

Prediktivno modelovanje kritičnog intervala sljeđenja zasnovano na tehnikama mašinskog učenja

DUNJA M. RADOVIĆ, Univerzitet u Istočnom Sarajevu,
Saobraćajni fakultet Doboj, Bosna i Hercegovina

MIRKO D. STOJČIĆ, Univerzitet u Istočnom Sarajevu,
Saobraćajni fakultet Doboj, Bosna i Hercegovina

Prethodno saopštenje

UDC: 004.85

DOI: 10.5937/tehnika2203354R

Zbog nemogućnosti direktnog mjerenja kritičnog intervala sljeđenja, razvijene su brojne metode i procedure za njegovo procjenjivanje. U ovom radu koristi se metoda maksimalne vjerodostojnosti za procjenu istog na pet kružnih raskrsnica, a na osnovu dobijenih rezultata i parova prihvaćenih i maksimalno odbačenih intervala, obučeno je i testirano više prediktivnih modela zasnovanih na tehnikama mašinskog učenja. Dakle, osnovni cilj istraživanja jeste kreiranje modela za predikciju (klasifikaciju) kritičnog intervala sljeđenja koji kao ulaze tj. nezavisne varijable koristi parove – prihvaćeni i maksimalno odbaćeni intervali. Osnovni zadatak modela jeste pridruživanje jedne od prethodno procijenjenih vrijednosti kritičnog intervala sljeđenja datom ulaznom paru intervala. Finalni prediktivni model se bira između više ponuđenih alternativa i to na osnovu tačnosti predikcije. Rezultati obuke i testiranja različitih modela zasnovanih na tehnikama mašinskog učenja u softveru IBM SPSS Modeler ukazuju da najveću tačnost predikcije pokazuje model stabla odlučivanja C5 (73,266%), koji je obučen i testiran na proširenom setu podataka dobijenim metodom augmentacije ili uvećanja skupa podataka (Data Augmentation - DA).

Ključne riječi: kritični interval sljeđenja, prediktivno modelovanje, mašinsko učenje, DA, stablo odlučivanja

1. UVOD

Pretpostavlja se da je kritični interval sljeđenja najkraći interval koji će vozač prihvatiti. U [1] kritični vremenski interval sljeđenja definiše se kao minimalno potrebna veličina intervala sljeđenja u glavnom toku koja omogućava jednom vozilu iz sporednog toka prolazak kroz središte raskrsnice. Vozači koji svojim vozilima namjeravaju da stupe u glavni (kružni) tok koriste svaki interval sljeđenja koji je jednak ili veći od kritičnog kako bi izvršili prolazak kroz raskrsnicu. Kritični interval sljeđenja ne može biti direktno izmjeren, a sve što je poznato je da je on manji od bilo kog prihvaćenog intervala, a veći od bilo kog odbijenog intervala. Tako se bar očekuje, jer ponekad vozači odbijaju jedan interval samo da bi kasnije prihvatili kraći interval. Međutim, smatra se da je ponašanje vozača

uglavnom dosljedno i jednolično i da će pažljivi i obazrivi vozači čekati veći interval kako bi stupili u glavni tok. U prethodno publikovanim istraživanjima dati su različiti modeli za procjenu vrijednosti kritičnog intervala sljeđenja, koji su zasnovani na terenskim istraživanjima, matematičkim metodama, statističkim metodama, a sve sa osnovnom pretpostavkom da će vrijednost kritičnog intervala sljeđenja biti u intervalu između minimalnog prihvaćenog i maksimalnog odbaćenog intervala. Zbog nemogućnosti direktnog mjerenja kritičnog intervala sljeđenja, razvijene su brojne metode i procedure za njegovo procjenjivanje. U ovom radu je korišćena metoda maksimalne vjerodostojnosti za njegovu procjenu, a na osnovu dobijenih rezultata i parova prihvaćenih i maksimalno odbačenih intervala za pet kružnih raskrsnica, obučeno je i testirano više prediktivnih modela zasnovanih na tehnikama mašinskog učenja. Kreirani modeli treba da izvrše klasifikaciju, tj. da datom paru intervala pridruže jednu od prethodno procijenjenih vrijednosti kritičnog intervala sljeđenja.

