

# Predviđanje nivoa vode u akumulaciji primjenom metoda mašinskog učenja

1<sup>st</sup> Milica Kašiković

Fakultet za proizvodnju i menadžment  
Trebinje

Univerzitet u Istočnom Sarajevu  
Trebinje, BiH

milica.kasikovic@fpm.ues.rs.ba

2<sup>nd</sup> Budimirka Marinović

Fakultet za proizvodnju i menadžment  
Trebinje

Univerzitet u Istočnom Sarajevu  
Trebinje, BiH

budimirka.marinovic@fpm.ues.rs.ba

3<sup>rd</sup> Aleksandar Anđelković

Fakultet tehničkih nauka  
Univerzitet u Novom Sadu

Novi Sad, Srbija

aleksandar.andjelkovic@gmail.com

**Abstract**— Efikasno upravljanje hidroelektranama direktno zavisi od tačnog predviđanja nivoa vode u akumulaciji što je ključno za optimizaciju proizvodnje i sigurnost sistema. U ovom radu istražena je primjena mašinskog učenja za kratkoročno predviđanje nivoa gornje vode. Razvijeni su modeli linearne regresije, regresionih stabala, metoda potpornih vektora i Gausovih procesa. Kao ulazni parametri korišteni su podaci o nivou vode, dotoku i padavinama za prethodni dan, dok je ciljna varijabla nivo vode za naredni dan. Performanse su evaluirane kroz koeficijent determinacije ( $R^2$ ) i statističke mjere greške MAE, MSE i RMSE. Rezultati su pokazali da svi modeli ostvaruju visoku tačnost, pri čemu je linearna regresija pružila najbolje performanse sa najmanjom greškom. Zaključuje se da jednostavniji regresioni modeli predstavljaju pouzdan alat za operativno upravljanje i planiranje u hidroenergetskim sistemima.

**Ključne riječi**—mašinsko učenje, model, nivo vode, akumulacija

## I. UVOD

U savremenim energetskim sistemima, hidroenergija predstavlja jedan od najznačajnijih obnovljivih izvora energije zbog svoje visoke pouzdanosti, fleksibilnosti i relativno niskog uticaja na životnu sredinu. Efikasno upravljanje hidroelektranama direktno zavisi od tačnog predviđanja hidroloških varijabli, među kojima je nivo vode u akumulaciji (kota gornje vode) od posebnog značaja. Precizne prognoze nivoa vode omogućavaju optimalno planiranje proizvodnje električne energije, povećanje efikasnosti, smanjenje rizika od poplava kao i očuvanje sigurnosti brana i prateće infrastrukture. U kontekstu sve izraženijih klimatskih promjena i povećane varijabilnosti hidroloških procesa, potreba za naprednim metodama predviđanja postaje sve izraženija.

Tradicionalni hidraulički i statistički modeli često imaju ograničenja u modeliranju nelinearnih i složenih odnosa između ulaznih varijabli kao što su dotok, padavine, isparavanje i prethodna stanja akumulacije. Zbog toga su metode mašinskog učenja u posljednjim decenijama privukle značajnu pažnju istraživača u oblasti hidroenergije [1].

Različite metode mašinskog učenja široko su primjenjivane za predviđanje različitih parametara hidroenergetskih sistema, uključujući proizvodnju električne energije [2,3,12,13], dotok u akumulaciju [4,5], nivo vode u akumulaciji [6,7], optimizaciju rada hidroelektrane [8] i druge ključne funkcionalnosti. Primjena ovih tehnika

omogućava poboljšanje efikasnosti upravljanja hidroelektranama, smanjenje operativnih rizika i povećanje pouzdanosti predviđanja u dinamičnim hidrometeorološkim uslovima.

U ovom radu razmatraju se različiti pristupi mašinskog učenja za kratkoročno predviđanje nivoa vode u akumulaciji hidroelektrane. Konkretno, analizirani su modeli linearne regresije, regresiona stabla, metode potpornih vektora (SVM) i modeli Gausovih procesa (GRP).

