

KOMPARATIVNA ANALIZA ALGORITAMA MAŠINSKOG UČENJA ZA PREDVIĐANJE BROJA AKTIVNIH BUNARA U SISTEMIMA VODOSNABDEVANJA

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR ACTIVE WELL PREDICTION IN WATER WORKS

OGNJEN IVIĆ¹

Originalni naučni rad
DOI: 10.5937/VIK24503I

Rezime: Specifičan domen ovog rada je predviđanje operativnih parametara u upravljanju vodosnabdevanjem, a fokus je na predviđanju neophodnog broja aktivnih bunara. Efikasno upravljanje brojem aktivnih bunara je ključno za optimizaciju resursa i održavanje stabilnosti u snabdevanju vodom, posebno u sistemima sa promenljivim potrošačkim zahtevima i sezonskim varijacijama. Ovaj rad istražuje primenu dva različita algoritma mašinskog učenja, K-Nearest Neighbors (KNN) i Decision Tree Regressor, za predviđanje broja aktivnih bunara na osnovu istorijskih podataka samog izvorišta. Performanse oba modela su ocenjene korišćenjem Mean Squared Error (MSE), Root Mean Squared Error (RMSE) i R-squared (R^2) metrika. Cilj rada je poređenje efikasnosti i tačnosti ovih algoritama u specifičnom domenu.

Ključne reči: vodosnabdevanje, mašinsko učenje, K-Nearest Neighbors (KNN), Decision Tree Regressor

Abstract: The specific domain of this study is the prediction of operational parameters in water supply management, with a focus on predicting the necessary number of active wells. Efficient management of the number of active wells is crucial for resource optimization and maintaining stability in water supply, especially in systems with variable consumer demands and seasonal variations. This study explores the application of two different machine learning algorithms, K-Nearest Neighbors (KNN) and Decision Tree Regressor, for predicting the number of active wells based on historical data from the water source. The performance of both models is evaluated using Mean Squared Error (MSE), Root Mean Squared Error

¹ Ognjen Ivić, JKP „Vodovod i kanalizacija“ Subotica, Trg Lazara Nešića 9/a, Subotica, ognjen@vodovodsu.rs, ORCID: 0009-0005-3005-512X

(RMSE), and R-squared (R^2) metrics. The aim of the study is to compare the effectiveness and accuracy of these algorithms in the specific domain.

Key Words: water supply, machine learning, K-Nearest Neighbors (KNN), Decision Tree Regressor

1. Uvod

Mašinsko učenje sve više pronalazi primenu u različitim domenima, uključujući optimizaciju i predviđanje u sistemima vodosnabdevanja. Efikasno predviđanje broja aktivnih bunara može značajno unaprediti upravljanje resursima i operativnu efikasnost sistema. U ovom radu upoređene su performanse dva popularna algoritma mašinskog učenja: K-nearest neighbors (KNN) i Decision Tree Regressor, za predviđanje neophodnog broja aktivnih bunara. Cilj je izvršiti komparativnu analizu ova dva modela, razmotriti njihove karakteristike, način učenja, tačnost predviđanja i efikasnost predviđanja u realnom vremenu.

Vodosnabdevanje grada Subotice se vrši sa dva izvorišta Vodozahvat 1 i Vodozahvat 2, odakle se voda eksploatiše, odnosno zahvata, vrši se prerada i potiskuje putem crpne stanice u sistem distributivne mreže. U ovom primeru biće obrađeni samo podaci dobijeni od zapisa sa „SCADE“ tokom skoro 2 godine (01.01.2023-12.08.2024, izostavljeni su datumi sa lošim zapisima) za svaki bunar pojedinačno, a podaci koji će se koristiti su podaci dobijeni samo sa Vodozahvata 1. Analiziran je broj aktivnih bunara u odnosu na zbirni protok sa bunarskog polja. Bunarsko polje raspolaže sa oko 27 aktivnih bunara, u okviru kojih se (svake sekunde) vrši zapis protoka, nivoa, pritiska i drugih parametara.

Analizirani su protoci Q (l/s) sa bunara i broja bunara koji su radili u tom trenutku, koji zavise od različitih faktora, kao što su potrošnja vode, pritisak u mreži, stanje infrastrukture, havarije i defekti, sezonske promene i drugo. Precizno predviđanje broja aktivnih bunara u odnosu na zadati protok, omogućava efikasnije planiranje i smanjenje operativnih troškova kao i racionalnije zahvatanje vode. Za vodosnabdevanje se ne koriste istovremeno svi bunari već se koriste po utvrđenom redu tako da se svi objekti koriste i drže u „formi“.