Do danas nije publikovan značajan broj radova koji se odnose na prediktivno modelovanje kritičnog intervala sljeđenja zasnovano na tehnikama mašinskog

Adresa autora: Dunja Radović, Univerzitet u Istočnom Sarajevu, Saobraćajni fakultet Doboj, Doboj, Vojvode Mišića 52, Bosna i Hercegovina

e-mail: dunja.radovic@sf.ues.rs.ba

Rad primljen: 17.02.2022.

Rad prihvaćen: 23.02.2022.

učenja. Autori su u radu [2] predstavili model za procjenu prostornog kritičnog intervala sljeđenja, koji je zasnovan na vektorima podrške (Support Vector Machines - SVM). Utvrđeno je da su procijenjene prostorne kritične praznine jako bliske onima koje su procijenjene korišćenjem binarnog logit modela. Takođe, rezultati su pokazali da SVM imaju vrlo dobar potencijal da budu alternativni alat za procjenu kritičnog intervala sljeđenja. U radu [3] pretpostavljeno je da ponašanje vozača, koje se odnosi na prihvatanje intervala na raskrsnicama regulisanim saobraćajnim znakom „obavezno zaustavljanje“, zavisi od interakcije brojnih geometrijskih, saobraćajnih i faktora okoline. Korišćeni su modeli vještačke neuronske mreže (Artificial Neural Network - ANN) za predikciju kritičnog intervala sljeđenja na raskrsnicama kontrolisanim saobraćajnim znakom „STOP“.

Autori su u radu [4] primijenili tehniku SVM za utvrđivanje vremenskih i prostornih kritičnih intervala sljeđenja kako na nekontrolisanim raskrsnicama, tako i na pješačkim prelazima između razdjelnih ostrva. Utvrđeno je da SVM metoda može biti uspješno korišćena za klasifikaciju i predviđanje prihvaćenih i odbačenih intervala na osnovu brzine i udaljenosti vozila iz konfliktnog toka. Modelovanje prihvatanja intervala pomoću logit modela je predstavljeno u [5]. U ovoj studiji procijenjeni su kritični intervali sljeđenja za manevre skretanja lijevo i desno i to za različite tipove vozila. Slično istraživanje sa primijenjenim logit modelom je predstavljeno i u [6]. Rad [7] se bavi primjenom tri široko korišćene tehnike mašinskog učenja za predikciju ponašanja vozača pri prihvatanju kritičnih intervala sljeđenja i to SVM, slučajne šume (Random Forest - RF) i stabla odlučivanja (Decision Trees - DT). Podaci o prihvatanju intervala u ovoj studiji su prikupljeni na četvorokrakim nesignalisanim raskrsnicama.

Osnovni cilj istraživanja predstavljenog u radu jeste kreiranje prediktivnog modela kritičnog intervala sljeđenja koji kao ulaze tj. nezavisne varijable koristi parove - prihvaćeni i maksimalno odbaćeni intervali. Zadatak modela jeste klasifikacija, tj. pridruživanje jedne od prethodno procijenjenih vrijednosti kritičnog intervala sljeđenja datom ulaznom paru intervala. Finalni prediktivni model se bira između više ponuđenih alternativa - modela zasnovanih na tehnikama mašinskog učenja i to na osnovu tačnosti predikcije.

Rad je strukturno podijeljen u četiri sekcije, a na kraju rada je dat pregled korišćene literature. Nakon uvodnog dijela, u drugoj sekciji su dati materijali i metode koje su primijenjene u istraživanju. Glavni fokus istraživanja jeste na trećoj sekciji u kojoj su prikazani najvažniji rezultati istraživanja. Zaključna razmatranja sa pravcima daljih istraživanja su data u četvrtoj sekciji.

