

PRIMENE SOFTVERSKIH ALATA U BIG DATA ANALIZI SENZORSKIH PODATAKA U SAOBRAĆAJU

Slađana Janković, Dušan Mladenović
Univerzitet u Beogradu - Saobraćajni fakultet,
s.jankovic@sf.bg.ac.rs, d.mladenovic@sf.bg.ac.rs

Rezime: Za prikupljanje podataka o drumskom saobraćaju koriste se statički senzori integrisani u putnoj infrastrukturi i senzori dostupni na vozilima i uređajima za navigaciju. Senzori različitih vrsta omogućavaju: brojanje i klasifikaciju vozila na saobraćajnicama, merenje prosečne brzine vozila, proveru zauzetosti saobraćajnih traka, kao i automatsku detekciju incidenata (vožnja u suprotnom smeru, objekti pali sa vozila, požar ili saobraćajna nezgoda). Senzorski podaci u drumskom saobraćaju poseduju sva četiri ključna Big Data obeležja, koja se odnose na: količinu, brzinu pristizanja novih podataka, raznovrsnost u pogledu struktuiranosti podataka i nepouzdanost. Ovaj rad istražuje mogućnosti primene softverskog alata Weka u obavljanju data mining poslova u Big Data analizi senzorskih podataka u drumskom saobraćaju. Weka omogućava: pripremu podataka, klasifikaciju, klasterovanje, selekciju relevantnih atributa, učenje pravilom asocijacije i vizuelizaciju podataka. U radu je demonstrirana primena tehnike klasifikacije u alatu Weka, na podacima koji predstavljaju karakteristike intenziteta saobraćajnih tokova, na izabranim deonicama puteva u Republici Srbiji, na mesečnom nivou.

Ključne reči: Big Data analitika, Weka, mašinsko učenje, senzorski podaci, automatski brojači saobraćaja

1. Uvod

Intelligentni transportni sistemi koriste Big Data tehnologije sa ciljem da povećaju energetske efikasnost, unaprede bezbednost saobraćaja, smanje zagađenje vazduha i ublaže saobraćajne gužve [1]. Različiti sistemi za prikupljanje podataka o saobraćajnim tokovima i Big Data tehnologije za skladištenje i obradu prikupljenih podataka, omogućili su ekspanziju primene neparametarskih metoda analize podataka u ovoj oblasti [2]. Senzorske tehnologije danas predstavljaju jedan od glavnih stubova na kojima počiva prikupljanje podataka o saobraćaju. Senzorski podaci poseduju sva ključna Big Data obeležja, koja se odnose na: količinu, brzinu pristizanja novih podataka, raznovrsnost u pogledu struktuiranosti podataka i nepouzdanost podataka. Stoga ova vrsta podataka zahteva primenu neke od tehnika Big Data analize. U ovom radu istraživane su mogućnosti primene jedne od neparametarskih metoda predikcije (nadgledanog

mašinskog učenja) za analizu podataka o saobraćaju, prikupljenih uz pomoć senzorskih tehnologija.

Predviđanje intenziteta protoka vozila smatra se ključnim elementom za uspešno projektovanje i razvoj inteligentnih transportnih sistema, posebno naprednih informacionih sistema u prevozu putnika, naprednih sistema za upravljanje saobraćajem, naprednih sistema javnog prevoza kao i operacija sa komercijalnim vozilima [3]. U ovom istraživanju cilj je bio da se, na raspoloživom skupu podataka o prosečnom mesečnom dnevnom saobraćaju (PMDS) na državnim putevima u Republici Srbiji, obuče i testiraju prediktivni modeli mašinskog učenja i izvrši predikcija PMDS za izabrani vremenski period u budućnosti. Imajući u vidu raspoloživu količinu podataka, prirodu definisanog problema i tehniku koja je izabrana za njegovo rešavanje, ovaj zadatak može se svrstati u zadatke *Big Data* analitike. U cilju rešavanja ovog zadatka istražene su mogućnosti *data mining* softverskog alata *Weka*. U *Weka* okruženju realizovane su sve faze procesa mašinskog učenja i kreirani različiti klasifikacioni i regresioni modeli za predikciju PMDS.