Algoritam koji je napisan koristi podatke koje sadrže istorijske zapise o protoku (Q l/s) sirove vode i broju aktivnih bunara (n), na osnovu kojih se „treniraju“ i ocenjuju modeli. Broj istorijskih zapisa koji su korišćeni su oko 500 podataka (svaki zapis predstavlja jedan dan 24 sata – 86.400 merenja). Algoritam koji je korišćen za potrebe učenja, evaluacije i predviđanja konstruisan je od strane autora, korišćenjem programa Python i korišćene su određene biblioteke za treniranje modela mašinskog učenja. Neke od ključnih biblioteka jesu: „pandas“, „numpy“, „scikit-learn“, „matplotlib“ i „joblib“.

2. Algoritmi mašinskog učenja

K-nearest neighbors (KNN) je jedan od najjednostavnijih algoritama za mašinsko učenje. Radi po principu traženja najbližih tačaka u skupu podataka (najbližih okolnih skupova) kako bi napravio predviđanje. Kada se koristi za regresiju, KNN predviđa vrednost na osnovu prosečne vrednosti skupa u odnosu na najbliži ulazni podatak odnosno vrednost. To je model sa nadzorom, koji ne pravi pretpostavke o distribuciji podataka, relativno je „lenj“ algoritam koji se oslanja na celokupan „trening“ skup za donošenje odluka. To znači da model ne uči eksplicitne parametre tokom obuke, već čuva „trening“ podatke i koristi ih direktno za predviđanje na osnovu udaljenosti između uzoraka kako bi klasifikovao novi uzorak. U ovom slučaju, kada koristimo samo jedan atribut, izlazni protok vode sa bunarskog polja, (-, *Izlazni_protok_sa_vz1.Qizl*), udaljenost između dva uzorka je jednostavna apsolutna razlika između njihovih vrednosti. Najčešće korišćen proračun za udaljenost je Evklidovska udaljenost [6].

Za dva uzorka X_1 i X_2 , Evklidovska udaljenost se računa kao: $d(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$, gde je n broj atributa (dimenzija). Nakon što izračuna udaljenosti između novog uzorka i svih uzoraka iz „trening“ skupa, KNN bira k najbližih susednih tačaka. U ovom primeru to je $k=5$, što znači da se uzimaju pet najbližih uzoraka. Iz svakog fajla koji služi kao istorijski podatak, koristi vrednosti iz kolone za ukupan protok kao ulazne karakteristike (Q l/s - features) da bi predvideo koliko bunara je aktivno bilo u tom trenutku (n - features).

Algoritam se mogao napisati tako da koristi jednostavni proračun tačnosti baziran na relativnoj toleranciji ($rtol=0.1$), koja proverava koliko su predviđene vrednosti blizu stvarnim vrednostima unutar određenog procenta (10% u ovom slučaju). Ako je predviđena vrednost unutar 10% od stvarne vrednosti, smatra se da je predviđanje tačno, ali za preciznije predviđanje i za dobijanje naprednijeg modela korišćeno je za ocenu performansi kriterijum greške kao što su MSE (Mean Squared Error), RMSE (Root Mean Squared Error) i R^2 (koeficijent determinacije). [3].

Decision Tree Regressor ili odlučujuće stablo, ovaj algoritam se bavi procenom broja aktivnih bunara koristeći model mašinskog učenja u vidu odlučujućeg stabla za regresiju koji se često koristi za predviđanje kontinuiranih vrednosti. Odlučujuće stablo pravi odluke kroz seriju pitanja na osnovu atributa podataka, što omogućava donošenje odluka bez potrebe za naprednim matematičkim formulama.

Postavljaju se karakteristike (feature set) „x“ koje uključuju kolonu vreme *Time* i *Ukupan_protok.Qsuma_1* kao ulazne varijable za model. DTR se koristi za predviđanje vrednosti broja aktivnih bunara koristeći kolonu *Broj_bunara_u_radu* (broj aktivnih bunara) na osnovu karakteristika u „x“. Evaluacija modela takođe koristi standardne proračune greške za regresiju, kao što su MAE, MSE, i R^2 Score.

