

# Korekcija navika vozača strojnim učenjem

---

**Kuzman, Jan**

**Undergraduate thesis / Završni rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Electrical Engineering and Computing / Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:168:945348>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-03-29**



*Repository / Repozitorij:*

[FER Repository - University of Zagreb Faculty of Electrical Engineering and Computing repository](#)



SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1577

# KOREKCIJA NAVIKA VOZAČA STROJNIM UČENJEM

Jan Kuzman

Zagreb, lipanj 2024.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 1577

# KOREKCIJA NAVIKA VOZAČA STROJNIM UČENJEM

Jan Kuzman

Zagreb, lipanj 2024.

## ZAVRŠNI ZADATAK br. 1577

Pristupnik: **Jan Kuzman (0036541188)**  
Studij: Elektrotehnika i informacijska tehnologija i Računarstvo  
Modul: Računarstvo  
Mentor: izv. prof. dr. sc. Vinko Lešić

Zadatak: **Korekcija navika vozača strojnim učenjem**

### Opis zadatka:

Transport je jedan od najbrže rastućih svjetskih potrošača energije i zagađivača okoliša stakleničkim plinovima. Stil vožnje i individualno ponašanje u prometu značajno se odražava na potrošnju energenata, a optimizacija profila brzine i ukazivanje na navike korisnika može značajno reducirati potrošnju. U radu je potrebno proučiti načine praćenja GPS podataka korisnika i vozila te pretražiti javno dostupne skupove podataka. Na temelju povijesnih podataka o poziciji i brzini vozila te mjereći vremena trajanja između različitih točaka puta, korisniku je potrebno prikazivati vlastiti stil vožnje u odnosu na trenutni zajednički prometni prosjek za odabranu dionicu gradske vožnje i unaprijed dostupnu bazu podataka iz navigacijskih servisa. Potrebno je zatim načiniti algoritam za predikciju potrošnje goriva zasnovane na strojnom učenju, te vizualno nagrađivati, ukazivati i poticati vozača na poboljšanja stila vožnje s ciljem smanjenja potrošnje goriva.

Rok za predaju rada: 14. lipnja 2024.



# Sadržaj

<b>1</b>	<b>Uvod .....</b>	<b>1</b>
<b>2</b>	<b>Trendovi u cestovnom prometu .....</b>	<b>3</b>
2.1	Važnost cestovnog prometa u brojkama.....	3
2.2	Tamna strana cestovnog prometa u brojkama .....	3
<b>3</b>	<b>Umjetna inteligencija.....</b>	<b>5</b>
3.1	Strojno učenje .....	5
3.1.1	Nadzirano učenje.....	6
3.1.2	Stabla odluke .....	8
3.1.3	Random forest.....	9
<b>4</b>	<b>Podaci voznih ciklusa .....</b>	<b>11</b>
<b>5</b>	<b>Obrada podataka.....</b>	<b>13</b>
<b>6</b>	<b>Analiza rezultata.....</b>	<b>15</b>
6.1	Predikcije brzine vozila.....	15
6.2	Korekcije vozača .....	18
6.3	Pogreške .....	19
<b>7</b>	<b>Zaključak .....</b>	<b>21</b>
	<b>Literatura.....</b>	<b>22</b>

# 1 Uvod

Transport predstavlja jedan od najdinamičnijih i jedan od najštetnijih sektora kada je u pitanju potrošnja energije i emisija stakleničkih plinova. Zbog stalnog rasta broja vozila u prometu, sve veće prosječne duljine vremena provedenog u automobilima te gušće naseljenosti urbanih sredina, doprinos transporta postaje sve značajniji kada je u pitanju globalno zagađenje. Zbog nezaustavnog trenda povećanja potražnje za prijevoznim sredstvima, smanjenje emisija i potrošnje energenata postaje kompliciraniji problem.

Često zanemarivan, a itekako važan faktor kod potrošnje goriva, a time i manjeg otpuštanja stakleničkih plinova je individualan stil vožnje. Način na koji vozač upravlja vozilom, gledajući akceleraciju, kočenje, brzinu te stupanj prijenosa, može značajno utjecati na potrošnju energenata te time i ispuštanje stakleničkih plinova. Stoga, optimizacija i edukacija vozača o benefitima ekonomične vožnje može biti korisnija i za okoliš i za smanjenje troškova vožnje. Za analiziranje obrazaca vožnje vozača može se koristiti raznim široko dostupnim tehnologijama kao što su GPS, senzori unutar telefona i sl. U praksi se inače koristi analiza GPS podataka koji omogućavaju izračun prosječne brzine između dvaju točaka i njihovu usporedbu s prosječnim brzinama ostalih vozača koje se kombiniraju sa sensorima kao što su akcelerometar i žiroskop te kao produkt dobivamo brojevi pokazatelj navika vozača. Kao alat za analizu ovih podataka koriste se algoritmi strojnog učenja koji nude mogućnost predikcije potrošnje goriva na temelju ulaznih podataka. Algoritmi strojnog učenja mogu primiti veliku količinu parametara te prepoznati veze i obrasce između njih te vratiti smisleni rezultat. Algoritmi strojnog učenja idealni su za ovakav zadatak jer količina prikupljenih podataka može biti izrazito velika te nam klasični deterministički i algoritamski pristup ovom zadatku ne bi bio od koristi.

Cilj ovog rada je analizirati javno dostupne podatke te istrenirati model strojnog učenja kako bi na temelju senzorskih očitavanja, vremenskih intervala i promjeni brzine vozila unutar intervala procijenio potrošnju goriva te upozorio vozača na određene elemente i

obrasce njegove vožnje. Vozač bi se tijekom vožnje „nagrađivao“ ili „kažnjavao“ u nadi smanjenja potrošnje goriva, sigurnijeg putovanja te jeftinijeg transporta.



## 2 Trendovi u cestovnom prometu

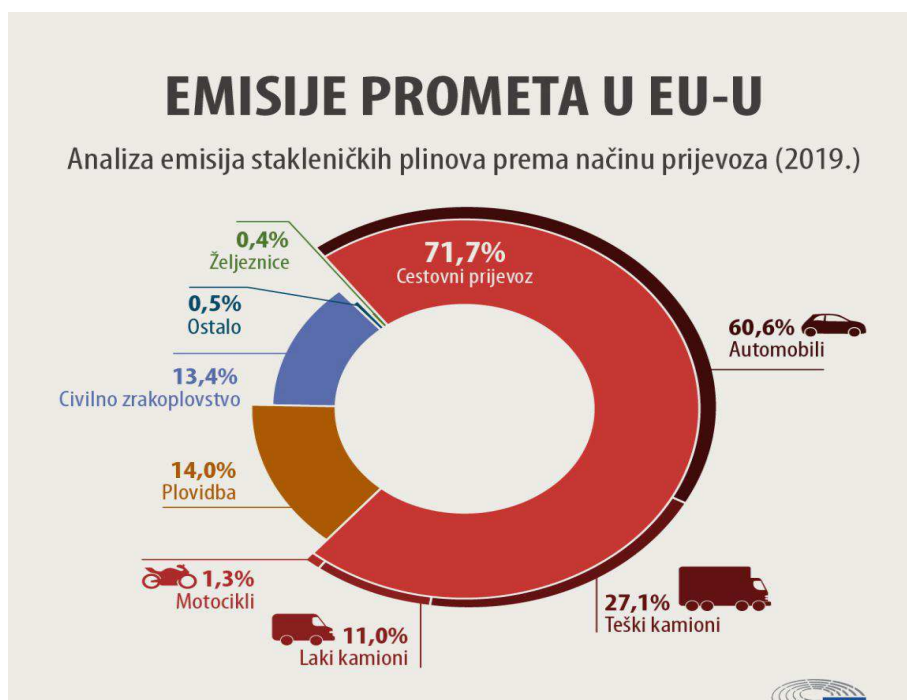
### 2.1 Važnost cestovnog prometa u brojkama

