слегнувањата за поединечните ќелии, означени со ознаките SC. Покрај нив, за секоја ќелија се користи и дополнителна колона `SC_num`, која ја содржи крајната теоретска вредност на слегнувањето добиена од претходна нумеричка анализа, за секоја ќелија соодветно.

За математичко опишување на временскиот развој на слегнувањето применета е експоненцијална асимптотска функција. Овој тип на функција е соодветен за процеси кај кои промената е поинтензивна во почетниот период, а потоа постепено се намалува и се приближува кон крајна стабилизирана вредност. Во кодот функцијата е дефинирана како зависност од времето „t“, кое е изразено во години, коефициентот „k“, почетното слегнување „s₀“ и крајното теоретско слегнување „s_{inf}“. Почетното слегнување „s₀“ се зема како првата вредност за дадената ќелија, додека „s_{inf}“ се презема од соодветната колона `SC_num`. На овој начин, моделот не ја бара произволно крајната вредност, туку ја прифаќа како однапред дефинирана теоретска граница, а преку оптимизација го одредува само параметарот „k“, кој ја контролира брзината со која слегнувањето се приближува кон таа крајна вредност.

За секоја ќелија се формира посебен сет на податоци, при што се отстрануваат редовите без валидни датуми или вредности на слегнување. Доколку за некоја ќелија има помалку од три податочни точки, таа не се обработува, бидејќи таквиот број на мерења не е доволен за стабилно оформување на кривата. Податоците потоа се подредуваат хронолошки, а времето се пресметува

како број на години поминати од првиот достапен датум на мерење. Со функцијата `curve_fit` од библиотеката `SciPy` се врши нелинеарно прилагодување на експоненцијалната функција кон достапните податоци, при што параметарот „k“ се ограничува да биде позитивен. Ова е важно бидејќи позитивниот параметар обезбедува физички логично однесување на моделот, односно постепено приближување на слегнувањето кон крајната вредност. За моделирање на развојот на слегнувањата е применет експоненцијален асимптотски модел од облик:

$$s(t) = s_0 + (s_{inf} - s_0) \cdot (1 - e^{-k \cdot t}) \quad (7.6.1)$$

, каде:

$s(t)$ – слегнување во време t ,

s_0 – почетно измерено слегнување,

s_{inf} – крајно теоретско слегнување,

k – параметар на брзина на консолидација,

t – време во години.

По успешното определување на параметарот „k“, моделот се користи за екстраполација на слегнувањето за идниот период. Во кодот се генерира временска серија од почетниот датум до 31.12.2100 година, со годишен чекор. За секој од овие датуми се пресметува предвиденото слегнување според прилагодена експоненцијална крива. Дополнително, за секоја ќелија се пресметуваат моментите кога слегнувањето достигнува 70%, 80% и 90% од крајното теоретско слегнување. Овие вредности се добиваат аналитички од експоненцијалниот модел, преку пресметка на времето потребно за достигнување на одреден процент од разликата помеѓу почетната и крајната вредност. За секој процент се пресметува соодветната година, датумот на достигнување, вредноста на слегнувањето и информација дали таа вредност се достигнува до 2100 година.

Резултатите од екстраполацијата се прикажуваат и графички. За секоја ќелија се исцртуваат измерените вредности, прилагодената и продолжена крива на слегнување, крајното теоретско слегнување, како и хоризонталните линии што ги означуваат 70%, 80% и 90% од крајната вредност на слегнувањата. Во графичкиот приказ слегнувањата се прикажани со негативен знак, што е вообичаено при инженерска интерпретација на вертикални деформации, бидејќи слегнувањето претставува поместување во насока на земјината тежа. На крај, за секоја ќелија се формира табела со предвидените вредности, а збирните резултати за сите ќелии се зачувуваат во Excel-датотека `settlement_reaching_times_summary.xlsx`. На овој начин, `DamResML` овозможува систематска, повторлива и транспарентна постапка за пресметка на долгорочниот развој на слегнувањата и за споредба на динамиката на стабилизација помеѓу различните мерни ќелии. Како што следи, дадени се коментари на излезните резултати.

7.6.5.2. Резултати

Во рамки на анализата на долгорочното однесување на браната Кнежево беше извршена екстраполација на слегнувањата за повеќе мерни ќелии со примена на Пајтон програмскиот код и нелинеарен експоненцијален модел на консолидација. Основната цел на анализата е да се процени идниот развој на слегнувањата и времето потребно за достигнување на стабилизирана состојба, односно процентуално достигнување на крајното теоретско слегнување.

Во анализата се обработени следните мерни ќелии: SC1, SC2, SC4, SC5, SC6, SC7, SC9, SC12, SC13 и SC14. За секоја ќелија се користени измерени вредности на слегнување, како и претходно дефинирана крајна теоретска вредност на слегнување добиена од нумерички анализи.

Како што следи даден е табеларен приказ на излезните резултати од екстраполацијата на слегнувањата во секоја ќелија, во Табела 58 и Табела 59. Фокусот е ставен на пресметката на екстраполираната крива на слегнувања и нејзиното надминување на прагот на 2100 година, т.е. дали крајните пресметани теоретски слегнувања ќе се случат во 21 век.

Табела 58. Табеларен приказ на дел од резултатите од пресметката на екстраполираните слегнувања во десетте ќелии, со вредности на времето на достигнување на 70% од пресметаните теоретски слегнувања.

Ќелија	Иницијално слегнување [cm]	Крајно теоретско слегнување [cm]	k [1/year]	Дата на достигнување на 70% од крајните слегнувања	Пресметана вредност на 70% од крајните слегнувања [cm]	Дали 70% од слегнувањата ќе се реализираат до 2100 година?
SC1	20.43	36.00	0.038	2050-11-29	31.328	Yes
SC2	25.68	41.70	0.056	2040-03-19	36.895	Yes
SC4	17.28	35.10	0.021	2075-09-11	29.753	Yes
SC5	23.10	42.60	0.064	2037-07-20	36.749	Yes
SC6	25.24	48.00	0.031	2057-07-28	41.173	Yes
SC7	22.85	37.40	0.030	2059-04-21	33.036	Yes
SC9	32.20	60.60	0.102	2030-08-31	52.081	Yes
SC12	25.94	43.30	0.053	2041-08-05	38.092	Yes
SC13	40.04	59.70	0.050	2043-02-20	53.802	Yes
SC14	47.31	66.70	0.000	Beyond calc. date	60.883	No

Табела 59. Табеларен приказ на дел од резултатите од пресметката на екстраполираните слегнувања во десетте ќелии, со вредности на времето на достигнување на 80% и 90% од пресметаните теоретски слегнувања.

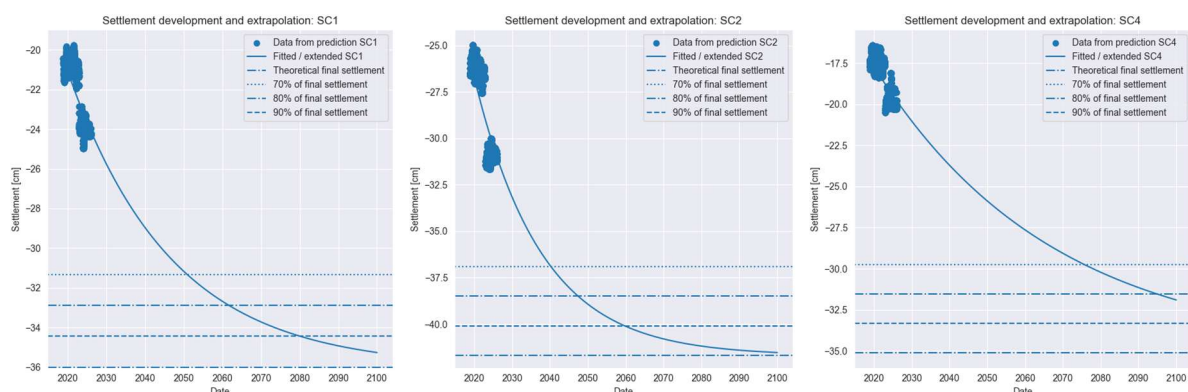
Ќелија	Дата на достигнување на 80% од крајните слегнувања	Пресметана вредност на 80% од крајните слегнувања [cm]	Дали 80% од слегнувањата ќе се реализираат до 2100 година?	Дата на достигнување на 90% од крајните слегнувања	Пресметана вредност на 90% од крајните слегнувања [cm]	Дали 90% од слегнувањата ќе се реализираат до 2100 година?
SC1	2061-09-17	32.885	Yes	2080-03-06	34.443	Yes
SC2	2047-06-01	38.497	Yes	2059-09-22	40.098	Yes
SC4	2094-11-04	31.535	Yes	2127-08-01	33.318	No
SC5	2043-11-08	38.699	Yes	2054-08-18	40.650	Yes
SC6	2070-08-14	43.449	Yes	2092-12-01	45.724	Yes
SC7	2072-12-06	34.491	Yes	2096-03-25	35.945	Yes
SC9	2034-08-25	54.921	Yes	2041-06-17	57.760	Yes
SC12	2049-04-05	39.828	Yes	2062-05-13	41.564	Yes
SC13	2051-04-29	55.768	Yes	2065-04-26	57.734	Yes
SC14	Beyond calculable date	62.822	No	Beyond calculable date	64.761	No

Коментари за секоја ќелија се дадени како што следи.

Кај мерната ќелија „SC1“, иницијалното слегнување изнесува 20.43 cm, додека крајното теоретско слегнување изнесува 36.00 cm. Добиениот параметар на консолидација изнесува $k = 0.038$ 1/година, што укажува на умерено бавен процес на стабилизација. Според резултатите од екстраполацијата, 70% од крајното слегнување, односно 31.328 cm, ќе се достигне на 29.11.2050 година. Осумдесет проценти од крајното слегнување, односно 32.885 cm, се очекува да се достигнат на 17.09.2061 година, додека 90% од крајното слегнување, односно 34.443 cm, ќе се достигнат на 06.03.2080 година. Резултатите укажуваат дека кај „SC1“ процесот на консолидација се одвива постепено, но стабилно, при што 90% од очекуваното слегнување ќе биде реализирано пред 2100 година. Графички приказ на екстраполираната крива за „SC1“ е прикажана во Прилог 117.

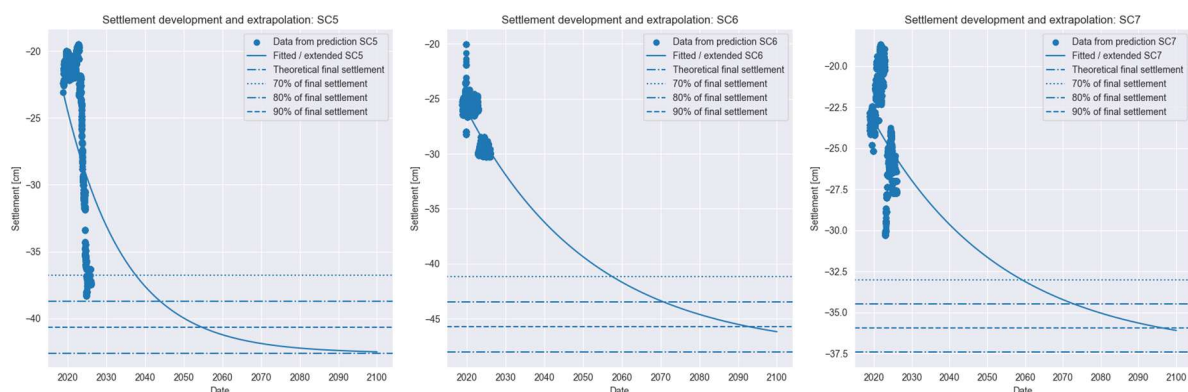
Кај „SC2“, иницијалното слегнување изнесува 25.68 cm, а крајното теоретско слегнување 41.70 cm. Вредноста на параметарот $k = 0.056$ 1/година укажува на релативно брз процес на консолидација. Според моделот, 70% од крајното слегнување, односно 36.895 cm, ќе се достигнат на 19.03.2040 година. Осумдесет проценти, односно 38.497 cm, ќе се достигнат на 01.06.2047 година, а 90% од крајното слегнување, односно 40.098 cm, ќе бидат реализирани на 22.09.2059 година. Овие резултати укажуваат дека „SC2“ покажува поволна динамика на стабилизација и

релативно брзо приближување кон крајната вредност.



Прилог 117. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерните ќелии „SC1“, „SC2“ и „SC4“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.

Кај „SC4“, иницијалното слегнување изнесува 17.28 cm, а крајното теоретско слегнување 35.10 cm. Добиениот параметар $k = 0.021$ 1/година покажува најбавен процес на консолидација помеѓу анализираните мерни ќелии. Според пресметките, 70% од крајното слегнување, односно 29.753 cm, ќе се достигнат на 11.09.2075 година, додека 80%, односно 31.535 cm, ќе се достигнат на 04.11.2094 година. Деведесет проценти од крајното слегнување, односно 33.318 cm, се очекува да бидат достигнати дури на 01.08.2127 година, односно по крајот на анализираниот период. Овој резултат укажува дека процесот на консолидација кај „SC4“ ќе продолжи и по 2100 година.



Прилог 118. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерните ќелии „SC5“, „SC6“ и „SC7“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.

Кај „SC5“, иницијалното слегнување изнесува 23.10 cm, а крајното теоретско слегнување 42.60 cm. Параметарот $k = 0.064$ 1/година претставува една од највисоките вредности во анализата, што укажува на брз процес на стабилизација. Според моделот, 70% од крајното слегнување, односно 36.749 cm, ќе се достигнат на 20.07.2037 година. Осумдесет проценти, односно 38.699 cm, ќе се достигнат на 08.11.2043 година, а 90%, односно 40.650 cm, ќе бидат реализирани на 18.08.2054 година.