2. MATERIJALI I METODE

Prvi korak istraživačkog procesa predstavlja prikupljanje podataka neophodnih za proračun kritičnog intervala sljeđenja. S obzirom na to da se prema metodi maksimalne vjerodostojnosti vrijednost kritičnog intervala sljeđenja nalazi između najvećeg odbaćenog intervala i prihvaćenog intervala, analizom video snimaka izdvojeni su prihvaćeni i odbaćeni intervali od strane vozača na sporednim prilazima na kružnim raskrsnicama. Detaljan postupak izdvajanja intervala sljeđenja je ilustrativno predstavljen u [8]. U ovom radu je primijenjen takozvani fotografski metod koji podrazumijeva analizu video snimaka realnog saobraćajnog toka na pet različitih lokacija u Republici Srpskoj (R_1, \dots, R_5). Video snimanje je vršeno sa pozicije koja omogućava dobar pogled na raskrsnicu, bez mnogo prepreka koje ulaze u kadar. Takođe pozicija kamere mora da pruži mogućnost da se u kadru obuhvate svi prilazi kružne raskrsnice kako bi se olakšalo izdvajanje podataka neophodnih za analizu. Poželjno je da kamera bude postavljena na mjestu koje nije vidljivo od strane vozača, kako oni ne bi mijenjali svoje ponašanje i kako bi vozili svakodnevnom rutinom. U skladu sa svim navedenim, sprovedeno je video snimanje pet kružnih raskrsnica u urbanom području gradova Bijeljina i Banja Luka. Za svaku raskrsnicu je izdvojen određen broj parova najvećeg odbaćenog intervala i prihvaćenog intervala koji je dat u tabeli 1.

Tabela 1. Lokacije snimanih raskrsnica i broj izdvojenih parova intervala

Lokacija snimane raskrsnice	Broj izdvojenih parova intervala
R1. Bijeljina, raskrsnica ulica Neznanih junaka, Gavrila Principa, Svetog Save i Filipa Višnjića	92
R2. Bijeljina, raskrsnica ulica Neznanih junaka, Ive Andrića, Stefana Dečanskog i Sremske ulice	32
R3. Bijeljina, raskrsnica ulica Neznanih junaka, Kulina bana i Dušana Baranina	73
R4. Banja Luka, raskrsnica Karadorđeve i Ulice Jovana Raškovića	85
R5. Banja Luka, raskrsnica ulica Teodora Kolokotronisa, Cara Lazara, Isaije Mitrovića i Patre	64
UKUPNO	346

Proračun kritičnog intervala sljeđenja je izvršen na osnovu prethodno prikupljenih vrijednosti parova intervala metodom maksimalne vjerodostojnosti za svaku raskrsnicu. Ova metoda se zasniva na procjeni parametara pretpostavljene distribucije vjerovatnoće. To se

postize maksimiziranjem funkcije vjerovatnoće tako da su, prema pretpostavljenom statističkom modelu, posmatrani podaci najvjerovatniji. Tačka u prostoru parametara koja maksimizira funkciju vjerovatnoće naziva se procjena maksimalne vjerovatnoće. Proračunate vrijednosti kritičnog intervala sljeđenja na ovaj način su predstavljene u tabeli 2.

Tabela 2. Procijenjene vrijednosti kritičnog intervala sljeđenja

Kružna raskrsnica	Kritični interval sljeđenja [s]
R ₁	4,19
R ₂	4,02
R ₃	4,66
R ₄	4,51
R ₅	4,57

Jedno od osnovnih i najvažnijih svojstava tehnika mašinskog učenja predstavlja upravo sposobnost modela da na osnovu trening skupa „nauči“ zavisnost izlaza od ulaza, te na taj način generiše izlaze, tj. izvrši predikciju, za do tada neviđene ulazne vrijednosti. Ovakav način obuke modela mašinskog učenja se naziva učenje sa učiteljem i predstavlja jednu od tri osnovne paradigme učenja u oblasti mašinskog učenja. Prema tome, potrebno je da set raspoloživih podataka bude strukturiran u ulazno-izlazne vektore. Ovo je izvršeno tako što je svakom izdvojenom paru najvećeg odbačenog intervala i prihvaćenog intervala, kao izlazna ili vrijednost zavisne varijable, pridružena pripadajuća proračunata vrijednost kritičnog intervala sljeđenja za određenu raskrnicu. Dakle, ovako dobijen osnovni skup podataka, prema tabeli 1., sastoji se od 346 ulazno-izlaznih vektora. Slika 1. prikazuje metodološke korake istraživačkog procesa.