Primjena različitih modela omogućava upoređivanje njihovih performansi u smislu tačnosti predviđanja, robusnosti i sposobnosti generalizacije na neviđene podatke, što je od suštinskog značaja za pouzdano upravljanje hidroenergetskim sistemom.

## II. METODE MAŠINSKOG UČENJA

Mašinsko učenje smatra se podskupom vještačke inteligencije, a ujedno i granom računarskih nauka koja analizira i interpretira podatke (što može biti u različitim oblicima, poput prepoznavanja obrazaca, reorganizacije, praćenja i traganja podataka) kako bi unaprijedila sistem i pomogla u donošenju odluka uz malo ili nimalo ljudske intervencije [9]. Osnovni koncept mašinskog učenja se ogleda kroz startovanje sa inicijalnim podacima, učenje iz tih podataka (treniranje algoritma nad tim podacima) i korišćenje rezultujućeg modela za predviđanje novog seta odgovarajućih podataka [10].

Kako bi se projektovao ili istrenirao model, u mašinskom učenju postoji mnogo vrsta algoritama i neophodno je izabrati odgovarajući algoritam u zavisnosti od primjene. Uopšteno, glavne vrste mašinskog učenja su nadgledano učenje (*supervised*), nenadgledano (*unsupervised*) učenje i učenje potkrepljivanjem (*reinforcement*). Nadgledano učenje se koristi za regresiju i klasifikaciju. Neke od metoda koje se koriste za regresiju u okviru nadgledanog učenja su: linearna regresija, regresiona stabla, metoda potpornih vektora i modeli Gausovih procesa.

### A. Linearna regresija

Linearna regresija se koristi kada se na osnovu jedne nezavisne promjenljive određuje vrijednost druge zavisne promjenljive vrijednosti. Višestruka linearna regresija (*Multi Linear Regression*) pretpostavlja linearnu vezu između nezavisnih i zavisnih promjenljivih. Zbog toga se najbolje prilagođavanje opisuje pravom linijom odnosa, pri čemu se pretpostavlja da su podaci normalno raspoređeni [11]. Opšta matematička formula (1) modela višestruke linearne regresije

za  $n$  nezavisnih promjenljivih zapisana je na sljedeći način [12]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \dots + \beta_n x_n + \varepsilon \quad (1)$$

gde je  $y$  zavisna promjenljiva,  $\beta_0$  predstavlja presjek sa  $y$ -osom,  $\beta_i$  je koeficijent nagiba  $i$ -te ulazne promjenljive  $x_i$ , a  $\varepsilon$  predstavlja grešku modela.

### B. Regresiono stablo

Regresiono stablo (*Regression Tree*) je algoritam stabla odlučivanja koji se koristi u regresionoj analizi. To je neparametarski statistički model koji se može koristiti za analizu složenih odnosa između ulaznih promjenljivih i kontinuirane ciljane promjenljive. Rezultat je model u obliku stabla koji se može koristiti za predviđanje ciljane promjenljive za nove podatke na osnovu vrijednosti ulaznih promjenljivih [13].

Stablo se konstruiše tako da listovi predstavljaju predviđene vrijednosti ciljane promjenljive, dok unutrašnji čvorovi predstavljaju ulazne promjenljive i njihova pravila grananja. Regresiona stabla se koriste kada je odnos između ulaznih promjenljivih i ciljane promjenljive nelinearan ili kada postoje interakcije između promjenljivih.

### C. Metoda potpornih vektora

Metode potpornih vektora (*Support Vector Machine* – SVM) su nadgledani modeli mašinskog učenja zasnovani na regresionom algoritmu koji može da se nosi sa nelinearnim podacima u svrhu predikcije. Ova metoda se ističe svojom robusnošću i visokom tačnošću predviđanja u uslovima velike dimenzionalnosti ulaznog prostora. Podaci za treniranje i testiranje koji se koriste u ovoj metodi pretpostavljaju se kao nezavisni i identično distribuirani, sa nepoznatom funkcijom vjerovatnoće.

