

# Analiza utjecaja meteoroloških prilika i stanja kolnika na poštivanje ograničenja brzine

---

**Batarilo, Zvonimir**

**Master's thesis / Diplomski rad**

**2024**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Transport and Traffic Sciences / Sveučilište u Zagrebu, Fakultet prometnih znanosti**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:119:127817>

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

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



*Repository / Repozitorij:*

[Faculty of Transport and Traffic Sciences - Institutional Repository](#)



Sveučilište u Zagrebu  
Fakultet prometnih znanosti

**Zvonimir Batarilo**

**ANALIZA UTJECAJA METEOROLOŠKIH PRILIKA I  
STANJA KOLNIKA NA POŠTIVANJE OGRANIČENJA BRZINE**

**DIPLOMSKI RAD**

**Zagreb, 2024.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET PROMETNIH ZNANOSTI**  
**POVJERENSTVO ZA DIPLOMSKI ISPIT**

Zagreb, 20. rujna 2024.

Zavod: **Zavod za inteligentne transportne sustave**  
Predmet: **Inteligentni transportni sustavi II**

**DIPLOMSKI ZADATAK br. 7692**

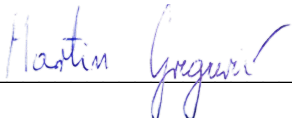
Pristupnik: **Zvonimir Batarilo (0135247828)**  
Studij: **Inteligentni transportni sustavi i logistika**  
Smjer: **Inteligentni transportni sustavi**

Zadatak: **Analiza utjecaja meteoroloških prilika i stanja kolnika na poštivanje ograničenja brzine**

Opis zadatka:

Temeljna svrha istraživanja je dobiti uvid u utjecaj analiziranih meteoroloških značajka i značajka stanja kolnika na brzinu prometnog toka u svrhu preporuka strategija uvođenja ITS usluga za povećanje sigurnosti prometa. Spomenute strategije se odnose na povećanje broja vozača koji poštuju ograničenje brzine u nepovoljnim meteorološkim i kolničkim prilikama. Cilj istraživanja je vizualizirati i analizirati podatke te kreirati model odabira značajki koji će, na osnovu metoda strojnog učenja, uspostaviti odnos značajki brzine vozila sa meteorološkim parametrima i parametrima stanja kolnika. Također, pružit će statistički osvrt na poštivanje ograničenja brzine u slučaju odabrane značajke koja postiže negativni utjecaj na promet.

Mentor:



doc. dr. sc. tech. Martin Gregurić

Predsjednik povjerenstva za diplomski ispit:

---

Sveučilište u Zagrebu  
Fakultet prometnih znanosti

**DIPLOMSKI RAD**

**ANALIZA UTJECAJA METEOROLOŠKIH PRILIKA I  
STANJA KOLNIKA NA POŠTIVANJE OGRANIČENJA BRZINE**

**Analysis Of Meteorological And Road Conditions On Speed  
Limit Obedience**

Mentor: doc. dr. sc. tech. Martin Gregurić

Student: Zvonimir Batarilo

JMBAG: 0135247828

Zagreb, rujan 2024.

## **Zahvale**

Zahvaljujem se svojem mentoru doc. dr. sc. tech. Martinu Greguriću što mi je pomogao napisati ovaj diplomski rad. Njegovo razumijevanje i beskrajno strpljenje bilo je nešto bez čega nisam mogao završiti ovaj rad. Zahvaljujem se svojoj obitelji i prijateljima za svu potporu tijekom mog studiranja.

## Sažetak

**Naslov:** Analiza utjecaja meteoroloških prilika i stanja kolnika na poštivanje ograničenja brzine

Znanost o podacima (engl. *Data science*) je računalna disciplina koja se bavi problemom izvlačenja znanja iz strukturiranih i nestrukturiranih podataka, uključujući velike skupove podataka. Jedna takva velika skupina podataka nudi uvid o brzinama i kategorijama vozila koja se kreću dionicom Zagreb-Karlovac na autocesti A1 u rasponu od 3 godine od 2018. godine do 2020. godine. Također, za isti period zabilježeni su i meteorološki podaci za tu istu dionicu. Podaci sadržavaju informacije kao što su: vrsta i intenzitet oborina; temperatura zraka i kolnika, točke ledišta kolnika; vidljivosti; debljini vodenog sloja; maksimalnoj brzini vjetra; opasnosti od poledice i stanju površine kolnika. Svaka ta meteorološka značajka i značajka o stanju kolnika utječe na brzinu vožnje motornih vozila na promatranoj dionici. Unutar diplomskog rada vizualizirat će se ti utjecaji kako bi se utvrdile one značajke koje najviše utječu na brzinu motornih vozila. Stvorit će se model korištenjem metoda strojnog učenja, odabir meteoroloških značajki i značajki stanja kolnika s najvećim utjecajem na brzinu vozila na autocesti. Zatim će se provesti analiza koja će dovesti u korelaciju poštivanje ograničenja brzine u odnosu na odabrane značajke. Rezultati će odgovarajuće vizualizirati i interpretirati u kontekstu mogućih intervencija u povećanju poštivanja ograničenja brzine u nepovoljnim prilikama.

**Ključne riječi:** meteorološki podaci, prosječna brzina, metoda strojnog učenja, odabrane značajke

## **Abstract**

**Title:** Analysis of meteorological and road conditions on speed limit obedience

**Data science** is a computational discipline that addresses the problem of extracting knowledge from structured and unstructured data, including large datasets. One such large dataset provides insight into the speeds and categories of vehicles traveling the Zagreb-Karlovac section of the A1 highway over a period of 3 years, from 2018 to 2020. Additionally, meteorological data for the same section were recorded during the same period. The data include information such as: type and intensity of precipitation; air and road surface temperatures; freezing points of the road surface; visibility; thickness of the water layer; maximum wind speed; risk of ice formation, and the condition of the road surface. Each of these meteorological features, as well as the road condition features, affects the speed of vehicles on the observed section. Within this thesis, these influences will be visualized to determine which features have the greatest impact on vehicle speed. A model will be created using machine learning methods to select the meteorological and road condition features with the greatest influence on vehicle speed on the highway. Following this, an analysis will be conducted to correlate speed limit compliance with the selected features. The results will be appropriately visualized and interpreted in the context of potential interventions to increase compliance with speed limits under adverse conditions.

**Keywords:** meteorological data, average speed, machine learning method, feature selection

# Sadržaj

<b>1. Uvod .....</b>	<b>1</b>
<b>2. Okruženje za analizu prometne sigurnosti – autocesta Zagreb-Karlovac .....</b>	<b>3</b>
2.1. Geografski položaj prometnog čvorišta Donja Zdenčina .....	3
2.2. Prikupljanje prometnih i meteoroloških podataka .....	4
2.3. Python okruženje i strojno učenje.....	6
<b>3. Predobrada i deskriptivna analiza prometnih podataka .....</b>	<b>7</b>
3.1. Deskriptivna analiza prometnih podataka.....	7
3.2. Predobrada prometnih podataka .....	9
<b>4..... Predobrada i deskriptivna analiza meteoroloških podataka i podataka o stanju kolnika na autocesti.....</b>	<b>12</b>
4.1. Predobrada meteoroloških podataka i podataka o stanju kolnika na autocesti. 12	
4.1.1. Grupiranje podataka.....	13
4.1.2. Kodiranje kategorijskih varijabli .....	14
4.2. Deskriptivna analiza meteoroloških podataka i podataka o stanju na kolniku . 15	
4.2.1. Statistički pregled osnovnih varijabli.....	15
4.2.2. Korelacijska analiza .....	17
<b>5. Primjena strojnog učenja u odabiru značajki iz skupa podataka.....</b>	<b>19</b>
5.1. Metode odabira značajki .....	20
5.1.1. Filtracijske metode.....	21
5.1.2. Omotne metode.....	21
5.2. Primjena metoda odabira značajki .....	22
<b>6. Odabir meteoroloških značajki i značajki o stanju kolnika s najvećim utjecajem na promjenu brzine na autocesti .....</b>	<b>28</b>
6.1. Rekurzivna eliminacija značajki (RFE).....	28
6.2. Logistička regresija.....	29



6.3. Primjena metode RFE na odabir najboljih meteoroloških značajki i značajki o stanju na kolniku .....	30
<b>7. Analiza poštivanja ograničenja s obzirom na odabrane značajke .....</b>	<b>35</b>
<b>8. Zaključak.....</b>	<b>40</b>
<b>Literatura.....</b>	<b>41</b>
<b>Popis slika .....</b>	<b>43</b>
<b>Popis tablica.....</b>	<b>44</b>

# 1. Uvod

Promet na cestama uvelike ovisi o raznim čimbenicima, a jedan od najznačajnijih je upravo prosječna brzina vozila. U modernim sustavima prometne sigurnosti, ključno je razumjeti što utječe na brzinu kretanja vozila kako bi se optimizirali uvjeti na cestama i poboljšala sigurnost svih sudionika u prometu. Među brojnim čimbenicima koji mogu utjecati na brzinu, meteorološke prilike i stanje kolnika zauzimaju posebno važno mjesto.

Različiti meteorološki uvjeti, poput oborina, temperature zraka, snijega i vjetra, mogu značajno utjecati na vozačke navike i sigurnost na cesti. Na primjer, kišne ili snježne padaline često uzrokuju smanjenje brzine kako bi se osigurala sigurnost vozila u teškim uvjetima. Stanje kolnika, koje može uključivati prisutnost vode, leda ili smanjenje trenja zbog loših vremenskih uvjeta, također igra ključnu ulogu u odabiru brzine kretanja.

Istraživanje utjecaja meteoroloških uvjeta i stanja na kolniku na prosječnu brzinu vozila važno je ne samo iz perspektive prometne sigurnosti, nego i za razvoj učinkovitih prometnih strategija. Na temelju ovih podataka moguće je optimizirati ograničenja brzine u realnom vremenu, prilagoditi sustave za nadzor prometa i pružiti vozačima relevantne informacije putem pametnih prometnih znakova.

Sadržaj diplomskog rada:

1. Uvod
2. Okruženje za analizu prometne sigurnosti – autocesta Zagreb-Karlovac
3. Predobrada i deskriptivna analiza prometnih podataka
4. Predobrada i deskriptivna analiza meteoroloških podataka i podataka o stanju kolnika na autocesti
5. Primjena strojnog učenja u odabiru značajki iz podatkovnih skupova
6. Odabir meteoroloških značajki i značajki o stanju kolnika s najvećim utjecajem na promjenu brzine na autocesti
7. Analiza poštivanja ograničenja brzine s obzirom na odabrane značajke
8. Zaključak

U drugom poglavlju predstaviti će se okruženje u kojem su snimljeni prometni i meteorološki podaci, kao i okruženje u kojem će se obrađivati ti isti podaci u sklopu diplomskog rada. U trećem poglavlju vršiti će se predobrada prometnih podataka, što uključuje sumiranje, konverziju podatkovnih tipova te grupiranje uz primjer vizualizacije radi potvrde ispravnosti rada. Četvrto poglavlje pokriva iste ciljeve kao i treće, samo što se obrađuju meteorološki podaci i podaci o stanju na kolniku.