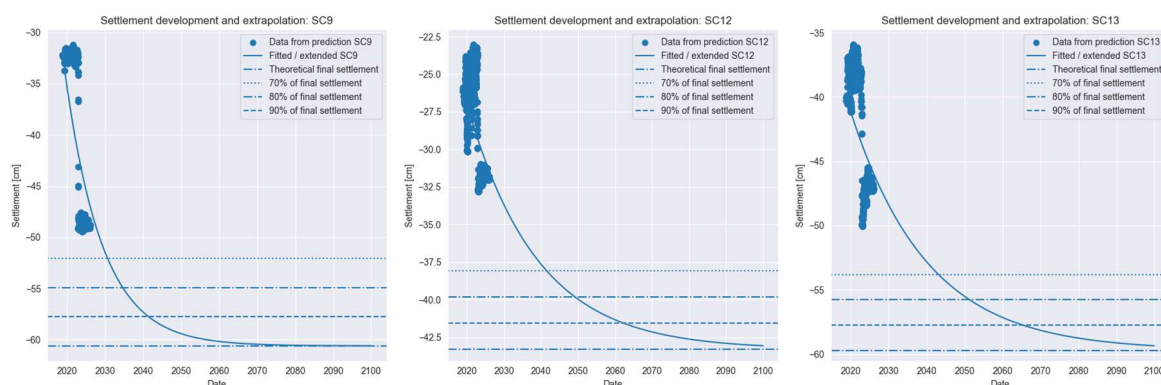
Кај „SC6“, иницијалното слегнување изнесува 25.24 cm, а крајното теоретско слегнување 48.00 cm. Вредноста на параметарот $k = 0.031$ 1/година укажува на умерено бавен процес на стабилизација. Според резултатите, 70% од крајното слегнување, односно 41.173 cm, ќе се достигнат на 28.07.2057 година. Осумдесет проценти, односно 43.449 cm, ќе се достигнат на 14.08.2070 година, а 90%, односно 45.724 cm, ќе се достигнат на 01.12.2092 година. Ова покажува дека процесот на консолидација кај SC6 ќе продолжи во текот на целиот XXI век, но 90% од слегнувањето сепак ќе се реализира пред 2100 година.

Кај „SC7“, иницијалното слегнување изнесува 22.85 cm, а крајното теоретско слегнување 37.40

см. Параметарот $k = 0.030$ 1/година укажува на релативно бавна динамика на консолидација. Според моделот, 70% од крајното слегнување, односно 33.036 см, ќе се достигнат на 21.04.2059 година. Осумдесет проценти, односно 34.491 см, ќе се достигнат на 06.12.2072 година, а 90%, односно 35.945 см, ќе се реализираат на 25.03.2096 година. Резултатите укажуваат дека „SC7“ има долгорочен процес на стабилизација, но најголемиот дел од слегнувањето ќе се реализира до крајот на XXI век.

Кај „SC9“, иницијалното слегнување изнесува 32.20 см, а крајното теоретско слегнување 60.60 см. Вредноста на параметарот $k = 0.102$ 1/година претставува највисока вредност помеѓу анализираниите ќелии, што укажува на многу брз процес на консолидација. Според моделот, 70% од крајното слегнување, односно 52.081 см, ќе се достигнат на 31.08.2030 година. Осумдесет проценти, односно 54.921 см, ќе се достигнат на 25.08.2034 година, додека 90%, односно 57.760 см, ќе се достигнат на 17.06.2041 година. Овие резултати покажуваат дека „SC9“ има најбрза динамика на стабилизација од сите анализирани мерни ќелии.

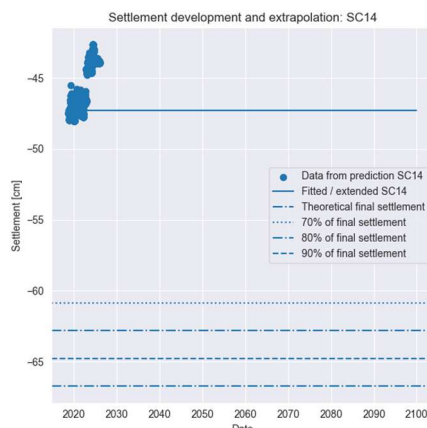
Кај „SC12“, иницијалното слегнување изнесува 25.94 см, а крајното теоретско слегнување 43.30 см. Добиениот параметар $k = 0.053$ 1/година укажува на стабилен и релативно брз процес на консолидација. Според пресметките, 70% од крајното слегнување, односно 38.092 см, ќе се достигнат на 05.08.2041 година. Осумдесет проценти, односно 39.828 см, ќе се достигнат на 05.04.2049 година, а 90%, односно 41.564 см, ќе се достигнат на 13.05.2062 година. Овие резултати укажуваат на поволна динамика на стабилизација.



Прилог 119. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерните ќелии „SC9“, „SC12“ и „SC13“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.

Кај „SC13“, иницијалното слегнување изнесува 40.04 см, а крајното теоретско слегнување 59.70 см. Вредноста на параметарот $k = 0.050$ 1/година покажува умерено брз процес на консолидација. Според моделот, 70% од крајното слегнување, односно 53.802 см, ќе се достигнат на 20.02.2043 година. Осумдесет проценти, односно 55.768 см, ќе се достигнат на 29.04.2051 година, а 90%, односно 57.734 см, ќе се реализираат на 26.04.2065 година. Резултатите покажуваат дека „SC13“ има стабилен и реалистичен тренд на долгорочна консолидација.

Кај „SC14“, иницијалното слегнување изнесува 47.31 см, а крајното теоретско слегнување 66.70 см. Меѓутоа, добиениот параметар $k \approx 0.000$ 1/година укажува дека моделот практично не успева да препознае реален процес на консолидација. Од графичкиот приказ на Прилог 120 и анализата на податоците се забележува дека измерените вредности покажуваат спротивен тренд од физички очекуваниот – наместо постепено зголемување на апсолутната вредност на слегнувањето, се забележува благо намалување на слегнувањето со текот на времето. Поради тоа, моделот не може да пресмета реални датуми за достигнување на 70%, 80% и 90% од крајното слегнување, па тие се означени како „Beyond calculable date“. Ова укажува дека податоците за „SC14“ најверојатно содржат мерна грешка, проблем со инструментот, промена на референтната кота или неконзистентност во обработката на податоците.



Прилог 120. Пресметана екстраполирана крива на слегнувања до 2100 година за слегнувања во мерната ќелија „SC14“ со приказ на теоретските пресметани вредности на слегнувањата кои одговараат на 70%, 80% и 90% од максималната пресметана вредност.

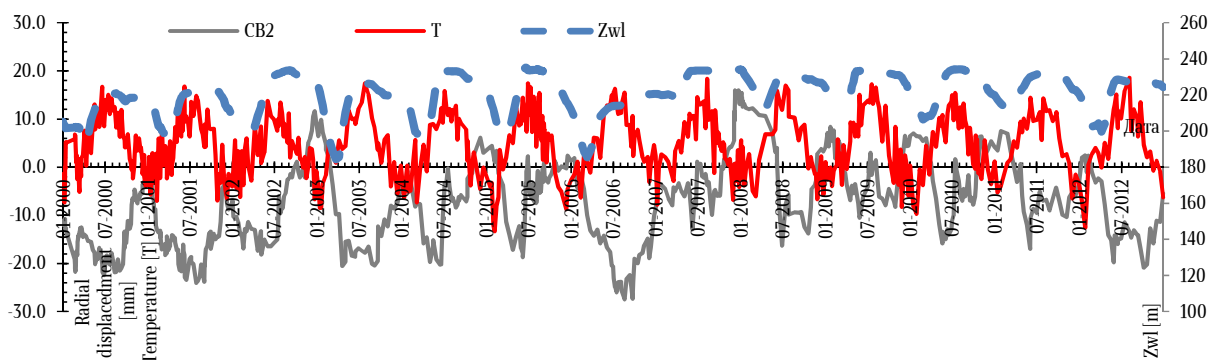
7.7. Анализа на случај 2.7 – $CB2=f(Zwl, T)$

7.7.1. Подготовка на податоците

Во последната анализа на случај од серијата студии на случај 2 ќе се анализира зависноста помеѓу радијалните поместувања „CB2“ во централниот блок на лачна брана во Франција во сопственост на EDF, во функција од нивото на водата во акумулацијата регистрирано на дневно ниво „Zwl“ и амбиенталната температура „T“ околу браната.

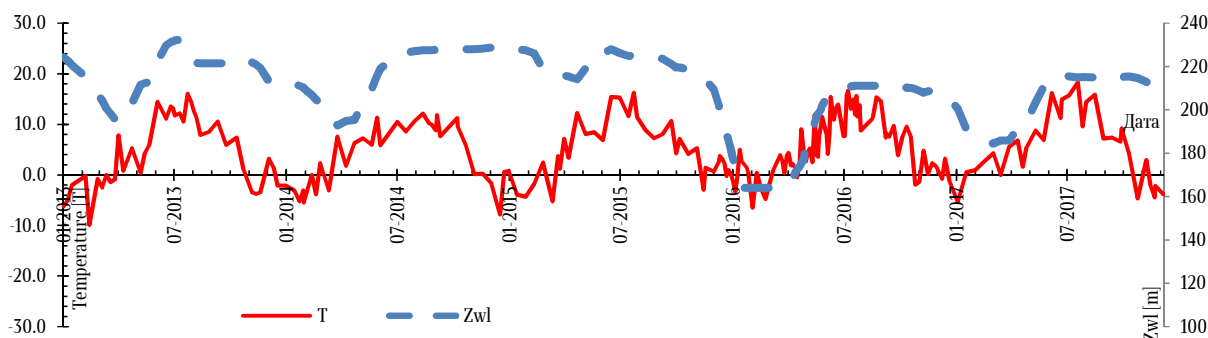
Целта на анализата на случај 2.7 ќе биде пресметка на радијалните поместувања „CB2“ за периодот помеѓу 27.01.2013 година до 18.12.2017 година кога истите не се познати, со помош на поучени модели со машинско учење кои ќе се обучуваат на сет на мерени податоци на целната и предиктор варијаблите од периодот 19.01.2020 година па сè до 27.01.2013 година.

Сетот на податоци на целната варијаблата „CB2“ и предиктор варијаблите „Zwl“ и „T“ кои ќе се применуваат во процесот на обучување на моделите, се состои од вкупно 699 вредности за периодот помеѓу 19.01.2020 година па сè до 27.01.2013 година. Мерењата се на дневно ниво, но истите се извршувани во неправилни временски интервали т.е. не се извршувани секојдневно. Вредностите на целната варијабла „CB2“ и предиктор варијаблите „Zwl“ и „T“ користени во процесот на обучување на моделите со машинско учење се прикажани во Прилог 121.



Прилог 121. Графички приказ на користените вредности на предиктор варијаблата „CB2“ и предиктор варијаблите „Zwl“ и „T“ во процесот на обучување на моделите со машинско учење за анализа на случај $CB2=f(Zwl, T)$.

Сетот на податоци на предиктор варијаблите „Zwl“ и „T“ кои ќе се користат во периодот на предвидување на целната варијабла „CB2“ се состои од вкупно 226 вредности за периодот помеѓу 27.01.2013 година па сè до 18.12.2017 година, исто така регистрирани во неправилни временски интервали, на дневно ниво. За овој период, целната варијабла „CB2“ е непозната. Истите се прикажани графички во Прилог 122.



Прилог 122. Графички приказ на предиктор варијаблите „Zwl“ и „T“ кои се користат во процесот на предвидување на целната варијабла „CB2“ за анализа на случај $CB2=f(Zwl,T)$.

Целната варијабла – радијалните поместувања „CB2“ во централниот блок, во анализите е во [mm]. Предиктор варијаблата „Zwl“ – ниво на вода во акумулацијата е во [m], додека амбиенталната температура „T“ е во мерна единица [K].

Целната варијабла „CB2“ може да има позитивни и негативни вредности како резултат на насоката на напредување на радијалните поместувања – возводно или низводно. Така, позитивните вредности на радијалните поместувања значат поместувања во низводен правец, додека негативните – во возводен правец.

7.7.2. Прелиминарна петкратна вкрстена валидација

По прегледување на низата на податоци и нивна поделба за период на обучување и предвидување, изработена е засебна .xls датотека до наслов „Train and test.xls“ со податоци кои ќе се вклучат во процесот на обучување на секој модел. Датотеката се повикува во DamResML кодот во првата ќелија, по што започнува процесот на петкратна вкрстена валидација со алгоритмите: (1) Случајна шума (анг. „Random Forest“), (2) Одлучувачко дрво (анг. „Decision tree“, (3) Перцепсија со потпомогнати вектори (анг. „Support Vector Regressor“), (4) K – најблиски соседи (анг. „K – nearest Neighbors“), (5) Гаусов процесен перцептор (анг. „Gaussian Process Regressor“), (6) XGBoost алгоритам за градиентно поткревање (анг. „XGBoost“, (7) Вештачки невронски мрежи (анг. „Artificial Neural Networks“, (8) Поткренати регресиони дрва (анг. „Boosted Regression Trees“) и (9) Генерализирана регресиона невронска мрежа (анг. „General Regression Neural Network“).

Податоците за целната варијабла се користени во оригинална форма, без примена на нормализација или стандардизација на податоците. Од друга страна, предиктор варијаблите се стандардизирани во рамки на секој сет од петкратната валидација, но само за модели кои се чувствителни на скалирање на податоците.

Резултатите од петкратната вкрстена валидација за предиктор варијаблата „CB2“ се дадени како што следи и истите се анализирани преку компарација на вредностите за метричките показатели MSE, RMSE, MAE и R^2 прикажани во Табела 60.

Табела 60. Резултати од прелиминарна петкратна вкрстена валидација за анализа на случај 2.7 – $CB2=f(Zwl,T)$.

Model	MSE	RMSE	MAE	R2
GPR	88.362	9.347	7.136	0.018
Decision Tree	62.127	7.859	5.451	0.301
KNN	42.469	6.511	4.833	0.524
XGBoost	42.317	6.499	4.679	0.526
GRNN	40.573	6.355	4.867	0.546
ANN	39.166	6.256	4.847	0.562
SVR	38.583	6.207	4.538	0.570
Boosted Trees	38.100	6.166	4.599	0.575
Random Forest	36.660	6.050	4.488	0.590
min:	36.660	6.050	4.488	0.018
max:	88.362	9.347	7.136	0.590

Резултатите од прелиминарната петкратна вкрстена валидација за анализа на случај 2.7 – $CB2=f(Zw1,T)$ покажуваат дека сите применети модели со машинско учење успеваат во одреден степен да ја репродуцираат зависноста помеѓу радијалните поместувања на централниот блок на лачната брана „CB2“ и предиктор варијаблите – нивото на водата во акумулацијата „Zw1“ и амбиенталната температура „T“. За разлика од дел од претходните анализи на случај во рамки на ова истражување, кај оваа анализа сите модели, освен Гаусовиот процесен регресор (GPR), остваруваат позитивни вредности на коефициентот на детерминација R^2 , што укажува дека моделите поседуваат способност да ја опишат варијабилноста на целната варијабла „CB2“ со прифатливо ниво на точност.

Од добиените резултати може да се забележи дека најдобри перформанси покажува моделот Случајна шума (Random Forest), кој остварува најниски вредности на грешките $MSE = 36.660$, $RMSE = 6.050$ mm и $MAE = 4.488$ mm, како и највисока вредност на коефициентот $R^2 = 0.590$. Ваквите резултати укажуваат дека Random Forest моделот има најголема способност за генерализација врз непознати податоци и најуспешно ја пресликува нелинеарната зависност помеѓу радијалните поместувања и влезните променливи. Дополнително, релативно ниските вредности на MAE и $RMSE$ укажуваат дека просечните отстапувања помеѓу мерените и предвидените вредности на „CB2“ се ограничени и прифатливи за инженерска интерпретација на однесувањето на браната.