SVM razvija linearni hiperplan koji transformiše višedimenzionalne ulazne vektore u izlazne vrijednosti, koje se zatim koriste za predviđanje budućih izlaznih vrijednosti. Za linearnu funkciju  $f$ , skup od  $n$  podataka dat je kao  $P(x_i, y_i)$  gdje je  $x_i$  ulazni vektor  $i$ -te tačke podataka, a  $y_i$  njena stvarna vrijednost, dok je hiperplan  $f(x)$  definisan izrazom (2) [14]:

$$f(x) = \omega x_i + b \quad (2)$$

gdje je  $\omega$  nagib, a  $b$  slobodni član.

### D. Modeli Gausovih procesa

U statističkoj analizi, Gausov proces je stohastički proces kod kog svaka konačna kolekcija slučajnih promjenljivih ima multivarijantnu normalnu distribuciju [15].

Regresija Gausovog procesa (Gaussian Process Regression – GPR) je neparametarska metoda i korisna je za rad sa malim skupovima podataka. Dodatna prednost je njena sposobnost da se nosi sa mjerenjem nesigurnosti predikcija. Gausov proces se označava na sledeći način (3):

$$f(x) = GP(m(x), k(x, \acute{x})) \quad (3)$$

gdje su  $m(x)$  i  $k(x, \acute{x})$  funkcija srednje vrijednosti i kovarijaciona funkcija, respektivno. Funkcija  $m(x)$  predstavlja očekivanu vrijenost funkcije  $f(x)$  u tački  $x$ , a konvrijaciona funkcija je mjera nivoa povjerenja u  $m(x)$ .

## III. OCJENA PERFORMANSI MODELA

Za ocjenu performansi formiranih modela korištene su sljedeće statističke mjere:

- $R^2$  – koeficijent determinacije
- MAE – srednja apsolutna greška
- MSE – srednja kvadratna greška i
- RMSE – korijen srednje kvadratne greške

Koeficijent determinacije ( $R^2$ ) je statistička mjera koja se koristi za testiranje pretpostavki u oblasti mašinskog učenja, Koeficijent determinacije se određuje na osnovu jednačine (4):

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

gdje su:  $\hat{y}$  – predviđena vrijednost promjenljive, a  $\bar{y}$  – srednja vrijednost promjenljive.

Srednja apsolutna greška (MAE - *Mean Absolute Error*) predstavlja prosječnu vrijednost apsolutne razlike između stvarnih i predviđenih vrijednosti u skupu podataka i određuje se na osnovu izraza (5):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (5)$$

Srednja kvadratna greška (MSE - *Mean Squared Error*) predstavlja prosječnu vrijednost kvadrata razlike između originalnih (stvarnih) i predviđenih vrijednosti u skupu podataka. MSE se određuje na osnovu izraza (6):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (6)$$

Korijen srednje kvadratne greške (RMSE - *Root Mean Squared Error*) je kvadratni korijen srednje kvadratne greške i određuje se na osnovu izraza (7):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (7).$$

## IV. RAZVOJ MODELA

Razvoj modela mašinskog učenja za predviđanje nivoa gornje vode za sljedeći dan zasniva se na analizi istorijskih hidrometeoroloških podataka i primjeni metoda nadgledanog učenja. Cilj razvijenih modela je da, na osnovu tri ulazne veličine izmjerene prethodnog dana — nivoa vode, dotoka u akumulaciju i količine padavina — izvrše predviđanje nivoa gornje vode za naredni dan.

Podaci korišteni za razvoj i testiranje modela prikupljeni su sa lokaliteta istočne Hercegovine, odnosno sa akumulacije Bilečko jezero, formirane izgradnjom brane Grančarevo, koja je u funkciji Hidroelektrane Trebinje 1. Bilečko jezero predstavlja jednu od najznačajnijih akumulacija u hidroenergetskom sistemu regiona, pri čemu maksimalna kota gornje vode iznosi 400 m, dok minimalna iznosi 351 m nadmorske visine.

Ulazni skup podataka formiran je kao trodimenzionalni vektor koji sadrži vrijednosti nivoa vode, dotoka i padavina za prethodni dan, dok izlaznu (ciljnu) varijablu predstavlja nivo gornje vode za sljedeći dan.