Osobni automobili su glavno prijevozno sredstvo današnjice. Tijekom prošlog stoljeća, globalni ekonomski rast je bio usko povezan sa cestovnim prometom kao posljedica globalizacije. Moderni načini transporta su omogućili eksponencijalni rast u broju veza između proizvođača i potrošača. U 2021. godini je transportni sektor pridonio 7 posto ukupnog svjetskog domaćeg proizvoda te je 5.6 posto, odnosno 193 milijuna ljudi, svjetske populacije bilo u njemu zaposleno. Tijekom 2021. godine je u transportnom sektoru potrošeno čak 113.4 exodžula (EJ) energije [12]. Prema podacima Državnog zavoda za statistiku, u 2023. godini, u Hrvatskoj prevezeno čak 84.5 milijuna putnika, od kojih je cestovnim prijevozom prevezeno čak 50-ak posto. Cestovnim prijevozom je prevezeno čak 88 milijuna tona robe što je čak 70-ak posto ukupnom prijevoza robe u Hrvatskoj [1]. Broj registriranih vozila u Hrvatskoj je tijekom 2023. godine iznosio otprilike 2.5 milijuna od kojih je otprilike 1.9 milijuna osobni vozila [2].

### 2.2 Tamna strana cestovnog prometa u brojkama

Cestovni promet je tijekom 2021. godine zaslužan za čak 78 posto utrošene energije od strane ukupnog transporta. Od svih primarnih sektora današnjice, transport ima najmanju stopu korištenja obnovljivih izvora energije. Čak 96 posto energije iskorištene za usluge prometa dolazi od neobnovljivih izvora energije, dok ostalih 4 posto dolazi od obnovljivih izvora energije. Kroz period od 2009. godine pa do 2019. godine, emisije uzrokovane cestovnim prometom su godišnje rasle za oko 4 posto. Na globalnoj skali, transportni sektor je 2021. godine ispustio čak 7.7 gigatona  $CO_2$  te oko 20 posto ukupnog globalnog ispuštenog  $CO_2$  u atmosferu. Transport za prijevoz putnika je najmanje efikasan oblik transporta koji iskoristi čak 2000 kilodžula energije po prevezenih tona-prijeđenih kilometara. Globalna energetska kriza koja je proizašla iz rata u Ukrajini te visoke cijene goriva su motivirale na uvođenje regulacija kako bi se potaknulo smanjenje korištenja fosilnih goriva. Mnoge države zacrtale su dugoročne ciljeve uvođenja prijevoznih sredstava pogonjena obnovljivim izvorima. Portugal je zacrtao rast sa 20

posto na 29 posto do 2030. godine, Nizozemska rast sa 14 posto na 28 posto do 2030. godine. Grad Curitiba zacrtao je da će do 2050. godine sav putnički transport biti pogonjen obnovljivim izvorima energije. Glavni energenti koji bi mogli zamijeniti fosilna goriva su struja, ugljik te razno razni oblici biogoriva. Električna vozila čine oko 15 – 18 posto, odnosno oko 10 milijuna, ukupno prodanih novih vozila u 2022.godini. Čak 5 milijuna električnih vozila je prodano u Kini. Na državnoj razini unutar Europe je najviše električnih vozila u postotku prodano u Norveškoj, odnosno čak 86 posto. Trenutke prepreke prelaska na električna vozila su njihova cijena, dugo punjenje baterije te nedostatak punionica. Druga najpopularnija opcija su vozila sa hidrogenskim pogonom, koja su manje efikasna od električnih, ali pružaju brže punjenje i veći domet od električnih vozila te su validna alternativa za smanjenje emisija. U 2021. godini je prodano oko 50 tisuća vozila sa hidrogenskim pogonom, što je i dalje nedovoljno, ali sa daljnjim napredovanjem te grane industrije se očekuje njihov značajan rast [12]. U Republici Hrvatskoj u 2023.godini je bilo 10.633 cestovnih prometnih nesreća u kojoj je stradala barem jedna osoba, što je za 6,3 posto više nego 2022. godine [2]. Prema podacima Europske Unije iz 2019. godine, promet zaslužan za otprilike četvrtinu ukupnih emisija  $CO_2$ , od čega 71,7 posto iznosi cestovni promet. Cestovna vozila su u prosjeku ispuštala 122 grama  $CO_2$  po kilometru [3].



Slika 2-1 Udjeli emisija stakleničkih plinova u cestovnom prometu [3]

## 3 Umjetna inteligencija

Inteligencija nema striktno definiranu definiciju, ali da bi netko/nešto bilo inteligentno mora se koristiti konceptima kao što su razumijevanje, učenje, apstraktno rasuđivanje, planiranje, rješavanje problema i sl. [4]. Umjetna inteligencija je dio računarstva koji se bavi razvojem sposobnosti računala da obavljaju zadaće za koje je potreban neki oblik inteligencije [5]. Može se reći da je umjetna inteligencija službeno nastala 1956. godine na Dartmouth College-u u Sjedinjenim Američkim Državama. Ocem umjetne inteligencije smatra se John McCarthy koji je radio na istoimenom koledžu, te je sa svojim kolegama organizirao dvomjesečnu radionicu na kojoj su glavna tema bili teorija automata, neuronske mreže i istraživanje inteligencije [6]. Danas se umjetna inteligencija smatra jednom od najvažnijih grana računarstva, te ju se smatra za prekretnicom čovječanstva.

### 3.1 Strojno učenje

Strojno učenje jest programiranje računala na način da optimiziraju neki kriterij uspješnosti temeljem podatkovnih primjera ili prethodnog iskustva. Strojno učenje raspolaže modelom koji je definiran do na neke parametre, a učenje se svodi na izvođenje algoritma koji optimizira parametra modela na temelju podataka ili prethodnog iskustva [7]. Glavna uloga strojnog učenja je generalizacija, odnosno na nekom neviđenom primjeru primijeniti istrenirani algoritam koji će nam dati neki odgovor. Strojno učenje se dijeli na tri skupine, a to su nadzirano (engl. supervised), nenadzirano (engl. unsupervised) i podržano/ojačano (engl. reinforcement). Nadzirano učenje se bavi podacima koji već imaju definirane ulaze i izlaze, nenadzirano učenje se bavi podacima koji nisu jasno označeni sa izlaznim vrijednostima, ali ono ih grupira (engl. clustering) ili pronalazi neke zajedničke obrasce kako bi dalo odgovor, dok se podržano učenje bazira na nagrađivanju i kažnjavanju ovisno o rezultatu algoritma. U ovom radu je korišteno nadzirano učenje te će ono biti detaljnije objašnjeno.