Многу блиски резултати на Random Forest покажуваат и моделите Поткренати регресиони дрва (Boosted Trees), Регресија со потпомогнати вектори (SVR), Вештачки невронски мрежи (ANN), Генерализирана регресиона невронска мрежа (GRNN), XGBoost и KNN. Овие модели остваруваат вредности на коефициентот R^2 помеѓу 0.524 и 0.575, што укажува на умерено добра способност за репродукција на радијалните поместувања. Особено е значајно што разликите помеѓу нивните метрички показатели се релативно мали, што укажува дека зависноста $CB2=f(Zw1,t)$ може успешно да се моделира со повеќе различни типови на алгоритми, без значително доминирање на само еден модел.

Одлучувачкото дрво (Decision Tree) покажува послаби резултати во однос на ансамбл моделите и останатите напредни алгоритми, со $R^2 = 0.301$ и значително повисоки вредности на грешките. Ова е очекувано, имајќи предвид дека единечното одлучувачко дрво често има ограничена способност за генерализација и е чувствително на варијации во податоците. Сепак, добиените резултати сепак покажуваат дека и овој модел успева делумно да ја опише зависноста помеѓу поместувањата и влијателните фактори.

Најслаби резултати покажува Гаусовиот процесен регресор (GPR), со највисоки вредности на грешките и најниска вредност на коефициентот на детерминација $R^2 = 0.018$. Иако вредноста е позитивна, таа укажува дека моделот има многу ограничена способност за објаснување на варијациите на целната варијабла. Ваквиот резултат може да биде последица на сложената нелинеарна природа на зависноста помеѓу радијалните поместувања, нивото на водата и температурата, како и чувствителноста на GPR моделот на изборот на кернел функцијата и хиперпараметрите.

Наместо заклучок, резултатите од петкратната вкрстена валидација покажуваат дека постои реална и препознатлива зависност помеѓу радијалните поместувања „CB2“ и предиктор варијаблите „Zw1“ и „T“. Вредностите на R^2 кои се движат до 0.590 укажуваат дека значителен дел од варијабилноста на радијалните поместувања може да се објасни преку промените на нивото на водата во акумулацијата и амбиенталната температура. Дополнително, добиените резултати потврдуваат дека ансамбл методите и нелинеарните модели имаат подобра способност за моделирање на сложеното структурно однесување на лачните брани во однос на поедноставните алгоритми.

Врз основа на добиените резултати од прелиминарната петкратна вкрстена валидација може да се заклучи дека моделите Random Forest, Boosted Trees и SVR претставуваат најсоодветни кандидати за понатамошниот процес на обучување и предвидување на непознатите вредности на радијалните поместувања „CB2“ за периодот помеѓу 27.01.2013 година и 18.12.2017 година.

7.7.3. Обучување на моделите

Согласно резултатите од петкратната вкрстена валидација, обучени се вкупно три модели со различни алгоритми за машинско учење, и тоа RF, BT и SVR, со цел комплетирање на анализа на случај 2.7 CB2=f(Zw1,T). Како што следи, даден е кус приказ на усвоените хиперпараметри за секој обучен модел кој е земен предвид во процесот на предвидување.

7.7.3.1. Модел на Случајна шума (RF)

Првиот обучен модел е моделот на Random Forest регресија, кој претставува ансамбл-метод базиран на повеќе одлучувачки дрвја.

За оптимизација на моделот е применет пристапот на мрежно пребарување (анг. „Grid Search“), при што се испитуваат различни комбинации на хиперпараметрите:

1) `n_estimators` (50, 100, 150) - Го дефинира бројот на дрвја во ансамблот. Поголем број обично води до подобра стабилност, но со поголема пресметковна сложеност.

2) `max_depth` (None, 10, 20) - Максималната длабочина на секое дрво. Контролира колку комплексен може да биде моделот.

3) `min_samples_split` (2, 5, 10) - Минимален број на примероци потребни за делење на јазол. Поголеми вредности ја намалуваат можноста за пренагласено прилагодување.

4) `max_features` ("sqrt", "log2", None) - Бројот на параметри што се разгледуваат при секое делење. Овој параметар воведува дополнителна случајност и ја подобрува генерализацијата. Со функцијата 'None' сите предиктор варијабли се вклучуваат во сите јазли, додека 'sqrt' и 'log2' земаат предвид само дел од предиктор варијаблите.

За добивање на објективна проценка на перформансите применета вгнездена n-кратна валидација, и тоа:

- Внатрешна трикратна вкрстена валидација - се користи за избор на најдобра комбинација на хиперпараметри преку GridSearchCV.
- Надворешна петкратна вкрстена валидација - се користи за евалуација на моделот, при што во секој сет се користи веќе оптимизиран модел од внатрешната валидација.

Овој пристап обезбедува непристрасна проценка на перформансите и ја намалува веројатноста за пренагласено прилагодување на моделот кон податоците. Перформансите на моделот се оценуваат преку метриците MSE, RMSE, MAE и R², што овозможува сеопфатна анализа на точноста и стабилноста. Дополнително, се избира финален модел врз основа на најдобриот сет од петкратната вкрстена валидација, кој потоа се обучува врз целокупното множество на податоци за обучување. За согледување на изборот на најдобри вредности на хиперпараметрите од процесот на надворешната петкратна валидација на моделот обучуван со RF, даден е табеларен приказ во Табела 61. Финалното обучување е направено согласно хиперпараметрите од петтиот сет од петкратната надворешна валидација, со вредности: 'max_depth': 10, 'max_features': sqrt, 'min_samples_split': 10, 'n_estimators': 150. Процесот на обучување траеше 38s и 138ms.

Табела 61. Табеларен приказ на најдобрите хиперпараметри од надворешната петкратна валидација за моделот со RF за студија на случај CB2=f(Zw1,T).

Fold	max_depth	max_features	min_samples_split	n_estimators
1	10	sqrt	10	100
2	NaN	sqrt	10	100
3	NaN	sqrt	10	150
4	NaN	sqrt	10	50
5	10	sqrt	10	150

Финалното обучување е направено согласно хиперпараметрите од петтиот сет од петкратната надворешна валидација, со вредности: 'max_depth': 10, 'max_features': sqrt, 'min_samples_split': 10, 'n_estimators': 150. Процесот на обучување траеше 38s и 138ms.

7.7.3.2. Модел со Поткрени регресиони дрва (BT)

Вториот модел за обучување на целната варијабла “CB2” се базира на алгоритмот на Boosted Regression Trees, односно Gradient Boosting Regressor, кој претставува ансамбл-метод базиран на последователно градење на одлучувачки дрвја.

Моделот е дефиниран со следните параметри:

1) `n_estimators = 100` - Го дефинира бројот на дрвја кои се додаваат последователно во моделот. Поголем број овозможува подобро учење, но ја зголемува пресметковната сложеност.

2) `learning_rate = 0.1` - Ја контролира брзината на учење, односно придонесот на секое поединечно дрво. Помали вредности водат до постабилно и постепено учење.

3) `random_state = 42` - Обезбедува репродуктивност на резултатите.

Бидејќи моделот е базиран на одлучувачки дрвја, не е чувствителен на скалирање на влезните податоци, па затоа не е применета дополнителна стандардизација.

За евалуација на перформансите на моделот применета е петкратна вкрстена валидација, при што моделот се обучува и тестира на различни сетови од податоците. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот. Нивната вредност е прикажана во Табела 62.

Табела 62. Вредности на метричките показатели од процесот на обучување на BT моделот по спроведена петкратна валидација за анализа на случај $CB2=f(Zw1,T)$.

Fold	MSE	RMSE	MAE	R2
1	37.334	6.110	4.592	0.581
2	33.522	5.790	4.311	0.569
3	43.765	6.616	5.076	0.516
4	39.498	6.285	4.555	0.606
5	36.381	6.032	4.460	0.601

Просечните вредности на метричките показатели, се: Avg MSE: 38.100, Avg RMSE: 6.166, Avg MAE: 4.599, Avg R^2 : 0.574. Потребното време за обучување беше 166ms.

7.7.3.3. Модел на Регресија со потпомогнати вектори (SVR)

Третиот обучен модел е моделот на Регресија со потпомогнати вектори или накратко SVR (анг. „Support Vector Regressor”). Моделот е дефиниран со следните параметри:

1) `kernel = "rbf"` - Се користи радијална базисна функција (Radial Basis Function), која овозможува моделирање на нелинеарни односи помеѓу влезните и излезните променливи.

2) `C = 100` - Регуларизациски параметар кој ја контролира рамнотежата помеѓу минимизирање на грешката и комплексноста на моделот. Поголема вредност на C значи дека моделот повеќе се обидува да ги минимизира грешките, но може да доведе до пренагласено прилагодување.

3) `epsilon = 0.1` - Дефинира зона на толеранција (ϵ -интервал) околу вистинските вредности, во која грешките не се земаат предвид. Ова овозможува моделот да биде поотпорен на шум во податоците.

Дополнително, моделот е вклучен во кодот заедно со StandardScaler, со што предиктор варијаблите се стандардизираат (средна вредност 0 и стандардна девијација 1) во рамки на секој сет од петкратна вкрстена валидација. Ова е особено важно за SVR, бидејќи моделот е чувствителен на размерот на влезните податоци.

За евалуација на моделот е применета петкратна вкрстена валидација, при што податоците се делат на пет множества, а моделот се обучува и тестира пет пати со различни комбинации на тренинг и тест податоци. Перформансите се оценуваат преку метриците MSE, RMSE, MAE и R^2 , што овозможува сеопфатна анализа на точноста и стабилноста на моделот, прикажани во Табела 63.

Табела 63. Табеларен приказ на метричките показатели од секој сет на петкратната вкрстена валидација во рамки на обучувањето на моделот со SVR, за анализа на случај CB2=f(Zwl,T).

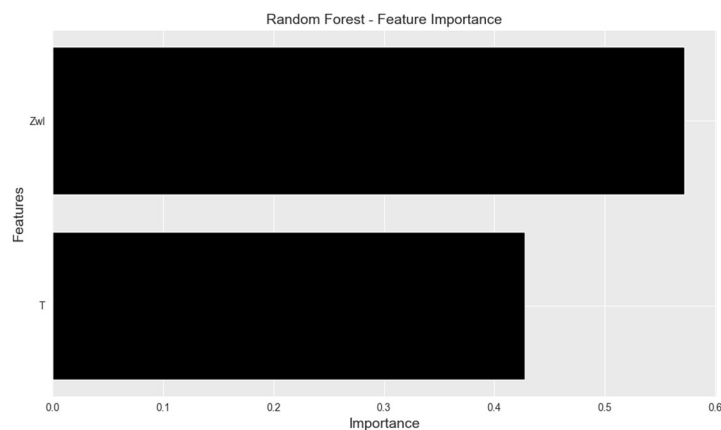
Fold	MSE	RMSE	MAE	R2
1	39.212	6.262	4.619	0.560
2	34.541	5.877	4.213	0.556
3	38.517	6.206	4.711	0.574
4	43.473	6.593	4.719	0.566
5	37.174	6.097	4.427	0.593

Средната вредност на метричките показатели по обучувањето на моделот изнесуваат: Avg MSE = 38.583, Avg RMSE = 6.207, Avg MAE = 4.537, Avg R² = 0.569. Процесот на обучување траеше 71ms.

7.7.4. Резултати по обучувањето на моделите

По обучување на три различни модели базирани на три различни алгоритми, како што следи презентирани се сумарни резултати од сите модели за претходната фаза во форма на табели и графици.

За моделот обучуван со RF, DamResML кодот е програмиран да прикаже график на важност на варијаблите (анг. „Feature importance“) во моделот. Важноста на варијаблите во моделот покажува колкаво е влијанието на предиктор варијаблите во предвидувањето на целната варијабла.

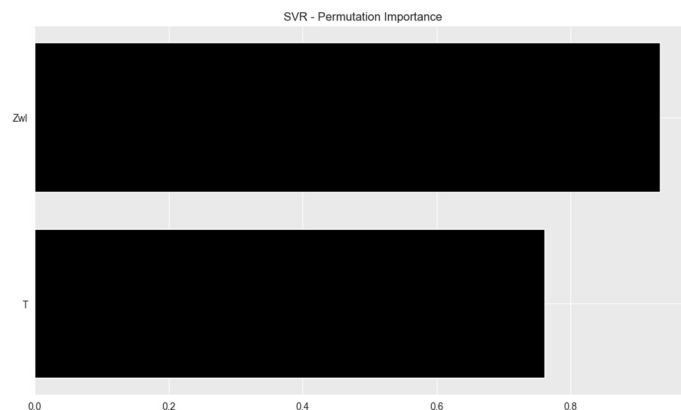


Прилог 123. График на важност на варијаблите (анг. „Feature importance“) во моделот со алгоритмот RF.

За моделот обучуван со RF, овој график е прикажан во Прилог 123. На графикот се забележува дека двата предиктор варијабли „Zwl“ и „T“ имаат блиско значење во предвидувањето на целната варијабла „CB2“, но сепак „Zwl“ – нивото на водата во акумулацијата се издвојува како варијабла со поголемо значење со вредност 0.57 наспроти значењето на „T“ – амбиенталната температура чие значење, согласно резултатите, е 0.42.

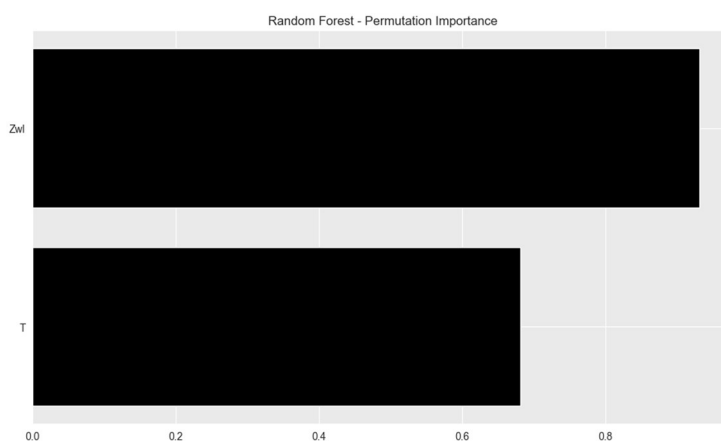
Дополнително, во DamResML кодот програмирано е извлекувањето на пермутациската важност (анг. „Permutation importance“) на моделите обучувани со SVR и RF. Пермутациската важност претставува метод за оценување на значењето на предиктор варијаблите врз основа на нивното влијание врз перформансите на моделот. Основната идеја се состои во тоа што вредностите на една променлива случајно се мешаат (пермутираат), со што се нарушува нејзината врска со целната варијабла. Потоа се мери промената во точноста на моделот. Доколку пермутацијата на одредена променлива доведе до значително зголемување на грешката (на пример RMSE), тоа укажува дека моделот во голема мера се потпира на таа променлива. Напротив, ако промената е мала, тогаш нејзиното влијание е ограничено.