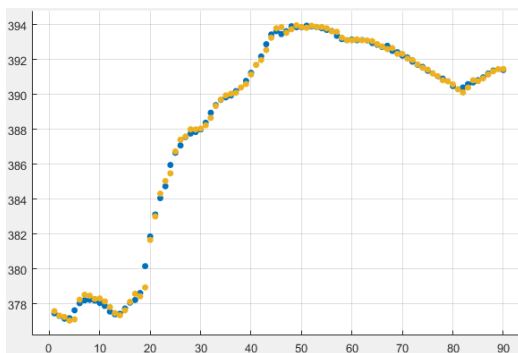
U okviru istraživanja razvijeno je više modela mašinskog učenja zasnovanih na različitim regresionim pristupima. Kao

bazni model korištena je linearna regresija koja omogućava jednostavnu interpretaciju odnosa između ulaznih varijabli i ciljne veličine. Pored toga, razvijeni su modeli regresionih stabala, koji omogućavaju hvatanje nelinearnih zavisnosti i interakcija između ulaznih parametara. Takođe su primjenjene metode potpornih vektora koje se odlikuju dobrom sposobnošću generalizacije u uslovima ograničene količine podataka. Konačno, razvijeni su i modeli bazirani na Gausovim procesima, koji pored predviđanja omogućavaju i procjenu nesigurnosti rezultata. Svi modeli formirani su pomoću programskog jezika Matlab.

Za razvoj modela korišteni su istorijski podaci prikupljeni tokom prvih tri mjeseca u godini. Ovaj vremenski period odabran je zbog izražene varijabilnosti hidroloških i meteoroloških uslova, što omogućava efikasnije učenje odnosa između nivoa vode, dotoka u akumulaciju i padavina, kao i njihovog uticaja na nivo gornje vode za sljedeći dan.

Formirani modeli linearne regresije su: linearni model, linearni model sa interakcijama, robustni linearni model i stepwise linearni model. Analizom performansi utvrđeno je da model linearne regresije sa interakcijama ostvaruje najmanju vrijednost RMSE, što ukazuje na njegovu najbolju tačnost među razmatranim linearnim modelima.

Na Slici 1 prikazano je poređenje stvarnih (izmjerenih) vrijednosti nivoa vode i vrijednosti dobijenih primjenom razvijenog modela linearne regresije. Kako bi se osigurala objektivnost procjene, model je testiran primjenom metode unakrsne validacije sa 5 preklopa (5-fold cross-validation). Na apscisi (x-osa) naveden je redni broj dana za koji je vršeno treniranje i validacija modela, čime je obuhvaćen vremenski kontinuitet procesa. Na ordinati (y-osa) prikazan je nivo vode izražen u metrima nadmorske visine. Stvarne vrijednosti nivoa vode označene su plavom bojom, dok su vrijednosti predviđene modelom prikazane narandžastom bojom.



Slika 1. Odnos stvarnih vrijednosti i vrijednosti modela u toku treniranja

Na osnovu regresionih stabala razvijeni su modeli: jednostavno stablo, srednje složeno stablo i složeno stablo. Analizom dobijenih rezultata utvrđeno je da složeno stablo daje najbolje rezultate, odnosno ostvaruje najveću tačnost predviđanja.

Na osnovu metoda potpornih vektora formirani su sljedeći modeli: linearni SVM, kubni SVM, fini Gausov SVM, srednji Gausov SVM i grubi Gausov SVM. Analiza rezultata pokazuje da linearni SVM ostvaruje najbolje

performanse, sa vrijednošću RMSE od 0,38, što ukazuje na njegovu najveću tačnost među razmatranim SVM modelima.

Na osnovu Gausovih procesa formirani su sljedeći modeli: racionalno-kvadratni GRP, kvadratno-eksponencijalni GRP, Matérn 5/2 GRP i eksponencijalni GRP. U poređnom analizom performansi utvrđeno je da kvadratno-eksponencijalni GRP ostvaruje najmanju vrijednost RMSE, čime pokazuje najbolju tačnost predviđanja među razmatranim modelima.