### 3.1.1 Nadzirano učenje

Nadzirano učenje se bavi dvama tipovima problema: klasifikacija i regresija. Kod klasifikacije hipoteza daje postupak izračuna kojoj je grupi iz podataka za treniranje najbliži novi ulazni podatak, primjerice je li pristigla elektronska pošta potencijalno malignog sadržaja. U ovom radu se koristi regresijski tip nadziranog učenja, odnosno konkretnije predviđa potrošnja goriva i koregiraju se navike vozača. Kod regresije, hipoteza daje postupak izračuna kontinuirane izlazne vrijednosti za neki novi ulazni podatak koji se najbolje uklapa u poznati skup podataka za treniranje [6]. Regresija se koristi za predviđanje cijena nekretnina, stanja tržišta, duljine života i sl. Cilj je dobiti predviđenu vrijednost koja će imati što manju pogrešku. Pogreška se računa uz pomoć određenih matematičkih funkcija, tzv. Funkcije pogreške.

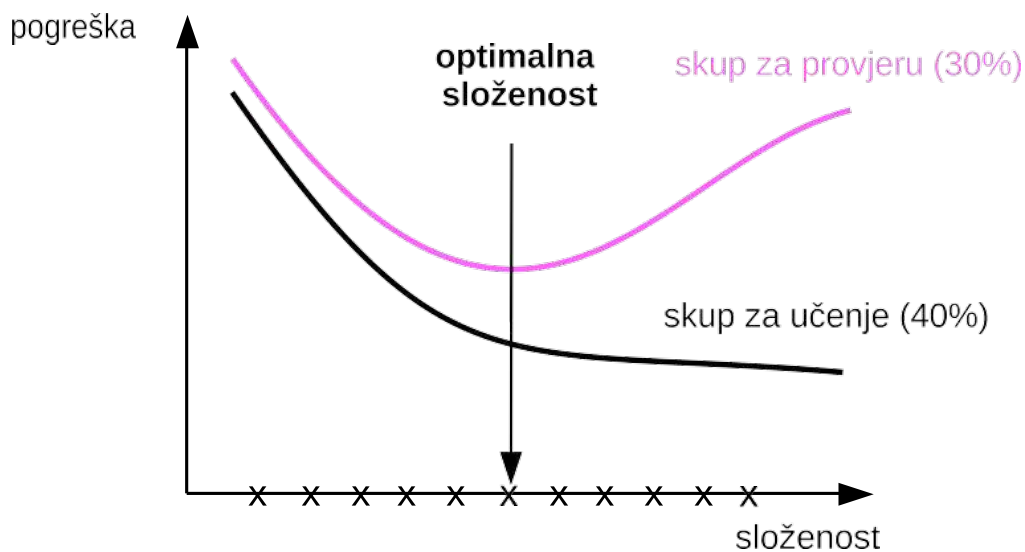
Najveća mana, odnosno slabost, strojnog učenja je mogućnost prenaučivosti algoritma. Ako je model vrlo složen, onda se može se vrlo dobro prilagoditi podacima i za te podatke davati gotovo savršene predikcije. Ako je model presložen, previše će se prilagoditi podacima na kojima je treniran, a davat će loše predikcije na neviđenim podacima čime on ne generalizira dobro na neviđenim podacima, tj. ne ispunjava svoju zadaću. Ta se pojava naziva pretreniranost jer je model jako dobro istreniran na primjerima koje je vidio, a radi jako loše na primjernima koje nikada nije vidio. Kako bi se spriječila pretreniranost modela u ovom se radu koristi grupiranje podataka na podatke za treniranje i podatke za testiranje. Postoji još jedan oblik podjele, a to je podjela na podatke za treniranje, na podatke za validaciju i podatke za testiranje. U slučaju prvog načina podjele podataka, obično se skupovi dijele u omjeru 80 (treniranje) : 20 (testiranje), a tako je napravljeno i u ovom radu. Skup namijenjen za testiranje stavljamo u algoritam treniranja modela strojnog učenja. Skup namijenjen za testiranje se stavlja u istrenirani algoritam, to on omogućuje procjenu koliko su izlazi algoritma dobri tj. koliko oni odudaraju od stvarnih izlaza koje model nikad nije vidio. Ova dva skupa moraju nužno biti disjunktna jer je ideja skupa za testiranje provjera koliko dobro algoritam radi na neviđenim podacima, koje skup za testiranje „imitira“. Drugi oblik podjele podataka koji skup dijeli na tri djela sadrži skup za validaciju (engl. validation set), te je njegova uloga unakrsna provjera tijekom

treniranja. Uobičajeno je da skup dostupnih podataka dijelimo u omjeru 40 (treniranje) : 30 (validacija) : 30 (testiranje) kako je prikazano na **Slika 3-1**.



*Slika 3-1 Podjela podataka na skup za treniranje, validaciju i testiranje [7]*

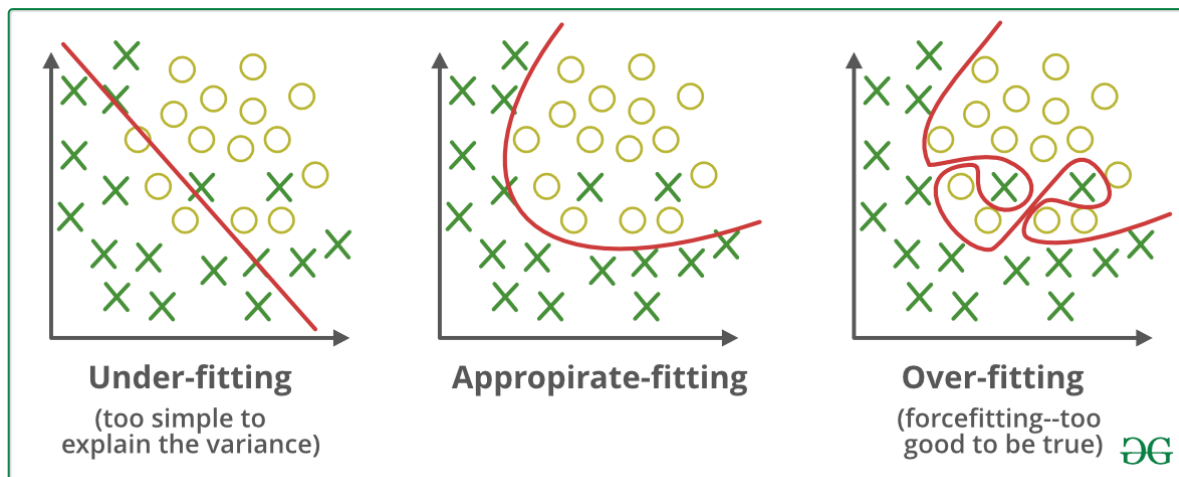
Taj način rada koristi se kod određenih modela koji imaju mogućnost ugađanja složenosti. Paralelno se testira na skupu za testiranje i na skupu za validaciju te kada vrijednosti pogreške na skupu za validaciju ponovno krenu rasti, prekidamo treniranje te smatramo da smo našli idealnu složenost algoritma.



*Slika 3-2 Traženje optimalne složenosti algoritma [7]*

Na **Slika 3-3** vidljiv je odnos pada pogreške i rasta složenosti, tj. što je algoritam složeniji, to pogreška više pada. Presloženim modelom, pogreška postaje minimalna i model je u potpunosti prilagođen podacima za treniranje koji mogu sadržavati šum, pa

time i dati djelomično krivu informaciji.



Slika 3-3 nedovoljno treniranje

dovoljno treniranje

pretreniranost [8]

Kroz treniranje algoritma cilj je pronaći približno dobar, idealno točan, interval u kojem je naš model dovoljno dobro istreniran da sa dobrom dozom pouzdanosti može donijeti odluku, ali ne i pretreniran kako ne bi loše radio na neviđenim primjerima.

