

## Ocena vozačevog stila upravljanja vozilom primenom nadgledanog mašinskog učenja

SLAĐANA R. JANKOVIĆ, Univerzitet u Beogradu,  
Saobraćajni fakultet, Beograd  
DAVOR B. VUJANOVIĆ, Univerzitet u Beogradu,  
Saobraćajni fakultet, Beograd  
STEFAN S. ZDRAVKOVIĆ, Univerzitet u Beogradu,  
Saobraćajni fakultet, Beograd  
MARKO A. STOKIĆ, Univerzitet u Beogradu,  
Saobraćajni fakultet, Beograd

Originalni naučni nrad  
UDC: 629.3.072.2  
629.048.3:004.85  
DOI: 10.5937/tehnika2202232J

*Nadgledano mašinsko učenje može biti efikasna metoda predikcije neke ciljne varijable, u zavisnosti od trenutnih vrednosti nezavisnih atributa, ukoliko su na raspolaganju istorijski podaci koji sadrže vrednosti nezavisnih atributa i ciljne varijable u prošlosti i ukoliko neki od algoritama mašinskog učenja daje dobre rezultate na raspoloživom skupu podataka. U ovom istraživanju cilj je bio da se pokaže da li se metoda nadgledanog mašinskog učenja može uspešno primeniti u ocenjivanju vozačevog stila vožnje sa aspekta potrošnje goriva, ako se kao nezavisni atributi, tj. parametri koji utiču na ocenu vožnje uzmu: broj obrtaja motora, procenat pritiska papučice gasa i ubrzanje vozila. Obučavanje, validacija, testiranje i primena modela mašinskog učenja obavljani su u softverskom alatu Weka. Na skupovima podataka za obučavanje i testiranje modela primenjeno je sledećih sedam algoritama mašinskog učenja: LinearRegression, MultilayerPerceptron, IBk (k-najbližih suseda), M5P, Random Forest, Random Tree i REPTree. Najbolje performanse pokazali su modeli bazirani na algoritmima IBk i Random Forest. Kao krajnji rezultat ovog istraživanja dobijene su prognozirane ocene vozačevog stila vožnje u intervalu od jedne sekunde. Finalna ocena vozačevog stila vožnje izračunata je kao aritmetička sredina prognoziranih ocena vožnje za svaku sekundu.*

**Ključne reči:** mašinsko učenje, stil vožnje vozača, Weka, k-najbližih suseda, predikcija

### 1. UVOD

Energetska efikasnost vozila je veoma važna tema u poslednje dve decenije pošto je transport postao sektor sa najvećom potrošnjom finalne energije [1]. Pri tome, drumski transport ima dominantno učešće u ukupnoj potrošnji energije u okviru transportnog sektora [2]. U tom smislu, sprovedena su mnoga istraživanja kako bi se smanjila potrošnja goriva, odnosno kako bi se povećala energetska efikasnost drumskih vozničkih parkova [3, 4, 5].

Stil vožnje vozača je jedan od faktora koji ima značajan uticaj na potrošnju goriva i na energetska efikasnost vozila [6]. Ponašanje vozača je veoma složen

stohastički sistem, koji su brojni autori pokušavali da modeliraju korišćenjem metode mašinskog učenja.

Autori u radu [7] su koristili tehniku klasterovanja, kako bi podelili ponašanja vozača na ona koja su povoljna po životnu sredinu (engl. Eco-friendly) i ona koja to nisu, pri čemu su takođe koristili i algoritme mašinskog učenja i algoritme dubokog učenja (engl. Deep Learning). Ponašanje vozača pratili su kroz sledeće atribute: brzina vozila, broj obrtaja motora, procenat položaja ventila (leptira) za protok goriva, temperatura motora, vrsta goriva koja se koristi u vozilu, nivo goriva u rezervoaru. U radu [8] za modelovanje ponašanja vozača korišćena je tehnika dubokog učenja, kao jedna od podoblasti mašinskog učenja. Ponašanje vozača opisano je pomoću parametra brzina vozila, s tim što je u model uključena i kategorija puta, kao nezavisni atribut. Rezultati MATLAB simulacije sprovedene u ovom istraživanju pokazali su da ovaj model ponašanja vozača dostiže tačnost od 90%.