U petom poglavlju vrše se statističke analize i provode se metode odabira parametara koji najviše utječu na prosječnu brzinu vozila. U šestom poglavlju također se vrše analize, ali putem nove metode u svrhu kasnijeg uspoređivanja rezultata koji će se prikazati u sedmom poglavlju.

## 2. Okruženje za analizu prometne sigurnosti – autocesta Zagreb-Karlovac

Za svrhu ovog diplomskog rada, te kao ogledan uzorak za analizu prometne sigurnosti na autocesti Zagreb-Karlovac, uzeto je prometno čvorište Donja Zdenčina.

### 2.1. Geografski položaj prometnog čvorišta Donja Zdenčina

Čvor Donja Zdenčina, jedan je od ključnih prometnih čvorišta na autocesti Zagreb-Karlovac, smješten približno 25 kilometara jugozapadno od Zagreba. Ovaj čvor povezuje lokalne i regionalne prometnice s autocestom A1, najdužom hrvatskom autocestom, te predstavlja važnu točku u prometnom sustavu, osobito za tranzitne vozače koji se kreću između glavnog grada Zagreba i južnih dijelova Hrvatske (Slika 1).



Slika 1: Prikaz čvora Donja Zdenčina

S obzirom na sve veće prometno opterećenje na ovoj dionici autoceste, čvor Donja Zdenčina igra značajnu ulogu u osiguravanju efikasnog protoka vozila, ali i u održavanju visoke sigurnosti. S obzirom na geografsku poziciju, ovaj čvor omogućava priključak lokalnog prometa

iz obližnjih mjesta i naselja na autocestu A1, što ga čini ključnim za stanovništvo koje svakodnevno migrira prema Zagrebu i obližnjim gospodarskim centrima.

## 2.2. Prikupljanje prometnih i meteoroloških podataka

Analiza prometa temelji se na prikupljanju i obradi različitih vrsta podataka koji omogućuju detaljan uvid u prometne obrasce, karakteristike vozila i potencijalne sigurnosne rizike. Podaci o prometu su ključni za donošenje informiranih odluka o upravljanju cestovnom infrastrukturom, sigurnosnim mjerama i dugoročnim planovima za unapređenje prometne mreže. Prikupljanje prometnih podataka na čvoru Donja Zdenčina vrši se vizualnim sustavom kamera, pod vlasništvom HAK-a, Hrvatskog autokluba,



Slika 2: Prikaz snimanja kamere na promatranom čvoru

Za svrhu diplomskog rada, upotrijebljen je set podataka koji se sastoji od broja i tipa vozila, te od prosječne brzine tih vozila u rasponu od 2018. godine do 2020. godine. Podaci su grupirani po pojedinačnom satu unutar ovog perioda, te podijeljeni na osnovu broja vozila, tipa vozila te određenog raspona brzina koja su zabilježena kod tih vozila, za oba smjera (Zagreb-Karlovac) te za oba prometna traka (vozni i pretjecajni trak).

Meteorološki uvjeti imaju značajan utjecaj na prometnu sigurnost. Podaci o vremenskim uvjetima, poput temperature, padalina, vjetra i vidljivosti, pružaju kontekst za analizu ponašanja vozača i njegovog vozila. Na primjer, kiša, magla, snijeg i niske temperature često uzrokuju

smanjenje brzine i povećanje rizika od nesreća zbog skliskih cesta, smanjene vidljivosti i neprilagođene brzine. Posljedično tome, oborine su meteorološki atribut koji se najviše istražuje u studijama o sigurnosti prometa. Dovoljno je malo kiše da stvori uvjete koji predstavljaju opasnost na cesti, ukoliko se umiješa i drugi čimbenik poput niske temperature, ta opasnost se drastično povećava te broj zabilježenih prometnih nesreća proporcionalno raste [1]. Meteorološki podaci za analizu unutar ovog diplomskog rada prikupljeni su pomoću meteorološke stanice Bosiljevo. Snimljeni su u obliku seta podataka u kojem su navedeni vremenski uvjeti i stanje na kolniku za čvor Donja Zdenčina za period od 2018. godine do 2020. godine, snimljeni su na način da svaka nova promjena predstavlja jedan redak unutar tog seta podataka (Tablica 1).

Tablica 1: *primjer jednog seta podataka (Intenzitet oborina)*

Rbr.	Datum	Senzor	Vrijednost
1	31.12.2018 10:33:00	čvor Donja Zdenčina - Intenzitet oborina	0 mm/h
2	31.12.2018 10:30:01	čvor Donja Zdenčina - Intenzitet oborina	0.10 mm/h
3	31.12.2018 09:14:06	čvor Donja Zdenčina - Intenzitet oborina	0 mm/h
4	31.12.2018 09:13:00	čvor Donja Zdenčina - Intenzitet oborina	0.10 mm/h
5	31.12.2018 09:12:00	čvor Donja Zdenčina - Intenzitet oborina	0.20 mm/h
6	27.12.2018 09:43:00	čvor Donja Zdenčina - Intenzitet oborina	0 mm/h
7	27.12.2018 09:42:00	čvor Donja Zdenčina - Intenzitet oborina	0.60 mm/h
8	20.12.2018 22:31:00	čvor Donja Zdenčina - Intenzitet oborina	0 mm/h
9	20.12.2018 22:30:00	čvor Donja Zdenčina - Intenzitet oborina	0.10 mm/h
10	20.12.2018 22:29:00	čvor Donja Zdenčina - Intenzitet oborina	0.20 mm/h
11	20.12.2018 22:28:00	čvor Donja Zdenčina - Intenzitet oborina	0.30 mm/h
12	20.12.2018 22:27:00	čvor Donja Zdenčina - Intenzitet oborina	0.10 mm/h
13	20.12.2018 20:10:00	čvor Donja Zdenčina - Intenzitet oborina	0 mm/h
14	20.12.2018 20:09:00	čvor Donja Zdenčina - Intenzitet oborina	0.10 mm/h
15	20.12.2018 20:08:00	čvor Donja Zdenčina - Intenzitet oborina	0.20 mm/h

Prikaz podataka o vremenskim uvjetima zajedno s prometnim podacima omogućuje identifikaciju vremenskih razdoblja u kojima je povećan rizik od prometnih incidenata. Ovakva analiza može pomoći u planiranju sezonskih sigurnosnih mjera, poput postavljanja privremenih ograničenja brzine ili povećane prisutnosti cestovnih službi tijekom zimskih mjeseci.

### 2.3. Python okruženje i strojno učenje

Python je jedan od najpopularnijih programskih jezika danas, prvenstveno zbog svoje jednostavne sintakse, fleksibilnosti i širokog spektra primjena. Koristi se u mnogim industrijama, od web razvoja i znanstvenih istraživanja do analize podataka, automatizacije i strojnog učenja. Njegova jednostavnost čini ga idealnim za početnike, dok njegova moćna biblioteka omogućava profesionalcima da razvijaju složene aplikacije i algoritme. Strojno učenje je područje računalne znanosti koje se bavi razvojem algoritama i modela koji omogućuju računalima da uče iz podataka i donose odluke bez eksplicitnog programiranja za svaki specifičan zadatak [2].

Python biblioteke koje su upotrijebljene za provođenje algoritama strojnog učenja su:

- **Pandas** – jedna od najpoznatijih i najkorištenijih biblioteka za analizu podataka u Pythonu, posebno kada je riječ o radu s tabličnim podacima. Pandas omogućuje jednostavno upravljanje i analizu podataka u obliku *DataFrame* objekta, koji je sličan tablicama u bazi podataka ili proračunskim tablicama [3]
- **Numpy** – osnovna Python biblioteka za numeričke izračune, koja je posebno učinkovita pri radu s velikim nizovima (arrays) i matricama. Osigurava snažne matematičke funkcije koje se koriste za obavljanje složenih znanstvenih izračuna, statističkih analiza i linearne algebre [4].
- **Sklearn** – pruža jednostavne, ali učinkovite alate za strojno učenje (eng. *Machine Learning*) i analizu podataka. Ova biblioteka omogućuje implementaciju različitih modela strojnog učenja, od jednostavnih do naprednih, uz minimalan otpor [5].
- **Matplotlib** – popularna za stvaranje statičkih, animiranih i interaktivnih vizualizacija. Omogućuje korisnicima stvaranje visokokvalitetnih vizualizacija koje se mogu prilagoditi različitim potrebama, što ga čini ključnim alatom za analizu podataka [6]
- **Seaborn** – oslanja se na Matplotlib biblioteku, pojednostavljuje stvaranje estetski ugodnih i informativnih grafova i olakšava rad sa statističkim podacima [7].

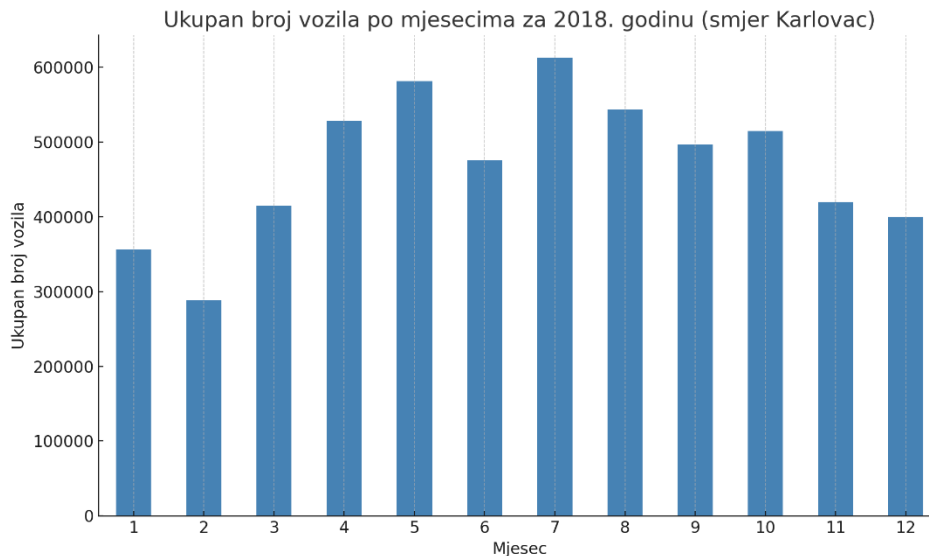
Ove biblioteke omogućuju brzu i učinkovitu analizu podataka, što Python čini savršenim alatom za analize prometnih i meteoroloških podataka.

### 3. Predobrada i deskriptivna analiza prometnih podataka

Predobrada podataka (eng. *Data Preprocessing*) je ključni korak u procesu strojnog učenja i analize podataka. Uključuje niz tehnika i procesa kojima se podaci pripremaju za analizu ili treniranje modela strojnog učenja. Pravilna predobrada pomaže poboljšati točnost, učinkovitost i performanse modela, jer stvarni podaci često dolaze s nedostacima kao što su nedostajući podaci, šum, nepodudaranje u formatima ili neprikladne skale.