На графициите во Прилог 124 и Прилог 125 прикажани се графици со пермутациска важност за моделите обучени со SVR и RF алгоритмите.



Прилог 124. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со алгоритмот SVR.

Сличен тренд на предиктор варијаблите се провлекува во сите двата графика на пермутациска важност – поголема пермутациска важност има предиктор варијаблата „Zwl“ со вредност над 0.9 додека предиктор варијаблата „T“ има помала важност со вредност 0.78 за моделот трениран со SVR и 0.70 за моделот трениран со RF.



Прилог 125. Пермутациска важност (анг. „Permutation importance“) за моделот обучуван со алгоритмот RF.

За секој обучуван модел, во DamResML кодот програмирано е да се прикаже график на резидуали (остатоци) – грешки во предвидувањата од секој модел. Сумарен график кој ги опфаќа сите 9 модели е прикажан во Прилог 126. Резидуалите се пресметуваат врз основа на кратко предвидување со обучените модели, спроведено на т.н. задржан сет на податоци (анг. „holdout“) кој не бил претходно вклучен во процесот на обучување на моделите, па така истиот е непознат за моделот. Ова издвојување на задржан сет настрана од сетот за обучување кодот го прави уште во првата ќелија од моделот.

Прикажаните графици ја претставуваат резидуалната анализа за три различни модели на машинско учење. Резидуалите ја претставуваат разликата помеѓу реалните и предвидените вредности, односно грешката на моделот. Во идеален случај, точките треба да бидат случајно распоредени околу хоризонталната нулта линија, без јасен тренд, со мала вертикална дисперзија. Таквата појава укажува дека моделот е стабилен, непристрасен и успешно ја опишува зависноста во податоците.

Во конкретната анализа, резидуалните дијаграми за моделите SVR, Random Forest и Boosted Trees покажуваат слично однесување и овозможуваат дополнителна проценка на квалитетот на моделите надвор од класичните метрички показатели. Кај сите три модели, најголем дел од резидуалите се распределени околу нултата линија, што укажува дека моделите генерално успеваат да ја репродуцираат зависноста помеѓу целната варијабла „CB2“ и предиктор варијаблите „Zwl“ и „T“ без изразена систематска пристрасност.

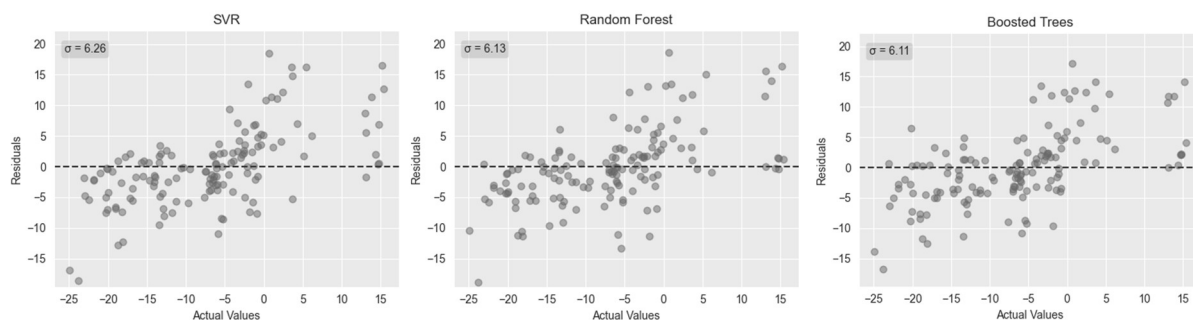
Кај моделот SVR може да се забележи умерена дисперзија на резидуалите околу нултата

вредност, при што дел од резидуалите кај повисоките позитивни вредности на целната варијабла достигнуваат и над ± 15 mm. Притоа, се забележува одредена тенденција на потценување на поголемите позитивни вредности на „CB2“, бидејќи кај повисоките реални вредности резидуалите почесто се позитивни. Ова укажува дека моделот има ограничена способност целосно да ги опфати екстремните варијации на радијалните поместувања. Стандардната девијација на резидуалите $\sigma = 6.26$ mm дополнително укажува на умерено ниво на расејување на грешките.

Кај Random Forest моделот распределбата на резидуалите е нешто покомпактна околу нултата линија, со $\sigma = 6.13$ mm, што укажува на најстабилни резултати од анализираниите модели. Резидуалите не покажуваат изразен тренд или систематска зависност од реалните вредности, што претставува позитивен показател за способноста на моделот за генерализација. Иако се присутни неколку отстапувања кај екстремните негативни и позитивни вредности на „CB2“, тие се релативно ограничени и не формираат јасна шема. Ова укажува дека Random Forest моделот успешно ја опишува нелинеарната зависност помеѓу влезните и излезната варијабла.

Кај моделот Boosted Trees резидуалите исто така се распределени околу нултата линија, со најниска стандардна девијација $\sigma = 6.11$ mm. Сепак, визуелно може да се забележи нешто поголема концентрација на позитивни резидуали кај повисоките реални вредности на „CB2“, што укажува на благо потценување на повисоките радијални поместувања. И покрај тоа, моделот покажува добра стабилност и способност за следење на трендот на податоците.

Сите три модели покажуваат прифатлива распределба на резидуалите без силно изразени нелинеарни шеми или систематски отклонувања, што укажува дека моделите успешно ја опишуваат основната зависност во податоците. Притоа, Random Forest и Boosted Trees покажуваат нешто подобра стабилност и помало расејување на грешките во однос на SVR, што е во согласност и со добиените метрички показатели од петкратната вкрстена валидација.



Прилог 126. Графици со резидуали за трите обучени модели.

7.7.5. Предвидување на целната варијабла

По завршување на процесот на обука на три модели со избраните три алгоритми за машинско учење, се извршува пресметка на предвидувања на целната варијабла „CB2“ – радијално поместување во централниот блок на анализираната лачна брана во Франција, во функција од предиктор варијаблите „Zwl“ – ниво на вода во акумулацијата и „T“ – амбиентална температура.

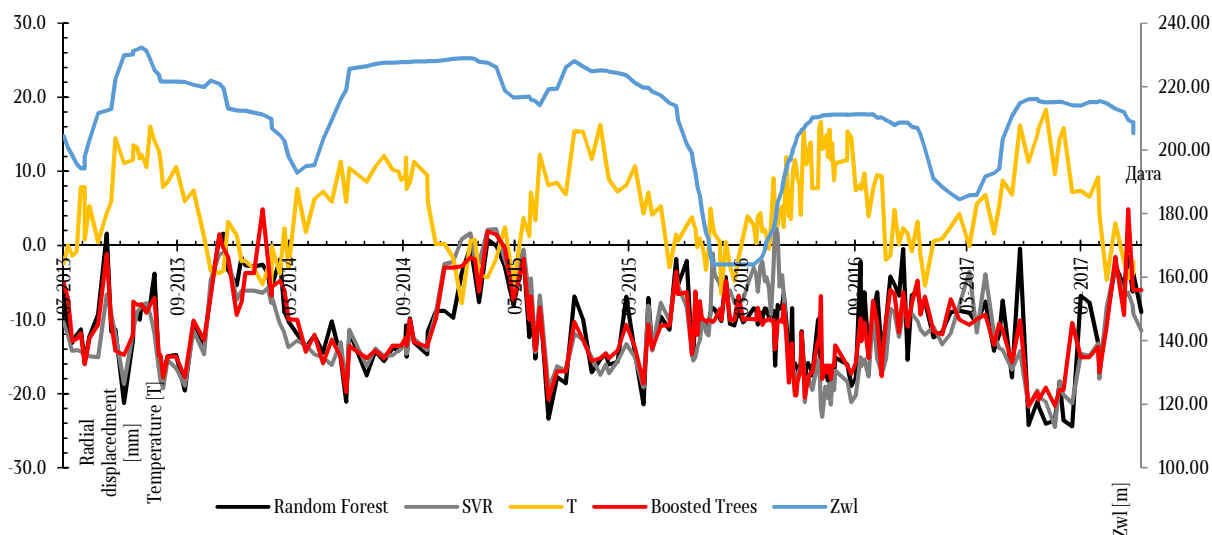
Во процесот на предвидување се користат податоци за предиктор варијаблите, и тоа вкупно 226 вредности кои одговараат на временскиот период помеѓу 27.01.2013 година па сè до 18.12.2017 година, регистрирани во неправилни временски интервали, на дневно ниво. За овој период, целната варијабла „CB2“ е непозната. Предиктор варијаблите се прикажани графички во Прилог 122.

Резултатите од предвидувањето се дадени како што следи.

На графикот во Прилог 127 може да се согледа дека предвидувањата од трите обучени модели се меѓусебно доста блиски, со вредности кои се движат од $CB2_{\min} = -24.517$ mm пресметано со SVR моделот, до $CB2_{\max} = 4.87$ mm пресметано со BT моделот (Табела 64).

Табела 64. Пресметани максимални и минимални радијални поместувања „CB2“ со секој од трите обучени модели со машинско учење за $CB2=f(Zwl,T)$.

	SVR	Random Forest	Boosted Trees
min	-24.517	-24.431	-21.652
max	3.250	1.596	4.870



Прилог 127. Графички приказ на пресметаните предвидувања на целната варијабла „CB2“ во функција од предиктор варијаблите „Zwl“ и „T“.

Со оглед дека зголемувањето на вредностите на радијалните поместувања означува поместување во низводен правец, може да се забележи дека повисоките вредности на нивото на водата во акумулацијата „Zwl“ генерално се поврзани со пораст на вредностите на „CB2“, односно со поизразено низводно поместување на централниот блок на браната. Ова е во согласност со физичкото однесување на лачните брани, каде што зголемениот хидростатички притисок предизвикува поместување на конструкцијата во низводен правец.

Од друга страна, влијанието на амбиенталната температура „T“ покажува спротивен ефект. Во периодите со повисоки температури, предвидените вредности на „CB2“ имаат тенденција на намалување, односно поместувањата се движат кон негативни вредности. Ова укажува на тенденција за возводно поместување како резултат на термичката експанзија на лачната конструкција. При загревање, бетонското тело на браната се шири, при што лачната геометрија предизвикува централниот дел да се придвижува кон возводната страна. Поради тоа, температурното влијание делумно го компензира низводното поместување предизвикано од хидростатичкиот притисок.

Предвидените временски серии покажуваат дека радијалните поместувања се резултат на комбинирано и нелинеарно дејство на нивото на водата и температурата. Во периодите кога нивото на водата е високо, а температурите се ниски, моделите предвидуваат највисоки вредности на „CB2“, што укажува на најизразено низводно поместување. Наспроти тоа, во периодите со повисоки температури, вредностите на „CB2“ се намалуваат, што укажува на намалување на низводното поместување или делумно придвижување во возводен правец.

Моделите со алгоритмите RF и BT покажуваат подинамични и поосетливи предвидувања со поизразени локални варијации, додека SVR обезбедува помазна временска серија со помалку нагласени екстремни вредности. Сепак, сите три модели покажуваат слична општа динамика и сезонско однесување, што претставува дополнителна потврда дека моделите успешно ја препознаваат физичката зависност помеѓу радијалните поместувања, нивото на водата и амбиенталната температура.

7.6. Сублимат од спроведената студија на случај 2

7.6.1. Согледувања од добиените метрички показатели од студија на случај 2

Со цел јасен приказ на метричките показатели на најдобрите обучени модели кои се коментирани и користени во анализите на случај дел од студија на случај 2, направен е табеларен приказ на метричките показатели на најдобрите модели со машинско учење, даден во Табела 65.

Табела 65. Табеларен приказ на метричките показатели на најдобрите модели во анализите на случај дел од студија на случај 2.

Ред. Бр.	Студија на случај	Конфигурација	Брана	ML	MSE	RMSE	MAE	R2
1	2.1	FP_14=f(Zwl)	Кнежево	RF	0.44	0.65	0.52	0.658
2		FP_15=f(Zwl)	Кнежево	BT	3.87	1.93	1.35	0.817
3	2.2	V_notch_left=f(Zwl, Q_inflow, FP_14, FP_15)	Кнежево	KNN	0.02	0.12	0.05	-4.670
4	2.3	FP_2=f(Zwl)	Кнежево	BT	0.85	0.86	0.59	-0.047
5		FP_3=f(Zwl)	Кнежево	GRNN	2.40	1.53	1.26	0.890
6	2.4	V_notch_right=f(Zwl, Q_inflow, FP_2, FP_3)	Кнежево	GRNN	25.98	4.73	2.68	-0.560
7	2.5	V_notch_middle=f(Zwl, Q_inflow)	Кнежево	RF	376.18	19.13	13.63	-0.020
8	2.6	SC1=f(Zwl,t)	Кнежево	RF	0.46	0.62	0.41	0.856
9		SC2=f(Zwl,t)	qarq	BT	0.10	0.31	0.24	0.982
10		SC4=f(Zwl,t)	Кнежево	BT	0.13	0.36	0.28	0.909
11		SC5=f(Zwl,t)	Кнежево	RF	0.88	0.93	0.66	0.963
12		SC6=f(Zwl,t)	Кнежево	BT	0.32	0.52	0.31	0.929
13		SC7=f(Zwl,t)	Кнежево	BT	0.76	0.85	0.63	0.900
14		SC9=f(Zwl,t)	Кнежево	RF	0.25	0.49	0.35	0.996
15		SC12=f(Zwl,t)	Кнежево	RF	1.61	1.25	0.79	0.851
16		SC13=f(Zwl,t)	Кнежево	RF	0.73	0.84	0.54	0.963
17		SC14=f(Zwl,t)	Кнежево	BT	0.18	0.41	0.29	0.907
18	2.7	CB2=f(Zwl,t)	EDF Франција	RF	36.66	6.05	4.49	0.590

Табела 65 дава сумиран преглед на најдобрите модели добиени за сите анализирани студии на случај во рамките на втората група истражувања, која се однесува на мониторингот на однесувањето на браните преку предвидување на поместувања, филтрациски протоци и слегнувања. Анализата покажува дека точноста на моделите значително зависи од природата на целната варијабла и достапните влезни податоци. Најдобри резултати се постигнати кај предвидувањето на слегнувањата (студија на случај 2.6), каде што за повеќето мерни точки се добиени високи вредности на коефициентот на детерминација, кои се движат од $R^2 = 0.851$ до $R^2 = 0.982$. Овие резултати укажуваат дека применетите модели успешно ја опишуваат зависноста помеѓу нивото на водата во акумулацијата, температурата и регистрираните слегнувања, при што најчесто како најдобар алгоритам се издвојува BT, а кај дел од мерните точки и RF.