Adresa autora: Slađana Janković, Univerzitet u Beogradu, Saobraćajni fakultet, Beograd, Vojvode Stepe 305  
e-mail: s.jankovic@sf.bg.ac.rs

Rad primljen: 07.02.2022.

Rad prihvaćen: 29.03.2022.

Autori u radu [9] daju sistematičan pregled 82 rada koja se bave primenom mašinskog učenja u predikciji ponašanja vozača. Prema podacima iz ovog rada, čak 72% od svih analiziranih istraživanja koristi jednu od sledećih vrsta algoritama mašinskog učenja: metode potpornih vektora (engl. Support Vector Machines), neuronske mreže (engl. Neural Networks), Bajesovi algoritmi (engl. Bayesian Algorithms) i ansambl metode (engl. Ensemble Methods). S druge strane, algoritmi tipa stabla odlučivanja (engl. Decision Trees) zastupljeni su u samo 6% radova, a Instance Based algoritmi u 7% radova.

Autori u radu [6] razvili su model fazi logike (engl. Fuzzy Logic) za dobijanje ocene vozačevog stila vožnje kod putničkih automobila, pri čemu su utvrdili da što je ocena vozača bolja to je specifična potrošnja goriva (u l/100 km) njihovih vozila manja, i obrnuto. U radu [10] predložen je model baziran na fazi logikom sistemu tipa 2 (engl. Fuzzy Logic System Type-2, FLS2) za utvrđivanje uticaja različitog stila vožnje profesionalnih vozača na potrošnju goriva, pri čemu su upotrebljeni sledeći parametri vožnje: broj obrtaja motora, ubrzanje/usporenje vozila i procenat pritiska papučiće gasa. Navedeni parametri vožnje su, pored ostalih, upotrebljeni i u radu [11] za vrednovanje uticaja terenskih i saobraćajnih uslova puta na vozačeve performanse vožnje i na energetska efikasnost vozila uz primenu mašinskog učenja. Pri tome, za ocenu vozačevog stila vožnje navedeni parametri snimljeni su samo na jednoj ruti.

Kako bi se potvrdilo da se metoda nadgledanog mašinskog učenja može sasvim verodostojno primeniti za vrednovanje vozačevog stila upravljanja vozilom, u ovom radu je upotrebljen izuzetno veliki skup podataka dobijen snimanjem parametara vožnje na tri različite rute. Za tu svrhu autori su upotrebili raspoložive snimljene podatke za parametre: broj obrtaja motora (BOM), procenat pritiska papučiće gasa (PPPG) i ubrzanje/usporenje vozila (UV), kao i izračunatu ocenu vozača primenom FLS2 metoda, iz rada [10].

Rad je strukturiran na sledeći način. U drugom delu rada opisana je metodologija za ocenjivanje vozačevog stila vožnje, koja je primenjena u ovom istraživanju. U trećem delu rada predstavljeni su rezultati predikcije ocene vozačevog stila vožnje, dobijeni primenom opisane metodologije. Četvrti deo rada sadrži zaključna razmatranja o mogućnostima korišćenja tehnike nadgledanog mašinskog učenja u ocenjivanju vozačevog stila vožnje.

## 2. METODOLOGIJA

Metodologija primenjena u ovom istraživanju pre svega je uslovljena zahtevima metode koja je izabrana za ocenjivanje vozačevog stila vožnje – nadgledanog

mašinskog učenja. Proces mašinskog učenja sastoji se od faza: priprema podataka, treniranje modela, validacija modela, testiranje modela i primena modela, tj. predikcija ciljne varijable [6]. Predikcija ocene vozačevog stila vožnje u svakoj sekundi vožnje realizovana je kroz sve gore navedene faze procesa mašinskog učenja, posebno za svakog od šest različitih vozača, na tri izabrane rute, pri čemu je svaki vozač na svakoj ruti upravljao istim vozilom tokom pet dana u nedelji. Prva ruta je prigradsko-gradska sa ravničarsko-brežuljkastim terenom i dužinom oko 120 km. Druga ruta je međugradska sa ravničarskim terenom i dužinom oko 260 km. Treća ruta je međugradska sa ravničarsko-brdovitim terenom i dužinom oko 330 km. Na kraju je, za svakog vozača, na osnovu predikcija ocena u svakoj sekundi izračunata njegova prosečna ocena vožnje.