### 3.1.2 Stabla odluke

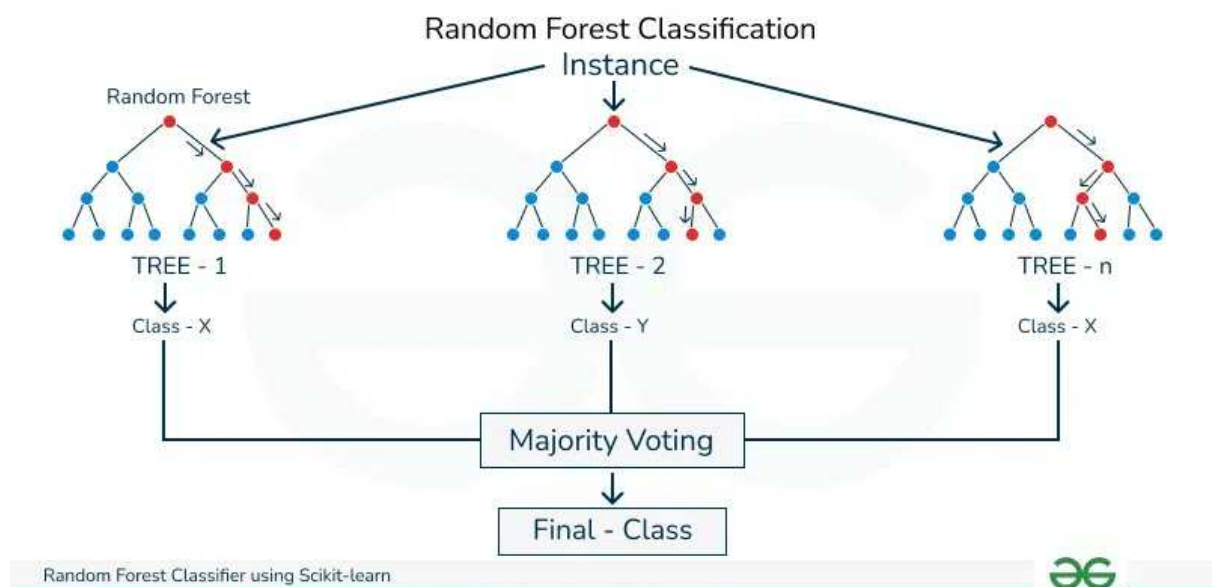
Stabla odluke su modeli strojnog učenja koji se koriste i kod problema klasifikacije i kod problema regresije. Nazivaju se stablima jer su građena od čvorova ili listova te su ti čvorovi najčešće spojeni značajkama (engl. features) koje se još i nazivaju granama. Prolaskom po izgrađenom stablu i dolaskom do listova, kao rezultat predikcije vraća se rezultat zapisan u listu stabla. Problem stabala kao i svih algoritama strojnog učenja je prenaučenosť. Stabla se primjerima za treniranje prilagode 100 posto što dovodi do prenaučenosťi. Za smanjivanje prenaučenosťi koriste se različite metode, poput: podrezivanja stabla, ograničavanja dubine stabla, ansambl metoda i sl.

Kod ansambl metoda koristi se više od jednog stabla odluke (100 i više), te se njihove predikcije kombiniraju kako bi procjena bila što točnija i bolje generalizirala. Stabla se međusobno ispravljaju te će rezultat gotovo uvijek biti isti i li bolji od rezultata samo jednog stabla. Postoje više pristupa algoritmima: *bagging* *boosting* *stacking* *voting* i sl.

S obzirom da se u ovom radu koristio random forest regression algoritam koji je bagging pristupa, on će biti detaljnije objašnjen.

### 3.1.3 Random forest

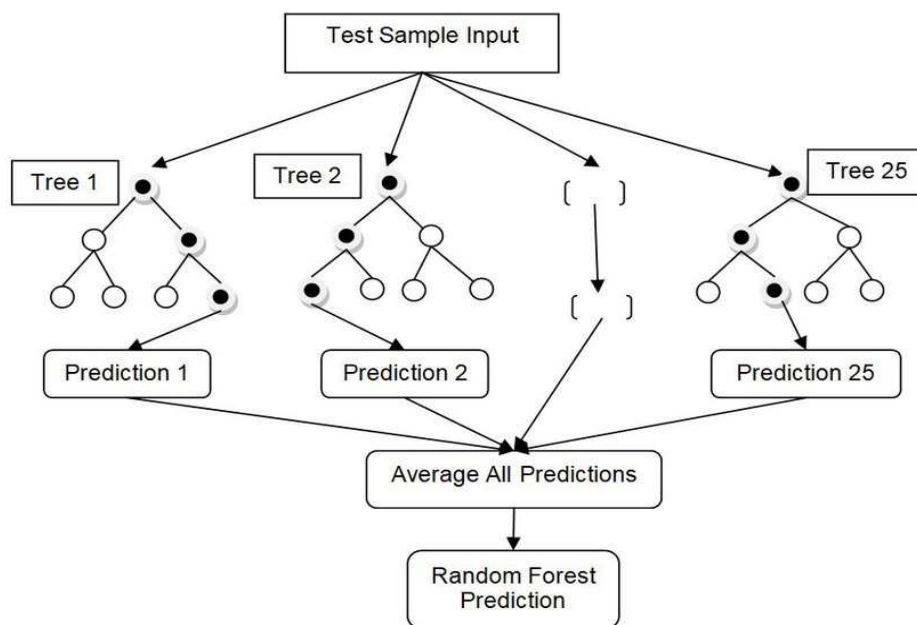
Algoritam je razvijen 2001. godine od strane profesora Lea Breimana koji je u to vrijeme radio i djelovao na Berkeley-u, California. Algoritam generira niz stabala odluke koristeći različite slučajne podskupove podataka za treniranje svakog stabla, što se naziva bootstrapping. Glavna razlika u odnosu na klasično stablo odluke je veća otpornost na prenaučenosť te bolja točnost predikcije. Za svako stablo, pri svakom dijeljenju čvora koristi se slučajni podskup značajki kako bi se dodatno povećala raznolikost stabala. Ovisno radi li se regresija ili klasifikacija algoritam će drugačije kombinirati rezultate stabala. Kod klasifikacije se koristi metoda „glasanja“ te će najčešća predikcija biti „izglasana“ i dodijeljena kako je prikazano na **Slika 3-4**.



**Slika 3-4** Rad algoritma Random Forest Classification [9]

Kod regresije se uzima prosjek svih predikcija te se to šalje kao rezultat predikcije. Random forest nije samo odličan kod predviđanja, nego i pruža uvid u važnost određenih

parametara kako bi se predikcije dodatno optimizirale. Može obrađivati velike količine podataka, manje je osjetljiv na nedostatke u podacima te ga neće zbuniti određeni neispravni podaci. Zbog toga je idealan za područja u kojima se obrađuju velike količine podataka kao što su bioinformatika, dijagnostika u medicini, financije i kretanje tržišta i sl.