Добри резултати се добиени и кај предвидувањето на пиезометриските нивоа „FP_14“ и „FP_3“, каде што се постигнати вредности $R^2 = 0.658$ и $R^2 = 0.890$, соодветно, што потврдува дека овие процеси можат успешно да се моделираат со примена на алгоритми на машинско учење. Од друга страна, предвидувањето на пиезометарот „FP_2“ покажува значително послаби резултати ($R^2 = -0.047$), што укажува дека зависноста помеѓу избраните предикторски варијабли и целната величина не е доволно изразена.

Најголеми ограничувања се забележуваат кај анализите на филтрациските протоци

„V_notch_left“, „V_notch_right“ и „V_notch_middle“, каде што сите модели постигнуваат негативни вредности на коефициентот R^2 . Ваквите резултати укажуваат дека расположливите податоци не обезбедуваат доволно информации за сигурно моделирање на овие процеси, што најверојатно е последица на малиот број мерења, ограничената варијабилност на податоците и комплексната природа на филтрацијата низ телото и темелот на браната. Слично, кај анализата на радијалното поместување на браната CB2 е постигната умерена точност ($R^2 = 0.590$), што покажува дека моделот успева да ја опише општата тенденција на однесувањето, но не и целосно сложената интеракција на факторите што влијаат врз деформациите.

Генерално, резултатите покажуваат дека методите на машинско учење се особено ефикасни за моделирање на процеси со јасно изразени и стабилни функционални зависимости, како што се слегнувањата и дел од пиезометриските мерења. Од друга страна, кај процесите на филтрација нивната успешност е ограничена од квалитетот и обемот на расположливите податоци, што укажува дека за овие случаи е потребно користење на дополнителни предикторски варијабли, подолги временски серии или ансамбл пристапи за подобрување на предиктивната способност.

7.6.2. Коментар на добиените резултати од студија на случај 2

Во рамки на Поглавје 7 беше извршена детална анализа на можностите за примена на модели со машинско учење при анализа на однесувањето на брани, со посебен акцент на предвидување на филтрациони протекувања кај насипни брани, пиезометриски нивоа, слегнувања и радијални поместувања кај различни типови на брани. Анализите беа спроведени преку седум засебни анализи на случај, при што беа вклучени податоци од различни брани, повеќе видови мерени параметри и беа разгледувани девет алгоритми од областа на машинското учење.

Добиените резултати покажаа дека моделите со машинско учење имаат значителен потенцијал за примена во системите за мониторинг и оцена на безбедноста на браните, особено во ситуации кога постојат нецелосни временски серии, ограничен број мерења или потреба од долгорочни прогнози на однесувањето на конструкцијата. Анализите потврдија дека, и покрај релативно малиот број достапни податоци кај одредени сетови, со правилен избор на алгоритам може да се постигнат задоволителни резултати.

Во анализите поврзани со пиезометрите „FP_14“, „FP_15“, „FP_2“ и „FP_3“ на брана „Кнежево“ беше утврдено дека постои јасна зависност помеѓу нивото на вода во акумулацијата и пиезометриските нивоа во фундаментот под браната. Особено добри резултати беа постигнати кај моделите RF, GRNN и BT, кои покажаа способност успешно да ги моделираат нелинеарните зависимости помеѓу променливите. Од друга страна, кај ANN моделите во повеќе анализи се јавија исклучително лоши резултати и нестабилни предвидувања, што укажува дека невронските мрежи не се најсоодветно решение за релативно мали сетови на податоци како што беа користени во ова истражување. Особено значајно е што кај пиезометрите поставени на возводната страна од асфалтбетонското јадро на брана „Кнежево“ се добија значително подобри метрички показатели, што е логично од инженерски аспект бидејќи тие директно го следат водното ниво во акумулацијата.

Кај анализите за предвидување на филтрационите текови преку дренажните колектори се потврди дека задачата за предвидување на процедурните води е значително посложена. Во повеќе случаи се добија негативни вредности на коефициентот R^2 , што укажува на ограничена способност на моделите за генерализација. Сепак, деталната анализа покажа дека ваквите резултати не мора директно да значат дека моделите се неупотребливи, туку дека самиот параметар R^2 не е доволно репрезентативен кога целната варијабла има многу мала варијабилност и тесен опсег на промени. Во тие случаи значително порелевантни показатели беа MSE, RMSE и MAE. Резултатите покажаа дека алгоритмите базирани на локална сличност, како GRNN, најуспешно ги следат малите варијации кај филтрационите текови и обезбедуваат најстабилни предвидувања.

Особено значаен придонес на истражувањето претставува примената на предвидените вредности за дефинирање на граници за предупредување (анг. „thresholds“) врз основа на статистичка анализа на предвидените целни варијабли. Со примена на стандардната девијација

од предвидените вредности беа дефинирани две нивоа на предупредување, што овозможува воспоставување на систем за рана детекција на потенцијално опасни промени во филтрациониот режим. Ваквиот пристап претставува значајна практична примена на машинското учење во системите за мониторинг на брани, бидејќи овозможува автоматизирано следење на однесувањето на конструкцијата и навремена идентификација на потенцијални аномалии.

Во анализата на вертикалните слегнувања кај насипната брана „Кнежево“ беше потврдено дека моделите со машинско учење можат успешно да се применат за проширување на временските серии на мерења и проценка на долгорочното стабилизирање на слегнувањата. Со екстраполација на предвидените вредности беше овозможена проценка на времето потребно за достигнување на 70%, 80% и 90% од теоретските пресметани финални слегнувања. Ваквиот пристап има особено значење кај брани со ограничени мерења и претставува корисна алатка за проценка на долгорочната стабилност на насипните конструкции.

Во последната анализа на случај, поврзана со радијалните поместувања на бетонска лачна брана во Франција, беше потврдено дека машинското учење успешно може да се примени и кај комплексни бетонски конструкции. Анализата покажа дека водното ниво во акумулацијата и амбиенталната температура имаат значително влијание врз радијалните поместувања на браната. Притоа беше потврдено дека зголемувањето на температурата доведува до термичка експанзија на браната и поместувања во возводна насока. Намалувањето на амбиенталната температура влијае кон поместувања во низводна насока. Поместувањата во низводна насока се предизвикани и од зголемени хидростатски притисоци претставени во моделите преку ниво на вода во акумулација, кои согласно истражената важност на варијаблите претставува доминантна предиктор варијабла. Добиените резултати покажаа добра усогласеност со физичкото однесување на лачните бетонски брани, што претставува дополнителна потврда за инженерската оправданост на применетите модели.

Генерално, истражувањето покажа дека не постои универзален модел кој е најдобар за сите анализирани проблеми. Изборот на оптимален алгоритам директно зависи од типот на податоците, бројот на мерења, варијабилноста на целната променлива и физичката природа на проблемот. Алгоритмите BT, RF и GRNN во најголем број случаи покажаа најдобар баланс помеѓу точност, стабилност и способност за моделирање на нелинеарни зависности.

Како финален заклучок може да се констатира дека моделите со машинско учење претставуваат моќна и перспективна алатка за анализа, прогноза и мониторинг на однесувањето на различни типови на брани. Нивната примена овозможува дополнување на нецелосни временски серии, идентификација на сложени нелинеарни зависности, предвидување на однесувањето на конструкцијата и воспоставување на интелигентни системи за рано предупредување. Ваквиот пристап претставува значаен чекор кон модернизација и дигитализација на системите за мониторинг на брани, со потенцијал за значително подобрување на нивната безбедност и стабилност во иднина.

“The important thing is not to stop questioning. Curiosity has its own reason for existing.”
Albert Einstein

8. ЗАКЛУЧОЦИ И ПРЕПОРАКИ ЗА ПОНАТАМОШНИ ИСТРАЖУВАЊА

Современото инженерство во областа на водните ресурси и инженерството за брани се соочува со сè покомплексни предизвици поврзани со управувањето со водостопанските системи, сигурноста на браните, климатските промени, растечките потреби за вода и енергија, како и потребата од континуирано следење и анализа на голем број мерени податоци. Во такви услови, примената на напредни методи од областа на вештачката интелигенција и машинското учење претставува логичен и неопходен чекор кон модернизација на инженерските методологии и процесите на анализа и управување.

Во рамки на оваа докторска дисертација беше извршено истражување насочено кон примена на алгоритми од машинско учење во два клучни сегменти од инженерството за брани: управување со водостопански системи и анализа на долготрајното однесување на брани врз основа на податоци од техничко набљудување. Истражувањето беше реализирано преку развој на сопствен програмски код во Пајтон средина именуван како DamResML, во кој беа имплементирани девет различни алгоритми од машинско учење, со можност за обработка на податоци, автоматизирана анализа, вкрстена валидација, споредба на модели, предвидување и визуелизација на резултати.

8.1. Општи заклучоци

Докторската дисертација претставува сеопфатно истражување на можностите за примена на методите од машинско учење во управувањето со водостопанските системи и анализата на однесувањето на брани. Преку две главни студии на случај, кои опфатија повеќе различни инженерски проблеми, беше развиена и верификувана единствена методологија за примена на машинското учење во хидротехничкото инженерство. Во продолжение се дадени главните заклучоци и научни согледувања произлезени од спроведеното истражување.

1. Основното сознание произлезено од спроведеното истражување е дека машинското учење не треба да се разгледува како замена за традиционалните инженерски методи, туку како нивно логично дополнување. Неговата примена овозможува целосно искористување на податоците од хидросистемите и системите за техничко набљудување на брани, подобро

- разбирање на сложените меѓузависности помеѓу анализираните параметри и обезбедување дополнителна поддршка при донесувањето инженерски одлуки.
2. Спроведеното истражување потврди дека методите од машинско учење успешно можат да се применат во две различни, но меѓусебно поврзани области на хидротехничкото инженерство – управувањето со водостопанските системи и анализата на однесувањето на брани. На тој начин беше покажано дека истата методологија може успешно да се адаптира за решавање на различни инженерски задачи, почнувајќи од управување со акумулации, па сè до анализа на податоците од системите за техничко набљудување.
 3. Во рамки на дисертацијата беше развиена сопствена методологија и програмски код во програмската околина Пајтон, со кој се овозможува интегриран процес на обработка на податоците, обучување и споредба на повеќе алгоритми од машинско учење, нивна валидација, инженерска интерпретација на резултатите и примена за предвидување и оптимизација. Ваквиот пристап претставува практично применлива и флексибилна основа за понатамошен развој на интелигентни инженерски алатки.
 4. Еден од најзначајните заклучоци од спроведеното истражување е дека не постои универзално најдобар алгоритам од машинско учење кој обезбедува најдобри резултати за сите анализирани инженерски проблеми. Изборот на најсоодветниот модел зависи од карактеристиките на расположливите податоци, нивниот обем, варијабилност и сложеноста на процесот што се моделира. Оттука, споредбената анализа на повеќе алгоритми претставува неопходен чекор при нивната инженерска примена.
 5. Истражувањето покажа дека квалитетот, обемот и репрезентативноста на расположливите податоци имаат пресудно влијание врз успешноста на моделите од машинско учење. Анализите потврдија дека поголеми и покомплетни множества на податоци овозможуваат подобра генерализација на моделите, додека ограничениот број мерења и малата варијабилност на целната варијабла претставуваат едни од најзначајните ограничувања при нивната примена.

8.2. Заклучоци од студија на случај 1 – примена на машинско учење во управување со водостопански системи

Во рамки на првата студија на случај беше истражена применливоста на методите од машинско учење при моделирање и анализа на процесите поврзани со управувањето на водостопанските системи. Истражувањето беше спроведено преку три засебни анализи, во кои беа разгледани различни аспекти од работењето на хидросистемите: предвидување на испуштањата од акумулацијата, предвидување на генерирана моќност во хидроцентрала и предвидување на водното ниво во акумулацијата. За секоја анализа беше развиена единствена методологија која опфаќаше обработка на податоците, обучување и споредба на девет алгоритми од машинско учење, избор на најуспешните модели, анализа на нивната точност преку повеќе метрички показатели, како и инженерска интерпретација на добиените резултати.

Покрај предвидувањето на целните варијабли, во рамки на сите анализи беше спроведена и анализа на важноста на предиктор варијаблите, со што беше овозможено подобро разбирање на влијанието на поединечните параметри врз анализираните процеси. Дополнително, најуспешните модели беа интегрирани со алгоритам за Хаос оптимизација, со што беше демонстрирана можноста машинското учење да се примени не само како алатка за предвидување, туку и како интелигентен систем за поддршка при донесување инженерски одлуки во управувањето со водостопанските системи.

Во продолжение се дадени главните заклучоци произлезени од секоја поединечна анализа, како и општите заклучоци кои ги сумираат најзначајните научни согледувања од првата студија на случај.

8.2.1. Научни согледувања од нализа на случај 1.1

1. Извршената анализа потврди дека методите од машинско учење успешно можат да ја моделираат сложената нелинеарна зависност помеѓу водното ниво во акумулацијата, волуменот на акумулираната вода, дотекот и испуштањето од акумулацијата. Со тоа се потврдува нивната применливост како дополнителна инженерска алатка при анализа и управување со водостопанските системи.
2. Споредбената анализа на девет алгоритми од машинско учење покажа дека не постои универзално најдобар модел за сите инженерски проблеми, туку изборот на алгоритмот зависи од природата на податоците, нивната варијабилност и комплексноста на зависностите помеѓу анализираните параметри.
3. Најдобри резултати беа постигнати со ансамбл алгоритмите Random Forest и Boosted Trees, кои покажаа најголема способност за генерализација и стабилност на резултатите. Ова укажува дека ансамбл методите претставуваат особено погоден избор за моделирање на сложени процеси во водостопанските системи.
4. Анализата на важноста на предиктор варијаблите потврди дека машинското учење не служи единствено за предвидување, туку претставува и корисна алатка за инженерска интерпретација на процесите, овозможувајќи идентификација на параметрите кои имаат најголемо влијание врз испуштањето од акумулацијата.
5. Примената на петкратна и вгнездена вкрстена валидација овозможи објективна процена на перформансите на моделите и значително го намали ризикот од претренирање, со што беше обезбедена поголема доверливост на добиените резултати.
6. Интеграцијата на моделите од машинско учење со алгоритмот за Хаос оптимизација покажа дека развиената методологија не е ограничена само на предвидување на идни вредности, туку може успешно да се примени и за определување на оптимални експлоатациони сценарија, што претставува значајна практична придобивка при донесување инженерски одлуки во процесот на управување со акумулациите.