U tabeli I su prikazane vrijednosti koeficijenta determinacije, kao i statističkih mjera greške (MAE, MSE i RMSE) za odabrane modele iz svake razmatrane kategorije. Na osnovu prikazanih rezultata može se uočiti da svi analizirani modeli ostvaruju veoma dobre performanse predviđanja na treniranom skupu podataka, što potvrđuje njihovu sposobnost da uspješno opišu zavisnost između ulaznih varijabli i ciljne veličine.

TABELA I PERFORMANSE MODELA U TOKU TRENIRANJA

Metoda	Performanse modela			
	$R^2$	MAE	MSE	RMSE
Stabla regresije	0,99	0,36	0,37	0,6
Linearna regresija	1,00	0,14	0,04	0,21
SVM	1,00	0,33	0,15	0,38
GPR	1,00	0,14	0,05	0,21

Razvijeni modeli predstavljaju pouzdan alat za kratkoročno predviđanje nivoa gornje vode i mogu imati značajnu primjenu u upravljanju akumulacijama, planiranju rada hidroenergetskih sistema i smanjenju rizika od poplava.

## V. TESTIRANJE MODELA

Za modele iz svake kategorije koji su pokazali najbolje performanse tokom treniranja, izvršeno je dodatno testiranje na podacima koji nisu korišteni u fazi treniranja. Za evaluaciju modela korišćen je skup ulaznih karakteristika prikupljenih tokom 15 dana četvrtog mjeseca, što je omogućilo objektivnu procjenu sposobnosti modela da generalizuju i precizno predviđaju nivo vode u akumulaciji u nevidenim uslovima.

TABELA II PERFORMANSE MODELA U TOKU TESTIRANJA

Metoda	Performanse modela			
	$R^2$	MAE	MSE	RMSE
Stabla regresije	0,7280	0,3635	0,1916	0,4377
Linearna regresija	0,9936	0,0574	0,0045	0,0673
SVM	0,9738	0,0991	0,0184	0,1359
GPR	0,9878	0,0839	0,0086	0,0928

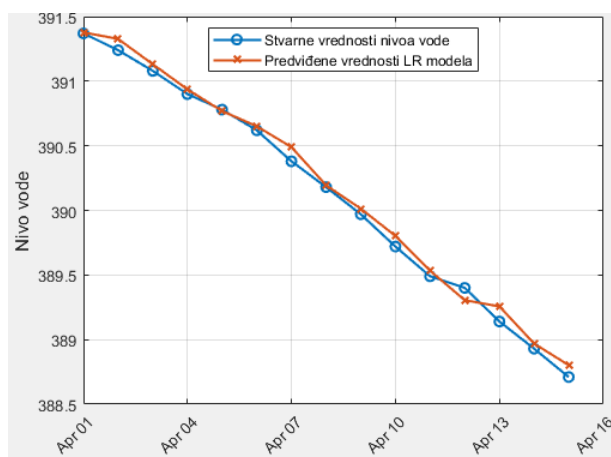
Analiza rezultata prikazanih u tabeli II pokazuje da model linearne regresije ostvaruje najmanju vrijednost RMSE, što ukazuje na njegovu visoku tačnost u predviđanju nivoa vode za testirani period. Model Gausovih procesa takođe pokazuje dobre performanse, sa RMSE vrijednošću nešto višom od linearne regresije, ali značajno nižom od RMSE vrijednosti regresionog stabla. Model baziran na metodama potpornih vektora ostvaruje srednje dobre

rezultate, dok regresiono stablo pokazuje najveću grešku predviđanja među analiziranim modelima.

Modeli Gausovih procesa pokazuju izuzetno precizno praćenje stvarnih vrijednosti. Numerički rezultati su vrlo bliski realnim, što ukazuje na to da ovaj model odlično modelira glatke prelaze i nelinearne zavisnosti u nivou vode.

SVM model pokazuje visoku pouzdanost i stabilnost. Rezultati SVM-a su konzistentni sa stvarnim trendom, uz minimalna odstupanja, što ga čini robusnim izborom za ovaj tip regresije.

Model baziran na stablima regresije pokazuje najlošije karakteristike jer generiše "stepenaste" rezultate (iste vrijednosti za više različitih dana). To ukazuje na to da model ne može da uhvati fine, kontinuirane promjene nivoa vode, već podatke grubo grupiše u kategorije.