#### 3.1. Deskriptivna analiza prometnih podataka

Priprema prometnih podataka, kao i priprema meteoroloških podataka, služi kako bi se postavili svi potrebni uvjeti za treniranje modela strojnog učenja. U tu svrhu, podaci, koji se nalaze u različitim setovima podataka, će biti obrađeni na način da se spoje u jednu glavnu tablicu koja će sadržavati sve atribute ili parametre koji utječu na prosječnu brzinu vozila. Prometni podaci su većim dijelom već sređeni bili i nije bilo potrebno mnogo toga predobraditi. Za svakih sat vremena svakog dana u periodu od 01.01.2018. do 31.12.2020. godine postoje podaci o broju vozila, njihovom tipu, njihovoj pojedinačnoj brzini te njihovoj prosječnoj brzini kako za voznu traku, tako i za pretjecajnu traku. Na slici 3, može se vidjeti usporedba broja vozila za svaki mjesec u 2018. godini za smjer Karlovac.

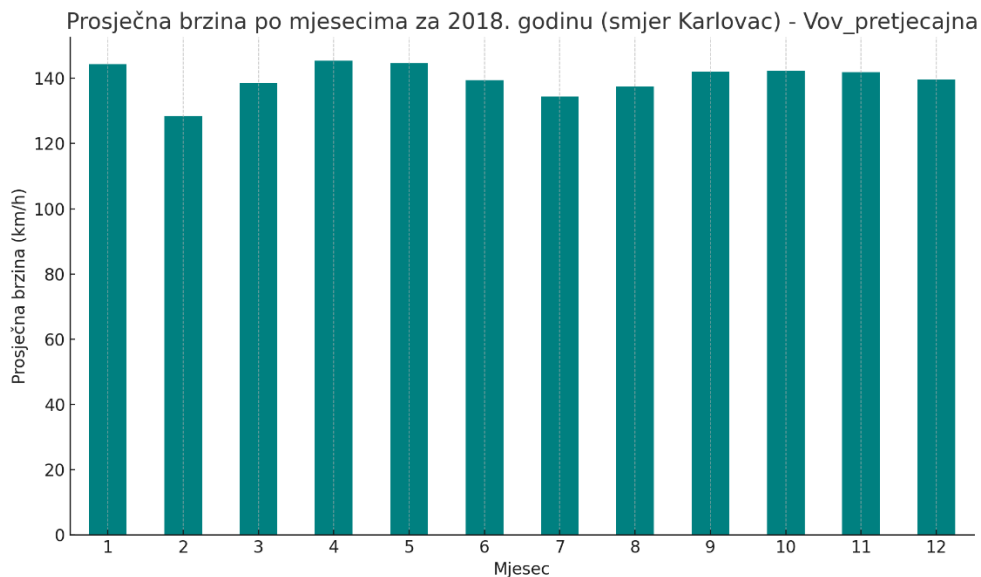


Slika 3: Ukupan broj vozila po mjesecima za 2018. godinu (smjer Karlovac)



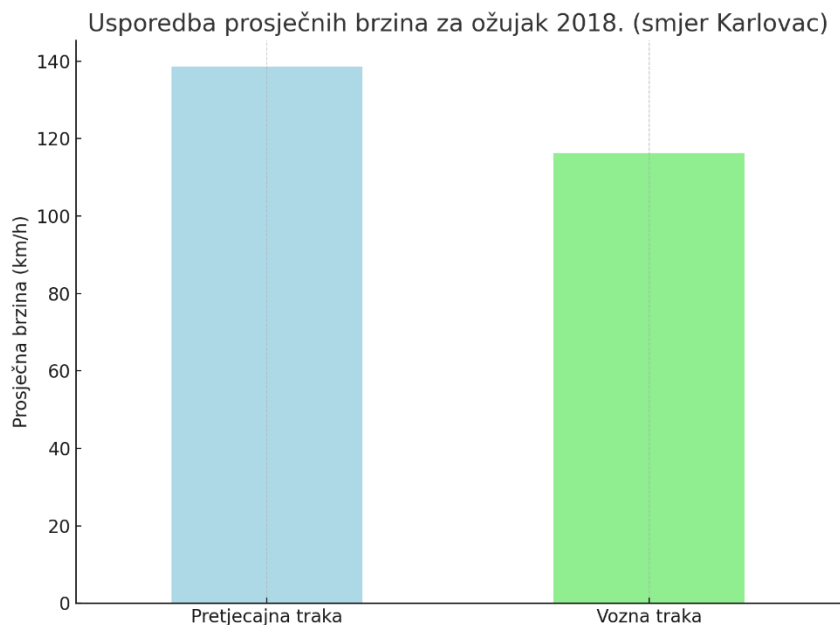
Na osnovu slike 3, može se zaključiti da postoji ogromna razlika u mjesečnom broju vozila kroz godinu. Već nakon ožujka, bilježi se značajan skok u prometu, što za Hrvatsku i nije neobično jer su to počeci sezonskog perioda.

Na slici 4, prikazane su prosječne brzine vozila u pretjecajnoj traci po mjesecima za 2018. godinu. Može se primijetiti povećan trend vožnje iznad dozvoljenog ograničenja od 100 km/h s obzirom da su vozila snimana na području prometnog čvorišta, a ne na otvorenoj autocesti gdje je ograničenje 130 km/h. Unatoč tome, vidljive su razlike u ljetnim mjesecima gdje je prosječna brzina pala ispod 130km/h zbog učestalih prometnih zagušenja koje su uzrokovane dolaskom znatnog broja turista, te najmanje je u veljači kad je vrijeme bilo najmanje povoljno.



Slika 4: Prosječna brzina po mjesecima za 2018. godinu (smjer Karlovac)

U nastavku, u ovom diplomskog radu, bit će iskorišteni podaci iz ožujka 2018. godine kao reprezentativni mjesec koji će poslužiti za treniranje modela strojnog učenja. Izabran je jer bilježi najmanje anomalija, što tijekom ljetnih, a što tijekom zimskih perioda.

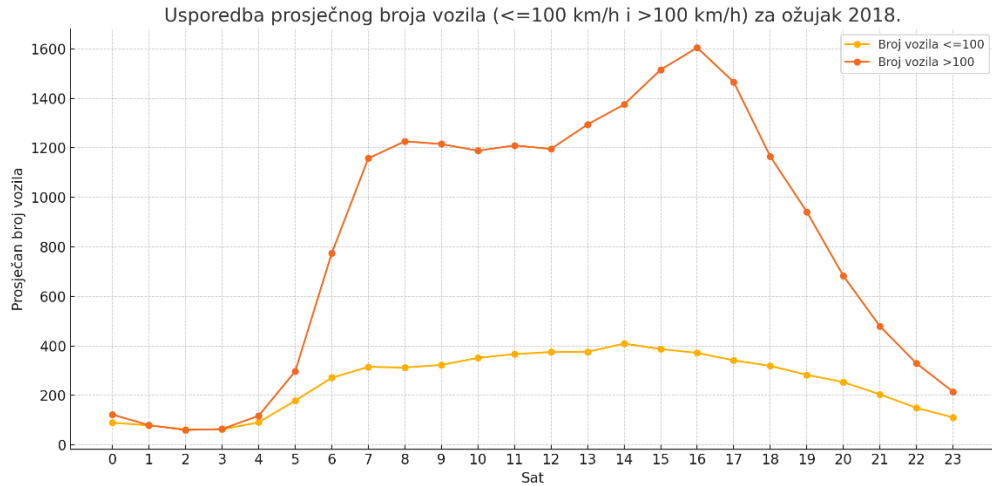


Slika 5: Usporedba prosječnih brzina po prometnim tracima

Na slici 5, uspoređene su prosječne vrijednosti brzina pretjecajne i vozne trake u smjeru Karlovca. Razlika je sasvim uočljiva i razumljiva jer je pretjecajna traka ipak brža od vozne, međutim, važno je naglasiti da se ni u voznoj traci vozila ne drže ograničenja od 100 km/h, što može upućivati prema lošijoj svijesti vozača o cestovnom ograničenju.

### 3.2. Predobrada prometnih podataka

Iako je većina podataka već bila poprilično sređena, podaci, koji su bili potrebni za spajanje u glavnu tablicu, još nisu pripremljeni do kraja. Za potrebe primjene strojnog učenja, napravljena je suma vozila koja su vozila po ograničenju ( $\leq 100$  km/h) te suma vozila koja je vozila iznad ograničenja ( $> 100$  km/h), za oba smjera (Zagreb-Karlovac), takav odnos prikazan je na slici 6. Na osnovu toga, dobivena je razlika između tih suma brojeva po svakom satu unutar mjeseca ožujka, kako bi se utvrdilo je li razlika pozitivna ili negativna.



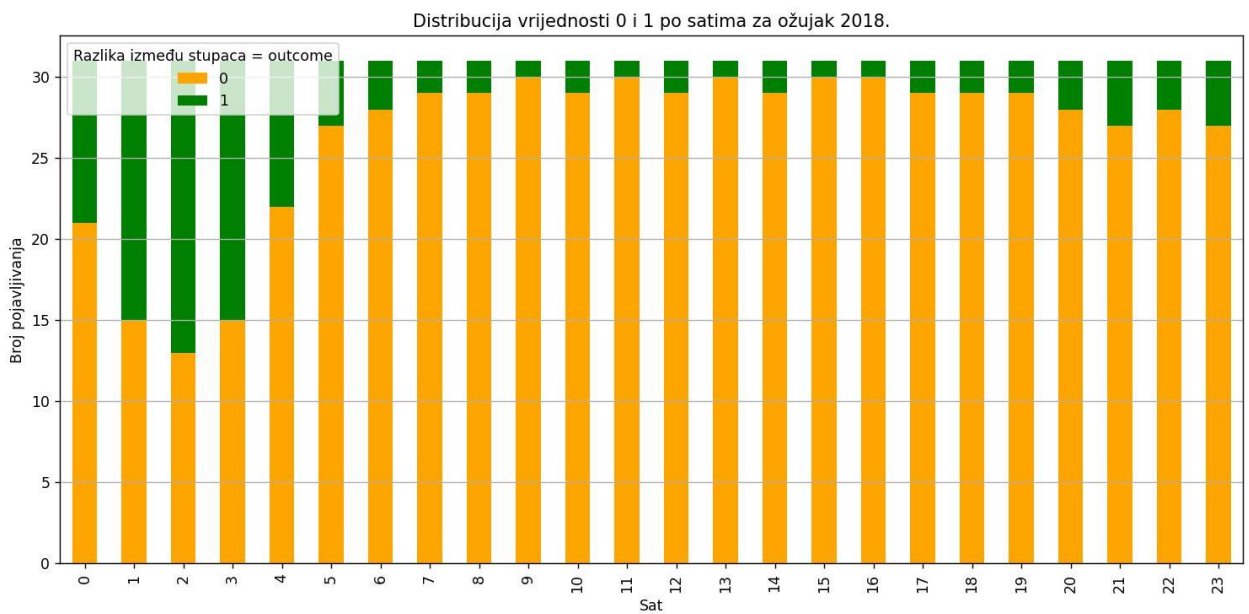
Slika 6: Usporedba prosječnog broja vozila po ograničenju

Jedan od ključnih koraka predobrade podataka jest transformacija podataka. To je korak koji uključuje prilagodbu podataka kako bi bili prikladni za algoritme strojnog učenja. Izračunata razlika među vozilima, poslužit će kao konačni rezultat (eng. *Outcome*) na kojeg utječu prometni i meteorološki čimbenici. Razlika između vozila koja su ispod i iznad ograničenja, prikazana je u stupcu “Razlika između stupaca = outcome”. Ti podaci u stupcu, pretvoreni su u binarne vrijednosti 1 ili 0, ovisno o tome je li razlika pozitivna ili negativna (Slika 7.)