U drugoj sekciji rada biće reči o ulozi senzora u savremenom saobraćaju. Treća sekcija rada sadrži kratak opis studije slučaja, realizovane u okviru ovog istraživanja. U četvrtoj sekciji rada analizirane su neke od mogućnosti primene softverskog alata *Weka* u *Big Data* analizi senzorskih podataka o saobraćaju. U petoj sekciji rada prikazani su i analizirani najznačajniji rezultati studije slučaja. Na kraju rada predstavljeni su zaključci o mogućnostima korišćenja softverskog alata *Weka* u *Big Data* analizi senzorskih podataka u saobraćaju.

2. Primena senzora u drumskom saobraćaju

Širom sveta gradske vlasti i preduzeća zadužena za upravljanje putnom infrastrukturom sve više koriste pametne tehnologije radi uvida i analize saobraćajne situacije u realnom vremenu. Ti podaci im omogućavaju da donose ispravne odluke, prilagođavaju saobraćaj realnoj situaciji na terenu i stalno unapređuju kvalitet života ljudi [4]. Jedna od mera, koja se već pokazala veoma efikasnom u mnogim gradovima, je instalacija saobraćajnih senzora za kontrolu rada svetlosnih saobraćajnih signala. Pametni senzori prikupljaju saobraćajne podatke u realnom vremenu, na raskrsnicama i glavnim ulicama koje predstavljaju takozvana "uska grla" u saobraćaju. Praćenjem saobraćaja na raskrsnicama i ulicama, šeme saobraćajnih signalizacija mogu se prilagoditi obimu saobraćaja, što za rezultat ima smanjenje vremena koje vozila provedu u praznom hodu. Na taj način povećava se protočnost vozila i pešaka i olakšava gradski saobraćaj. Takođe, detektori prisustva pešaka mogu da omoguće pešacima odgovarajuću dužinu trajanja zelenog signala, kao i da isti signal ne aktiviraju ako niko ne prilazi pešačkom prelazu.

Važna uloga saobraćajnih senzora je automatska detekcija incidenata, kao što su: vožnja u suprotnom smeru, objekti pali sa vozila, požar ili saobraćajna nezgoda. Posebno je izražena upotreba pametnih senzora za automatsku detekciju incidenata na mostovima i u tunelima. Detekcija rizičnog ponašanja učesnika u saobraćaju u realnom vremenu omogućava nadležnim službama u kontrolnim centrima da pravovremeno reaguju i primene odgovarajuće mere, na primer - da zaustave voz.

Saobraćajni senzori se takođe mogu koristiti za sakupljanje različitih podataka o saobraćaju (brojanje i klasifikacija vozila, merenje prosečne brzine, zauzetost saobraćajnih traka). Podaci o saobraćaju koriste se u:

- dimenzionisanju putnih deonioca, priključaka i raskrsnica,
- dimenzionisanju kolovoznih konstrukcija,
- dimenzionisanju mera za zaštitu okoline (npr. mere zaštite od buke),
- ekonomskom i finansijskom vrednovanju projekata,
- planiranju upravljanja i održavanja putne infrastrukture [5].

U različitim fazama koriste se različite kategorije podataka o saobraćaju. Tako na primer, u fazi planiranja putno saobraćajne infrastrukture, koristi se pokazatelj koji se zove prosečan godišnji dnevni saobraćaj (PGDS) po vrstama vozila, odvojeno za putnički i teretni saobraćaj.

3. Studija slučaja

Na mreži državnih puteva I kategorije u Republici Srbiji instaliran je 391 automatski brojač saobraćaja. Automatski brojači saobraćaja vrše detekciju i klasifikaciju vozila u realnom vremenu, 24 sata u toku dana, 365 dana u godini, u svim vremenskim uslovima, pomoću induktivnih petlji koje se postavljaju u isečen i zaliven utor na asfaltnom sloju kolovozne konstrukcije. Jedan takav brojač saobraćaja prikazan je na slici 1.