8.2.2. Научни согледувања од нализа на случај 1.2

1. Резултатите од анализата покажаа дека методите од машинско учење успешно ја моделираат зависноста помеѓу генерираната моќност и анализираните хидролошки и експлоатациони параметри, при што беше постигната висока точност на предвидувањето. Ова ја потврдува нивната применливост при анализа и планирање на работата на хидроенергетските системи.
2. Споредбата на девет различни алгоритми покажа дека повеќе модели постигнуваат високо ниво на точност, што укажува дека при добро дефинирани и квалитетни податоци машинското учење претставува сигурен пристап за моделирање на процесите на производство на електрична енергија.
3. Анализата покажа дека количината на енергетски обработениот проток, дотекот, волуменот на акумулацијата и водното ниво претставуваат меѓусебно поврзан систем на параметри чии нелинеарни интеракции можат успешно да се опфатат со современите алгоритми од машинско учење, без потреба од експлицитно математичко дефинирање на нивната функционална зависност.
4. Анализата на важноста на предиктор варијаблите овозможи подобро инженерско разбирање на влијанието на поединечните параметри врз производството на електрична енергија, што претставува значајна предност на предложената методологија при поддршка на процесите на управување и експлоатација на хидросистемите.
5. Примената на алгоритмот за Хаос оптимизација, во комбинација со најуспешните модели од машинско учење, покажа дека развиената методологија може успешно да се користи за определување на оптимални работни состојби со цел максимизирање на производството на електрична енергија и генерираната моќност, при истовремено почитување на зададените експлоатациони ограничувања.

6. Добиените резултати потврдуваат дека интеграцијата на методите од машинско учење и оптимизациските алгоритми претставува современ пристап за интелигентно управување со водостопанските системи, со потенцијал за примена како систем за поддршка при донесување оперативни одлуки во реални услови.

8.2.3. Научни согледувања од нализа на случај 1.3

1. Извршената анализа потврди дека методите од машинско учење успешно можат да ја моделираат динамиката на промените на водното ниво во акумулацијата, земајќи ги предвид истовремените влијанија на хидролошките и експлоатационите параметри. Со тоа се потврдува нивната применливост при следење и планирање на режимот на работа на акумулациите.
2. Резултатите покажаа дека водното ниво во акумулацијата претставува комплексна целна варијабла, чија вредност е резултат на заемното дејство на природните хидролошки процеси и управувачките активности во системот. Машинското учење овозможи успешно моделирање на овие нелинеарни меѓузависности без потреба од нивно експлицитно математичко дефинирање.
3. Споредбата на алгоритмите покажа дека повеќе модели постигнуваат висока точност при предвидувањето на водното ниво, што укажува дека современите методи од машинско учење можат да обезбедат сигурна поддршка при оперативното управување со акумулациите и планирањето на нивниот режим на експлоатација.
4. Анализата на важноста на предиктор варијаблите овозможи квантифицирање на влијанието на поединечните влезни параметри врз промените на водното ниво, со што предложената методологија добива дополнителна инженерска вредност преку подобро разбирање на процесите кои го определуваат режимот на работа на акумулацијата.
5. Примената на алгоритмот за Хаос оптимизација покажа дека обучените модели можат успешно да се користат не само за предвидување на идните состојби, туку и за определување на оптимални комбинации на управувачките параметри, што создава основа за развој на интелигентни системи за поддршка при донесување одлуки во управувањето со водостопанските системи.
6. Добиените резултати потврдуваат дека развиената методологија има потенцијал за практична примена при секојдневното управување со акумулациите, бидејќи овозможува навремено предвидување на промените на водното ниво, подобро планирање на експлоатацијата и намалување на неизвесноста при донесување оперативни инженерски одлуки.

8.2.4. Заклучоци од студија на случај 1

1. Студијата на случај потврди дека методите од машинско учење успешно можат да се применат за моделирање на различни процеси во водостопанските системи, вклучувајќи предвидување на испуштањата од акумулацијата, производството на електрична енергија и промените на водното ниво. Со тоа се потврдува нивната широка применливост како дополнителна инженерска алатка при управувањето со сложени хидросистеми.
2. Споредбената анализа на девет алгоритми од машинско учење покажа дека не постои единствен алгоритам кој обезбедува најдобри резултати за сите анализирани инженерски проблеми. Изборот на најсоодветниот модел зависи од карактеристиките на податоците, структурата на зависностите помеѓу варијаблите и природата на анализираниот процес, што ја нагласува потребата од споредбена анализа пред практична имплементација.
3. Развиениот програмски код во програмската околина Пајтон овозможи воспоставување на единствена методологија за обработка на податоците, обучување, валидација, споредба и примена на повеќе алгоритми од машинско учење. Ваквиот пристап обезбедува значително поголема флексибилност, транспарентност и можност за понатамошно надградување во споредба со користењето на поединечни комерцијални

- софтверски решенија.
4. Анализата на важноста на предиктор варијаблите покажа дека методите од машинско учење, покрај високата точност на предвидувањето, претставуваат и корисна алатка за подобро инженерско разбирање на процесите во водостопанските системи, овозможувајќи идентификација и квантификација на влијанието на поединечните параметри врз анализираните појави.
 5. Интеграцијата на моделите од машинско учење со алгоритмот за Хаос оптимизација ја прошири примената на развиената методологија од класично предвидување кон интелигентна оптимизација на режимот на работа на хидросистемите. На овој начин беше покажано дека машинското учење може успешно да се користи како составен дел од системите за поддршка при донесување инженерски одлуки.
 6. Врз основа на сите извршени анализи може да се заклучи дека предложената методологија претставува современ, флексибилен и практично применлив пристап за поддршка на процесите на управување со водостопанските системи. Нејзината примена овозможува подобро искористување на расположливите податоци, повисока сигурност при предвидувањето и создава основа за развој на интелигентни системи за управување со водните ресурси.

8.3. Заклучоци од студија на случај 2 – примена на машинско учење во анализа на однесувањето на брани

Во втората студија на случај беше анализирана применливоста на методите од машинско учење во процесот на техничко набљудување и анализа на однесувањето на различни типови брани. Истражувањето беше насочено кон моделирање на повеќе параметри кои се следат во рамките на системите за оскултација, со цел да се испита можноста за нивно сигурно предвидување, подобро разбирање на нивните меѓусебни зависности и навремено препознавање на потенцијални отстапувања во однесувањето на браните.

Во рамки на студијата беа анализирани повеќе инженерски проблеми, кои опфатија предвидување на водните нивоа во пиезометрите, филтрационите количества, вертикалните слегнувања и радијалните поместувања на брани. За сите анализирани случаи беше применета единствена методологија која опфаќаше обработка и анализа на реални податоци од системите за техничко набљудување, обучување и споредба на девет алгоритми од машинско учење, процена на нивната точност преку повеќе метрички показатели, анализа на важноста на предиктор варијаблите и инженерска интерпретација на добиените резултати.

Особено внимание беше посветено на оценување на практичната применливост на машинското учење како дополнителна инженерска алатка за поддршка на процесот на техничко набљудување на брани. Покрај предвидувањето на анализираните параметри, добиените модели овозможија подобро разбирање на факторите кои влијаат врз однесувањето на браните, идентификација на нивните меѓусебни зависности, како и воспоставување на основа за идно развивање на интелигентни системи за следење и рано предупредување.

Во продолжение се прикажани главните заклучоци од секоја поединечна анализа, како и општите заклучоци кои ги обединуваат најзначајните научни согледувања од втората студија на случај.

8.3.1. Научни согледувања од анализа на случај 2.1

1. Спроведеното истражување потврди дека методите од машинско учење успешно можат да ја моделираат зависноста помеѓу водното ниво во акумулацијата и нивото на водата во пиезометрите FP_14 и FP_15. Добиеените резултати ја потврдуваат применливоста на предложената методологија за анализа на филтрационите процеси во фондментот на браната.
2. Анализата покажа дека и покрај користењето на една предиктор варијабла – водното

ниво во акумулацијата дополнета со изведените пресметани предиктор варијабли поврзани со водното ниво во акумулацијата, обучените модели беа способни успешно да ја опфатат основната зависност помеѓу анализираните параметри. Ова укажува дека машинското учење може да обезбеди сигурни резултати и во случаи кога расположливите влезни податоци се ограничени.

3. Споредбата на различните алгоритми покажа дека нивната успешност значително зависи од карактерот на податоците и сложеноста на анализираниот процес. Најуспешните модели обезбедија стабилни и конзистентни предвидувања, што ја потврдува потребата од споредбена анализа при изборот на алгоритам за конкретна инженерска примена.
4. Добиените резултати укажуваат дека моделите од машинско учење можат успешно да се применат за дополнување на временски низи со недостасувачки мерења, како и за предвидување на идните вредности на водните нивоа во пиезометрите, што претставува значајна практична придобивка во процесот на техничко набљудување на браната.
5. Предвидувањето на водните нивоа во пиезометрите овозможува воспоставување на референтни очекувани вредности, со кои можат да се споредуваат реално измерените податоци. На тој начин машинското учење може да претставува дополнителна алатка за навремено препознавање на потенцијални аномалии и промени во однесувањето на браната.
6. Извршената анализа потврди дека предложената методологија претставува практично применлив пристап за поддршка на интерпретацијата на податоците од техничкото набљудување, при што машинското учење не го заменува инженерското расудување, туку обезбедува дополнителни информации кои придонесуваат кон посигурна процена на состојбата на браната.

8.3.2. Научни согледувања од анализа на случај 2.2

1. Спроведената анализа покажа дека методите од машинско учење имаат ограничена примена во моделирање на филтрационите количества во левиот дренажен колектор на браната, при што беше воспоставена функционална зависност помеѓу филтрацијата и анализираните хидролошки и експлоатациони параметри.
2. Иако статистичките показатели укажаа на ограничена способност на моделите за генерализација, што пред сè се должи на малиот број достапни мерни податоци и ниската варијабилност на целната варијабла, обучените модели обезбедија корисна основа за понатамошна инженерска интерпретација на процесот на филтрација. Овој резултат покажува дека оценката на успешноста на моделите не треба да се базира исклучиво на метричките показатели, туку и на нивната практична применливост во конкретниот инженерски контекст.
3. Предложената методологија за определување на граници за предупредување врз основа на предвидените вредности и нивната стандардна девијација овозможи воспоставување на очекуван опсег на нормално однесување на филтрацијата. На тој начин машинското учење беше искористено како основа за развој на систем за рано предупредување при техничкото набљудување на браните.
4. Споредбата на реално измерените и предвидените филтрациони количества покажа дека предложениот пристап може успешно да се користи за идентификација на мерења кои отстапуваат од очекуваното однесување. Ваквиот пристап претставува дополнителна поддршка при навремено препознавање на потенцијални промени во филтрациониот режим на браната.
5. Извршеното истражување покажа дека машинското учење има значително поголема вредност од класичното предвидување на поединечни параметри. Со интегрирање на предвидените вредности со статистички дефинирани граници за предупредување се создава практична алатка која може да се примени во системите за техничко набљудување и да придонесе кон пообјективна процена на состојбата на браната.

8.3.3. Научни согледувања од анализа на случај 2.3

1. Спроведеното истражување потврди дека методите од машинско учење успешно можат да го моделираат однесувањето на водните нивоа во пиезометрите FP_2 и FP_3, при што воведувањето на овие варијабли понатаму како предиктор варијабли во следната анализа на случај овозможи попрецизно опишување на процесите на филтрација во фундаментот на браната.
2. Споредбената анализа покажа дека користењето на дополнителни информации, како што се водното ниво во акумулацијата, нивото во соседниот пиезометар и дотекот во акумулацијата, значително придонесува за подобрување на точноста на предвидувањето. Ова потврдува дека однесувањето на филтрационите процеси е резултат на заемното дејство на повеќе меѓусебно поврзани фактори.
3. Анализата покажа дека предвидувањето на водните нивоа во пиезометрите може успешно да се искористи за дополнување на временски низи со недостасувачки мерења, како и за процена на очекуваните вредности во идните периоди на набљудување. Ваквиот пристап претставува значајна практична придобивка при управувањето со податоците од техничкото набљудување.
4. Споредбата на повеќе алгоритми повторно потврди дека изборот на најсоодветен модел зависи од карактеристиките на конкретниот инженерски проблем, при што моделите со најдобри перформанси обезбедија висока точност и стабилност на предвидувањата. Ова ја нагласува потребата секоја анализа да биде проследена со објективна споредба на повеќе алгоритми, наместо однапред да се избира еден метод.

8.3.4. Научни согледувања од анализа на случај 2.4

1. Спроведената анализа покажа дека методите од машинско учење можат успешно да се применат за моделирање на филтрационите количества во десниот дренажен колектор, при што беа воспоставени функционални зависности помеѓу филтрацијата и анализираните хидролошки и експлоатациони параметри.
2. И покрај ограничениот број расположливи мерни податоци и изразената варијабилност на филтрационите количества, предложената методологија овозможи добивање на предвидени вредности кои претставуваат основа за понатамошна инженерска анализа на однесувањето на филтрациониот режим. Добиените резултати уште еднаш ја нагласуваат важноста од обезбедување подолги и поконтинуирани временски низи за обучување на моделите од машинско учење.
3. Предложениот пристап за дефинирање на граници за предупредување врз основа на статистичката обработка на предвидените вредности овозможи воспоставување на референтен опсег на очекувано однесување на филтрацијата. Со тоа се создава можност за автоматизирана споредба на реално измерените вредности со очекуваниот режим на работа на браната.
4. Анализата покажа дека отстапувањето на измерените вредности надвор од дефинираните граници за предупредување може да претставува индикатор за промена на филтрациониот режим, поради што предложената методологија има потенцијал да се примени како дополнителна алатка за рано предупредување при техничкото набљудување на брани.
5. Добиените резултати потврдуваат дека практичната вредност на машинското учење не се ограничува единствено на постигнување висока точност на предвидувањето, туку се огледува и во можноста за воспоставување на објективни критериуми за интерпретација на податоците од системите за техничко набљудување.

8.3.5. Научни согледување од анализа на случај 2.5

1. Спроведената анализа потврди дека предложената методологија може успешно да се примени и за моделирање на филтрационите количества во средниот дренажен колектор, со што беше дополнително потврдена нејзината применливост за различни

- мерни локации во рамките на истиот систем за техничко набљудување.
2. Добиените резултати покажаа дека машинското учење успешно ги идентификува основните закономерности во промената на филтрационите количества, при што точноста на моделите останува условена од квалитетот, бројот и варијабилноста на расположливите мерни податоци. Ова уште еднаш ја нагласува важноста на континуираното и систематско прибирање на податоци од системите за оскултација.
 3. Воспоставувањето на граници за предупредување врз основа на предвидените вредности овозможи дефинирање на очекуваниот режим на филтрација и создаде основа за континуирана споредба со реално измерените вредности. На тој начин предложената методологија обезбедува дополнителна поддршка при интерпретацијата на податоците од техничкото набљудување.
 4. Анализите на трите дренажни колектори (лев, десен и среден) потврдија дека предложениот пристап не е ограничен на една конкретна мерна локација, туку претставува унифицирана и практично применлива методологија за следење на филтрационите процеси кај насипни брани. Со тоа се потврдува можноста за нејзина поширока примена во системите за техничко набљудување и рано предупредување.