Slika 7: Prosječna razlika po satu za ožujak 2018.

Na slici 7, prikazana je prosječna vrijednost po satu unutar mjeseca ožujka 2018. godine, na kojoj se može primijetiti da vozači bolje voze u noćnim satima kada je manja vidljivost i veći oprez. Tamo gdje je razlika pozitivna, vrijednost je 1, a ako je negativna, vrijednost je 0. Tijekom dana, prosječna vrijednost više gravitira prema nuli, te s time naglašava da sve manje vozača poštuje ograničenje na ovom prometnom čvorištu. Na slici 8, prikazano je koliko puta je zabilježena vrijednost 0 i koliko 1 za svaki sat. Ovaj prikaz puno jasnije elaborira odnos u kojem je negativna razlika značajno izraženija upućujući na negativan trend nepoštovanja ograničenja brzine.



Slika 8: Distribucija vrijednosti 0 i 1 po satima

## 4. Predobrada i deskriptivna analiza meteoroloških podataka i podataka o stanju kolnika na autocesti

U ovom poglavlju analiziraju se meteorološki podaci i podaci o stanju kolnika na autocesti, s ciljem utvrđivanja njihovog utjecaja na prometnu sigurnost. Proces započinje predobradom podataka, uključujući rukovanje nedostajućim vrijednostima i normalizaciju. Nakon toga provodi se deskriptivna analiza kako bi se statistički opisale ključne varijable poput temperature, oborina i stanja na kolniku, te njihov međusobni odnos kroz vrijeme.

### 4.1. Predobrada meteoroloških podataka i podataka o stanju kolnika na autocesti

Meteorološki podaci i podaci o stanju kolnika prikupljeni su putem meteorološke stanice Bosiljevo koja je prikupila podatke za dionicu autoceste kod prometnog čvorišta Donja Zdenčina. Podaci obuhvaćaju razdoblje od 01. siječnja 2018. godine do 31. prosinca 2020. godine, te uključuju informacije o temperaturi zraka, temperaturi kolnika, temperaturi smrzavanja, brzini vjetra, intenzitetu i vrsti oborina, vidljivosti, kao i stanju površine kolnika, debljini vodenog sloja na kolniku i opasnosti od poledice. Svaki parametar imao je vlastiti set podataka koji je grupiran za period od ožujka 2018. godine u jednu veliku *master* tablicu (Tablica 2) u svrhu izvršavanja modela strojnog učenja.

Tablica 2: *Master tablica sa svim grupiranim atributima*

F	G	H	I	J	K	L	M	N	O
Vidljivost	Vrste oborina	Max. brzina vjetra	Intenzitet oborina	Debljina vodenog sloja	Opasnost od poledice	Stanje površine kolnika	Temperatura kolnika	Temperatura zraka	Temperatura smrzavanja
0	0	1.31	0.082352941	0.03	0	0	-10.3875	-15.04137931	-0.3
0	0	1.316666667	0.285	0.024	0	0	-10.75	-15.67407407	-0.52
0	0	1.3	0.307894737	0.074615385	0	0	-10.2125	-15.21666667	-0.625
0	0	1.376923077	0.8825	0.053333333	0	0	-9.525	-14.81	-0.25
0	0	1.208333333	0.262962963	0.026666667	0	0	-9.677777778	-14.38421053	-0.4
0	0	1.477777778	0.665116279	0.190869565	0	0	-9.033333333	-13.6	-0.3
0	0	1.52	0.490909091	0.1656	0	0	-8.57	-12.43157895	-0.4
0	0	1.82	0.321875	0.305357143	0	0	-7.3	-11.3	-0.3
0	6	1.491666667	0.270731707	0.234516129	0	0	-5.704545455	-10.35238095	-0.25
0	6	0.9625	0.082352941	0.03	2	0	-3.540740741	-9.721428571	-0.25
0	6	1.166666667	0.285	0.024	2	3	-0.50625	-8.525	-4.990625
0	6	1.52	0.307894737	0.074615385	0	0	-0.528571429	-7.736842105	-5.786956522
0	6	1.342857143	0.8825	0.053333333	0	3	-1.659259259	-7.4625	-4.125
0	0	1.44	0.262962963	0.026666667	0	3	-2.278571429	-7.44	-5.251515152
0	0	1.49	0.665116279	0.190869565	0	0	-2.136666667	-6.922727273	-3.579069767
0	0	1.788888889	0.490909091	0.1656	0	2	-1.907407407	-6.581481818	-2.00952381
0	6	0.861538462	0.321875	0.305357143	0	4	-3.209090909	-6.8	-1.160714286
0	6	0.544444444	0.270731707	0.234516129	0	2	-4.192857143	-7.007142857	-5.181818182
0	6	1.73	0.360606061	0.24	0	4	-4.685714286	-6.806666667	-1.053846154
0	0	1.63	0.10625	0.291153846	0	0	-4.784615385	-6.746153846	-0.52
0	0	1.116666667	0.1125	0.215263158	0	0	-4.86	-6.788235294	-0.625
0	0	1.27	0.205	0.395151515	0	0	-4.888888889	-6.75	-0.428571429
0	0	0.846153846	0.432608696	0.365384615	0	0	-4.85	-6.7	-1.076923077
0	0	1.355555556	0.142105263	1.491935484	0	0	-4.99	-6.65	-1
0	6	1.055555556	0.155555556	1.028648649	0	3	-4.976923077	-6.536363636	-2.709677419

Meteorološki podaci prikupljeni su u realnom vremenu, najčešće na način da su se bilježili promjenom vrijednosti određenog parametra. Na primjer, svaka promjena u iznosu temperature zraka zabilježena je u novom retku seta podataka. Svaka varijabla ima specifičnu ulogu u procjeni prometne sigurnosti. Tako da, u slučaju niske temperature može se očekivati i veća opasnost od poledice na kolniku, u slučaju jačeg intenziteta oborina smanjit će se vidljivost odnosno povećat će se akumulacija vode na cesti. Takvi podaci i njihove korelacije mogu se iskoristiti za izradu modela i statističku analizu, s ciljem boljeg razumijevanja kako različiti vremenski uvjeti utječu na sigurnost prometa [8].

#### 4.1.1. Grupiranje podataka

Za predobradu, potrebno je bilo grupirati podatke iz deset različitih setova meteoroloških podataka te podataka o stanju na kolniku. S obzirom da ti setovi podataka sadrže različite vrijednosti, nije ih moguće grupirati na isti način. Za vrijednosti iz tablica o temperaturama, debljini vodenog sloja kolnika, intenzitetu oborina, maksimalnoj brzini vjetera te vidljivosti, izračunate su prosječne vrijednosti kao reprezentativan uzorak za svaki sat unutar mjeseca ožujka. Na slici 9, prikazan je kod koji je grupirao podatke iz seta podataka o atributu temperaturi zraka. U svakom setu podataka postoje ćelije s praznim vrijednostima koje su također uzete u obzir kako bi svaki sat dobio svoj redak u master tablici. Te prazne vrijednosti konvertirane su kasnije u vrijednost 0 kako bi model strojnog učenja mogao iskoristiti to za usporedbu i analizu.

```
7 # Pretvaranje 'Datum' stupac u datetime format
8 df['Datum'] = pd.to_datetime(df['Datum'], format='%d.%m.%Y %H:%M:%S', errors='coerce')
9
10 # Osiguravanje da su podaci sortirani po datumu i vremenu
11 df = df.sort_values(by='Datum').reset_index(drop=True)
12
13 # Uklanjanje "mm/h" iz stupca 'Vrijednost' i pretvaranje u numerički format
14 df['Vrijednost'] = df['Vrijednost'].str.replace(' °C', '').astype(float)
15
16 # Kreiranje dodatnog stupca koji sadrži samo sat (bez minuta i sekundi)
17 df['Hour'] = df['Datum'].dt.floor('H')
18
19 # Grupiranje podataka po satu i izračunavanje prosječne vrijednosti unutar svakog sata, uključujući prazne sate
20 aggregated_data = df.groupby('Hour').agg({'Vrijednost': 'mean'}).reset_index()
21
22 # Kreiranje potpune liste satnih vremenskih oznaka za cijeli mjesec
23 all_hours = pd.date_range(start=df['Hour'].min(), end=df['Hour'].max(), freq='H')
24
25 # Reindeksiranje agregiranih podataka kako bi uključili sve sate, popunjavajući nedostajuće vrijednosti s NaN
26 aggregated_data = aggregated_data.set_index('Hour').reindex(all_hours).reset_index()
27 aggregated_data.columns = ['Hour', 'Vrijednost']
```

Slika 9: Grupiranje podataka o temperaturi zraka

### 4.1.2. Kodiranje kategorijskih varijabli

Meteorološki podaci o vrsti oborina, opasnosti od poledice te stanju površine kolnika opisani su tekstualnim vrijednostima, a ne brojčanim. Ti podaci podijeljeni su po kategorijskim vrijednostima koje je nužno konvertirati iz teksta u brojčanu vrijednost. Za konverziju, vrši se metoda za konverziju (eng. *Label encoding*). Nakon konverzije, potrebno je poredati kategorije tako da najviša kategorija ima najviši prioritet u odabiru reprezentativnog uzorka. Ukoliko je u jednom satu zabilježeno više različitih kategorijskih varijabli, reprezentativna varijabla, za taj sat, je ona s najvišim prioritetom. Na slici 10, prikazan je kod koji je grupirao podatke ovih atributa, pretvorio vrijednosti u brojčane i prioritzirao kategorije s najvišim prioritetom.

```
7 # Pretvaranje 'Datum' stupca u datetime format
8 df['Datum'] = pd.to_datetime(df['Datum'], errors='coerce')
9
10 # Filtriranje podataka za Ožujak 2018
11 march_data = df[(df['Datum'].dt.year == 2018) & (df['Datum'].dt.month == 3)].copy()
12
13 # Definiranje prioriteta za agregaciju (najniži broj je najviši prioritet)
14 priority = {
15     'Snijeg': 1,
16     'Solika': 2,
17     'Kiša': 3,
18     'Oborine': 4,
19     'Rosulja': 5,
20     'Nema oborina': 6
21 }
22
23 # Dodavanje stupca s prioritetom
24 march_data['Priority'] = march_data['Vrijednost'].map(priority)
25
26 # Grupiranje podataka po satu i zadržavanje retka s najvišim prioritetom (najmanji broj)
27 march_data['Hour'] = march_data['Datum'].dt.floor('H')
28 aggregated_data = march_data.loc[march_data.groupby('Hour')['Priority'].idxmin()]
29
30 # Zadržavanje relevantnih stupaca
31 aggregated_data = aggregated_data[['Hour', 'Vrijednost']]
```