### 2.1. Skupovi podataka

Za potrebe ovog istraživanja upotrebljeni su snimljeni podaci o parametrima vožnje i izračunate ocene šest vozača na tri različite rute primenom FLS2 modela iz rada [10]. Dakle, skupovi podataka koji su korišćeni za obučavanje, testiranje i primenu modela mašinskog učenja sastoje se od sledećih atributa:

- broj obrtaja motora (BOM),
- procenat pritiska papučiće gasa (PPPG),
- ubrzanje vozila (UV) i
- ocena vozačevog stila vožnje (OCENA).

OCENA je zavisna (ciljna) varijabla, tj. atribut čije će vrednosti biti projektovane uz pomoć prediktivnih modela. Parametri BOM, PPPG i UV su nezavisni atributi prediktivnih modela, tj. atributi od kojih zavisi ciljna varijabla.

U skupovima podataka za predikciju, pre primene prediktivnog modela, kolona ciljne varijable u svakom redu ima vrednost „?“ prema zahtevima softvera *Weka* u kojem je predikcija rađena. Kada se na skup podataka za predikciju primeni prediktivni model koji je izabran kao najbolji, kolona ciljne varijable biće ispunjena projektovanim vrednostima ciljne varijable.

Za svakog od šest vozača formirana su po tri skupa podataka: za obučavanje modela, za testiranje modela i za primenu modela mašinskog učenja (predikciju OCENE), tako da je pripremljeno i korišćeno ukupno 18 skupova podataka. U tabeli 1 prikazan je broj instanci u svakom od skupova podataka korišćenih u ovom istraživanju. Ukupan broj instanci korišćenih u ovom istraživanju je 1 189 912.

Skupovi podataka pripremljeni su u relacionom sistemu za upravljanje bazama podataka Microsoft Access 2016. Za svakog vozača realizovan je proces mašinskog učenja, počev od pripreme tri skupa podataka, zatim izgradnje i validacije grupe modela mašinskog učenja baziranih na različitim algoritmima, preko

testiranja modela do izbora najboljeg modela i njegove primene u predikciji ciljne varijable.

Tabela 1. Kvantitativni opis skupova podataka

	Broj instanci u skupu podataka za obučavanje	Broj instanci u skupu podataka za testiranje	Broj instanci u skupu podataka za predikciju
Vozač 1	113 704	37 901	37 902
Vozač 2	123 607	41 202	41 204
Vozač 3	132 688	44 229	44 231
Vozač 4	99 833	33 277	33 279
Vozač 5	110 662	36 887	36 888
Vozač 6	133 450	44 483	44 485

## 2.2. Izgradnja i validacija modela mašinskog učenja

Izgradnja, validacija, testiranje i primena modela mašinskog učenja obavljani su u softverskom alatu Weka (Waikato Environment for Knowledge Analysis) [12]. Za predikciju ciljne varijable (OCENA), na skupovima podataka za trening za svih šest vozača, u softverskom alatu Weka, primenjeno je sledećih sedam algoritama mašinskog učenja: LinearRegression (linearna regresija), MultilayerPerceptron (višeslojna neuronska mreža), IBk (k-najbližih suseda), M5P (modifikovan M5 algoritam), Random Forest (model slučajne šume kombinuje više stabala odlučivanja), Random Tree (stablo odlučivanja) i REPTree (stablo odlučivanja bazirano na C4.5 algoritmu). Za validaciju modela primenjena je unakrsna validacija (engl. cross validation), implementirana u softveru Weka.