Slika 2. Poređenje stvarnih vrijednosti i vrijednosti dobijenih modelom

Na dijagramu (slika 2.) je prikazano poređenje stvarnih vrijednosti nivoa vode (izraženih u metrima nadmorske visine) u akumulaciji i vrijednosti predviđenih modelom linearne regresije za testni vremenski period. Manja lokalna odstupanja između stvarnih i predviđenih vrijednosti mogu se pripisati činjenici da model linearne regresije ne uzima u obzir nelinearne procese koji utiču na vodni bilans akumulacije, kao što su promjene dotoka, padavine, ispuštanje vode ili operativno upravljanje branom. Zbog toga model nije u stanju da precizno reprodukuje kratkoročne oscilacije nivoa vode koje su prisutne u stvarnim podacima.

Dobijeni rezultati ukazuju da, za kratkoročno predviđanje nivoa vode u akumulaciji, jednostavniji regresioni modeli, kao što je linearna regresija, mogu biti efikasniji od kompleksnijih modela koji se oslanjaju na nelinearne zavisnosti, posebno u situacijama kada su ulazni podaci stabilni i linearno povezani sa ciljnim parametrom. S druge strane, modeli poput Gausovih procesa pružaju dodatnu prednost u smislu procjene nesigurnosti predviđanja, što može biti korisno u operativnom planiranju rada hidroenergetskih sistema.

Važno je naglasiti da su modeli u ovom radu trenirani na podacima prikupljenim tokom prvih tri mjeseca u godini, odnosno u relativno ograničenom vremenskom prozoru. Iako ovaj period karakteriše određena varijabilnost hidroloških parametara, on ne obuhvata kompletan godišnji ciklus, niti izražene sezonske ekstreme koji se mogu javiti tokom

prolječnih i jesenjih smjena ili u periodima intenzivnih padavina i suša.

Ukoliko bi se za treniranje modela koristio drugačiji vremenski okvir, naročito onaj koji uključuje izraženije sezonske oscilacije nivoa vode, moguće je da bi performanse linearnog modela bile drugačije. Linearna regresija pretpostavlja približno linearnu zavisnost između ulaznih i izlaznih varijabli, te u uslovima stabilnih i umjereno promjenljivih podataka može dati veoma precizne rezultate, što je potvrđeno u ovom istraživanju. Međutim, u slučaju izraženih nelinearnih promjena vodnog bilansa – usljed naglog porasta dotoka, intenzivnih padavina ili promjena u režimu upravljanja akumulacijom – može se očekivati povećanje greške linearnog modela, jer on nije u stanju da adekvatno opiše složene nelinearne odnose.

Sa druge strane, modeli kao što su regresiona stabla, SVM i Gausovi procesi imaju veću fleksibilnost u modeliranju kompleksnih zavisnosti i potencijalno bi mogli pokazati prednost pri treniranju na dužim vremenskim serijama koje obuhvataju različite hidrološke režime. Veća količina podataka, posebno ako uključuje sezonske i ekstremne vrijednosti, omogućila bi ovim modelima da bolje nauče obrasce ponašanja sistema i ostvare robusniju generalizaciju.

Razlog zbog kojeg su u ovom istraživanju kompleksniji modeli pokazali veću grešku u odnosu na linearni model može se objasniti relativno malim skupom podataka. U uslovima ograničenog broja uzoraka, modeli sa većim brojem parametara i većom fleksibilnošću imaju veći rizik od preprilagođavanja (*overfitting*), odnosno učenja specifičnih oscilacija prisutnih u trenirajućem skupu, što može dovesti do slabije generalizacije na testne podatke. Linearna regresija, kao model sa manjim brojem parametara i nižom složenosti, pokazuje veću stabilnost i bolju sposobnost generalizacije u takvim uslovima.

Na osnovu navedenog može se zaključiti da izbor optimalnog modela u značajnoj mjeri zavisi od dužine i strukture vremenske serije, kao i od izraženosti nelinearnih i sezonskih efekata. U slučaju većeg i raznovrsnijeg skupa podataka koji obuhvata cjelokupan godišnji ciklus, očekivano je da bi nelinearni modeli mogli ostvariti komparativnu prednost, dok u uslovima ograničenog i relativno homogenog skupa podataka jednostavniji linearni modeli mogu predstavljati efikasnije i pouzdanije rješenje.