Slika 1. Automatski brojač saobraćaja baziran na induktivnim petljama

Brojači tipa QLTC-10C kontinuirano broje i klasifikuju vozila u deset kategorija, dok brojači tipa QLTC-8C klasifikuju vozila u osam kategorija. QLTC-10C brojači, npr. razvrstavaju vozila u sledeće kategorije: A0 - Motocikli, A1 - Putnički automobili i Putnički automobili sa prikolicom, A2 - Kombinovana vozila i Kombinovana vozila sa prikolicom, B1 - Laka teretna vozila i Laka teretna vozila sa prikolicom, B2 - Srednje teška teretna vozila, B3 - Teška teretna vozila, B4 - Teška teretna vozila sa prikolicom, B5 - Tegljači sa poluprikolicom, C1 - Autobusi, C2 - Zglobni autobusi, X - Nekategorisana (ostala) vozila.

Za svako vozilo koje detektuje, brojač evidentira: datum, vreme, smer kretanja vozila, redni broj vozila u tom danu za posmatrani smer, saobraćajnu traku, kategoriju vozila i brzinu kretanja vozila. Dobijeni podaci se skladište na memorijske SD (*Secure Digital*) kartice.

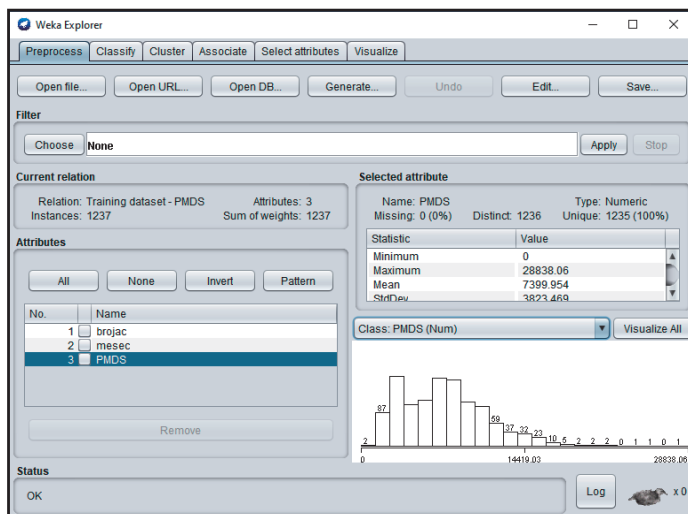
U studiji slučaja korišćeni su podaci dobijeni automatskim brojanjem saobraćaja na državnim putevima u Republici Srbiji, na 21 brojačkom mestu, u periodu od 2011. do 2018. godine. Izabrana brojačka mesta imaju sledeće oznake, odnosno nazive: 1025

(Kraljevo 2), 1026 (Trstenik), 1027 (Pojate), 1046 (Vodice), 1050 (Prijanovci), 1052 (Pridvorica), 1057 (Prijepolje), 1156 (Mojsinje), 1157 (Mrčajevci), 1183 (Trupale Bg-Ni), 1191 (Ineks), 1193 (Kneževići), 1194 (Zlatibor), 1195 (Kokin Brod 2), 1196 (Nova Varoš), 1198 (Gorjani), 1202 (Međuvršje), 1207 (Prijepolje 2), 1208 (Velika Župa), 1225 (Lučina) i 1270 (Preljina). Cilj studije slučaja bio je da se na osnovu istorijskih podataka o prosečnom mesečnom dnevnom saobraćaju (PMDS) na izabranim brojačkim mestima, izvrši predikcija ovog istog pokazatelja intenziteta saobraćaja za narednu kalendarsku godinu.

Za obučavanje modela mašinskog učenja korišćeni su podaci koji se odnose na period od 2011. do 2015. godine, dok su za testiranje modela korišćeni preostali raspoloživi podaci, koji se odnose na period od 2016. do 2018. godine. Za neke brojače nisu bili raspoloživi podaci o PMDS za svaki mesec u posmatranom periodu, tako da se skup podataka za trening sastojao od 1237 instanci, a skup podataka za testiranje od 765 instanci. Atributi korišćenih skupova podataka bili su: brojač (oznaka brojača) i mesec, dok je ciljna varijabla (klasa) bila PMDS.