Unakrsna validacija koristi samo podatke za trening, a sastoji se od sledećih faza:

- Raspoloživi skup podataka za treniranje modela deli se na K jednakih delova-podskupova (engl. folds).
- Model se trenira na K-1 podskupova podataka (npr. na prvih K-1 podskupova).
- Model se ocenjuje na jedinom preostalom (K-tom) podskupu podataka.
- Koraci 2 i 3 ponavljaju se K puta. U svakoj iteraciji uzima se jedan deo podataka za potrebe validacije modela, dok se ostatak (K-1 delova) koristi za učenje. Bira se uvek različit podskup koji će se koristiti za validaciju modela.
- Performanse modela izračunavaju se kao aritmetičke sredine performansi dobijenih u K iteracija.

Validacija (ocenjivanje) modela vrši se za svaki izabrani skup vrednosti hiperparametara modela. Neka su projektovane vrednosti ciljne promenljive, dobijene za skup instanci za validaciju modela:  $p_1, p_2, \dots, p_n$ ;

dok su stvarne vrednosti ciljne promenljive:  $a_1, a_2, \dots, a_n$ .  $\bar{a}$  je prosek stvarnih vrednosti ciljne varijable, dok je  $\bar{p}$  prosek projektovanih vrednosti ciljne varijable. Za ocenjivanje uspešnosti numeričke predikcije korišćeno je nekoliko različitih mera [13], koje su definisane u nastavku.

Srednja kvadratna greška – SKG (1) je glavna i najčešće korišćena mera.

$$SKG = \frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n} \quad (1)$$

Srednja apsolutna greška – SAG (2) je prosečna veličina individualnih grešaka bez uzimanja u obzir njihovog znaka. Srednja kvadratna greška ima tendenciju da preuveličava efekat izuzetaka.

$$SAG = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n} \quad (2)$$

Kvadratni koren srednje kvadratne greške – KKSJKG (3) izračunava se na očigledan način.

$$KKSJKG = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}} \quad (3)$$

Relativna kvadratna greška – RKG (4) uzima ukupnu kvadratnu grešku i normalizuje je tako što je deli ukupnom kvadratnom greškom podrazumevanog prediktora. Podrazumevani prediktor je prosek stvarnih vrednosti ciljne varijable iz podataka za trening, a označava se sa  $\bar{a}$ .

$$RKG = \frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2} \quad (4)$$

Kvadratni koren relativne kvadratne greške – KKRKKG (5) izračunava se na očekivani način.

$$KKRKG = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}} \quad (5)$$

Relativna apsolutna greška – RAG (6) je ukupna apsolutna greška, sa istom vrstom normalizacije.

$$RAG = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{|a_1 - \bar{a}| + \dots + |a_n - \bar{a}|} \quad (6)$$

Poslednja mera tačnosti predikcije je koeficijent korelacije (7), koji meri statističku korelaciju između vrednosti  $a$  i  $p$ .

$$\text{Koeficijent korelacije} = \frac{S_{PA}}{\sqrt{S_P S_A}} \quad (7)$$

gde je:

$$S_{PA} = \frac{\sum_{i=1}^n (p_i - \bar{p})(a_i - \bar{a})}{n-1} \quad (8)$$

$$S_P = \frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n-1} \quad (9)$$

$$S_A = \frac{\sum_{i=1}^n (a_i - \bar{a})^2}{n-1} \quad (10)$$

## 2.3. Testiranje modela mašinskog učenja

Unakrsnom validacijom dobijen je po jedan „naj-

bolji“ model za svaki algoritam, a u tom procesu korišćeni su samo podaci za trening. Da bi se predvidele performanse modela na novim podacima, potrebno je proceniti mere njihovih performansi na skupu podataka koji nije igrao nikakvu ulogu u formiranju modela.

Ovaj nezavisni skup podataka naziva se skupom podataka za testiranje i on mora imati identičnu strukturu i format podataka, kao skup podataka za trening. Testiranje i izbor najboljeg prediktivnog modela odvija se kroz sledeće faze:

- predviđanje ciljane varijable na skupu podataka za testiranje, uz pomoć svakog od modela;
- izračunavanje mera tačnosti predikcije svakog modela, koristeći dobijena predviđanja i vrednosti ciljane varijable iz skupa za testiranje;
- poređenje performansi modela dobijenih na skupu podataka za testiranje sa performansama dobijenim na skupu podataka za trening. Ako model radi veoma dobro na podacima za trening, ali slabo na podacima za testiranje, onda postoji problem prevelikog podudaranja (engl. overfitting);

- izbor najboljeg modela. Najbolji model bi trebalo da ima najbolje performanse na skupu podataka za testiranje, da je dobar prema različitim merama performansi i da ima najbolji ili jedan od najboljih rezultata unakrsne validacije na skupu podataka za trening.