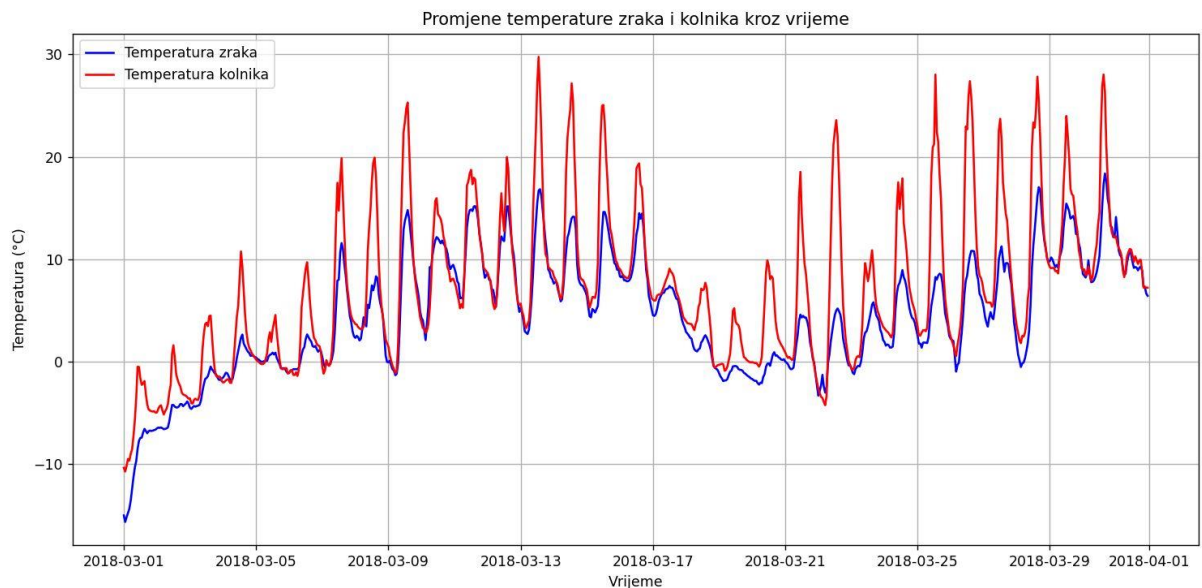
Slika 10: Predobrada tekstualnih podataka

## 4.2. Deskriptivna analiza meteoroloških podataka i podataka o stanju na kolniku

Deskriptivna analiza provedena je kako bi se dobio jasan statistički pregled meteoroloških podataka i podataka o stanju kolnika na autocesti. Fokus ove analize bio je na razumijevanju osnovnih karakteristika podataka i identifikaciji obrazaca koji mogu biti relevantni za sigurnost prometa.

### 4.2.1. Statistički pregled osnovnih varijabli

U sklopu predobrade, uzeti su neki parametri za vizualnu interpretaciju kako bi se dokazala vjerodostojnost podataka prije razvijanja algoritama strojnog učenja. Na slici 11, prikazan je graf s podacima o promjeni temperature zraka i temperature kolnika kroz mjesec ožujak.

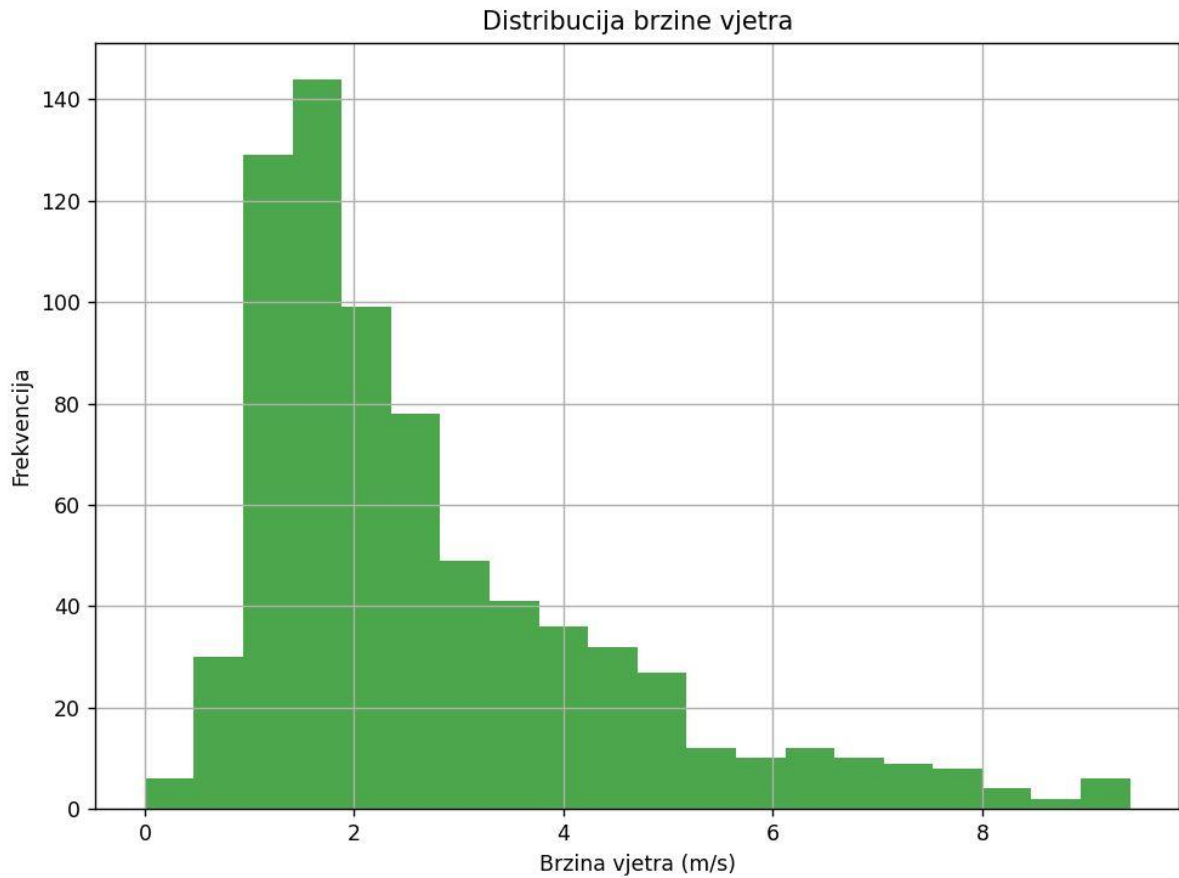


Slika 11: *Promjene temperature zraka i kolnika kroz vrijeme*

Postoji jasno vidljiva korelacija između ova dva parametra gdje temperatura kolnika ovisi o temperaturi zraka, no ipak bilježi više vrijednosti. To bi se moglo objasniti utjecajem vozila koja prolaze po tom istom kolniku povećavajući temperaturu zagrijavanjem gumama ili pak utjecajem sunca koje grije cestu.

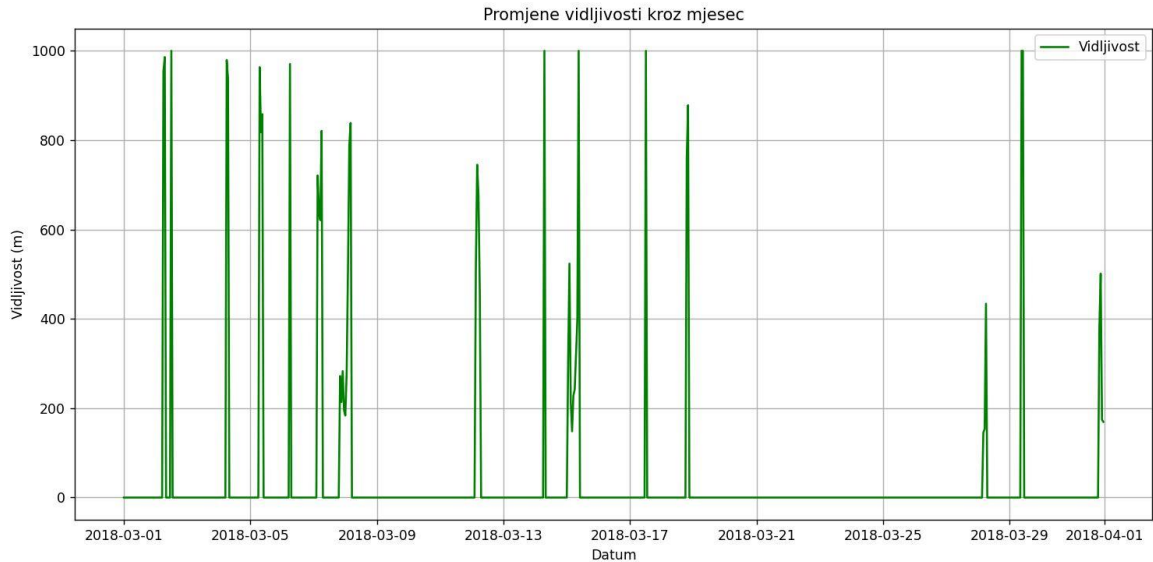


Na slici 12, opisana je distribucija brzine vjetra kroz mjesec ožujak. Najviše puta zabilježen je vjetar do 2 m/s, a najviša zabilježena vrijednost je oko 9 m/s. Može se zaključiti da maksimalna brzina vjetra nije bila mnogo jaka te da nije mnogo ugrožavala sigurnost vozača u prometu.



Slika 12: *Distribucija brzine vjetra*

Od ovih deset atributa, postoje i neki koji nemaju dovoljno dobre podatke jer nisu dovoljno puta bili zabilježeni. Jedan od tih atributa je “Vidljivost”, kojem nedostaju cijeli dani s informacijama o vidljivosti na cesti (Slika 13.). Vidljivost je značajan atribut koji možda i najviše utječe na sigurnost u prometu, međutim, takvi podaci nisu toliko opširni kao drugi.