4. Big Data analiza u softverskom alatu Weka

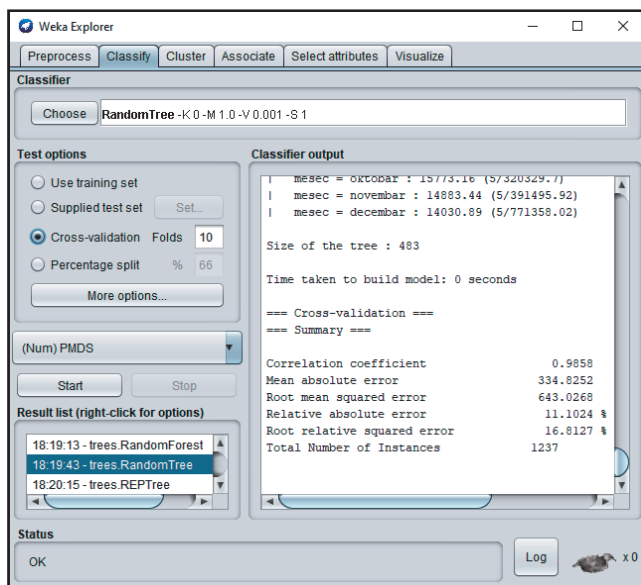
Softverski alat *Weka* (*Waikato Environment for Knowledge Analysis*) je softver otvorenog koda razvijen na Univerzitetu Waikato na Novom Zelandu. *Weka* omogućava obavljanje različitih *data mining* poslova, kao što su: priprema podataka za analizu, klasifikacija, regresiona analiza, klasterovanje, učenje pravilom asocijacije, selekcija relevantnih atributa i vizuelizacija podataka [6]. Svaki od navedenih poslova obavlja se u zasebnom prozoru grafičkog korisničkog interfejsa softvera *Weka* (*Weka Explorer*), a otvara se izborom odgovarajuće kartice *Weka Explorer*-a (slika 2). Prozor *Preprocess*, prikazan na slici 2, omogućava učitavanje i pripremu raspoloživog skupa podataka za kasniju analizu.



Slika 2. Grafički korisnički interfejs softverskog alata Weka – prozor za pripremu podataka

Weka omogućava da se podaci koji će biti predmet analize učitaju iz datoteke, sa URL (*Uniform Resource Locator*) adrese ili iz baze podataka. Osnovni format datoteka koje se mogu učitati u *Weka* okruženje je ARFF (*Attribute-Relation File Format*). To je ASCII (*American Standard Code for Information Interchange*) tekstualna datoteka u kojoj je jedan skup instanci opisan istim atributima. Osim ARFF datoteka i CSV (*Comma-Separated Values*) datoteke pogodne su za učitavanje podataka u *Weka* softver. Učitani skup podataka može se i editovati u prozoru *Preprocess*. Editovanje može da se obavi ručno ili automatizovano, tj. primenom nekog od ugrađenih filtera. Filteri se mogu primenjivati na atribute ili na instance skupa podataka. Filteri atributa omogućavaju različita konvertovanja vrednosti atributa, iz jedne forme u drugu, kao npr.: *NominalToBinary*, *DateToNumeric*, *StringToNominal*, *ChangeDateFormat*, *Discretize*, itd. Filteri instanci omogućavaju sortiranje i filtriranje instanci prema različitim kriterijumima, kao npr.: *Randomize*, *RemovePercentage*, *RemoveRange*, *RemoveWithValues*, *RemoveFrequentValues*, *SubsetByExpression*, itd. U prozoru *Preprocess* atributi raspoloživog skupa podataka mogu se i ručno isključivati i ponovo uključivati. Na slici 2 prikazan je prozor *Preprocess* nakon što je u *Weka Explorer* učitani skup podataka za trening, korišćen u studiji slučaja. Na ovom prozoru mogu se videti nazivi i tipovi atributa skupa podataka, ukupan broj instanci skupa podataka, osnovni statistički podaci o vrednostima svakog atributa, kao i histogram raspodele vrednosti svakog atributa. U ovom prozoru može se izabrati atribut koji će predstavljati zavisnu varijablu ili klasu (engl. *class*) ili se može izabrati da posmatrani skup podataka nema klasu, tj. da je neoznačen.

Prozor *Classify*, prikazan na slici 3, omogućava obučavanje i testiranje modela mašinskog učenja baziranih na različitim klasifikacionim i regresionim algoritmima.