Slika 1 - Metodološki koraci istraživačkog procesa

S obzirom na to da je izlazna, tj. zavisna varijabla u ovom istraživanju kategorijska, u softveru SPSS Modeler osnovni set podataka je doveden na ulaz čvora

Auto Classifier. Čvor Auto Classifier omogućava ispitivanje više modela zasnovanih na različitim tehnikama mašinskog učenja u jednom ciklusu modeliranja. Ovaj čvor ispituje svaku moguću kombinaciju opcija, rangira svaki model kandidat na osnovu tačnosti predikcije, tj. procenta ispravno klasifikovanih ulaza, te čuva najbolje modele za korišćenje u daljoj analizi.

Imajući u vidu da je prikupljeni osnovni set podataka od 346 ulazno-izlaznih vektora jako mali za primjenu tehnika mašinskog učenja, to uzrokuje nezadovoljavajući nivo tačnosti predikcije, tj. klasifikacije. Ipak, performanse prediktivnih modela je moguće unaprijediti povećavanjem količine podataka, tj. generisanjem novih podataka na osnovu postojećih. Tehnika vještačkog proširenja skupova podataka za obuku modela poznata je kao uvećavanje podataka (Data Augmentation). U ovom istraživanju se u tu svrhu koristi Autoenkoder neuronska mreža koja proširuje osnovni skup podataka sa modifikovanim kopijama podataka na ulazu. Autoenkoder je vještačka neuronska mreža koja po paradigmi učenja bez učitelja uči kako efikasno kompresovati i kodirati podatke, a zatim uči kako rekonstruisati podatke iz ovakvog komprimovanog i kodiranog prikaza u reprezentaciju koja je što bliža originalnom ulazu. Strukturu korišćenog Autoenkodera, koji je prikazan na slici 2, čine dva ulaza (dvije nezavisne varijable), dva čvora u skrivenom sloju i dva čvora u izlaznom sloju, dok funkcije aktivacije imaju sigmoidni oblik. Na ovaj način, dovođenjem određenog broja ulaznih vektora na ulazni sloj mreže, isti broj vektora se dobija na izlazu. Prema tome, jednom opisanom iteracijom skup podataka koji se dovede na ulaz autoenkodera se udvostruči. Generisanim kopijama ulaznih vektora pridružuju se odgovarajuće postojeće vrijednosti kritičnog intervala sljeđenja, čime se kompletira novi set ulazno-izlaznih vektora. U ovom istraživanju korišćena je implementacija Autoenkoder mreže u MATLAB programskom paketu [9].



Slika 2 - Autoenkoder neuronska mreža za augmentaciju podataka

U cilju komparacije performansi predikcije pri različitom procentu dodatno generisanih podataka metodom DA, uvećanje količine podataka je izvršeno postepeno, u tri koraka. U prvom koraku na osnovni skup od 346 vektora dodato je 115 vektora generisanih pomoću Autoenkoder neuronske mreže. U drugom

koraku osnovnom skupu je dodato još 115, što znači da je broj „vještačkih“ vektora jednak 230. Konačno, u posljednjem koraku je generisan dvostruko veći broj vektora u odnosu na osnovni skup, što znači ukupno 692 ulazno-izlazna vektora.

Za svako proširenje, tj. za svaki od navedenih koraka u kojima je primijenjena metoda DA, ponovljena je procedura kreiranja prediktivnih modela u SPSS Modeler okruženju uz pomoć metode automatskog modelovanja čvorom Auto Classifier. Za svaki od tri slučaja, softver kao rezultat prikazuje listu kreiranih prediktivnih modela rangiranih prema tačnosti predikcije. Kao finalni korak istraživačkog procesa, izabran je model koji pri testiranju ima najveći procenat ispravno klasifikovanih ulaza.

3. REZULTATI I DISKUSIJA

3.1. Inicijalni rezultati modelovanja

Na bazi osnovnog seta podataka od 346 ulazno-izlaznih vektora izvršeno je inicijalno prediktivno modelovanje kritičnog intervala sljedećeg. Kao rezultat ovog procesa, IBM SPSS Modeler prikazuje sljedeću listu od tri najbolja modela rangirana prema tačnosti predikcije:

- C5 (tačnost predikcije = 50,578%);
- C&R Tree (tačnost predikcije = 39,595%);
- Bayesian Network (tačnost predikcije = 36,416%).