#### 2.4. Primena prediktivnog modela

Skup podataka koji se koristi za predikciju mora imati strukturu i format podataka potpuno identičan strukturi i formatu podataka u skupu za obučavanje i skupu za testiranje modela.

### 3. REZULTATI

Performanse modela predikcije ocene stila vožnje Vozača 1, merene na skupu podataka za trening, prikazane su u tabeli 2, dok su performanse modela predikcije za istu ciljnu varijablu, merene na skupu podataka za testiranje, prikazane u tabeli 3. Podaci u tabelama 2 i 3 pokazuju da najbolje mere performansi ima model baziran na algoritmu IBk (k-najbližih suseda), pa je predikcija ocene stila vožnje Vozača 1 izvršena korišćenjem ovog modela.

Tabela 2. Performanse modela predikcije ocene stila vožnje Vozača 1 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 [%]
LinearRegression	0.8798	1.0796	1.3277	44.2413	47.5346
MultilayerPerceptron	0.9531	0.6196	0.8555	25.3899	30.6307
IBk	0.9998	0.02	0.0571	0.8209	2.0444
M5P	0.9992	0.0688	0.1105	2.8209	3.9572
Random Forest	0.9998	0.0336	0.0533	1.3764	1.9098
Random Tree	0.9994	0.064	0.099	2.6227	3.5454
REPTree	0.9993	0.0704	0.106	2.8847	3.794

Tabela 3. Performanse modela predikcije ocene stila vožnje Vozača 1 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 [%]
LinearRegression	0.8744	1.1068	1.3356	45.7949	48.6155
MultilayerPerceptron	0.9762	0.5087	0.649	21.0459	23.6235
IBk	0.9998	0.0187	0.0559	0.7747	2.0364
M5P	0.9992	0.069	0.1122	2.8567	4.084
Random Forest	0.9998	0.0344	0.059	1.4213	2.1481
Random Tree	0.9993	0.0612	0.101	2.5324	3.6782
REPTree	0.9991	0.0709	0.114	2.9344	4.1479

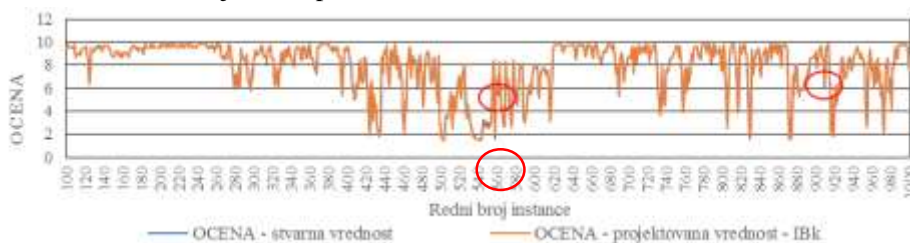
Na slici 1 prikazan je odnos stvarne i projektovane ocene stila vožnje Vozača 1, na uzorku od 900 uzastopnih instanci iz skupa podataka za testiranje. Projektovane vrednosti ocene dobijene su primenom

prediktivnog modela koji je izabran kao najbolji (model baziran na algoritmu IBk – k-najbližih suseda). Na grafikonu koji prikazuje slika 1 može se videti da je plava linija koja prati stvarne vrednosti ocena u