8.3.6. Научни согледувања од анализа на случај 2.6

1. Спроведеното истражување потврди дека методите од машинско учење успешно можат да го моделираат процесот на вертикални слегнувања кај насипните брани, при што беа успешно опфатени временските промени на слегнувањата во повеќе мерни ќелии. Со тоа се потврдува нивната применливост за анализа на долготрајното однесување на браните.
2. Анализата на десетте мерни ќелии покажа дека предложената методологија овозможува индивидуално моделирање на однесувањето на секоја мерна локација, при што беа утврдени разлики во динамиката на развојот на слегнувањата долж телото на браната. Ова овозможува подетална инженерска интерпретација на процесите на консолидација и стабилизација на насипот.
3. Добиените резултати покажаа дека машинското учење може успешно да придонесе за идентификација на атипични мерни серии. Анализата на мерната ќелија SC14 укажа на однесување кое значително се разликува од останатите мерни точки, што може да биде последица на мерна грешка, локални геотехнички услови или специфични конструктивни влијанија. Овој резултат ја потврдува можноста машинското учење да претставува дополнителна алатка за контрола на квалитетот на податоците од техничкото набљудување.
4. Екстраполацијата на пресметаните криви на слегнување овозможи процена на идниот развој на деформациите и определување на временските периоди во кои се очекува достигнување на 70%, 80% и 90% од конечните слегнувања. На овој начин беше демонстрирана можноста машинското учење да се користи не само за краткорочно предвидување, туку и за долгорочна анализа на однесувањето на браните.
5. Анализата потврди дека комбинирањето на податоците од техничкото набљудување со современите методи од машинско учење овозможува подобро разбирање на процесот на стабилизација на браната и обезбедува дополнителни информации кои можат да бидат од значење при планирањето на идните активности за следење и одржување на објектот.
6. Извршеното истражување покажа дека предложената методологија претставува современ и практично применлив пристап за анализа на долготрајното однесување на насипните брани. Покрај високата применливост за предвидување на слегнувањата, методологијата овозможува идентификација на аномалии, процена на идните трендови и обезбедување дополнителна поддршка при инженерската интерпретација на податоците од техничкото набљудување.

8.3.7. Научни согледувања од анализа на случај 2.7

1. Спроведеното истражување потврди дека методите од машинско учење успешно можат да го моделираат однесувањето на бетонските лачни брани преку предвидување на нивните радијални поместувања. Добиените резултати ја потврдуваат применливоста на предложената методологија и кај овој тип на брани, што значително ја проширува областа на нејзина практична примена.
2. Анализата го потврди фактот дека водното ниво во акумулацијата и температурата на воздухот претставуваат доминантни фактори кои влијаат врз радијалните поместувања на анализираниот централен блок на браната. Моделите од машинско учење успешно ја опфатија нелинеарната интеракција помеѓу овие параметри, овозможувајќи реалистично моделирање на однесувањето на конструкцијата.
3. Добиените резултати потврдија дека обучените модели успешно ги следат сезонските промени во радијалните поместувања, при што беа веродостојно опфатени како периодите на поголеми температурни варијации, така и промените предизвикани од осцилациите на водното ниво во акумулацијата. Ова укажува на нивната способност за моделирање на сложени временски зависности кај бетонските брани.
4. Анализата на важноста на предиктор варијаблите овозможи подобро инженерско разбирање на влијанието на поединечните фактори врз однесувањето на браната, со што машинското учење повторно се покажа не само како алатка за предвидување, туку и како средство за инженерска интерпретација на процесите кои влијаат врз деформациите на конструкцијата.
5. Предвидувањето на очекуваните радијални поместувања создава можност за континуирана споредба помеѓу реално измерените и пресметаните вредности, што претставува значајна поддршка при навремено откривање на потенцијални отстапувања во однесувањето на браната и подобрување на процесот на техничко набљудување, овозможувајќи побрзо препознавање на евентуални промени во однесувањето на конструкцијата и обезбедувајќи дополнителна поддршка при процената на безбедноста на браната.
6. Извршеното истражување покажа дека предложената методологија успешно може да се примени кај различни конструктивни типови брани. Со тоа се потврдува дека машинското учење претставува универзална дополнителна инженерска алатка која може да придонесе за анализа, предвидување и интерпретација на податоците од техничкото набљудување, независно од конструктивните карактеристики на браната.

8.3.8. Заклучоци од студија на случај 2

1. Спроведената студија на случај потврди дека методите од машинско учење успешно можат да се применат при анализа на различни параметри кои го опишуваат однесувањето на браните, вклучувајќи водни нивоа во пиезометри, филтрациони количества, вертикални слегнувања и радијални поместувања. Со тоа се потврдува нивната широка применливост како дополнителна инженерска алатка во процесот на техничко набљудување на брани.
2. Извршените анализи покажаа дека точноста и доверливоста на моделите од машинско учење во голема мера зависат од квалитетот, обемот и репрезентативноста на расположливите мерни податоци. Истражувањето потврди дека подолгите временски низи и поголемата варијабилност на податоците придонесуваат за подобро генерализирање на моделите и повисока точност на предвидувањата, додека ограничените множества на податоци можат да ја намалат нивната предиктивна способност.
3. Примената на машинското учење овозможи не само успешно предвидување на параметрите од техничкото набљудување, туку и идентификација на атипични мерни серии, дополнување на недостасувачки податоци, анализа на долгорочните трендови и воспоставување на очекувани граници на однесување. На тој начин беше покажано дека

предложената методологија има значително поширока инженерска примена од класичното предвидување на поединечни варијабли.

4. Анализата на различни конструктивни типови брани покажа дека предложената методологија не е ограничена на еден конкретен тип конструкција. Нејзината успешна примена кај насипна и бетонска лачна брана потврдува дека методите од машинско учење можат успешно да се интегрираат во различни системи за техничко набљудување, независно од конструктивните карактеристики на објектот.
5. Спроведеното истражување потврди дека машинското учење не претставува замена за инженерската процена, туку современа дополнителна алатка која овозможува пообјективна анализа на големи множества податоци, подобро разбирање на процесите кои влијаат врз однесувањето на браните и обезбедување дополнителни информации при донесувањето инженерски одлуки.
6. Врз основа на сите извршени анализи може да се заклучи дека предложената методологија претставува сигурен, флексибилен и практично применлив пристап за анализа на податоците од системите за техничко набљудување на брани. Нејзината примена овозможува унапредување на процесите на следење, интерпретација и рано препознавање на потенцијални промени во однесувањето на браните, со што се создаваат услови за поефикасно управување со нивната безбедност и долгорочна експлоатација.

8.4. Препораки за идни истражувања

Врз основа на резултатите од спроведеното истражување и согледаните можности за понатамошен развој на предложената методологија, се предлагаат следните насоки за идни истражувања:

1. Проширување на примената на предложената методологија врз поголем број водостопански системи и брани со различни конструктивни карактеристики, хидролошки услови и режими на експлоатација, со цел дополнителна верификација на нејзината универзалност и применливост.
2. Вклучување на поголеми и подолги временски низи од податоци, како и дополнителни следени параметри, со цел подобрување на точноста и робусноста на моделите од машинско учење, особено кај анализите со ограничен број достапни мерења.
3. Истражување на можностите за примена на современи методи од областа на длабокото учење (анг. „Deep Learning“), како што се LSTM, GRU и трансформер архитектурите, кај задачи поврзани со предвидување на временски серии во водостопанските системи и техничкото набљудување на брани.
4. Развој на хибридни модели кои ќе ги комбинираат физички базираните инженерски модели со методите од машинско учење, со цел подобрување на интерпретабилноста на резултатите и зголемување на доверливоста на предвидувањата.
5. Развој на интелигентни системи за континуирано следење на состојбата на браните преку интеграција на моделите од машинско учење со автоматски системи за прибирање на податоци, ‘Internet of Things – IoT’ технологии и дигитални платформи за техничко набљудување.
6. Унапредување на предложената методологија преку примена на методи за процена на неизвесноста на предвидувањата и техники од областа на појаснетата вештачка интелигенција (анг. „Explainable Artificial Intelligence – XAI“), со цел подобра инженерска интерпретација на резултатите и поголема доверба во донесувањето инженерски одлуки.
7. Проширување на примената на алгоритмите за оптимизација во комбинација со моделите од машинско учење за решавање на сложени повеќекритериумски задачи поврзани со управувањето со акумулациите, производството на електрична енергија, заштитата од поплави и рационалното користење на водните ресурси.
8. Верификација на предложената методологија преку примена во реални инженерски

системи и нејзина интеграција во системите за поддршка при донесување инженерски одлуки, со цел оценување на нејзината практична примена во секојдневното управување со водостопанските системи и безбедноста на браните.

“Civil engineering is the art of shaping the world around us.”
Anonymus

КОРИСТЕНА ЛИТЕРАТУРА

- [1] A. Ahmed, A., Sayed, S., Abdoulhalik, A., Moutari, S., & Oyedele, L. (2024). Applications of machine learning to water resources management: A review of present status and future opportunities. *Journal of Cleaner Production*, 441. doi:10.1016/j.jclepro.2024.140715
- [2] Agnew, C., & Woodhuse, P. (2011). *Water resources and development*. New York: Routledge.
- [3] Aleksova Dosevska, E. (2025). Utilization of hydropower potential through a small hydropower plant at the Knezhevo dam. *Proceedings of the 6th Congress on dams* (pp. 55-74). Skopje: Macedonian Committee on Large Dams (MACOLD).
- [4] Altman, N. S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, 46(3), 175-185. doi:10.1080/00031305.1992.10475879
- [5] Alwosheel, A., van Cranenburgh, S., & G. Chorus, C. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. *Journal of Choice Modelling*, 167-182.
- [6] Amin Hariri-ARdebili, M., Salazar, F., Pourkamali-Anaraki, F., Mazza, G., & Mata, J. (2023). Soft Computing and Machine Learning in Dam Engineering. *Water*, 15(917). doi:10.3390/w15050917
- [7] Ampère, A.-M. (1834). *Essai sur la philosophie des sciences*.
- [8] Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall Ltd. ISBN 9781614277651.
- [9] Aspray, W. (1990). John von Neumann and the origins of modern computing. *The Massachusetts Institute of Technology*.
- [10] Atieh, M., Mehlretter, S., Gharabaghi, B., & R., R. (2015). Integrative neural networks model for prediction of sediment rating curve parameters for ungauged basins. *Journal of Hydrology*, 531(3), 1095-1107. doi:10.1016/j.jhydrol.2015.11.008
- [11] Babbage, C. (1834). The calculating engine. *The Edinburgh Review* no. CXX.
- [12] Beljkas, Z., Krulanovic, I., Rutesic, S., Lazarevska, M., & Zafirovski, Z. (2025). Data-Driven Water Demand Prediction: Evaluating Artificial Intelligence Models. *Advances in Civil Engineering*. doi:10.1155/adce/2370659

- [13] Bhattacharya, B., Price, R., & Solomatine, D. P. (2007). A data mining approach to modelling sediment transport. *Journal of hydraulic engineering*. doi:10.1061/(ASCE)0733-9429(2007)133:4(440)
- [14] Biswas, A. K. (1975). Systems approach to water management. In A. K. Biswas, Systems approach to water management. McGraw-Hill Inc.
- [15] Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 245-271. doi:10.1016/S0004-3702(97)00063-5
- [16] Bogner, K., & Padron S., R. (2026). Improving sub-seasonal hydrological forecasts utilizing the randomness in Deep Learning models. *Stochastic Environmental Research and Risk Assessment*, 40(1). doi:10.1007/s00477-025-03138-2
- [17] Bozorg, O. H.; ed. (2021). *Essential Tools for Water Resources Analysis, Planning, and Management*. Singapore: Springer Nature Singapore Pte Ltd.
- [18] Brankovic Markovic, J., & Markovic, M. (2026). Enhancing Multi-Horizon Probabilistic Water Level Forecasting Using Horizon- and Event-Aware Deep Learning Models. *SN Applied Sciences*, 16(10). doi:10.3390/app16105004
- [19] Breiman, L. (1996). Bagging Predictors. *Machine learning*, 26(2), 123-140.
- [20] Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- [21] Breiman, L., Friedman, J., Olshen, R., & J. Stone, C. (1984). *Classification and Regression Trees*. Taylor and Francis. doi:10.1201/9781315139470
- [22] Brooks, H., & Tucker, N. (2015). Electrospinning predictions using artificial neural networks. *Polymer*, 58(10), 22-29. doi:10.1016/j.polymer.2014.12.046
- [23] Bugliarello, G., & Gunther, F. (1974). *Computer systems and water resources*. Amsterdam: Elsevier Scientific Publishing Company.
- [24] Bui, D., Hoang, N., Martínez-Alvarez, F., Ngo, P., Hoa, P., Pham, T., . . . Costache, R. (2020). A Novel Deep Learning Neural Network Approach for Predicting Flash Flood Susceptibility: A Case Study at a High Frequency Tropical Storm Area. *Science of The Total Environment*, 701. doi:10.1016/j.scitotenv.2019.134413
- [25] Casalino, G., Facchini, F., Mortello, M., & Mummolo, G. (2016). ANN modelling to optimize manufacturing processes: the case of laser welding. *IFAC-PapersOnLine*, 49(12), 378-383. doi:10.1016/j.ifacol.2016.07.634
- [26] Chau, K. (2006). Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. *Journal of Hydrology*, 329(3), 363-367. doi:10.1016/j.jhydrol.2006.02.025
- [27] Chau, K. (2007). A split-step particle swarm optimization algorithm in river stage forecasting. *Journal of Hydrology*, 346(3), 131-135. doi:10.1016/j.jhydrol.2007.09.004
- [28] Chen, B., Zhang, L., Qian, Q., Dou, Yanhong, & Ji, Z. (2017). Research on the Seepage Safety Monitoring Indexes of the High Core Rockfill Dam. *World Journal of Engineering and Technology*, 5(3B). doi:10.4236/wjet.2017.53B006
- [29] Chen, S. H. (2015). *Hydraulic structures*. Heidelberg: Springer-Verlag Berlin.
- [30] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *22nd ACM SIGKDD Int. Conf. on Knowledge Discovery (pp. 785-794)*. New York: ACM.
- [31] Cimen, M. (2008). Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal*, 53(3), 656-666. doi:10.1623/hysj.53.3.656
- [32] Cortes, C., & Vapnik, V. (1995, September). Support-vector networks. *Machine Learning*, pp. 273-297.
- [33] Couffignal, L. (1956). *Essai d'une définition générale de la cybernétique* . The First

International Congress on Cybernetics. Namur, Belgium.