*Slika 3-5 Rad algoritma Random Forest Regression [10]*

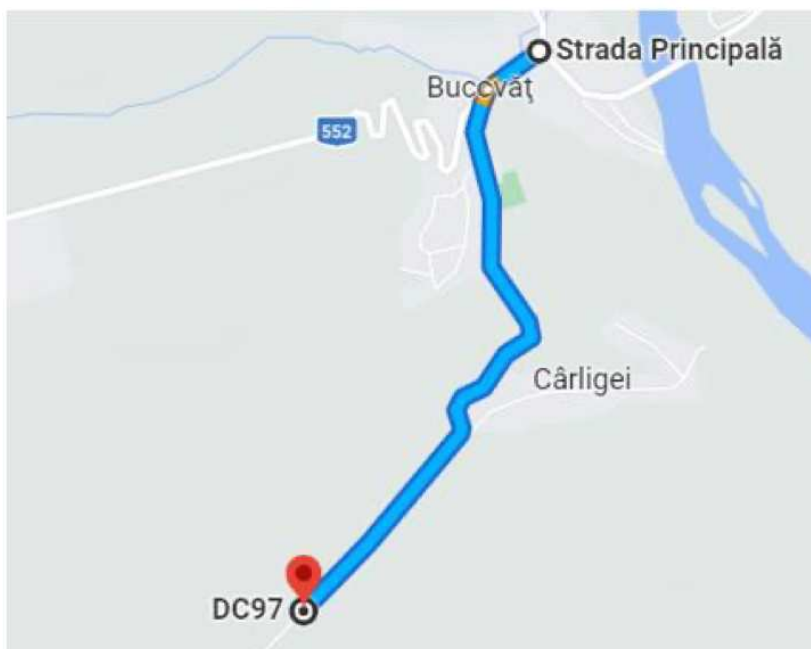


## 4 Podaci voznih ciklusa

Skup podataka (engl. dataset) koj je korišten je preuzet sa stranice pod imenom Kaggle te su autori ovog rada i skupa podataka Paul-Stefan Popescu i Ion Cojocararu [11].

Podaci su se prikupljali u trima različitim načinima vožnje, a to su agresivni (naglo skretanje, brzo kretanje), normalni (prema ograničenjima brzine) i spori (sporije od ograničenja, spori ulasci u zavoje). Od senzora su korišteni akcelerometar i žiroskop koji se nalaze u svakom modernom mobilnom uređaju. Pametni uređaj Samsung Galaxy S10 je postavljen na stabilni automobilski držač za mobilni uređaj te su za prikupljanje podataka korišteni njegovi senzori, točnije akcelerometar i žiroskop. Senzori su prikupljali podatke otprilike svakih 0.57 sekundi, odnosno otprilike 2 zapisa u sekundi. Kao prijevozno sredstvo korištena je Dacia Sandero 1.4 MPI sa 75 konjskih snaga. Ovo vozilo je odabrano kako bi podaci bili što bliži široj populaciji ljudi, odnosno imali više koristi u ljudskoj svakodnevici. Skup podataka se sastoji od 3084 zapisa te svaki zapis ima sljedeće značajke (engl. feature):

1. vrijednosti senzora akcelerometra u x, y i z osima (3)
2. vrijednosti žiroskopa u x, y i z osima
3. oznaku je li mjerenje rađeno u agresivnom, normalnom ili sporom načinu vožnje (1)
4. vremensku oznaku (engl. timestamp) (1). Spori način vožnje sveukupno ima 1273 zapisa, normalan način vožnje 997, a agresivan način vožnje 814 zapisa. Sva tri testa rađena su na istoj ruti te agresivni stil vožnje ima najmanje mjerenja zbog bržeg prolaska po označenoj ruti [11].



*Slika 4-1 Ruta na kojoj je napravljen dataset [11]*

Tijekom izračuna određeni podaci su izbačeni zbog pogreška ili šuma u mjerenju kako bi dobili najmjerodavnije rezultate. Nakon izbacivanja ostalo je 628 zapisa za agresivnu vožnju, 885 zapisa za normalnu vožnju i 1177 zapisa za sporu vožnju što dovodi do sveukupne sume od 2690 zapisa. Sveukupno je izbačeno 394 zapisa od kojih je najveći dio iz agresivnog skupa. Zapisi iz svake kategorije su podijeljeni u omjeru 80 (treniranje) : 20 (testiranje). Za treniranje modela koji predviđa potrošnju kod agresivne vožnje korišteno je 502 zapisa, a za testiranje istog njih 126. Kod normalne vožnje za treniranje korišteno je 708 zapisa, a za testiranje 177 i naposljetku za sporu vožnju model je treniran na 941 zapisu, a testiran na njih 236. Svi podaci za testiranje i treniranje su objedinjeni te je model istreniran na 2152 zapisa, a testiran na njih 538.

## 5 Obrada podataka

Svi navedeni podaci pohranjeni su u bazu podataka, a kao alat za upravljanjem bazom podataka korišten je pgAdmin 4. Na korištenom skupu podataka prvo je nadodan stupac id (engl. identification) koji sadrži jedinstvenu oznaku za taj zapis (prirodan broj od 1 do N). Zatim se podaci iz baze učitavaju sortirani prema vremenskoj oznaci u Python kod u kojem se rade izračuni. Podaci se prvo sortiraju u podatke u tri različite liste, svaka za jedan stil vožnje, te se prvi podatak u svakoj listi označi kao stanje mirovanja iz kojeg kreće gibanje, te se početna brzina postavlja na 0. Brzine kretanja su ograničene na 18 m/s (64.8 km/h) za sporu vožnju, 24 m/s (86.4 km/h) za normalnu vožnju i 30 m/s (108 km/h) za agresivnu vožnju. Za dobivanje brzine za svaku os kretanja, korištena je formula (1).

$$v_{os} = \int_0^t a_{os} dt + v_{0(os)} \quad (1)$$

Potom se iz brzine po svakoj osi izračuna prijeđena udaljenost uz pomoć formule (2).

$$d_{os} = \int_0^t v_{os} dt \quad (2)$$

Time se dobiva prijeđena udaljenost po svakoj osi kretanja te se one vektorski zbroje čime se dobiva ukupna prijeđena udaljenost  $d_{uk}$ , prikazana u formuli (3).

$$d_{uk} = \sqrt{d_x^2 + d_y^2 + d_z^2} \quad (3)$$

Prijeđenu udaljenost skalira se sa vremenskim intervalom od 0.57 sekundi kako bi se dobila ukupna prijeđena kilometraža u vremenskom intervalu od jednog sata. Prijedeći kilometri pretvoreni su u potrošnju na 100 prijedećih kilometara preko formule (4).

$$\frac{d_{uk} * potrošnja}{100 km} \quad (4)$$

Korištena potrošnja goriva iznosi 6.5 L / 100 km jer je to prosječna kombinirana (otvorena i gradska vožnja) potrošnja goriva za korišteno prijevozno sredstvo. U bazu je nadodan stupac koji označava prosječnu potrošnju za svaki interval, odnosno za svaki zapis baze podataka. Potrošnje za svaki način vožnje su zbrojene te je izračunata aritmetička sredina formulom (5).

$$\frac{\sum_1^{n_{broj\ intervala}} potrošnja\ intervala}{n_{broj\ intervala}} \quad (5)$$

Osim potrošnje u bazu su nadodani stupci koji predstavljaju početnu i krajnju brzinu dostignutu u tom intervalu. Krajnja brzina je dobivena vektorskim zbrojem brzina po svim osima formulom (6).

$$v_{uk} = \sqrt{v_x^2 + v_y^2 + v_z^2} \quad (6)$$

Početna brzina ovog intervala je uzeta kao krajnju brzinu prošlog intervala formulom (7).

$$v_{0[i]} = v_{krajnja[i-1]} \quad (7)$$

Izračunati podaci su podijeljeni u prethodno objašnjene skupove. Od svakog stila vožnje je uzeto prvih 80 posto podataka, grupiranih u jednu cjelinu, koji su poslužili kao materijal za treniranje modela. Ostalih 20 posto podataka od svakog načina vožnje je grupirano i kasnije se na njima testirala kvaliteta generalizacije istreniranog modela.