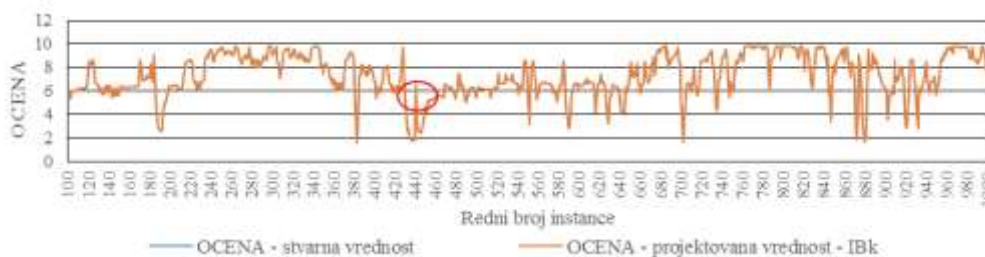
potpunosti preklapljenom narandžastom linijom koja prati projektovane vrednosti ocena stila vožnje Vozača 1. Mala odstupanja između stvarnih i projektovanih ocena, u posmatranom uzorku od 900 instanci, postoje samo na tri mesta. Ta tri mesta na grafikonu su obeležena crvenim elipsama.

Na slici 2 prikazan je odnos stvarne i projektovane ocene stila vožnje Vozača 2, na uzorku od 900 uzastopnih instanci iz skupa podataka za testiranje. Projektovane vrednosti ocene dobijene su primenom

prediktivnog modela koji je izabran kao najbolji (IBk). Na grafikonu koji prikazuje slika 2 može se videti da je plava linija koja prati stvarne vrednosti ocena u potpunosti preklapljenom narandžastom linijom koja prati projektovane vrednosti ocena stila vožnje Vozača 2. Malo odstupanje između stvarne i projektovane ocene postoji samo na jednom mestu, koje je obeleženo crvenom elipsom. Slični dijagrami kreirani su i za sve ostale vozače, ali u radu nisu prikazani zbog nedostatka prostora.



Slika 1 - Stvarna i projektovana ocena stila vožnje Vozača 1 (uzorak od 900 uzastopnih instanci)



Slika 2 - Stvarna i projektovana ocena stila vožnje Vozača 2 (uzorak od 900 uzastopnih instanci)

Tabela 4. Stvarne i projektovane ocene stila vožnje Vozača 1 (uzorak od 20 uzastopnih instanci iz skupa podataka za testiranje)

Redni broj instance	Stvarna OCENA	Projektovana OCENA - IBk
1	5.76	5.727
2	4.62	4.607
3	3.82	3.833
4	3.7	3.74
5	3.68	3.66
6	4.25	4.24
7	4.25	4.21
8	4.17	4.193
9	4.22	4.193
10	4.45	4.457
11	4.38	4.377
12	4.37	4.4
13	6.74	6.723
14	7.12	7.257
15	7.28	7.28
16	5.17	5.2
17	5.05	5.107
18	7.78	7.773
19	9.05	9.05
20	9.15	9.156

U tabeli 4 prikazane su stvarne i projektovane ocene stila vožnje Vozača 1 na uzorku od 20 uzastopnih instanci iz skupa podataka za testiranje.

Iz podataka prikazanih u tabeli 4 vidi se da su stvarna i projektovana ocena u celobrojnom delu uvek identične, a kod većine instanci identične su i na prvoj decimali.

Prosečna stvarna ocena stila vožnje Vozača 1 dobijena na skupu podataka za testiranje je 6.6742 dok je prosečna projektovana ocena istog vozača, dobijena na skupu podataka za testiranje, primenom prediktivnog modela baziranog na algoritmu IBk 6.6759.

Podudaranje stvarne i projektovane prosečne ocene, čak na dva decimalna mesta, takođe pokazuje koliko je uspešan prediktivni model.

Na isti način izvršena je predikcija ocena stila vožnje za preostalih pet vozača.

U tabeli 5 prikazane su stvarne i projektovane prosečne ocene stila vožnje, dobijene primenom metode mašinskog učenja na skupovima podataka za testiranje, za svih šest vozača. Za pet vozača najbolje rezultate dao je algoritam IBk (k-najbližih suseda), tako da je za te vozače predikcija ocena stila vožnje izvršena primenom ovog algoritma. Za vozača pod rednim brojem 3 najbolje performanse imao je model

baziran na algoritmu Random Forest, pa je za ovog vozača predikcija izvršena primenom ovog algoritma. U tabeli 5 može se videti da se prosečne stvarne i projektovane ocene svih šest vozača podudaraju na jednu decimalu, a kod četiri vozača i na dve decimale.