Kod MAE i MSE, niže vrednosti ukazuju na bolji model. Idealno, MAE i MSE bi trebali biti što manji, a za R^2 visoka vrednost (bliža 1) ukazuje da model dobro objašnjava varijaciju u podacima. Na osnovu istorijskih podataka, karakteristike „x“ predstavljaju ulazne podatke za sam model, dok su ciljevi „y“, broj aktivnih bunara. Model koristi podatke za kreiranje stabla koje deli podatke u različite grane na osnovu odluka koje maksimizuju razlike u ciljevima između grana. Na svakom čvoru stabla, model postavlja pitanje o karakteristikama (vreme i protok) kako bi se odlučilo koja grana treba da se prati. Ove odluke su bazirane na merama kao što su Gini indeks, entropija (u slučaju klasifikacije) ili varijanse (u slučaju regresije). U samom algoritmu, model odlučujućeg stabla za regresiju koristi varijansu kao meru za odluke o granama koje treba pratiti. Varijansa se koristi da bi se odredilo koliko su podaci „raspršeni“ u određenim čvorovima. Manja varijansa ukazuje na to da su predviđanja u tom čvoru bliža stvarnim vrednostima, što je poželjno [5].

Varijansa se izračunava kao: $V = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$, gde je y_i stvarna vrednost za uzorak i , \bar{y} srednja vrednost svih stvarnih vrednosti u čvoru, a N broj uzoraka u čvoru. Početni čvor sadrži sve uzorke podataka, izračunava se varijansa ciljne vrednosti (broj bunara) za sve uzorke u ovom čvoru. Model razmatra sve moguće granice koje se mogu postaviti na osnovu vrednosti karakteristika za vreme i protok. Deljenje čvora, gde se podaci dele u dva nova čvora na osnovu granica, zatim se izračunava varijansa za nova dva čvora, i bira najbolja granica.

Odabir granice koja najbolje smanjuje varijansu, se radi na osnovu smanjenja varijanse, a koristeći formulu smanjenje varijanse: $S_V = v_{PRE} - \left(\frac{N_L}{N} v_L + \frac{N_R}{N} v_R \right)$, gde je V_{pre} varijansa pre deljenja, V_L i V_R su varijanse u leve i desne grane nakon deljenja, N_L i N_R su broj uzoraka u levoj i desnoj grani nakon deljenja, a N je ukupan broj uzoraka pre deljenja. Ovaj postupak se ponavlja za svaku novu granu dok se ne postignu kriterijumi za zaustavljanje (npr. minimalna dubina stabla ili minimalni broj uzoraka u čvoru). U ovom modelu odlučujućeg stabla za regresiju, odluke o tome koje grane treba pratiti bazirane su na smanjenju varijanse. Smanjenje varijanse omogućava modelu da kreira predviđanja koje su što preciznije i koje bolje objašnjavaju stvarne ciljne vrednosti broja aktivnih bunara. Varijansa se koristi kao mera za izbor najboljih odluka i deljenje podataka unutar stabla [4].

3. Upporedna komparativna analiza modela

KNN (K-nearest Neighbors) – osnovni princip je instancijski model, pripadajući, koji se oslanja na pristup „sličnosti“ za predviđanje. U osnovi, model analizira sve obučene podatke i identifikuje K najbližih tačaka prema matrici udaljenosti (Euklidskoj). Za regresiju, KNN koristi srednju vrednost ili ponderisanu srednju vrednost izlaza najbližih K tačaka. Obuka KNN-a je relativno brza jer se sastoji samo

u skladištenju obučavajućih podataka, nema posebne faze obuke gde se model parametarski podešava. Decision Tree Regressor je model koji pravi odluke bazirane na vrednostima atributa u stablu odluka. Obuka Decision Tree uključuje konstrukciju stabla donošenja odluka kroz proces deljenja podataka na osnovu atributa koji najbolje razdvaja podatke u različite grupe. Ovo se radi pomoću kriterijuma poput varijanse ili entropije za regresiju. KNN i DTR (Decision Tree Regressor) su različiti u svojoj prirodi i pristupima obuci. KNN je jednostavan, fleksibilan i neparametarski, dok je Decision Tree Regresor parametarski i koristi pravila odlučivanja koja mogu biti lakše interpretirana. Izbor između njih zavisi od specifičnih zahteva problema, veličine i prirode podataka, kao i od željenih karakteristika modela.