Rađeno je sa dvama tipovima podataka, „originalnima“ i podacima s dodanim šumom. U podatke za treniranje i testiranje je nadodan šum (engl. noise) kako bi se simulirale smetnje mjerenja koje se svakodnevno pojavljuju.

```
noise1 = np.random.normal(0, 0.5, xTrain.shape)
xTrain = xTrain + noise1
noise2 = np.random.normal(0, 0.5, xTest.shape)
xTest = xTest + noise2
```

*Slika 5-1 Dodavanje šuma u podatke za treniranje i testiranje*

Podaci su ubačeni u funkciju pod imenom RandomForestRegressor koja je sastavni dio Python-ove Sklearn biblioteke. Kako algoritam radi sa varijabilnim brojem stabala koji se može predefinirati, analiziran je rad algoritma sa zadanih 100.

```
model = RandomForestRegressor(n_estimators=x)
model.fit(xTrain, yTrain)
yprediction= model.predict(xTest)
```

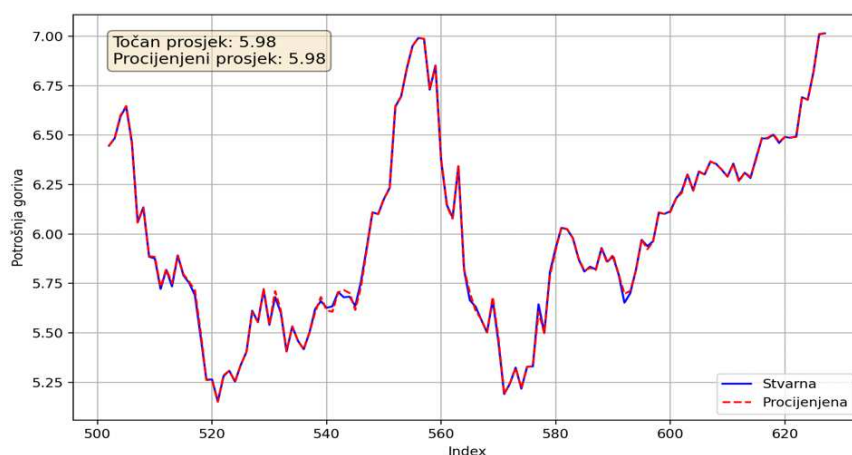
*Slika 5-2 Kreiranje instance modela, treniranje modela, testiranje modela*

## 6 Analiza rezultata

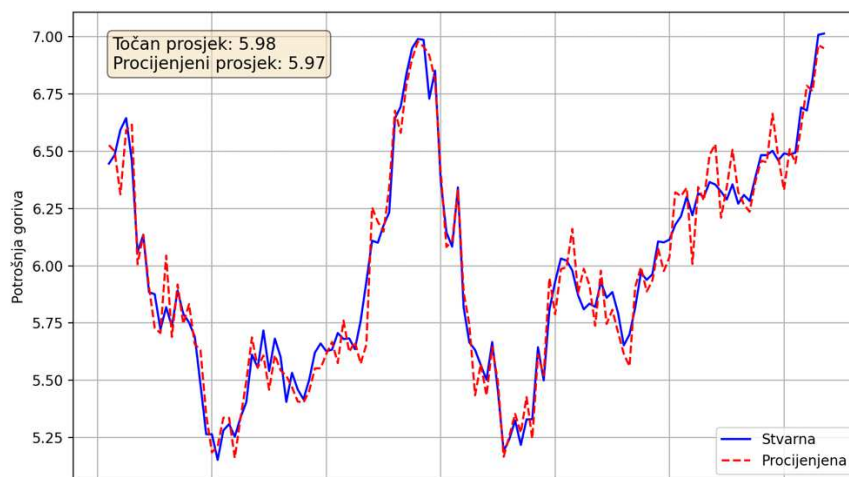
### 6.1 Predikcije brzine vozila

Kako bi se rezultati najbolje prikazali izabran je grafički prikaz koji sadržava stvarnu potrošnju i predviđenu potrošnju. Rezultati na slikama **Slika 6-** i **Slika 6-** su rezultati koji se dobiju kada RandomForestRegressor radi s ansamblom od 100 stabala.

Na **Slika 6-** su korišteni podaci agresivnog načina vožnje bez nadodanog šuma. Algoritam gotovo savršeno predviđa potrošnju goriva. Na **Slika 6-** korišteni su podaci agresivnog načina vožnje sa nadodanim šumom. Algoritam malo pogriješi ali procijenjeni prosjek ostaje praktički isti. Prosjek je mjerjen u litrama na 100 kilometara.

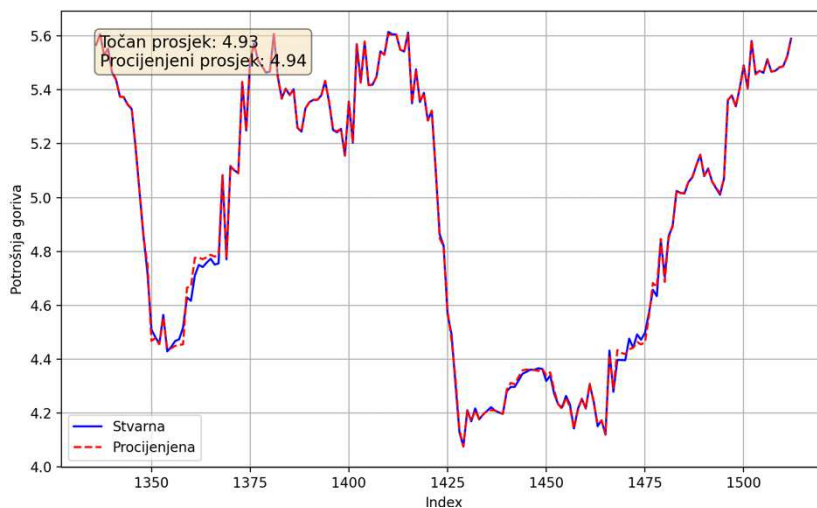


**Slika 6-1** Agresivna vožnja, bez šuma



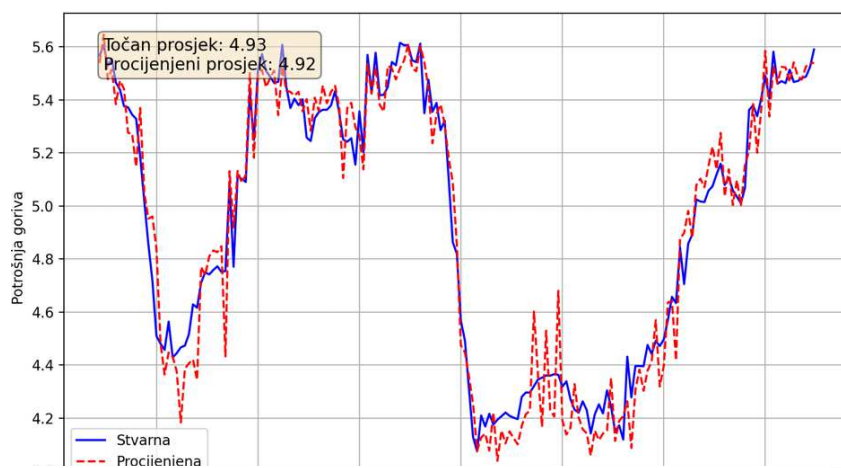
**Slika 6-2** Agresivna vožnja, sa nadodanim šumom

Sljedeći grafovi prikazuju vrijednosti potrošnje kod normalnog načina vožnje.