Slika 3. Prozor za izgradnju, validaciju, testiranje i primenu modela mašinskog učenja u softverskom alatu *Weka*

Klasifikacija se odnosi na diskretnu predikciju, odnosno klasifikovanje podataka u definisan, konačan broj klasa. Regresija se, sa druge strane, odnosi na kontinualno predviđanje. Kod regresije, algoritam uči i predviđa kako će se funkcija ponašati u budućnosti, na osnovu istorijskih podataka. Veoma važna funkcionalnost *Weka* softvera je da on prepoznaje i nudi algoritme koji se mogu primeniti na učitanoj skupu podataka. Raspoloživost algoritma uslovljena je pre svega prirodom ciljne promenljive, ali i tipovima atributa skupa podataka. Neki algoritmi mogu se primeniti i na nominalne i na numeričke klase. Kod jednog broja algoritama tolerišu se nedostajuće vrednosti klase, dok kod drugih algoritama za svaku instancu mora postojati vrednost klase. Za svaki algoritam koji je implementiran u softveru *Weka*, jasno su definisana ograničenja koja se odnose na primenu posmatranog algoritma. Algoritmi koji se ne mogu primeniti na učitanoj skupu podataka, ne mogu se ni izabrati u grafičkom korisničkom interfejsu *Weka* softvera (onemogućeni su u prozoru *Classify*). Zbog svega navedenog prozor *Classify* ne može se ni otvoriti dok se ne učita skup podataka u prozoru *Preprocess*.

Svi raspoloživi klasifikatori (algoritmi mašinskog učenja) u *Weka* okruženju podeljeni su na sledeće kategorije: *bayes*, *functions*, *lazy*, *meta*, *misc*, *rules*, *trees*. Kada je u pitanju predikcija bilo kog pokazatelja intenziteta saobraćaja, izraženog u broju vozila u jedinici vremena, softverski alat *Weka* nudi različite algoritme uz pomoć kojih se može pokušati predikcija ove numeričke ciljne varijable (broj vozila). Neki od tih algoritama su: *GaussianProcesses*, *LinearRegression*, *MultilayerPerceptron*, *SMOreg* (iz kategorije *functions*), *IBk*, *KStar*, *LWL* (iz kategorije *lazy*), *M5P*, *RandomForest*, *RandomTree*, *REPTree* (iz kategorije *trees*) [7]. Kakvi će biti rezultati predikcije to se ne može znati unapred. Tek kada se na raspoloživom skupu podataka obuče i testiraju modeli mašinskog učenja i utvrde njihove performanse, postaje jasno da li je neki od raspoloživih *Weka* algoritama upotrebljiv za predikciju konkretne ciljne varijable. Najbolje je da se nad istim skupom podataka primene različiti algoritmi i uporede performanse njima odgovarajućih modela mašinskog učenja [8]. Na osnovu performansi modela donosi se zaključak da li je neki algoritam primenljiv i koji algoritam daje najbolje rezultate.

Nakon izbora algoritma mašinskog učenja, u prozoru *Classify* bira se jedna od četiri ponuđene metode verifikacije modela: korišćenjem celokupnog skupa podataka za obučavanje modela, korišćenjem posebno pripremljenog skupa podataka za testiranje, unakrsna validacija i procentualna podela raspoloživog skupa podataka na skup podataka za obučavanje i skup podataka za testiranje.

Klikom na komandno dugme *Start* u prozoru *Classify* pokreće se proces mašinskog učenja. Rezultati tog procesa prikazuju se u *Classifier output* segmentu prozora *Classify*, a sastoje se od: reprezentacije izgrađenog modela, predikcije dobijene primenom tog modela i performansi modela. Za prediktivne modele kod kojih je ciljna varijabla numerička, kao što je slučaj kod predikcije broja vozila, performanse modela su: koeficijent korelacije, srednja apsolutna greška, kvadratni koren srednje kvadratne greške, relativna apsolutna greška i kvadratni koren relativne kvadratne greške (slika 3).

Kod modela iz kategorije *trees* *Weka* okruženje nudi i mogućnost vizuelizacije stabla odlučivanja.

Važna funkcionalnost softverskog alata *Weka* je mogućnost čuvanja obučenog modela mašinskog učenja na medijumu eksterne memorije. Sačuvani model može se

kasnije učitati u *Weka* okruženje kako bi se izvršila njegova re-evaluacija na skupu podataka za testiranje. Ova mogućnost unapređuje efikasnost procesa mašinskog učenja.