Prvi, najbolji kreirani model, ima tačnost predikcije jednaku 50,578%, što se može smatrati nedovoljno dobrom performansom. Ovaj model koristi C5.0 algoritam za kreiranje stabla odlučivanja ili skupa pravila. C5.0 model dijeli uzorak na osnovu polja koje pruža maksimalan dobitak informacija (Information Gain). Svaki poduzorak definisan prvom podjelom se zatim ponovo dijeli, a proces se ponavlja sve dok se svi poduzorci ne mogu dalje dijeliti. Konačno, podjele najnižeg nivoa se preispituju, a one koje ne doprinose značajno vrijednosti modela se uklanjaju.

Performansa tačnosti predikcije drugog modela iznosi 39,595%. Klasifikaciono i regresiono stablo (C&R), slično kao C5.0, koristi rekursivno particioniranje da podijeli podatke za obuku u segmente sa sličnim vrijednostima izlaznog polja. Kao rezultat podjele formiraju se dvije podgrupe, od kojih se svaka potom dijeli na još dvije podgrupe i tako dalje, sve dok se ne dostigne jedan od kriterijuma zaustavljanja.

Treći po redu rangirani model ima tačnost predikcije ili klasifikacije jednaku 36,416%. Bayesova mreža je probabilistički grafički model za predstavljanje znanja u domenu neizvjesnosti gdje svaki čvor odgovara slučajnoj varijabli i svaki link predstavlja uslovnu vjerovatnoću za odgovarajuće slučajne varijable [10].

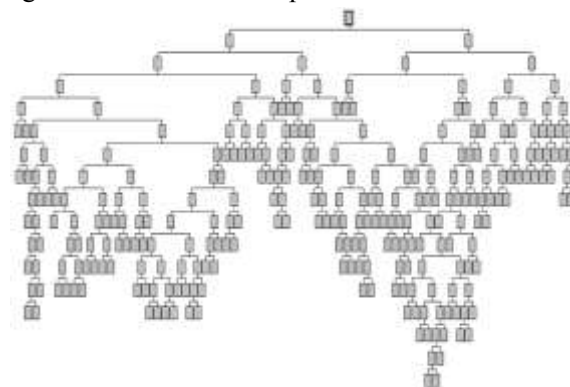
3.2. Rezultati modelovanja na osnovu proširenog skupa podataka

Kao što je prethodno prikazano, performanse predikcije modela kreiranih na osnovu osnovnog skupa podataka nisu zadovoljavajuće. Zbog toga je dalje modelovanje orijentisano na skup podataka koji je proširen podacima generisanim metodom DA. Tabela 3. daje uporedni prikaz tačnosti predikcije (klasifikacije) za tri modela kreirana nakon svakog koraka proširenja osnovnog skupa, tj. za 461, 576 i 692 ulazno-izlazna vektora. Na osnovu podataka prikazanih u Tabeli 3., zaključuje se da najveću tačnost predikcije pokazuje model C5 koji je obučen i testiran na dvostruko većem broju ulazno-izlaznih vektora (692) u odnosu na osnovni skup (346). Prema tome, ovaj model je izabran kao finalni.

Tabela 3. Tačnost predikcije modela kreirana nakon svakog koraka proširenja osnovnog skupa

Model	Tačnost predikcije – klasifikacije (%)		
	Dodatnih 115 vektora ukupno:461	Dodatnih 230 vektora ukupno:576	Dodatnih 346 vektora ukupno:692
C5	57,484	68,229	73,266
C&R Tree	42,299	39,931	41,040
Bayesian Network	42,950	39,583	35,549

Slika 3. prikazuje strukturu stabla odlučivanja predstavljenog izabranim modelom C5. Osim početnog čvora na najvišem hijerarhijskom nivou, koji definiše izlaznu promjenljivu kritični interval sljedećeg, stablo se grana u 16 nivoa sa ukupno 276 čvorova.