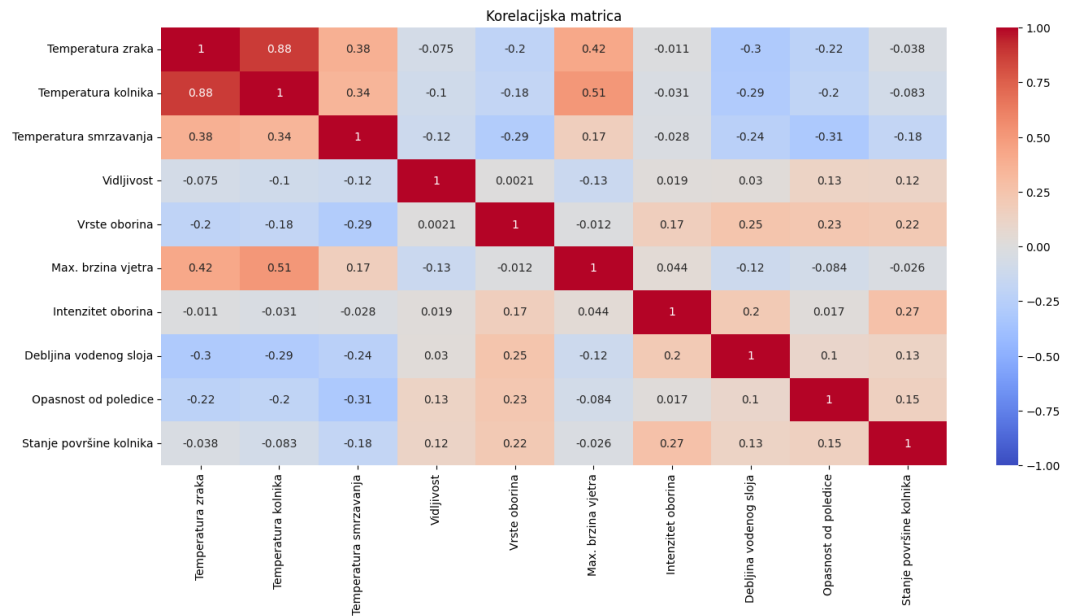


Slika 13: Promjena vidljivosti kroz mjesec

#### 4.2.2. Korelacijska analiza

Korelacijska analiza koristi se za mjerenje snage i smjera linearne veze između dviju varijabli. Rezultat korelacije izražava se korelacijskim koeficijentom koji varira između -1 i 1, gdje vrijednost 1 označava savršenu pozitivnu korelaciju, a -1 savršenu negativnu korelaciju. Korelacija 0 znači da nema linearne povezanosti između varijabli. Korelacijska analiza omogućuje bolje razumijevanje odnosa između varijabli, što je korisno u predviđanju i analizi podataka. Često se koristi u analizi podataka kako bi se identificirale varijable koje su međusobno povezane, a zatim vizualizira kroz korelacijske matrice i scatter plotove [9].

Na slici 14, napravljena je korelacijska analiza svih meteoroloških parametara i parametara o stanju na kolniku u svrhu identificiranja atributa koji su međusobno povezani.



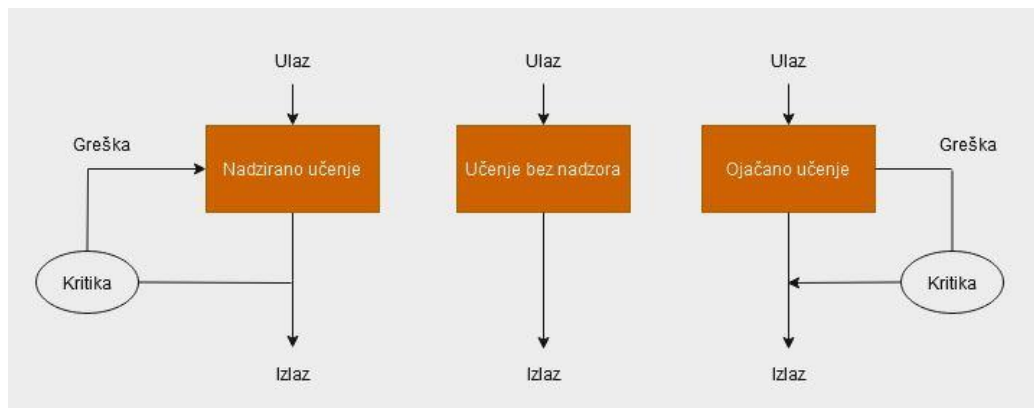
Slika 14: Korelacijska matrica svih atributa

Može se primijetiti da najviše koreliraju temperature između sebe, no, s temperaturama najviše korelira maksimalna brzina vjetra. Osim toga, značajna je i korelacija između vrste oborina, debljine vodenog sloja, opasnosti od poledice, stanju površine kolnika te intenziteta oborina. Međutim, treba napomenuti da tipovi podataka nisu isti i da se drugačije ponašaju, neki su cijeli brojevi, a neki decimalni. To bi objasnilo negativnu korelaciju temperature smrzavanja i opasnosti od poledice koji bi u teoriji morali biti jako dobro međusobno povezani.

## 5. Primjena strojnog učenja u odabiru značajki iz skupa podataka

Strojno učenje (eng. *Machine learning, ML*) predstavlja jednu od najvažnijih grana umjetne inteligencije (eng. *Artificial intelligence, AI*) koja se bavi razvojem algoritama i modela sposobnih za učenje iz podataka. Za razliku od klasičnih algoritama koji izvršavaju zadatke prema unaprijed definiranom skupu pravila, strojno učenje omogućava računalima da "nauče" obrasce iz podataka i prilagode svoja ponašanja kako bi izvršili određene zadatke s visokom razinom točnosti. S obzirom na brzinu kojom se podaci danas generiraju i potrebu za njihovom obradom, strojno učenje postaje ključna tehnologija u raznim industrijama, uključujući financije, zdravstvo, transport i mnoge druge [10].

Postoje tri glavne kategorije učenja u strojnim sustavima: nadzirano, nenadzirano i potkrijepljeno učenje (Slika 15).



Slika 15: Kategorije učenja u strojnim sustavima

Izvor: [11]

- **Nadzirano učenje** - proces gdje su podaci za treniranje strojnog modela jasno označeni, tj. svaki primjer ima pridruženu izlaznu vrijednost. Algoritmi nadziranog učenja uče iz tih podataka kako bi predvidjeli ishode na neoznačenim podacima. Primjeri nadziranih algoritama uključuju regresiju, stablo odlučivanja i neuronske

mreže. Klasifikacija, što je tema ovog rada, također je vrsta nadziranog učenja gdje je cilj predvidjeti diskretne izlazne vrijednosti [12].

- **Nenadzirano učenje** - za razliku od nadziranog učenja, radi s podacima koji nemaju unaprijed definirane oznake. Algoritmi nenadziranog učenja otkrivaju skrivene obrasce u podacima, a najpoznatiji pristupi uključuju klasteriranje (npr. K-srednji algoritam) i smanjenje dimenzionalnosti [12].
- **Učenje s potkrijepljenjem** - koristi koncept nagrade i kazne kako bi algoritmi naučili izvršavati zadatke. Ova metoda je posebno korisna u dinamičkim okruženjima gdje sustav mora naučiti optimalne akcije kroz interakcije s okolinom [12].

Strojno učenje je postalo integralni dio mnogih industrijskih sustava. U zdravstvu se koristi za dijagnostiku i personalizirano liječenje pacijenata, dok je u financijama primijenjeno u algoritamskom trgovanju i analizi rizika. U autonomnoj vožnji, strojno učenje omogućuje računalima da donose odluke u realnom vremenu na temelju podataka senzora i vizualnih informacija. Osim toga, prepoznavanje govora, filtriranje e-pošte i personalizacija preporuka također koriste strojnoučnu tehnologiju za poboljšanje korisničkog iskustva.

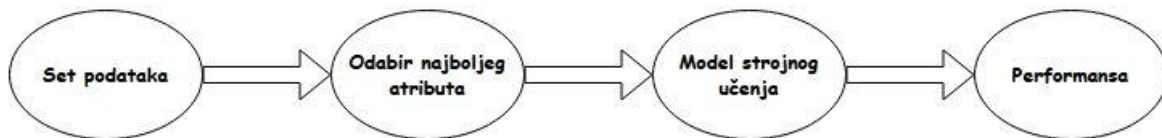
## 5.1. Metode odabira značajki

Jedan od najvažnijih koraka u procesu razvoja modela strojnog učenja je odabir značajki (eng. *feature selection*). Značajke predstavljaju ulazne podatke modela i izravno utječu na njegovu točnost, brzinu treniranja i interpretabilnost. U praksi, skup podataka često sadrži velik broj značajki koje nisu sve relevantne za postizanje optimalnih performansi modela. Uklanjanje irelevantnih ili redundantnih značajki može značajno poboljšati efikasnost modela smanjujući složenost problema. Osim toga, smanjeni broj značajki olakšava tumačenje rezultata, što je posebno važno u primjenama gdje su potrebna objašnjenja za donošenje odluka, kao što su zdravstveni sustavi ili sudske analize [13].

Kroz proces odabira značajki, strojno učenje koristi različite tehnike za identificiranje najrelevantnijih značajki. Te tehnike mogu biti filtracijske, omotne ili metode ugrađene unutar samih modela. Filtracijske metode oslanjaju se na statističke testove kako bi odabrale značajke, omotne metode treniraju modele za različite podskupove značajki, dok ugrađene metode koriste algoritme strojnog učenja koji automatski procjenjuju važnost značajki tijekom treniranja modela [13].

### 5.1.1. Filtracijske metode

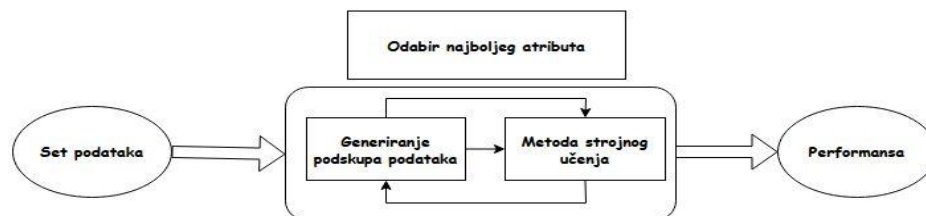
Filtracijska metoda oslanja se na opću jedinstvenost podataka koji se evaluiraju i odabire podskup značajki, ne uključujući nikakav algoritam rudarenja podataka. Ova metoda koristi osnovni kriterij rangiranja i primjenjuje metodu rangiranja za odabir varijabli. Razlog korištenja metode rangiranja je njezina jednostavnost te sposobnost da proizvede izvrsne i relevantne značajke. Metoda rangiranja će filtrirati irelevantne značajke prije nego što započne proces klasifikacije. Filtracijska metoda općenito se koristi kao korak u predobradi podataka. Na slici 16, opisan je process rada ove metode. Koristeći cijeli set podataka, odabire najbolje attribute i priprema ih za izvršavanje statističkih analiza po potrebi [14].



Slika 16: *Metoda filtriranja*

### 5.1.2. Omotne metode

Omotne metode su pristupi koji koriste model strojnog učenja kako bi procijenili kvalitetu podskupova značajki. Za razliku od filtracijskih metoda, koje procjenjuju značajke na temelju inherentnih svojstava podataka, omotne metode uključuju treniranje modela više puta s različitim podskupovima značajki kako bi se odredilo koji podskup daje najbolje performanse (Slika 17). Ove metode direktno optimiziraju performanse modela, čineći ih preciznijima, ali i računalno skupljima [14].



Slika 17: *Omotna metoda*

## 5.2. Primjena metoda odabira značajki

Da bi se primijenila filtracijska metoda odabira značajki na ovom skupu podataka unutar master tablice, mogu se koristiti statističke metode poput *Pearsonovog* koeficijenta korelacije, koji je već prethodno napravljen, *ANOVA* ili *chi-squared* test kako bi se se moglo procijeniti koje su značajke najvažnije za predviđanje ciljne varijable (eng. *Outcome*). S obzirom da su neke od ovih metoda osjetljive na vrstu podataka (npr. *Chi-squared* test zahtijeva nenegativne vrijednosti) nužno je odlučiti se za metodu koja odgovara strukturi podataka.