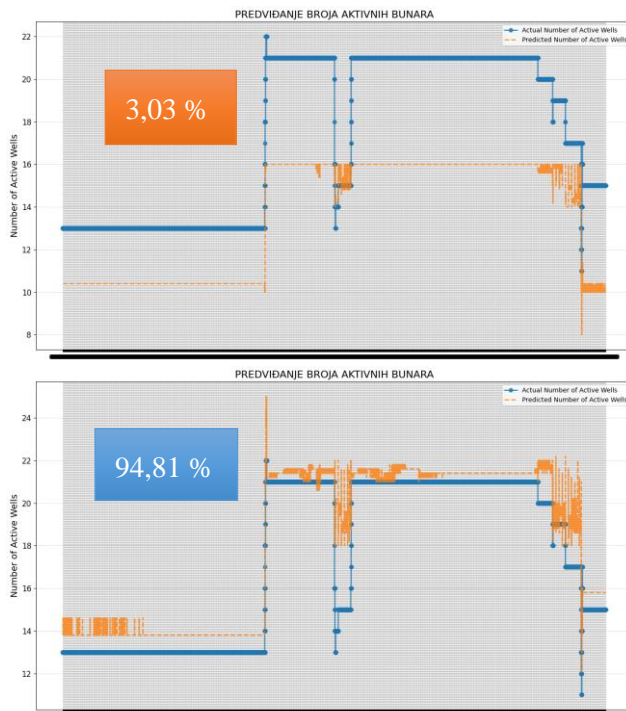
Ako su odnosi između protoka i broja aktivnih bunara jednostavni i linearni, KNN može biti dobra opcija, posebno ako su podaci čisti i nema previše problema sa „šumom“. Šum u podacima odnosi se na nepravilnosti koje se javljaju u podacima zbog različitih faktora. KNN je jednostavan za implementaciju i može raditi dobro sa malim do srednjim veličinama podataka. Ako su odnosi nelinearni i kompleksni, Decision Tree Regressor može biti bolji izbor jer može da obuhvati složene odnose i pruži interpretaciju kroz pravila odlučivanja. Takođe, može biti korisno koristiti tehnike za obuku stabla kako bi se sprečilo prekomerno obučavanje, kao što su ograničavanje dubine stabla ili korišćenje ensemble metoda kao što je Random Forest.

4. Rezultati korišćenja KNN i DTR modela

Python algoritam koji je napisan za ocenu performansi KNN modela (K-nearest Neighbors), trenira se na skupu istorijskih podataka oko 525 fajlova, odnosno dana (svaki fajl je jedan dan i poseduje oko 86.400 zapisa o protoku i broju aktivnih bunara, ukupno oko 45 miliona zapisa). Zatim se vrši prikupljanje karakteristika i ciljne promenljive, „x“ koja sadrži protok i „y“ koja sadrži ciljnu promenljivu, podaci se dele na obučavanje i testiranje, gde za trening model KNN koristi 5 najbližih susdenih tačaka. Model predviđa vrednosti za testiranje, izračunavaju se performanse modela i čuva najbolji model od svih modela prema željenom kriterijumu MSE, RMSE ili R^2 , gde se na kraju dobija model odnosno datum sa najboljim performansama. Tokom evaluacije i inicijalizacije proveravaju se podaci koji se nalaze u samim fajlovima, filtriraju se na neragularnost („NaN“).

Izvršeno je testiranje, proračun, a kao rezultat dobijen je model sa datumom 26.04.2024. najbolji u pogledu MSE i RMSE kriterijuma (sa ocenom performansi od 0.003 i 0.0612), a model u pogledu R^2 kriterijuma najbolji je bio dana 05.08.2023 sa ocenom performansi od 0.9998. Zatim se za predviđanje broja aktivnih bunara na temelju novih podataka Q (l/s) protoka, koji nisu viđeni tokom obuke modela, gde je kao primer uzet protok za datum 17.08.2024. To znači da ako se primeni model

za dan 26.04.2024. (MSE, RMSE) i model na dan 05.08.2023. (R^2) i iskoristi za predviđanje broja aktivnih bunara i uporede dobijena predviđanja sa stvarnim realnim, slika 1, dobijaju se sledeći rezultati:

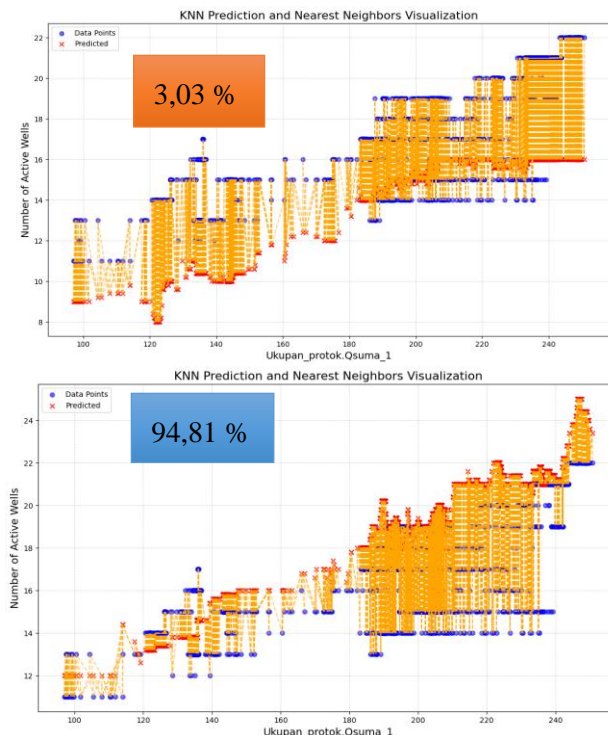


Slika 1. Prikaz rezultata predviđanja i preklapanja broja aktivnih bunara tokom 24 sata, sa 3.03% tačnost predviđanja (MSE i RMSE kriterijum na dan 26.04.2024.), i 94,81% tačnost predviđanja (R^2 kriterijum na dan 05.08.2023.)