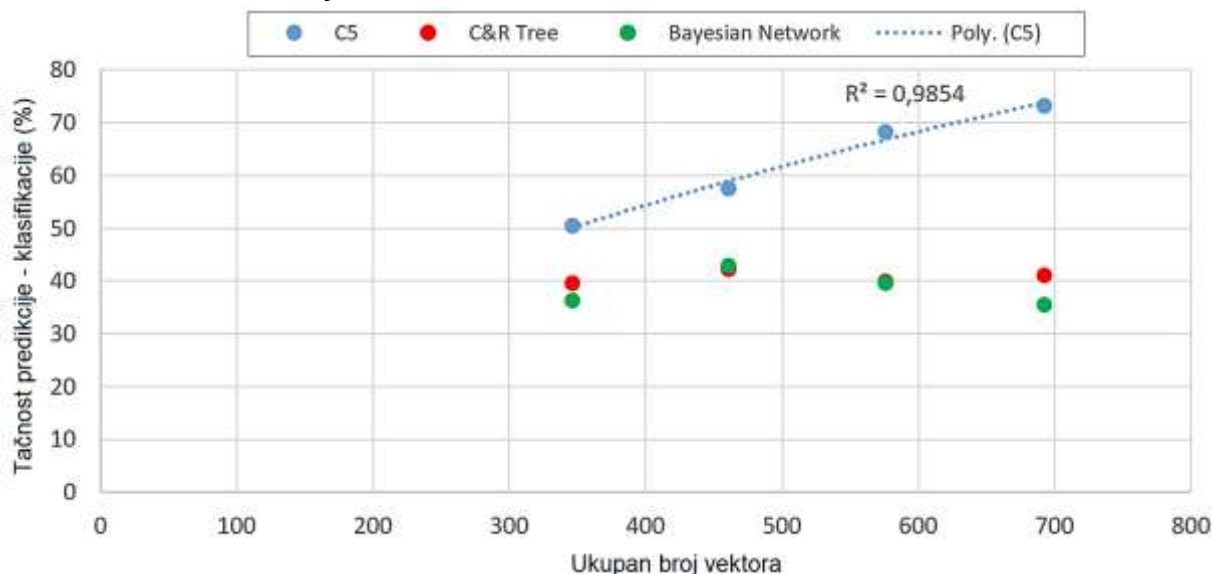
Slika 3 - Struktura izabranog finalnog modela C5

Na slici 4. je prikazan dijagram, koji je konstruisan na osnovu podataka datih u tabeli 3, kao i inicijalnih rezultata modelovanja. Svaka tačka dijagrama je određena ukupnim brojem vektora i tačnošću predikcije kreiranog modela na osnovu takvog skupa. Tačnost predikcije raste sa proširenjem skupa podataka za model C5, što se može iskazati regresionim modelom koji predstavlja kvadratnu funkciju sljedećeg oblika:

$$y = -4 \cdot 10^{-5} x^2 + 0,1064x + 17,682 \quad (1)$$

gdje y predstavlja tačnost predikcije, a x veličinu skupa podataka. Koeficijent determinacije, kao pokazatelj kvaliteta modela, u ovom slučaju iznosi $R^2=0,9854$ što

predstavlja jako visoku vrijednost. Na osnovu toga, moguće je odrediti broj vektora sa kojima je potrebno proširiti osnovni skup da bi se dostigla ciljna vrijednost tačnosti.



Slika 4 - Dijagram tačnosti predikcije u zavisnosti od ukupnog broja vektora

Takođe, na osnovu slike 4. se zaključuje da kod modela C&R Tree i Bayesian Network, pri povećanju seta podataka, ne dolazi do povećanja tačnosti predikcije. Nasuprot tome, dolazi čak do blagoga pada.

4. ZAKLJUČAK

U radu je kreiran model za predikciju – klasifikaciju kritičnog intervala sljeđenja na osnovu ulaznih parova prihvaćenih i maksimalno odbačenih intervala. Metodom augmentacije, osnovni skup podataka je proširen sa svojim „grubim“ kopijama, što je uticalo na povećanje performansi tačnosti modela. Između više alternativa izabran je model zasnovan na stablu odlučivanja (C5), koji ima najveći procenat ispravno klasifikovanih ulaza (73,266%). Iako ovakav rezultat nije izvanredan, u radu je obrazložen način postizanja ciljne vrijednosti tačnosti. Za izabrani model C5, tačnost predikcije raste sa porastom broja vještački generisanih podataka, što je prikazano pomoću regresionog modela. Ipak, potrebno je imati na umu mogućnost da bi se rezultati dobijeni na ovakav način mogli razlikovati od rezultata modela obučenog i testiranog na skupu realnih podataka, iste veličine.