Tabela 5. Stvarne i projektovane prosečne ocene stila vožnje svih šest vozača dobijene na skupu podataka za testiranje

Redni broj vozača	Prosečna stvarna OCENA	Prosečna projektovana OCENA	Primenjeni algoritam mašinskog učenja
1	6.6742	6.6759	IBk
2	6.6635	6.6652	IBk
3	7.8366	7.8377	Random Forest
4	5.0278	5.0306	IBk
5	6.6199	6.6210	IBk
6	7.3061	7.3062	IBk

U radu [11] objedinjeni su podaci koji opisuju vožnje 12 različitih vozača na jednoj ruti, pa je obučavanje modela mašinskog učenja izvršeno nad jednim skupom podataka od 122 559 instanci, testiranje nad jednim skupom podataka od 61 279 instanci, dok je za predikciju ocene vozačevog stila vožnje korišćen jedan skup podataka od 48 588 instanci. U pomenutom radu obučavano je i testirano šest različitih modela, a najbolji rezultat pokazao je model baziran na algoritmu Random Forest, pa je predikcija ocene vozačevog stila vožnje izvršena uz pomoć ovog modela.

U ovom istraživanju korišćeno je znatno više podataka, odnosno ukupno 1.189.912 instanci, za šest vozača, na tri različite rute. Zbog tehničkog ograničenja softverskog alata Weka nije se moglo raditi sa skupovima podataka koji bi objedinjavali sve podatke za svih 6 vozača, već su za svakog pojedinačnog vozača kreirana po tri skupa podataka: za obučavanje, testiranje i predikciju (tabela 1). Za svakog vozača obučavano je i trenirano po sedam modela mašinskog učenja, baziranih na sedam različitih algoritama, što je ukupno 42 modela. Algoritam Random Forest dao je najbolji rezultat samo kod jednog vozača, dok su za ostalih pet vozača najbolje performanse pokazali modeli bazirani na algoritmu IBk (k-najbližih suseda). Iz ovoga se može zaključiti da algoritam Random Forest bolje modelira različitosti u stilovima vožnje različitih ljudi, dok algoritam IBk bolje opisuje stil vožnje pojedinačnog vozača kada se upotrebi veći uzorak podataka.

#### 4. ZAKLJUČAK

U ovom radu predložena je i verifikovana metodologija za vrednovanje vozačevog stila vožnje prime-

nom nadgledanog mašinskog učenja. Skupovi podataka na kojima su obučavani i testirani modeli mašinskog učenja opisuju stilove vožnje šest vozača na tri različite rute. Kao parametri kojima su opisani stilovi vožnje vozača u svakoj sekundi, odnosno nezavisni atributi u modelima mašinskog učenja, izabrani su: broj obrtaja motora, procenat pritiska papučice gasa i ubrzanje vozila, dok je ciljna varijabla ocena vozačevog stila vožnje. Za svakog od šest vozača, u softverskom alatu Weka, obučavani su i testirani modeli mašinskog učenja, bazirani na sledećim algoritmima: Linear Regression, Multilayer Perceptron, IBk (k-najbližih suseda), M5P, Random Forest, Random Tree i REPTree. Svi modeli pokazali su dobre rezultate, dok je pet modela pokazalo odlične rezultate (koeficijent korelacije veći od 0.99). Kod pet vozača najbolje performanse pokazali su modeli bazirani na algoritmu IBk (k-najbližih suseda), dok je kod jednog vozača najbolje rezultate dao algoritam Random Forest.

Istraživanje je pokazalo da se tehnika nadgledanog mašinskog učenja može uspešno primeniti za vrednovanje vozačevog stila vožnje, ako se kao nezavisni atributi koriste broj obrtaja motora, procenat pritiska papučice gasa i ubrzanje vozila, za svaku sekundu vožnje.

Buduća istraživanja biće usmerena na polje primene nenadgledanog mašinskog učenja, kroz tehniku klasterovanja, u cilju otkrivanja paterni, tj. obrazaca u stilovima vožnje vozača.