Za primjenu filtracijske metode, koristit će se metoda *SelectKBest* u Python okruženju, koja će rangirati značajke na temelju odabranog statističkog testa. Na slici 18 i slici 19, prikazan je kod kojim je napravljen odabir 3 najbolje značajke koje najviše doprinose cilju.

```
# Uvoz potrebnih biblioteka
import pandas as pd
from sklearn.feature_selection import SelectKBest, f_classif
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt

# Učitavanje podataka
file_path = r'C:\Users\zvoni\OneDrive\Desktop\Faks\diplomski\Master_tablica_v6.xlsx'
df = pd.read_excel(file_path)

# Izbacivanje atributa koji ne doprinose cilju
df_cleaned = df.drop(columns=["Broj vozila <=100", "Broj vozila >100"])
df_cleaned.rename(columns={"Razlika između stupaca = outcome": "outcome"}, inplace=True)

# Definiranje značajke (X) i ciljne varijable (y)
X = df_cleaned.drop(columns=["Datum", "outcome"])
y = df_cleaned["outcome"] # Ciljna varijabla

# Podjela podataka na trening i testni skup
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Primjena SelectKBest (ANOVA test) za rangiranje značajki
select_kbest = SelectKBest(f_classif, k='all')
select_kbest.fit(X_train, y_train)

# Prikupljanje rezultata i rangiranje značajki
scores = select_kbest.scores_
feature_names = X.columns
results = pd.DataFrame({"Feature": feature_names, "Score": scores})
results = results.sort_values(by="Score", ascending=False)
```

Slika 18: Prvi dio koda za odabir najbolje 3 značajke

```

# Odabir 3 najbolje značajke
top_features = results["Feature"].head(3).values
X_train_selected = X_train[top_features] # Selekcija 3 najbolje značajke
X_test_selected = X_test[top_features]

# Treniranje Random Forest modela na 3 najbolje značajke
model = RandomForestClassifier(random_state=42)
model.fit(X_train_selected, y_train)

# Predikcija i evaluacija modela
y_pred = model.predict(X_test_selected)
print("Točnost modela:", accuracy_score(y_test, y_pred))
print("Izvešće o klasifikaciji:\n", classification_report(y_test, y_pred))

# Vizualizacija važnosti značajki
plt.figure(figsize=(10, 6))
plt.barh(top_features, model.feature_importances_, color='skyblue')
plt.xlabel('Važnost značajki')
plt.title('Važnost 3 najbolje značajke u Random Forest modelu')
plt.gca().invert_yaxis() # Značajke s većom važnošću na vrhu
plt.show()

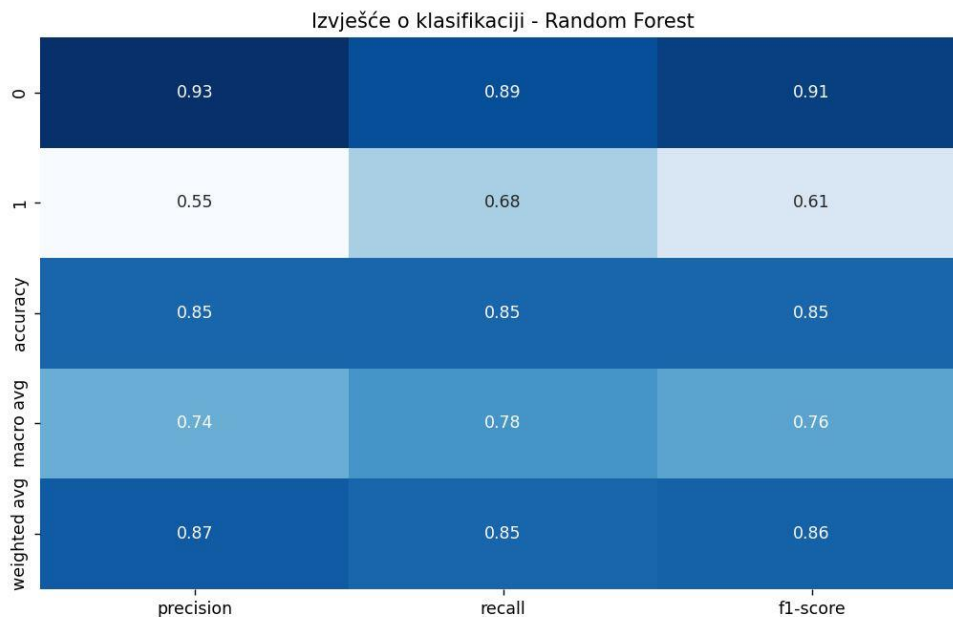
```

Slika 19: Drugi dio koda za odabir 3 najbolje značajke

Nakon primjene metode *SelectKBest* uz ANOVA (eng. *Analysis of Variance*) test, odabrane su tri najbolje značajke koje su pokazale najveći utjecaj na ishod. Upotrijebljen je *Random Forest* model na te tri značajke i evaluirana je njegova točnost na testnom skupu podataka. Točnost modela daje uvid u to koliko dobro model predviđa ishod koristeći samo odabrane značajke. Na slici 20, predstavljeno je izvješće o klasifikaciji u kojoj se mogu vidjeti preciznost, osjetljivost i F1-mjera.

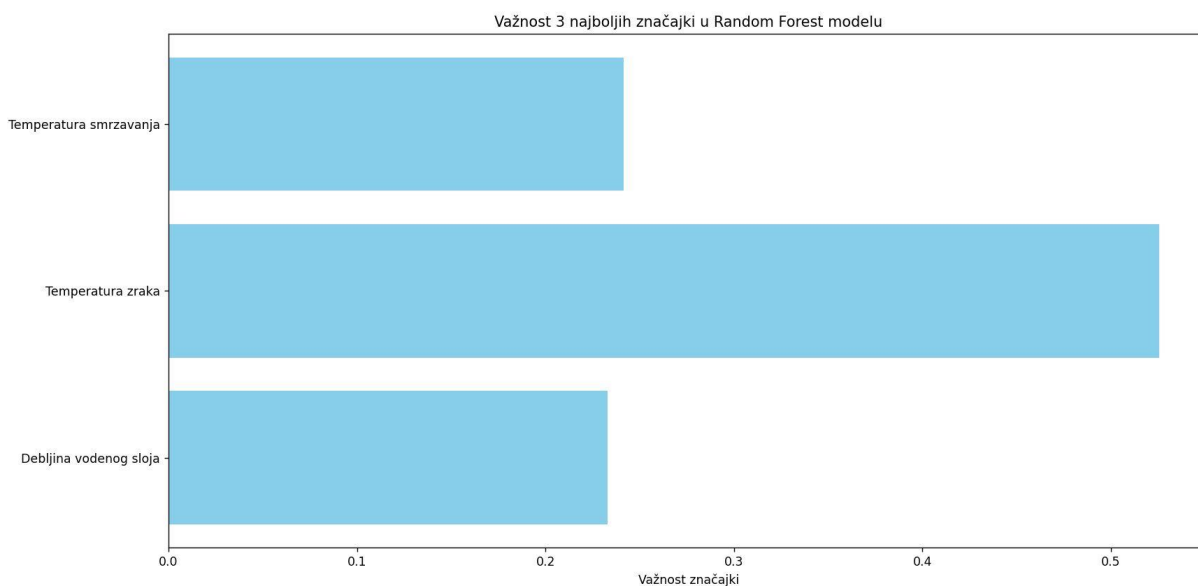
Preciznost pokazuje koliko su točne pozitivne predikcije modela, to je omjer točno predviđenih pozitivnih slučajeva i svih predviđenih pozitivnih slučajeva. Osjetljivost (eng. *Recall*) mjeri koliko je model uspješan u pronalaženju svih stvarnih pozitivnih primjera, omjer točno predviđenih pozitivnih slučajeva i svih stvarno pozitivnih slučajeva. F1-mjera je harmonijska sredina preciznosti i osjetljivosti. Koristi se kao ukupna mjera kada je važno imati ravnotežu između preciznosti i osjetljivosti [13].





Slika 20: Izvešće o klasifikaciji za 3 najbolje značajke

Nakon odabira značajki, na slici 21 vizualiziran je grafikon koji prikazuje važnost značajki u treniranom *Random Forest* modelu koristeću metodu *SelectKBest*. Na osnovu toga, može se donijeti odluka i odabrati koje značajke imaju najbolji utjecaj kako bi se pristupilo daljnjim analizama.