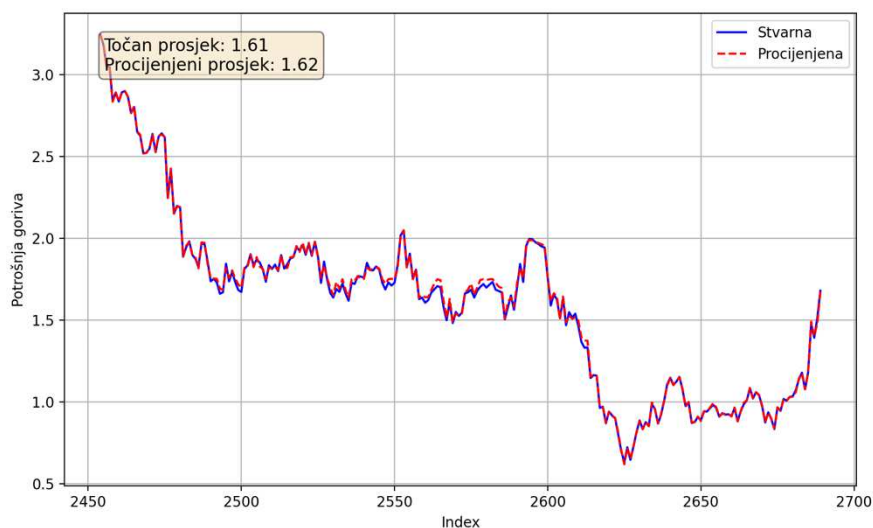
**Slika 6-1** Normalna vožnja, bez šuma

Na **Slika 6-1** su korišteni podaci normalnog načina vožnje bez nadodanog šuma. Algoritam ponovno praktički savršeno predviđa potrošnju goriva. Na **Slika 6-2** su korišteni podaci normalnog načina vožnje sa nadodanim šumom. Uz obraćanje pozornosti na pojedinačne točke, vidljivo je da algoritam nešto više griješi nego kod agresivne vožnje.



**Slika 6-2** Normalna vožnja, sa nadodanim šumom

Sljedeći grafovi prikazuju vrijednosti potrošnje goriva kod sporog načina vožnje.



**Slika 6-3** Spora vožnja, bez šuma



**Slika 6-4** Spora vožnja, sa nadodanim šumom

Algoritam i kod spore vožnje bez nadodanog šuma radi gotovo savršeno, dok kod spore vožnje sa nadodanim šumom značajno odstupa. S obzirom da je spora vožnja imala najviše testnih primjera pogreška svake točke se akumulira, pa ona ispadne nešto veća. Vidljivo je da što su mjerenja lošija, odnosno sa većom količinom šuma, tako će i rezultati procjene biti lošiji. Ako su mjerenja napravljena u savršenim uvjetima, algoritam će gotovo savršeno predvidjeti potrošnju goriva.

## 6.2 Korekcije vozača

Kako bi se smanjile emisije ispušnih plinova, nije dovoljno samo procijeniti potrošnju, već je potrebno i dati vozaču do znanja što krivo radi i što bi mogao promijeniti. U slučaju kada vozač tri (ili neki višekratnik broja 3) vremenska intervala vozi sa prosječnom potrošnjom većom od 6 litara na 100 km, vozaču se ispiše poruka koja ga upozorava da smanji brzinu kako bi optimizirao potrošnju goriva. Primjer rada sustava ukazivanja na pogreške prikazan je na **Slika 6-6**. Također osim ukazivanja na probleme kod vozača postoji i sustav vizualnog nagrađivanja vozača. U slučaju kada vozač 5 (ili neki višekratnik broja 5) vremenskih intervala vozi sa prosječnom manjom od 4.5 litara na 100 kilometara, vozaču se ispiše poruka koja ga pohvaljuje za ekonomičnu vožnju i motivira ga da i nastavi sa istom. Primjer rada sustava nagrađivanja također je vidljiv na **Slika 6-5**.

```
4.487087036373303
4.789030064632562
4.303373480302354
4.1296904706264055
4.025582193841328
4.102572175053439
4.134812195114606

CONGRATULATIONS
You are driving economically for some time, keep up with the good work!

4.142504020599216
4.176957910432902
4.2486467467916444
4.107559363394374
4.239414979537461

CONGRATULATIONS
You are driving economically for some time, keep up with the good work!

4.114744528866962
4.124602124420712
4.1210072604729735
4.236931554806464
```

*Slika 6-5 Poruke nagrade*



```

5.464298435287196
5.531484564647872
5.687759536532122
5.587582705452038
5.4750700994566985
5.634479831106504
5.750274033885237
5.584580061848651
5.542052405871259
5.724118928843933
6.070969571372654
6.043009255014895
5.894843719156716
6.142920902508735
6.282515334710947
6.5990889531694386

WARNING!
Slow down for the more economic ride!

6.76557033032477
6.8710600757525455
6.969551533315915

WARNING!
Slow down for the more economic ride!

6.906545313757142
6.935611996828295
6.77640371559153

```

*Slika 6-6* Poruke upozorenja

### 6.3 Pogreške

Kako bi provjerili kvalitetu rada modela korištene su dvije metode izračuna pogreške. Prva korištena mjera je srednja kvadratna pogreška (engl. mean squared error). Mean squared error, prikazan formulom (8), se računa kao suma kvadrata razlike koji su podijeljeni sa brojem ukupno sumiranih elemenata. Uloga kvadrata je da više kažnjava veću pogrešku. Što je MSE manji to je algoritam precizniji.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

Druga korištena mjera preciznosti je R-kvadratna vrijednost (engl. R2 score) koja je poznata kao koeficijent determinacije prikazana je formulom (9). Njezina uloga je također procjena koliko dobro su predviđene vrijednosti slične onim stvarnima. Vrijednosti variraju između 0 i 1 te će vrijednost biti bliže 1 što je algoritam precizniji u svojoj procjeni.

$$R^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (9)$$

Maloprije spomenuta Sklearn biblioteka ima svoje definirane funkcije te su one korištene u ovom radu.

```

mse = mean_squared_error(yTest, yprediction)
r2 = r2_score(yTest, yprediction)

```

*Slika 6-8* Prikaz Sklearn-ovih metoda za računanje navedenih pogrešaka

Pogreške su izračunate na temelju cijelog skupa podataka s kojima je rađeno. Na **Slika 6-** vidimo pogreške koje se odnose na cijeli skup podataka bez pridodanog šuma.

```
RESULTS  
Mean squared error: 0.00022978822734805157  
R2 score: 0.9999408996058067
```

*Slika 6-8* Rezultati kod podataka bez šuma

Kao što je i prethodno objašnjeno, algoritam je bolji što je pogreška manja u slučaju srednje kvadratne pogreške (engl. MSE) dok je kod R-kvadratne vrijednosti cilj da dobivena vrijednost bude što bliže broju 1. Vidljivo je da je srednja kvadratna pogreška prilično malog iznosa te R-kvadratna vrijednost skoro 1 što je dodatni pokazatelj kvalitete predikcije algoritma. Na **Slika 6-** prikazane su pogreške koje se odnose na cijeli skup podataka na koje je nadodan šum. Ovdje su rezultati malo lošiji, što je i logično s obzirom da podaci sadrže.

```
RESULTS  
Mean squared error: 0.014983491772274949  
R2 score: 0.9961463201124202
```

*Slika 6-9* Rezultati kod podataka sa nadodanim šumom

šum koji kvari treniranje i procjenu, ali su procjene i dalje jako dobre i mjerodavne. Vidljivo je da je srednja kvadratna pogreška kod podataka sa šumom veća otprilike 65 puta, dok je R-kvadratna vrijednost manja otprilike 0.3 posto.