## VI. ZAKLJUČAK

U ovom radu analizirana je primjena različitih metoda mašinskog učenja za kratkoročno predviđanje nivoa gornje vode u akumulaciji hidroelektrane, sa ciljem unapređenja upravljanja hidroenergetskim sistemima. Korišćenjem istorijskih hidrometeoroloških podataka i pristupa nadgledanog učenja razvijeni su modeli linearne regresije, regresionih stabala, metoda potpornih vektora i Gausovih procesa, koji su upoređeni na osnovu standardnih statističkih mjera performansi.

Rezultati dobijeni tokom faze treniranja pokazali su da svi razmatrani modeli mogu veoma uspješno opisati zavisnost između ulaznih varijabli (nivo vode, dotok i padavine) i ciljne veličine. Međutim, testiranje modela na nevidenim podacima jasno je ukazalo na razlike u njihovoj sposobnosti generalizacije. Model linearne regresije ostvario je najbolje performanse u fazi testiranja, sa najmanjom

vrijednošću RMSE, što potvrđuje da u uslovima relativno stabilnih i približno linearnih odnosa jednostavniji modeli mogu nadmašiti složenije nelinearne pristupe. Modeli Gausovih procesa takođe su pokazali visoku tačnost predviđanja, uz dodatnu prednost procjene nesigurnosti rezultata, što ih čini posebno pogodnim za donošenje operativnih odluka u uslovima povećane hidrološke neizvjesnosti.

Na osnovu sprovedene analize može se zaključiti da razvijeni modeli mašinskog učenja predstavljaju pouzdan i efikasan alat za kratkoročno predviđanje nivoa gornje vode u akumulaciji. Njihova primjena može doprinijeti optimalnom planiranju proizvodnje električne energije, smanjenju operativnih rizika i unapređenju bezbjednosti hidroenergetskih objekata.

Dobijeni rezultati potvrđuju da performanse modela u značajnoj mjeri zavise od obima i strukture raspoloživog skupa podataka, pri čemu u uslovima ograničenog i relativno homogenog vremenskog intervala jednostavniji modeli mogu ostvariti bolju generalizaciju. U slučaju proširenja vremenske serije na cjelokupan godišnji ciklus sa izraženim sezonskim i ekstremnim hidrološkim događajima, očekivano je da bi nelinearni modeli mogli pokazati dodatne prednosti. Stoga izbor optimalnog pristupa treba uskladiti sa karakteristikama podataka i dinamikom sistema. Dalja istraživanja mogu biti usmjerena ka proširenju ulaznog skupa podataka, kako u pogledu dužine vremenske serije, tako i uključivanjem dodatnih hidrometeoroloških i operativnih parametara radi unapređenja tačnosti i robusnosti predviđanja.

#### REFERENCES

- [1] B. Marinović, M. Kašiković, ARTIFICIAL NEURAL NETWORKS MODELS IN ENERGY FORECASTING – A REVIEW. *Journal of Engineering and Management*, 3(1), 2025. <https://jem.fpm.ues.rs.ba/index.php/journal/article/view/142>
- [2] S.I. Ramarope, O.S. Fatoba, T.C. Jen, Hydro-power generation forecast in South Africa based on Machine Learning (ML) models, *Scientific African*, Volume 22, 2023. <https://doi.org/10.1016/j.sciaf.2023.e01981>
- [3] B. Aksoy, Estimation of Energy Produced in Hydroelectric Power Plant Industrial Automation Using Deep Learning and Hybrid Machine Learning Techniques. *Electric Power Components and Systems*, 49(3), 2021 213–232. <https://doi.org/10.1080/15325008.2021.1937401>
- [4] B.J. Kim, Y.T. Lee, B.H. Kim, A Study on the Optimal Deep Learning Model for Dam Inflow Prediction. *Water* 2022, 14, 2766. <https://doi.org/10.3390/w14172766>
- [5] B. Marinović, M. Kašiković, A. Anđelković, Hydropower Reservoir Inflow Forecasting using Time Series Neural Networks, 17th International Conference on Accomplishments in Mechanical and Industrial Engineering – DEMI 2025
- [6] M. Sapitang, M.W. Ridwan, K.F. Kushiar, N. Ahmed, A. El-Shafie, A. Machine Learning Application in Reservoir Water Level