5. Rezultati i analiza rezultata

U softverskom alatu *Weka* realizovan je proces mašinskog učenja i predikcija prosečnog mesečnog dnevnog saobraćaja (PMDS) na izabranim brojačkim mestima, za narednu kalendarsku godinu.

U tabeli 1 prikazane su performanse sedam modela predikcije PMDS merene na skupu podataka za trening. Lako se uočava da modeli bazirani na algoritmima iz kategorije *functions* (*Linear Regression* i *SMOreg*) imaju značajno slabije performanse od ostalih pet kreiranih i obučenih modela. Poređenja radi, koeficijent korelacije ova dva modela iznosi oko 0.4, dok koeficijent korelacije ostalih modela iznosi preko 0.9. Takođe, sve mere greške ova dva modela imaju vrednost za red veličine veću od vrednosti mera greške ostalih modela. Na taj način, za testiranje su kao najbolji izabrani modeli koji na skupu podataka za trening imaju koeficijent korelacije veći od 0.9. To su modeli bazirani na sledećim algoritmima: *IBk*, *M5P*, *Random Forest*, *Random Tree* i *REPTree*. U tabeli 2 prikazane su njihove performanse merene na skupu podataka za testiranje.

Tabela 1. Performanse sedam modela predikcije merene na skupu podataka za trening

Algoritam	Koeficijent korelacije	Srednja apsolutna greška	Kvadratni koren srednje kvadratne greške	Relativna apsolutna greška (%)	Kvadratni koren relativne kvadratne greške (%)
Linear Regression	0.4097	2885.822	3487.1333	95.6906	91.175
SMOreg	0.3891	2883.9394	3538.0019	95.6281	92.505
IBk	0.9843	337.5301	676.6501	11.1921	17.6918
M5P	0.9052	1011.7437	1669.4497	33.5483	43.6496
Random Forest	0.9852	374.6306	657.77	12.4223	17.1981
Random Tree	0.9858	334.8252	643.0268	11.1024	16.8127
REPTree	0.9102	931.0314	1582.7682	30.8719	41.3832

Ako se uporede performanse modela merene na skupu podataka za testiranje, sa performansama modela merenim na skupu podataka za obučavanje, uočava se da kod algoritama *IBk*, *Random Forest* i *Random Tree* nešto bolje performanse imaju modeli za trening. Ovakvi rezultati pokazuju da kod tih modela mašinskog učenja ne postoji problem prevelikog podudaranja (eng. *overfitting*) i da se mogu koristiti za predikciju. Za modele bazirane na algoritmima *M5P* i *REPTree* može se uočiti da je prisutan problem prevelikog podudaranja, tako da se oni ne mogu koristiti za predikciju.

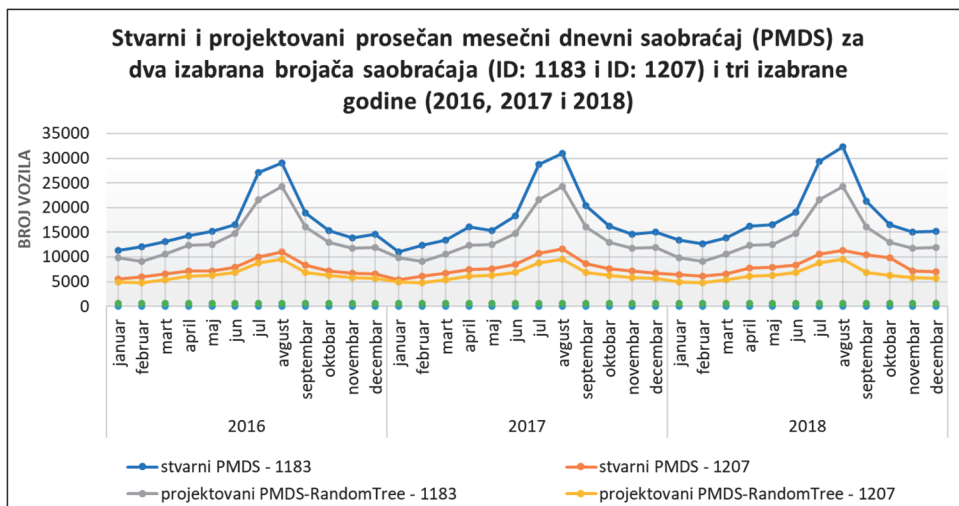
Analizirajući performanse modela *IBk*, *Random Forest* i *Random Tree* merene na skupu podataka za testiranje (tabela 2) uočava se da model baziran na algoritmu *Random Tree* ima najbolje performanse. Naime, ovaj model ima koeficijent korelacije 0.9812, kao i model baziran na algoritmu *IBk*, ali su njegove mere greške nešto manje od