#### 5. ZAHVALNICA

Ovaj rad delimično je podržan od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije, u okviru projekta pod brojem 036012.

#### LITERATURA

- [1] European Commission, Directorate-General for Mobility and Transport, EU transport in figures: statistical pocketbook 2021, Publications Office, 2021, [citirano 25.03.2022]. Dostupno na: <https://op.europa.eu/en/publication-detail/publication/14d7e768-1b50-11ec-b4fe-01aa75ed71a1/language-en>
- [2] Davis SC, et al. *Transport Energy Data Book, Report of Oak Ridge National Laboratory*, Edition 35, U.S. Department of Energy, Washington D.C., USA, 2016.
- [3] Ang-Olson J, Schroeder W. Energy Efficiency Strategies for Freight Trucking, Potential Impact on Fuel Use and Greenhouse Gas Emissions, *Transport Research Record*, Vol. 1815, No. 02-3877, pp. 11-18, 2002.

- [4] Stokić M, et al. A new comprehensive approach for efficient road vehicle procurement using hybrid danp-topsis method, *Sustainability*, Vol. 12, No. 10, 4044, pp. 1-16, 2020.
- [5] Vujanović D, et al. A hybrid multi-criteria decision making model for the vehicle service center selection with the aim to increase the vehicle fleet energy efficiency, *Thermal Science*, Vol. 22, No. 3, pp. 1549-1561, 2018.
- [6] Stokic M, et. al. Evaluation of driver's eco-driving skills based on fuzzy logic model – A realistic example of vehicle operation in real-world conditions, *Journal of Applied Engineering Science*, Vol. 17, No. 2, pp. 217-223, 2019.
- [7] Peppes N, et. al. Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data, *Sensors*, Vol. 21, No. 14, pp. 4704, 2021.
- [8] Liu J, Jia Y and Wang Y. Development of Driver-Behavior Model Based on WOA-RBM Deep Learning Network, *Journal of Advanced Transportation*, Vol. 2020, Article ID: 8859891, 2020.
- [9] Ellassad ZEA, Mousannif H, Moatassime HA, Kar-kouch A. The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review, *Engineering Applications of Artificial Intelligence*, Vol. 87, 2020.
- [10] Zdravković S, et. al. Evaluation of professional driver's eco-driving skills based on type-2 fuzzy logic model, *Neural Computing and Applications*, Vol. 33, No. 18, pp. 11541-11554, 2021.
- [11] Vujanović et al. Evaluation of the influence of terrain and traffic road conditions on the driver's driving performances by applying machine learning, *Thermal Science, on-line first*, 2022.
- [12] Bouckaert RR, et. al. WEKA Manual for Version 3-9-3, University of Waikato, Hamilton, New Zealand, 2018.
- [13] Witten IH, et. al. *Data mining: practical machine learning tools and techniques Fourth Edition*, Elsevier, Cambridge, United States, 2017.

## SUMMARY

### ASSESSMENT OF THE DRIVER'S DRIVING STYLE USING SUPERVISED MACHINE LEARNING

*Supervised machine learning can be an effective method of predicting a target variable, depending on current values of independent attributes, if historical data containing values of independent attributes and target variables in the past is available and if some machine learning algorithm gives good results on the available data set. In this research, the aim was to show whether the method of supervised machine learning can be successfully applied in assessing the driver's driving style in terms of fuel consumption, if as independent attributes, i.e. parameters that affect the assessment of driving take: engine speed, percentage of accelerator pedal pressure and vehicle acceleration. Training, validation, testing and application of machine learning models were performed in the Weka software tool. The following seven machine learning algorithms were applied to the data sets for model training and testing: LinearRegression, MultilayerPerceptron, IBk (k-nearest neighbors), M5P, Random Forest, Random Tree and REPTree. The best performance was shown by models based on the IBk and Random Forest algorithms. As a final result of this research, predicted scores of the driver's driving style in the interval of one second were obtained. The final score of the driver's driving style was calculated as the arithmetic mean of the predicted driving scores for each second.*

**Key Words:** machine learning, driver's driving style, Weka, k-nearest neighbors, prediction