Figure 1. Display of prediction results and overlap of the number of active wells over 24 hours, with 3.03% prediction accuracy (MSE and RMSE criteria as of April 26, 2024) and 94,81% prediction accuracy (R^2 criteria as of August 5, 2023).

Različiti kriterijumi za ocenjivanje kao što su MSE i RMSE (3,03%) mere prosečnu kvadratnu grešku predviđanja i stvarnih vrednosti, što može biti osetljivo na velike greške, odnosno da je moguće da je prilikom izrade modela došlo do grešaka na nekoliko tačaka. Dok R^2 (94,81%) može pokazivati bolji opšti trend, jer meri koliko varijabilnosti u zavisnoj varijabli objašnjavaju model, i može biti korisnije za ocenu modela kada se procenjuju generalni šabloni i odnosi u podacima. Predstavljanje predviđanja i načina na koji KNN mašinsko učenje vrši predviđanje korišćenjem modela za gore navedene datume, 26.04.2024. i 05.08.2023, moguće je prikazati u vidu linjskog dijagrama.

Na dijagramu plave tačke predstavljaju stvarne podatke, crveni markeri koji predstavljaju predviđanja, a isprekidane linije predstavljaju linije koje povezuju tačke predviđanja sa najbližim susedom odnosno tačkama u prostoru podataka slika 2.



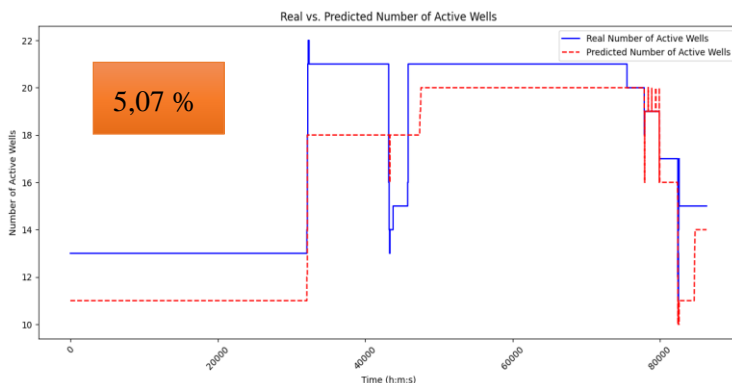
Slika 2 Vizuelni prikaz učenja i testiranja KNN modela

Figure 2. Visual representation of training and testing a KNN model

Kako bi mogli steći bolju sliku o tome kako bi svi modeli predvideli broj aktivnih bunara, od svih naučenih modela 524 modela (jedan dan je imao grešku „NaN“ vrednosti) korišćenu su za predviđanje broja bunara na primeru protoka na dan 17.08.2024. Kao rezultat toga, tačnost predviđanja je sledeća (za proveru tačnost modela korišćen je metod preklapanja jer je bolji pokazatelj koliko su predviđanja blizu stvarnim): tačnost predviđanja od 0-25% - 194 modela, tačnost od 25-50% - 146 modela, tačnost od 50-75% -102 modela, a oko 82 modela mogu da predvide broj bunara sa tačnošću od 75-100%. Što znači da od svih modela 15,6% (82 od 524 modela) ima tu mogućnost predviđanja sa tačnošću od preko 75%.

Kada je u pitanju DTR (Decision Tree Regressor), isti princip rada algoritma je primenjen, prikupljeni su istorijski podaci oko 524 fajlova, odnosno dana, obrađeno i izvršena je evaluacija koristeći ponovo MSE, RMSE i R2 kriterijume. Kako bi se