- Forecasting for Sustainable Hydropower Generation Strategy. *Sustainability* 2020, 12, 6121. <https://doi.org/10.3390/su12156121>
- [7] I. Rehamnia, A.M. Meymand, Advancing Reservoir Water Level Predictions: Evaluating Conventional, Ensemble and Integrated Swarm Machine Learning Approaches. *Water Resour Manage* 39, 779–794 (2025). <https://doi.org/10.1007/s11269-024-03990-x>
  - [8] J. Bernardes, M. Santos, T. Abreu, L. Prado, D. Miranda, R. Julio, P. Viana, M. Fonseca, E. Bortoni, G.S. Bastos, G.S., Hydropower Operation Optimization Using Machine Learning: A Systematic Review. *AI* 2022, 3, 78–99. <https://doi.org/10.3390/ai3010006>
  - [9] K.S. Priya, Linear Regression Algorithm in Machine Learning through MATLAB, Vol 9, I. XII, December 2021., pp. 989–995. <https://doi.org/10.22214/ijraset.2021.39410>
  - [10] S.. Obradović, B. Milošević, and P. Štrbac, “MATLAB i mašinsko učenje,” *Proc. in Infoteh-Jahorina* Vol. 13, , March 2014, pp. 969–974 <https://infoteh.etf.ues.rs.ba/zbornik/2014/radovi/RSS-7/RSS-7-4.pdf>
  - [11] G.K. Uyanik and N. Güler, A study on multiple linear regression analysis, *Procedia-Social and Behavioral Sciences*. (2013) 106, pp. 234–240 <https://doi.org/10.1016/j.sbspro.2013.12.027>
  - [12] P. Ekanayake, L. Wickramasinghe, J.M. Jeevani, W. Jayasinghe, U. Rathnayake, Regression-Based Prediction of Power Generation at Samanalawewa Hydropower Plant in Sri Lanka Using Machine Learning, *Mathematical Problems in Engineering*, July, 2021. <https://doi.org/10.1155/2021/4913824>
  - [13] J. Jayasinghe, P. Ekanayake, O. Panahatipola, C.I. Madhushani, U. Rathnayake, Forecasting the power generation at renewable power plants in Sri Lanka using regression trees, *Results in Engineering*, Volume 22, 2024 <https://doi.org/10.1016/j.rineng.2024.102111>
  - [14] M. Wauters M. Vanhoucke, Support vector machine regression for project control forecasting, *Automation in Construction*. (2014) 47, 92–106 <https://doi.org/10.1016/j.autcon.2014.07.014>
  - [15] D. Kong, Y. Chen, N. Li, Gaussian process regression for tool wear prediction, *Mechanical Systems and Signal Processing*. (2018) 104, 556–574 <https://doi.org/10.1016/j.ymssp.2017.11.021>

#### ABSTRACT

Efficient management of hydropower plants directly depends on the accurate prediction of reservoir water levels, which is crucial for production optimization and system safety. This paper researches the application of machine learning methods for short-term prediction of upper water levels. Models based on linear regression, regression trees, support vector machines (SVM), and Gaussian processes (GPR) were developed and compared. Data on water level, inflow, and precipitation from the previous day were used as input parameters, while the target variable was the water level for the following day. Model performance was evaluated using the coefficient of determination ( $R^2$ ) and statistical error measures: MAE, MSE, and RMSE. The results demonstrate that all analyzed models achieve high predictive accuracy, with linear regression showing the best performance and the smallest prediction error. It is concluded that simpler regression models can serve as effective and reliable tools for operational management and planning in hydropower systems.

Key words: machine learning, model, water level, reservoir.

#### FORECASTING RESERVOIR WATER LEVEL USING MACHINE LEARNING METHODS

Milica Kašiković, Budimirka Marinović, Aleksandar Anđelković