greška *IBk* modela. Stoga je za predikciju vrednosti ciljne varijable u budućnosti izabran model baziran na algoritmu *Random Tree*.

Tabela 2. Performanse pet najboljih modela predikcije merene na skupu podataka za testiranje

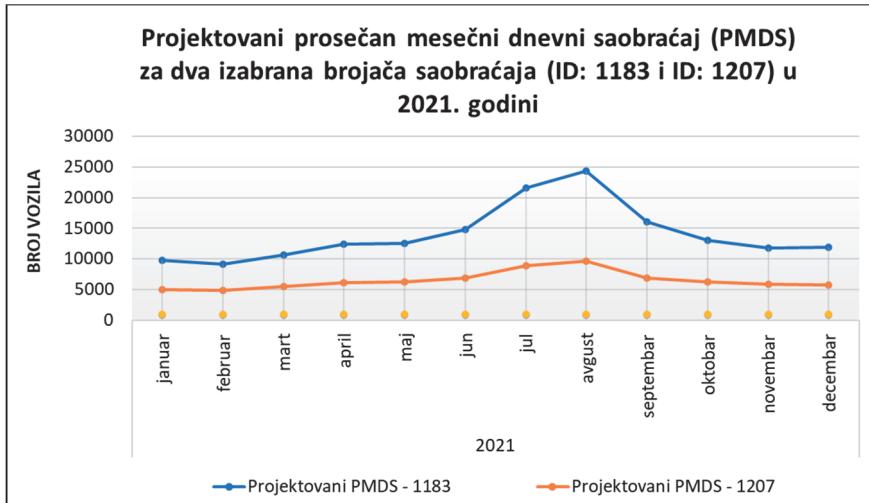
Algoritam	Koeficijent korelacije	Srednja apsolutna greška	Kvadratni koren srednje kvadratne greške	Relativna apsolutna greška (%)	Kvadratni koren relativne kvadratne greške (%)
IBk	0.9812	1267.7229	1609.0712	35.7986	34.9354
M5P	0.9477	1608.0937	2222.224	44.8492	47.0715
Random Forest	0.9807	1256.0444	1623.694	35.0307	34.3934
Random Tree	0.9812	1253.33	1607.7056	34.955	34.0547
REPTree	0.9561	1481.1124	1888.083	41.3077	39.9937

Na grafikonu prikazanom na slici 4 može se videti odnos stvarnog i projektovanog PMDS za dva izabrana brojača saobraćaja (1183 i 1207) i period od 2016-2018. Projekcija PMDS izvršena je korišćenjem modela baziranog na algoritmu *Random Tree*. Na grafikonu se jasno vidi da predikcija PMDS izvršena na skupu podataka za testiranje dobro prati stvarne vrednosti PMDS u posmatranom periodu (slika 4).



Slika 4. Stvarni i projektovani prosečan mesečni dnevni saobraćaj (PMDS) za dva izabrana brojača saobraćaja i tri izabrane godine

Rezultati predikcije PMDS na dva brojačka mesta (1183 i 1207) za 2021. godinu prikazani su na slici 5.