Kao poseban doprinos ovog rada, može se istaći činjenica da nije publikovan značajan broj sličnih istraživanja. Takođe, kao najvažnije novine predstavljene u ovom radu se mogu izdvojiti: obučeni i testirani različiti modeli od kojih su prikazana tri najbolja, novi set ulaznih varijabli (dvije varijable), te primjena metode DA u cilju popravke performansi tačnosti predikcije.

Budući pravci istraživanja mogu biti orijentisani na kreiranje prediktivnih modela na osnovu većeg seta realnih podataka. Osim toga, potrebno je uporediti i analizirati rezultate dobijene na takav način sa onima dobijenim metodom DA.

LITERATURA

- [1] Kuzović Lj. & Bogdanović V, *Teorija saobraćajnog toka*. Fakultet tehničkih nauka, Novi Sad, 2010.
- [2] Pawar D. S. & Patil G. R, Analyzing variations in spatial critical gaps at two-way stop controlled intersections using parametric and non-parametric techniques. *Journal of traffic and transportation engineering (English edition)*, 8(1), 129-138, 2021.
- [3] Pant P. D. & Balakrishnan P, Neural network for gap acceptance at stop-controlled intersections. *Journal of transportation engineering*, 120(3), pp. 432-446, 1994.
- [4] Pawar D. S, Patil G. R. Chandrasekharan A. & Upadhyaya S, Classification of gaps at uncontrolled intersections and midblock crossings using support vector machines. *Transportation research record*, 2515(1), 26-33, 2015.
- [5] Harwood D. W, Mason J. M. & Brydia R. E, Sight distance for stop-controlled intersections based on gap acceptance. *Transportation Research Record* 1701(1), pp. 32–41, 2000.
- [6] Zohdy I, Sadek S. & Rakha H. A, Empirical analysis of effects of wait time and rain intensity on driver left-turn gap acceptance behavior. *Transportation research record*, 2173(1), pp. 1-10, 2010.

- [7] Nagalla R, Pothuganti P. & Pawar D. S, Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests. *Procedia Computer Science*, 109, 474-481, 2017.
- [8] Bogdanović V, & Radović D, Procjena kritičnog intervala sljeđenja vozila na kružnim raskrscima. *Put i saobraćaj*, 66(2), 7-13, 2020.
- [9] Stackoverflow <https://stackoverflow.com/questions/39265746/data-augmentation-techniques-for-general-datasets>
- [10] Yang X. S, Introduction to algorithms for data mining and machine learning. Academic press; 2019. <https://doi.org/10.1016/C2018-0-02034-4>

SUMMARY

PREDICTIVE MODELING OF CRITICAL HEADWAY BASED ON MACHINE LEARNING TECHNIQUES

Due to the impossibility of directly measuring of critical headway, numerous methods and procedures have been developed for its estimation. This paper uses the maximum likelihood method for estimating the same at five roundabouts, and based on the obtained results and pairs of accepted and maximum rejected headways, several predictive models based on machine learning techniques were trained and tested. Therefore, the main goal of the research is to create a model for the prediction (classification) of the critical headway, which as inputs, i.e. independent variables use pairs - accepted and maximum rejected headways. The basic task of the model is to associate one of the previously estimated values of the critical headway with a given input pair of headways. The final predictive model is chosen from several offered alternatives based on the accuracy of the prediction. The results of training and testing of various models based on machine learning techniques in IBM SPSS Modeler software indicate that the highest prediction accuracy is shown by the C5 decision tree model (73.266%), which was trained and tested on an extended data set obtained by augmentation or data set augmentation (Data Augmentation - DA).

Key Words: *critical headway, predictive modeling, machine learning, DA, decision tree*