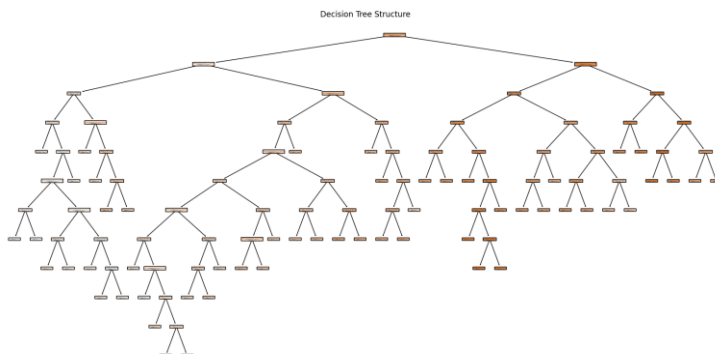
od svih raspoloživih modela, odabrali samo modeli sa najboljim performansama odnosno kriterijuma, odabran je model sa datumom 09.05.2023., sa rezultatima MSE=0,001, RMSE=0,0108 i $R^2=1$. Ovako niske vrednosti gotovo savršene, može ukazati na prilagođenost, pogotovo ako podaci imaju šum ili su složeni. Koristeći ovaj model za isti datum na primeru protoka (dan 17.08.2024) utvrđeno je da je procenat preklapanja dosta loš svega 5,07% za sva tri kriterijuma. Kako bi se stekla bolja slika o tome kako DTR modeli predviđaju broj bunara na osnovu svih kreiranih modela, dobijeni su sledeći rezultati: sa tačnošću preklapanja od 0-25% - 403 modela, sa tačnošću od 25-50%-104 modela, od 50-75%-14 modela, a svega 3 modela ima tačnost predviđanja od preko 75-100%. Što znači da od svih modela svega 0,6% (3 modela) ima tu mogućnost predviđanja sa tačnošću preko 75%, slika 3 (plava boja prave vrednosti broja bunara, a crvenom bojom predviđeni broj). Preprilagođavanje se dešava kada model uhvati čak i najsitniji šum i detalj, zbog toga se treba uvesti „kros-validacija“ kako bi se uverili da model dobro generalizuje na neviđene podatke [1].



Slika 3. Linijski dijagram preklapanja predviđenog i stvarnog broja bunara – procenat preklapanja (5,07%)

Figure 3. Line diagram showing the overlap between the predicted and actual number of wells – percentage of overlap (5.07%)

Na osnovu analize raspoloživih podataka, može se zaključiti sledeće, da je korišćenjem mašinskog učenja Decision Tree, predviđanja korišćenjem najboljeg modela sa kriterijumom performansi R^2 za svih 524 dana: broj dana sa tačnošću predviđanja od 0-25% je 397 dana, od 25-50% tačnost predviđanja ima čak 107 dana, od 50-75% svega 17 dana, a najmanja tačnost je bila za tačnost 75-100% svega 3 dana. Što znači da prilikom mašinskog učenja i izrade modela, podaci su imali određene nepravilnosti i veliki broj šumova, pa je zato model imao slabije performanse. Prikaz strukture stabla, uključujući čvorove, grane i listove prikazano je na slici 4, kako je model doneo odluke na osnovu ulaznih podataka.