Slika 5. Projektovani prosečan mesečni dnevni saobraćaj (PMDS) za dva izabrana brojača saobraćaja u 2021. godini

6. Zaključak

U sprovedenoj studiji slučaja, u softverskom alatu *Weka*, obučavani su modeli mašinskog učenja za predikciju prosečnog mesečnog dnevnog saobraćaja, bazirani na algoritmima: *Linear Regression*, *SMOreg*, *IBk*, *M5P*, *Random Forest*, *Random Tree* i *REPTree*. Na skupu podataka za trening, modeli bazirani na regresionim stablima odlučivanja i model baziran na *IBk* algoritmu pokazali su značajno bolje performanse od modela iz kategorije *functions* (*Linear Regression* i *SMOreg*). Stoga su na skupu podataka za testiranje testirani samo ovi modeli. Kod modela baziranih na algoritmima *IBk*, *Random Forest* i *Random Tree* nije uočen problem prevelikog podudaranja, dok je kod modela *M5P* i *REPTree* ovaj problem prisutan. Među zadovoljavajućim modelima (*IBk*, *Random Forest* i *Random Tree*) najbolje performanse imao je model baziran na algoritmu *Random Tree*, tako da je predikcija prosečnog mesečnog dnevnog saobraćaja izvršena korišćenjem ovog modela.

Softverski alat *Weka* pokazao je sledeće značajne osobine: posedovanje ugrađenih filtera koji omogućavaju različite transformacije podataka u fazi pripreme podataka; automatsko prepoznavanje algoritama koji se mogu koristiti na raspoloživom skupu podataka i efikasna primena obučenog modela na novim skupovima podataka. Na taj način nametnuo se zaključak da se *Weka* može efikasno koristiti u *Big Data* analizi senzorskih podataka u saobraćaju.

Zahvalnica

Ovaj rad delimično je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije.

Literatura

- [1] D. Ni, "Traffic Sensing Technologies - Characteristics, Experimental Methods, and Numerical Techniques", *Traffic Flow Theory*, Waltham, 2015, pp. 14-29.
- [2] S. Janković, S. Zdravković, D. Mladenović, S. Mladenović, A. Uzelac, "Predikcija obima saobraćaja korišćenjem regresionih stabala odlučivanja", *XLVII Simpozijum o operacionim istraživanjima - SYM-OP-IS '20, 20-23. septembar 2020, Beograd, Srbija, Zbornik radova*, str. 287-292, Univerzitet u Beogradu – Saobraćajni fakultet
- [3] Y. Lv, Y. Duan, W. Kang, Z. Li and F. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach", *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865-873, 2015.
- [4] eKapija, "Pametni senzori unapređuju kvalitet saobraćaja", 07.06.2018., Available at: <https://www.ekapija.com/news/2139946/pametni-senzori-unapredjuju-kvalitet-saobracaja-uredjaji-kompanije-flir-systems-analiziraju-podatke-i>
- [5] Javno preduzeće Putevi Srbije, *Priručnik za projektovanje puteva u Republici Srbiji*, Javno preduzeće Putevi Srbije, Beograd 2012.
- [6] I. Witten, E. Frank, M. Hall, C. Pal, *Data Mining, Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Publishers, 2016.
- [7] Y. Wang, and I. H. Witten, *Induction of model trees for predicting continuous classes*, The University of Waikato, Hamilton, New Zealand, 1996.
- [8] Y. Xu, Q. Kong, and Y. Liu, "Short-term traffic volume prediction using classification and regression trees", *Proceedings of the 2013 IEEE Intelligent Vehicles Symposium (IV)*, 2013, Gold Coast, Australia, pp. 493-498.

Abstract: *Static sensors integrated in the road infrastructure, sensors located on vehicles and navigation devices are used to collect road traffic data. Various types of sensors enable counting and classification of vehicles on roads, measuring the average speed of vehicles, checking the occupancy of traffic lanes, as well as automatic detection of incidents (driving in the opposite direction, objects falling from vehicles, fire or traffic accident). Sensory data in the road traffic has all four key Big Data features, relating to: volume, speed of generating new data, diversity in terms of data structure and unreliability. This paper explores the possibilities of applying the Weka software tool in performing data mining operations in Big Data analysis of sensor data in the road traffic. Weka enables: data preparation, classification, clustering, selection of relevant attributes, association rule learning and data visualization. The paper demonstrates the application of the classification technique in the Weka tool, on data that represent the characteristics of traffic flow intensity, on selected sections of roads in the Republic of Serbia, on a monthly basis.*

Keywords: *Big Data analytics, Weka, machine learning, sensory data, automatic traffic counters*

APPLICATIONS OF SOFTWARE TOOLS IN BIG DATA ANALYSIS OF SENSOR DATA IN TRAFFIC

Sladana Janković, Dušan Mladenović