- [34] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. doi:10.1109/TIT.1967.1053964
- [35] Davie, T. (2008). *Fundamentals of hydrology*. New York, USA: Routledge.
- [36] Deo, M., & Naidu, C. (1999). Real time wave forecasting using neural networks. *Ocean engineering*, 26(3), 191-203. doi:10.1016/S0029-8018(97)10025-7
- [37] Djordjevic, B. (1990). *Vodoprivredni sistemi*. Belgrad: Naucna knjiga.
- [38] Dosevska Aleksova, E. (2023). Rehabilitation and upgrade of 'Zletovica' Hydrosystem. 13th Conference on Water Economy and Hydrotechnics (pp. 35-42). Skopje: Macedonian Committee on Large Dams (MACOLD).
- [39] Flach, P. (2012). *Machine learning The art and science of algorithms that make sense of data*. New York, USA: Cambridge University Press.
- [40] Flood, I. (1990). Simulating the construction process using neural networks. *Proceedings of the 7th ISARC*, 374-382. doi:10.22260/ISARC1990/0048
- [41] Flood, I., & Kartam, N. (1994). *Neural networks in civil engineering II: Systems and application. Computing in Civil Engineering*.
- [42] Flood, I., & Paul, C. (1996). Modeling construction processes using artificial neural networks. *Automation in Construction*, 4(4), 307-320.
- [43] Grafton, Q., & Hussey, K. (2011). *Water resources planning and management*. Cambridge, UK: Cambridge University Press.
- [44] Grigg, N. (2016). *Integrated Water Resource Management: An Interdisciplinary Approach*. UK: Palgrave Macmillan.
- [45] Gupta, P., & Kulkarni, N. (1993). An Introduction of Soft Computing Approach over Hard Computing. *International Journal of latest Trends*.
- [46] Gurney, K. (1997). *An Introduction to Neural Networks*. University College London (UCL) Press.
- [47] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining, Concepts and Techniques*. Elsevier.
- [48] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Nature Link.
- [49] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning, Textbook*. Springer Nature Link.
- [50] Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. USA: Prentice Hall PTR.
- [51] Hodges, A. (2012). *Alan Turing: The Enigma*. Princeton University Press, Princeton and Oxford.
- [52] Hosseini, K., Nodoushan, E. J., Barati, R., & Shahheydari, H. (2016). Optimal design of labyrinth spillways using meta-heuristic algorithms. *KSCE Journal of Civil Engineering*, 20, 468-477. doi:10.1007/s12205-015-0462-5
- [53] Hosseini, H. (2021). A deep learning model for predicting river flood depth and extent. *Environmental Modelling & Software*, 145. doi:10.1016/j.envsoft.2021.105186
- [54] Huang, W., Cheng, J., Yang, Y., & Guo, G. (2019). An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis. *Neurocomputing*, 359. doi:10.1016/j.neucom.2019.05.052
- [55] Ibrahim, D. (2016). An Overview of Soft Computing. *Procedia Computer Science*, 102, 34-38.
- [56] Izonin, I., K. Kazantzi, A., Tkachenko, R., & Mitoulis, S.-A. (2025). GRNN-based cascade ensemble model for non-destructive damage state identification: small data approach.

Engineering with Computers, 41, 723-738. doi:10.1007/s00366-024-02048-1

- [57] J. Martyr, A., & R. Rogers, D. (2021). Engine Testing. doi:<https://doi.org/10.1016/B978-0-12-821226-4.00014-0>
- [58] Jain, S. K. (2001). Development of Integrated Sediment Rating Curves Using ANNs. *Journal of Hydraulic Engineering*, 127(1). doi:10.1061/(ASCE)0733-9429(2001)127:1(30)
- [59] Jain, S. K., & Singh, V. P. (2003). *Water resources systems planning and management*. Amsterdam: Elsevier Science B. V.
- [60] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer Nature Link.
- [61] Jermar, M. K. (1987). *Water resources and water management*. Munich: ELSEVIER.
- [62] Kauffman, L. (2007). CYBCON discussion group.
- [63] Kecman, V. (2001). *Learning and soft computing, Support Vector Machines, Neural Networks and Fuzzy Logic*. MIT Press.
- [64] Liu, Z., Zhou, J., & Yang, X. (2024). Research on Water Resource Modeling Based on Machine. *Water*, 16(472). doi:10.3390/w16030472
- [65] Loucks, D. P., & Beek, E. V. (2017). *Water resources Systems Planning and Management*. DELTARES and UNESCO_IHE.
- [66] Loucks, D. P., & Van Beek, E. (1981). *Water Resources Systems Planning and Analysis*. New Jersey, USA: Prentice-Hall, Englewood Cliffs.
- [67] M. Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Springer Nature Link.
- [68] Markovic, M., Brankovic Markovic, J., Stosovic Andrejevic, M., & Zivkovic, S. (2021). A New Method for Pore Pressure Prediction on Malfunctioning Cells Using Artificial Neural Networks. *Water Resources Management*, 35(3). doi:10.1007/s11269-021-02763-0
- [69] Markovska, M. (2020). *Real-time classification of voltage disturbances in the power grid*, PhD Thesis. Skopje: Ss. Cyril and Methodius University in Skopje, Faculty of electronics and infomatics technologies. Retrieved from https://feit.ukim.edu.mk/wp-content/uploads/2020/11/mmarkovska_phd.pdf
- [70] Marsland, S. (2009). *Machine learning An algorithmic perspective*. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Frabcis Group.
- [71] Matyjaszek, M., Riesgo Fernandez, P., Krzemien, A., Wodarski, K., & Fidalgo Valverde, G. (2019). Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory. *Resources policy*, 61, 283-292. doi:10.1016/j.resourpol.2019.02.017
- [72] Mays, L. (2005). *Water resources engineering*. USA: John Wiley and Sons, Inc.
- [73] Merrett, S. (1997). *Introduction to the economics of water resources*. London, UK: UCL Press Limited.
- [74] Mindell, D. A. (2000). *Knowledge domains in Engineering systems*.
- [75] Mitchel, M. T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- [76] Modab, A., & Reza, G. (2025). Attention-driven deep learning models for multivariate time series forecasting of reservoir water levels. *Water Science and Technology*, 92(12). doi:10.2166/wst.2025.182
- [77] Mossalam, A., & Arafa, M. (2018). Using artificial neural networks (ANN) in projects monitoring dashboards' formulation. *HBRC Journal*, 14(3), 385-392. doi:10.1016/j.hbrcj.2017.11.002
- [78] Nagy, H., Watanabe, K., & Hirano, M. (2002). Prediction of Sediment Load Concentration in

- Rivers using Artificial Neural Network Model. *Journal of Hydraulic Engineering*, 128(6). doi:10.1061/(ASCE)0733-9429(2002)128:6(588)
- [79] Nam Hoang, B., Tran Thu, P., Hoang Duc, V., & Le Van, N. (2025). Assessing LSTM algorithm performance for daily runoff prediction at Hoa Duyet hydrological station, Vietnam. *The Transport and Communications Science Journal*. doi:10.47869/tcsj.76.4.9
 - [80] Novak, P., Moffat, A. I., Nalluri, C., & Narayanan, R. (2007). *Hydraulic structures*. London: Taylor & Francis Group, London.
 - [81] Oscar, T. (2017). Neural Network Model for Thermal Inactivation of Salmonella Typhimurium to Elimination in Ground Chicken: Acquisition of Data by Whole Sample Enrichment, Miniature Most-Probable-Number Method. *Journal of Food Protection*, 80(1), 104-112. doi:10.4315/0362-028X.JFP-16-199
 - [82] Palisade Corporation. (2013). *NeuralTools Version 6 User Guide*. Retrieved from <https://www.scribd.com/document/188548979/Manual-Neuraltools6-En>
 - [83] Paul Muller, J., & Massaron, L. (2021). *Machine Learning for Dummies* (2nd ed.). Wiley.
 - [84] Pidd, M. (2004). *Computer Simulation in Management Science*. West Sussex, England: John Wiley & Sons, Ltd.
 - [85] Qiu, J., Su, H., Fang, Z., & Qi, Y. (2026). A hybrid prediction model for the displacement of concrete dams considering temperature hysteresis effects. *Engineering Structures*, 359. doi:10.1016/j.engstruct.2026.122719
 - [86] Ren, T., Liu, X., Niu, J., Lei, X., & Zhang, Z. (2020). Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *Journal of Hydrology*, 585. doi:10.1016/j.jhydrol.2020.124783
 - [87] Richard Wexelblat, ed. (1981). *History of programming languages*. New York: Academic Press Inc.
 - [88] Roser, M., & Ritchie, H. (2023). *Our world in data*. Retrieved from [Our world in data: https://ourworldindata.org/population-growth-over-time](https://ourworldindata.org/population-growth-over-time)
 - [89] Salazar, F., Irazabal, J., & Conde, A. (2024). SOLDIER: SOLution for Dam behavior Interpretation and safety Evaluation with boosted Regression trees. *SoftwareX*, 25. doi:10.1016/j.softx.2023.101598
 - [90] Sammut, C., & Webb, G. (2011). *Encyclopedia of machine learning*. Springer reference Springer Science+Business Media, LLC.
 - [91] Saqr, M. A. (2025). Improving the Accuracy of Groundwater Level Forecasting by Coupling Ensemble Machine Learning Model and Coronavirus Herd Immunity Optimizer. *Water Resources Management*. doi:10.1007/s11269-025-04210-w
 - [92] Sayed Ahmed, M. M., Khursheed, M. Z., Alshameri, B., Hamed, A. K., & Elshaarawy, M. K. (2026). Estimating seepage in heterogeneous earthfill dams on permeable foundations using explainable machine learning. *Nature Research*, 16(1). doi:10.1038/s41598-026-45048-5
 - [93] Schnitter, N. (1994). *A History of Dams: The usefull pyramids*. Rotterdam: A. A. Balkema.
 - [94] Senapati, U., Bera, A., Dutta, L., Pal, S. K., & Das, T. K. (2026). Leveraging machine learning for the optimal selection of surface rainwater harvesting sites in semi-arid regions for sustainable water resource management. *Physics and Chemistry of the Earth*, 144. doi:10.1016/j.pce.2026.104490
 - [95] Shahra, E., Wu, W., & Romano, M. (2019). Considerations on the deployment of heterogeneous IoT devices for smart water networks. 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing. Internet of People and Smart City Innovation, (pp. 791-796). Leicester.

- [96] Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge Univ. Press.
- [97] Shao, H., & Deng, X. (2018). AdaBoosting Neural Network for Short-Term Wind Speed Forecasting Based on Seasonal Characteristics Analysis and Lag Space Estimation. *Computer Modeling in Engineering & Sciences*, 277-293. doi:10.3970/cmcs.2018.114.277
- [98] Shiha, A., M. Dorra, E., & Nassar, K. (2020). Neural Networks Model for Prediction of Construction Material Prices in Egypt Using Macroeconomic Indicators. *Journal of Construction Engineering and Management*, 146(3). doi:10.1061/(ASCE)CO.1943-7862.0001785
- [99] Smith, N. (1971). *A History of dams*. London: Peter Davies.
- [100] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 199-222. doi:10.1023/B:STCO.0000035301.49549.88
- [101] Snousy, G. M. (2025). Enhancing the prediction of groundwater quality index in semi-arid regions using a novel ANN-based hybrid arctic puffin-hippopotamus optimization model. *Journal of Hydrology: Regional Studies*, 59. doi:10.1016/j.ejrh.2025.102424
- [102] Song, R., Ding, W., Wu, J., Liu, A., Zhou, H., & Chu, Z. (2020). Flash Flood Forecasting Based on Long Short-Term Memory Networks. *Water*. doi:10.3390/w12010109
- [103] Specht, D. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6), 568-576. doi:10.1109/72.97934
- [104] Special edition of 'Water'. (2023). *Soft Computing and Machine Learning in Dam Engineering*. MDPI.
- [105] Sun, J., Yu, T., & Dong, P. (2023). Three-dimensional soil slope dynamic stability assessment using minimum potential energy approach. *Soil Dynamics and Earthquake Engineering*, 168.
- [106] Tanchev, L. (2014). *Dams and appurtenant hydraulic structures*. London: CRC Press, Taylor & Francis Group.
- [107] Thomas Bergin, Jr.; Richard Gibson, Jr.; ed. (1996). *History of programming languages - II*. New York: Association for Computing Press & Addison-Wesley Publishing Company.
- [108] Tsai, C.-P., Lin, C., & Shen, J. (2002). Neural network for wave forecasting among multi-stations. *Ocean Engineering*, 29, 1683-1695. doi:10.1016/S0029-8018%2801%2900112-3
- [109] Turing, A. (1936). On computable numbers, with an application to the entscheidungsproblem.
- [110] Votruba, L., & Broža, V. (1989). *Water management in reservoirs*. Amsterdam: Elsevier Science Publishers B. V.
- [111] Wang, L., Yang, C., & Wang, M. (2016). *Advances in water resources management. Handbook of Environmental Engineering*. Switzerland: Springer International Publishing.
- [112] Wang, M.-X., Huang, D., Wang, G., & Li, D.-Q. (2020). SS-XGBoost: A Machine Learning Framework for Predicting Newmark Sliding Displacements of Slopes. *J. Geotech. Geoenviron. Eng.*, 146(9). doi:10.1061/(ASCE)GT.1943-5606.0002297
- [113] Water Science School. (2019). How much water is there on Earth? USGS. Retrieved September 9, 2025, from <https://www.usgs.gov/water-science-school/science/how-much-water-there-earth%20>
- [114] Wiener, N. (1948). *Cybernetics: or, Control and Communication in the Animal and the Machine*.
- [115] Woherem, E. E., Odeyemi, J. K., Ugwoke, P. O., Ayoola, O. S., Akinyemi-Ayoola, A., Ayeni, B., & Ubah, I. D. (2026). Machine learning based predictive model for forecasting and managing the economic impact of cross border dam water release in Nigeria. *Discover Artificial Intelligence*, 6(1). doi:10.1007/s44163-026-01003-x

- [116] Zadeh, L. (1993). Fuzzy logic, neural networks and soft computing. University of California at Berkeley.
- [117] Zeinalie, M., Bozorg-Haddad, O., & B., C. (2021). Modeling and Simulation in Water Resources Management. In O. Bozorg-Haddad, & ed., Essential Tools for Water Resources Analysis, Planning, and Management. Singapore: Springer Nature Singapore Pte Ltd.
- [118] Zhang, J., Yue, H., Basirifard, M., Cao, J., & Yang, T. (n.d.). A Mamba-type of deep state space model for reservoir release simulation with a large-scale verification over 441 dams across CONUS. *Journal of Hydrology*, 662. doi:10.1016/j.jhydrol.2025.134145
- [119] Кондински, И. (2025). Елаборат за анализа и оценка на стабилноста и функционалноста на браната „Стрежево“ - Битола со придружните објекти и теренот околу акумулацијата за 2024 година. Битола: ЈП „Стрежево“ - Битола.
- [120] Собрание, Р. Македонија. (2008). Закон за води, Службен весник на Р.М бр. 87 од 15.07.2008 година. Скопје.