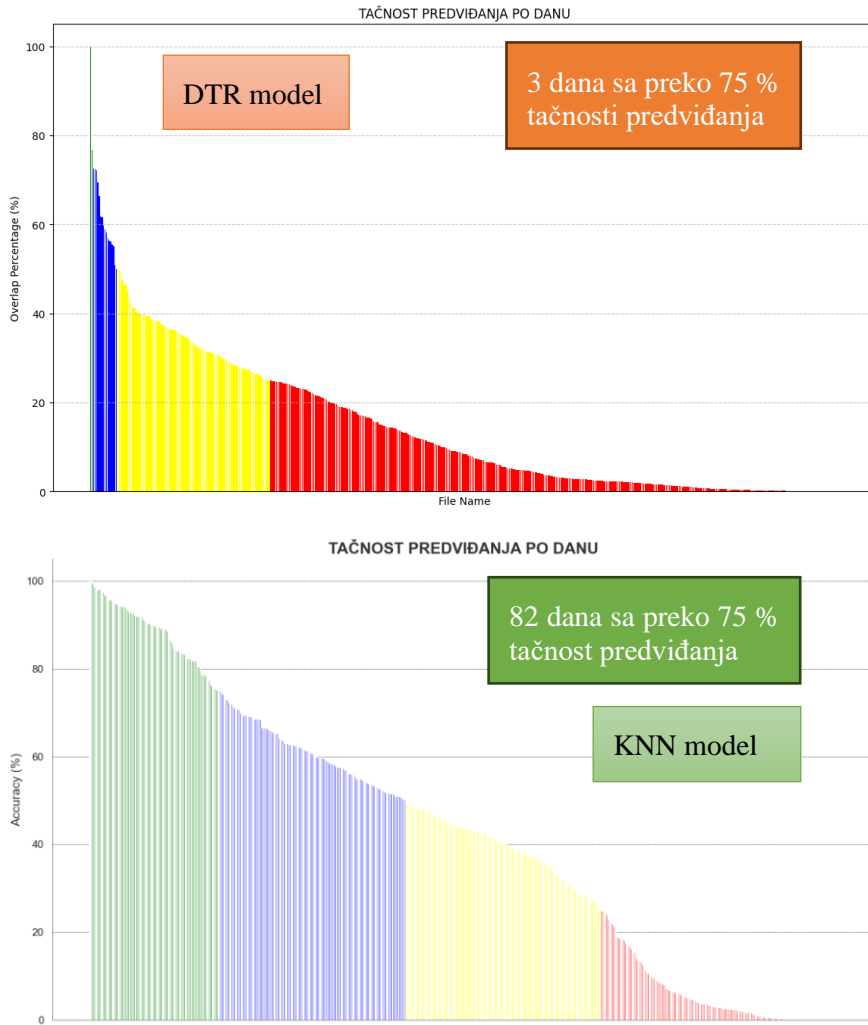


Slika 4. Vizuelni prikaz učenja i testiranja najboljeg Decision Tree modela
Figure 4. Visual representation of training and testing for the best Decision Tree model

5. Zaključak

Ako se uporede obe metode mašinskog učenja Decision Tree i KNN svaka ima svoje prednosti i mane. Decision Tree je relativno lak za interpretaciju, jer predstavlja odluke u obliku stabla, ali može biti previše osetljiv na šum u podacima, što može dovesti do preteranog prilagođavanja modela sa podacima. Kada je KNN u pitanju, on je otporniji na šum u podacima, jer zasniva se na predviđanja na sličnostima sa najbližim susedom. Obzirom da se radi o istorijskim podacima, koji imaju tragove šuma, tokom provere i tokom performansi modela može se zaključiti da je KNN bolji izbor mašinskog učenja za pravljenje modela, ako se podesi broj suseda („k“ npr. 5). KNN ima tendenciju da smanji uticaj šuma putem agregacije rezultata iz susednih tačaka. Ukoliko bi se ovi istorijski podaci mogli kontrolisati ili se podaci mogu dodatno prečistiti i uvođenjem npr. „kros-validacije“, tada bi mogli imati bolje performanse predviđanja (od 80-100% preklapanja). Obzirom da je reč o manipulaciji radom bunara od strane operatera procesa proizvodnje vode, odnosno čoveka, on može dosta otežati predviđanje, jer može uticati na nezakonitost ili pojavu šuma u podacima. Ljudski faktor ima značajan uticaj na rad bunara, pa bi trebalo razmotriti mogućnost uključivanje dodatnih karakteristika koje mogu uhvatiti informacije ili koristiti naprednije modele koji mogu bolje da generalizuju u takvim situacijama [7].

Kako bi se bolje prikazali rezultati predviđanja oba modela, za model KNN i Decision Tree uzet je model sa najboljim R^2 kriterijumom i predviđajući broj aktivnih bunara, za svaki dan pojedinačno. Ovo je urađeno sa ciljem da se vidi u kom procentu i sa kojom tačnošću bi modeli predvideli broj aktivnih bunara tokom perioda od 524 dana, slika 5.



Slika 5. Rezultati predviđanja korišćenjem najboljeg R^2 modela na svim istorijskim podacima (dana 524)

Figure 5. Prediction results using the best R^2 model on all historical data (524 days)

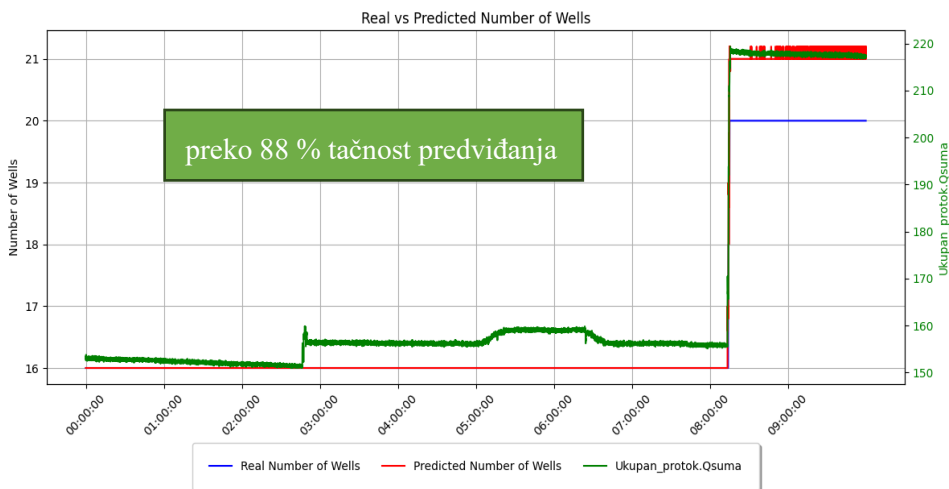
KNN model može biti koristan alat za predviđanje broja bunara i optimizaciju upravljanja vodom, ali uspeh zavisi od kvaliteta podataka, broja dostupnih podataka, i pažljivog podešavanja modela. Korišćenje podataka u realnom vremenu (kumulativni protok sa bunarskog polja) za predviđanje broja bunara omogućava dinamičko i proaktivno upravljanje sistemom za snabdevanje vodom. To može