## 7 Zaključak

Prikazom i detaljnom analizom rezultata uz pomoć mjera pogrešaka, utvrđeno je da se strojnim učenjem može unaprijediti predikcija potrošnje goriva kod različitih stilova vožnje. Random Forest Regression model se pokazao izvrsnim alatom za procjenu potrošnje goriva kod svih stilova vožnje – agresivnu, normalnu i sporu. Implementirani sustav upozorenja i nagrađivanja vozača pokazao je potencijal za poticanje ekonomičnije vožnje.

Predstavljeni model je primjenjiv za sustave upravljanja flotom vozila i aplikacije za vozače koji žele optimizirati svoju potrošnju goriva. Prostora za nadogradnju ima u vidu dodatnih senzora koji bi u stvarnom vremenu dobavljali GPS podatke ili dobavljali podatke o trenutnim vremenskim prilikama i neprilikama kako bi se algoritam mogao njima prilagoditi. Također daljnji rad bi mogao biti usmjeren ka izradi modela za autonomnu vožnju.

## Literatura

[1] Snježana Kos, Korana Šakić Pokrivač, Lidija Pilat, Ninoslava Podnar i Laura Ćelić, *Transport u četvrtom tromjesečju 2023.*, (2024, veljača). Poveznica: <https://podaci.dzs.hr/2023/hr/58579>; pristupljeno 1. lipnja 2024.

[2] Snježana Kos i Ninoslava Podnar, *Registrirana cestovna vozila i cestovne prometne nesreće u 2023.*, (2024, travanj). Poveznica: <https://podaci.dzs.hr/2024/hr/77312>; pristupljeno 1. lipnja 2024.

[3] Nepoznat autor, *Emisije CO2 u prometu EU-a: Činjenice i brojke*, (2019, ožujak). Poveznica: <https://www.europarl.europa.eu/topics/hr/article/20190313STO31218/emisije-co2-u-prometu-eu-a-cinjenice-i-brojke>; pristupljeno 1. lipnja 2024.

[4] Prof. dr. sc. Jan Šnajder i izv. prof. dr. sc. Marko Čupić. Uvod u umjetnu inteligenciju, 1. Uvod u umjetnu inteligenciju, prezentacija – slajd 1 – 75, 2024.

[5] Nepoznat autor, *umjetna inteligencija*, (2024, siječanj). Poveznica: <https://www.enciklopedija.hr/clanak/umjetna-inteligencija>; pristupljeno 6. lipnja 2024.

[6] Prof. dr. sc. Darko Stipaničev, izv. prof. dr. sc. Ljiljana Šerić, doc. dr. sc. Maja Braović. Uvod u umjetnu inteligenciju (2021. studeni). Poveznica: <https://ai.fesb.hr/knjiga/AI-knjiga-FINAL.pdf>; pristupljeno 2. lipnja 2024.

[7] Prof. dr. sc. Bojana Dalbelo Bašić i izv. prof. dr. sc. Jan Šnajder. Uvod u umjetnu inteligenciju, 10. Strojno učenje, prezentacija – slajd 1 – 91, 2020.

- [8] Nepoznat autor , *ML | Underfitting and Overfitting*, (2024, ožujak). Poveznica: <https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/>; pristupljeno 2.lipnja 2024.
- [9] Nepoznat autor, *Random Forest Classifier using Scikit-learn* (2024, siječanj). Poveznica: <https://www.geeksforgeeks.org/random-forest-classifier-using-scikit-learn/>; pristupljeno 2.lipnja 2024.
- [10] Amita Sharma, *Assessment of the Machine Learning Models for Prediction of Cluster Bean (*Cyamopsis tetragonolova* L. Taub.) Yield* (2020, kolovoz). Poveznica: [https://www.researchgate.net/figure/Sample-random-forest-regression-tree\\_fig1\\_343992982](https://www.researchgate.net/figure/Sample-random-forest-regression-tree_fig1_343992982); pristupljeno 3.lipnja 2024.
- [11] Paul-Stefan Popescu, Ion Cojocaru, *Driving Behaviour* (2022, kolovoz). Poveznica: <https://www.kaggle.com/datasets/outofskills/driving-behavior>; pristupljeno 4.ožujka 2024.
- [12] REN21, *Global Status Report, Energy demand* (2023). Poveznica: [https://www.ren21.net/wp-content/uploads/2019/05/GSR2023\\_Demand\\_Modules.pdf](https://www.ren21.net/wp-content/uploads/2019/05/GSR2023_Demand_Modules.pdf); pristupljeno 12.lipnja 2024.

# Korekcija navika vozača strojnim učenjem

## Sažetak

Potrošnja goriva u vozilima je pod velikim utjecajem stila vožnje, koji može varirati od sporog do agresivnog. Koristeći strojno učenje nad prikupljenim podacima iz akcelerometra i žiroskopa, ovaj rad istražuje kako se potrošnja goriva mijenja ovisno o stilu vožnje. Razvijen je model za predikciju potrošnje goriva korištenjem algoritma Random Forest Regression. Podaci su prikupljeni iz senzora ugrađenih u mobilnim uređajima, a obuhvaćaju razdoblje vožnje te su podijeljeni u tri kategorije – agresivna, normalna i spora vožnja. Model je treniran na segmentiranim podacima, a njegova točnost je testiranja na odvojenom i još neviđenom skupu podataka. Postignuta je visoka, skoro savršena, točnost za sve stilove vožnje. Uspješno su implementirani sustavi upozorenja i nagrađivanja vozača koji mogu pomoći u poticanju ekonomičnije vožnje, smanjujući potrošnju goriva i emisije stakleničkih plinova.

**Ključne riječi:** strojno učenje; Random Forest Regression; predikcija potrošnje goriva; korekcija navika

# Driver behaviour correction using machine learning

## Abstract

Fuel consumption in vehicles is greatly influenced by driving style, which can vary from slow to aggressive. Utilizing machine learning on collected data from accelerometers and gyroscopes, this paper investigates how fuel consumption changes depending on the driving style. A model for predicting fuel consumption was developed using the Random Forest Regression algorithm. The data was collected from sensors embedded in mobile devices, covering the driving period and categorized into three categories – aggressive, normal and slow driving. The model was trained on segmented data, and its accuracy was tested on separate and previously unseen dataset. High, nearly perfect accuracy was achieved for all driving styles. Successfully implemented driver warning and reward system can help promote more economical driving, reducing fuel consumption and greenhouse gas emissions.

**Keywords:** machine learning; Random Forest Regression; fuel consumption prediction; habit correction