poboljšati efikasnost, smanjiti troškove, povećati pouzdanost, i omogućiti bržu reakciju na promene i probleme. Ova proaktivna i adaptivna strategija može značajno doprineti stabilnosti i efikasnosti sistema za proizvodnju vode. Ako je potrebno koristiti veći broj bunara za postizanje istog kapaciteta, a modeli pokazuju loša predviđanja, to može značiti da se kapacitet pojedinih bunara smanjuje. Preporučuje se redovno ažuriranje modela i analiza stanja bunara kako bi se preciznije predviđao broj bunara potreban za određeni protok.

Takođe, važno je pratiti promene u kapacitetu bunara i koristiti te informacije za poboljšanje tačnosti modela. Velika odstupanja između predviđanja i stvarnih vrednosti mogu biti indikator da model ispravno prepoznaje potrebu za većim brojem bunara, dok stvarni kapacitet bunara opada. U tom slučaju, potrebno je ažurirati model i uzeti u obzir promene u kapacitetu bunara kako bi se poboljšala tačnost predviđanja i optimizovala upravljanja resursima [2].

U svakom slučaju, implementacija modela može doprineti poboljšanju efikasnosti i smanjenju troškova u proizvodnji vode. Primer prikaza rezultata korišćenjem najboljeg KNN modela, sa greskom R^2 , prikazano je na slici 6.

Na tom dijagramu predviđa se broj bunara na osnovu podataka u trenutku na dan 22.08.2024. (od 00:00-9:30 sati), i predviđa se broj aktivnih bunara u odnosu na realan broj bunara u trenutku (plavom linijom realan broj bunara, crvenom bojom predviđen broj bunara, a zelenom bojom kretanje protoka Q (l/s).



Slika 6. Rezultati predviđanja u realnom vremenu korišćenjem najboljeg R^2 KNN modela – tačnost predviđanja 88,30%

Figure 6. Prediction results in real time using the best R^2 KNN model – prediction accuracy: 88.30%

6. Literatura

- [1] Donghyun Kim, Sijung Choi, Sungkyu i Kuiseong Noh. Studija o razvoju modela predviđanja i klasifikacije potrebe za vodom na osnovu veštačke inteligencije za stanicu za prijem vode u Guryeu, MDPI (online), *Water* 2023, 15(23), 4160, 2023.
- [2] Isha Bansal. K-Nearest Neighbors (KNN) in Python. Posećeno dana : 04.08.2024. Preuzeto sa: <https://www.digitalocean.com/community/tutorials/k-nearest-neighbors-knn-in-python>, 2022.
- [3] Jovanović Jelena *Mašinsko učenje*. Posećeno dana: 10.08.2024. https://ai.fon.bg.ac.rs/wp-content/uploads/2016/10/ML_intro_2016.pdf, 2020.
- [4] Jocović Vladimir, *Inteligentni sistemi*. Posećeno dana : 09.08.2024. http://ri4es.etf.rs/materijali/vezbe/IS_Masinsko_ucenje.pdf, 2023.
- [5] Nihal Desai & Vatsal Patel (2021) Linear Decision Tree Regressor: Decision Tree Regressor Combined with Linear Regressor. Posećeno dana: 10.08.2024. Preuzeto sa: https://www.researchgate.net/publication/353546682_Linear_-Decision_-Tree_Regressor_Decision_Tree_Regressor_Combined_with_Linear_Regressor, 2021.
- [6] Pranay Rishith, *Understanding K-Nearest Neighbors: A simple approach to classification and regression*. Posećeno dana: 12.08.2024. Preuzeto sa: <https://towardsai.net/p/l/understanding-k-nearest-neighbors-a-simple-approach-to-classification-and-regression>, 2023.
- [7] Zsolt Nagy, *Osnove veštačke inteligencije i mašinskog učenja*. Beograd, Kompjuter biblioteka, 2019.