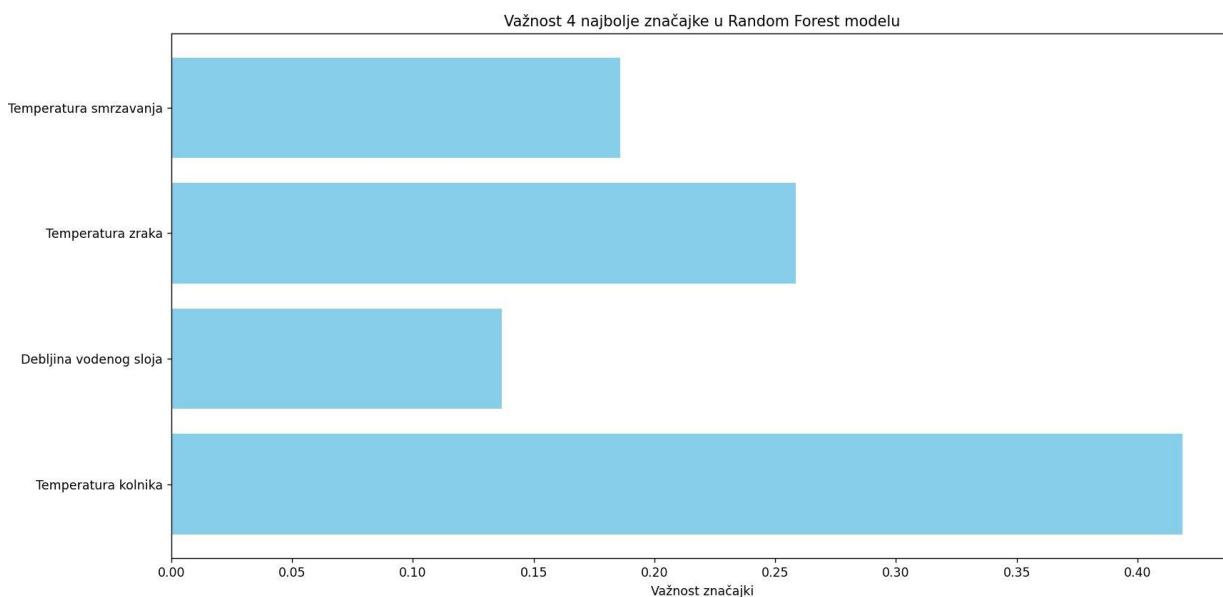


Slika 21: Vizualizacija 3 najbolje značajke u Random Forest modelu

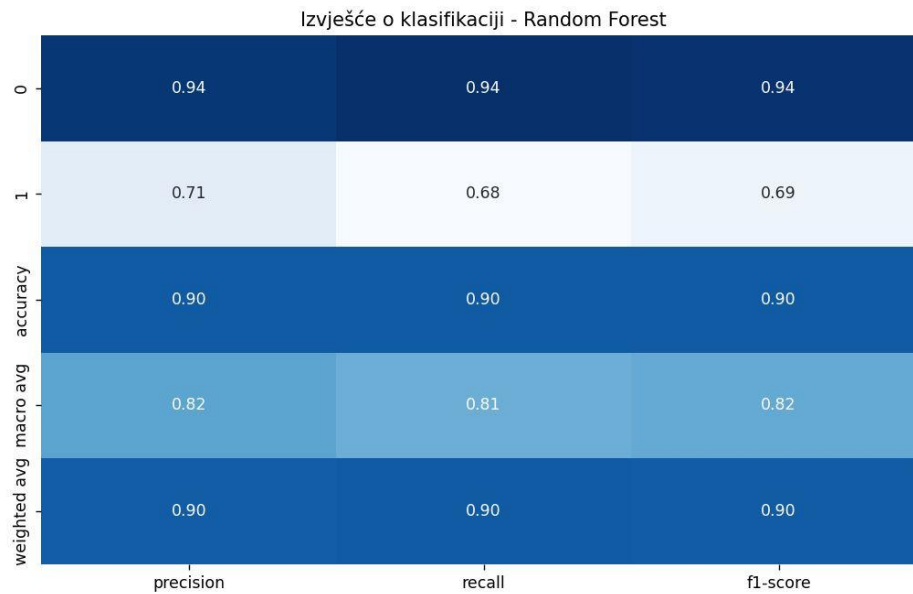
Horizontalna osa prikazuje vrijednosti važnosti značajki. Što je veća vrijednost, to je značajka važnija za predikciju ciljne varijable. Vertikalna osa predstavlja poredane značajke po važnosti, s najvažnijom na vrhu. A najvažnija značajka koja utječe na cilj jest temperatura zraka, odnosno prosječna brzina vozila po satu. Druge dvije značajke su debljina vodenog sloja i temperatura smrzavanja.

Važnost značajki u kontekstu *Random Forest* modela odnosi se na to koliko ta značajka pomaže u smanjenju pogreške prilikom donošenja odluka unutar stabala koja čine šumu. *Random Forest* koristi višestruka stable odlučivanja, i pri svakom čvoru stable procjenjuje koja značajka najbolje dijeli podatke na način da smanjuje pogrešku modela.

Nakon izvršavanja odabira tri značajke, napravljeno je isto za 4 najbolje značajke. Na slici 22, prikazana je vizualizacija atributa, a na slici 23 i izvješće o klasifikaciji. Četvrta najbolja značajka koja utječe na doprinos cilju jest temperatura kolnika.

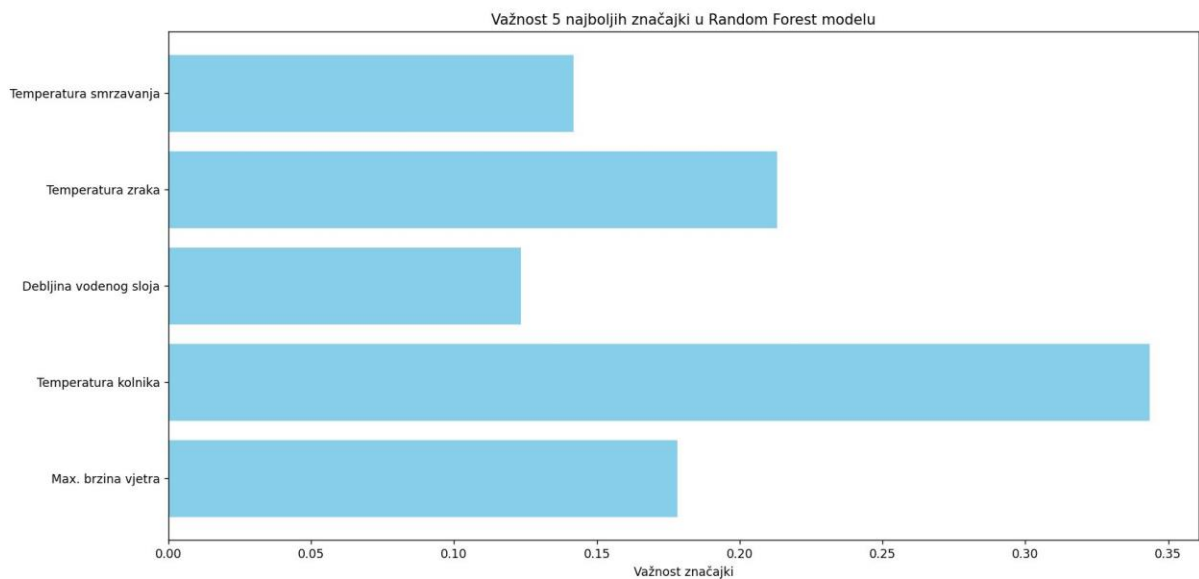


Slika 22: Vizualizacija važnosti 4 najbolje značajke

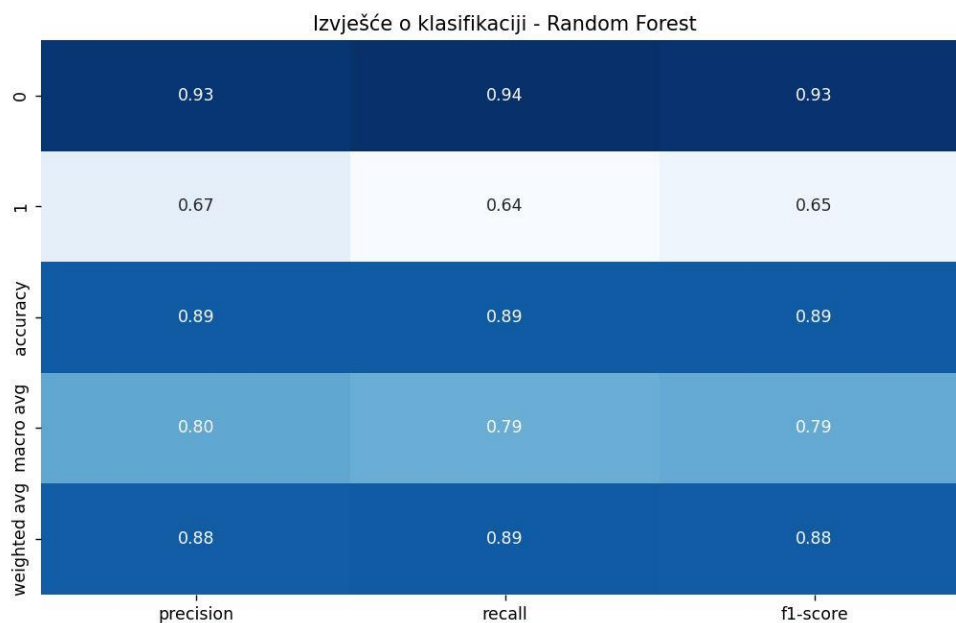


Slika 23: *Izvešće o klasifikaciji za 4 atributa*

Izvršen je još jedan slijed naredbi koji potražuje 5 najboljih značajki. Na slici 24 je napravljena vizualizacija najboljih 5 značajki koje najviše doprinose cilju, a na slici 25 jest izvješće o klasifikaciji.



Slika 24: *Vizualizacija važnosti 5 najboljih značajki*



Slika 25: Izvešće o klasifikaciji za 5 najboljih atributa

Peta značajka koja je dobivena koristeći model *Random Forest* je maksimalna brzina vjetra. Može se primijetiti da je po važnosti na drugom mjestu iza prosječne brzine vozača.

## 6. Odabir meteoroloških značajki i značajki o stanju kolnika s najvećim utjecajem na promjenu brzine na autocesti

Strojno učenje, kao ključna komponenta modernih analitičkih metoda, zahtijeva precizno upravljanje značajkama (feature) kako bi se postigli optimalni rezultati. Odabir značajki jedan je od najvažnijih koraka u izgradnji modela, jer su značajke osnova na kojoj model temelji svoje predikcije. Međutim, nije svaka značajka jednako korisna, a previše značajki može negativno utjecati na performanse modela. U tom kontekstu, tehnike odabira značajki, poput Rekurzivne eliminacije značajki (eng. *Recursive Feature Elimination*), igraju ključnu ulogu u osiguravanju da model koristi samo najrelevantnije podatke [14].

### 6.1. Rekurzivna eliminacija značajki (RFE)

Rekurzivna eliminacija značajki je popularna tehnika odabira značajki koja koristi rekurzivni pristup kako bi iterativno uklanjala najmanje važne značajke sve dok ne ostanu one najvažnije. Cilj RFE-a je smanjiti složenost modela uklanjanjem redundantnih ili irelevantnih značajki, što može rezultirati boljim performansama modela i smanjenom mogućnošću prekomjernog prilagođavanja [15].

RFE funkcionira tako što:

1. Trenira model na svim dostupnim značajkama
2. Izračunava važnost svake značajke na temelju njezina doprinosa točnosti modela. Značajke se mogu procijeniti različitim modelima, kao što su logistička regresija, linearna regresija ili Random Forest.
3. Uklanja značajku koja najmanje doprinosi modelu.
4. Ponovno trenira model na preostalim značajkama i ponavlja proces dok ne ostane željeni broj značajki.

RFE osigurava da model postepeno uklanja manje važne značajke, istovremeno čuvajući one koje najviše doprinose točnosti predikcija. Ova metoda je posebno korisna kod modela koji su osjetljivi na prevelik broj značajki, kao što su linearni modeli i logistička regresija.

## 6.2. Logistička regresija

Logistička regresija je jedan od temeljnih modela u klasifikaciji, koji se koristi za predviđanje binarnih izlaznih varijabli (npr. 0 ili 1, istina ili laž, pozitivno ili negativno). Iako nosi naziv "regresija", koristi se u svrhu klasifikacije, a ne predviđanja kontinuiranih vrijednosti. Logistička regresija temelji se na linearnim kombinacijama ulaznih značajki, ali koristi posebnu sigmoidnu funkciju koja omogućuje predikciju vjerojatnosti da će se neki događaj dogoditi (ili da će pripadati jednoj od dviju klasa) [16].

Logistička regresija ima nekoliko važnih karakteristika [16]:

- Jednostavnost i interpretabilnost - koeficijenti logističke regresije omogućuju jednostavno tumačenje – pozitivne vrijednosti koeficijenata povećavaju vjerojatnost pozitivnog ishoda, dok negativne smanjuju tu vjerojatnost.
- Primjena na klasifikaciju - Za razliku od linearne regresije koja predviđa kontinuirane vrijednosti, logistička regresija predviđa diskretne ishode, što je čini korisnom za klasifikacijske zadatke
- Fleksibilnost - Logistička regresija može se proširiti na više klasa koristeći metode poput "One-vs-Rest" ili "Softmax" klasifikacije.

U kontekstu RFE, logistička regresija često se koristi kao osnovni model za procjenu značajki. Zbog jednostavne strukture logističke regresije i njezine sposobnosti da jasno odredi koje značajke najviše doprinose klasifikaciji, RFE može precizno identificirati koje značajke treba zadržati, a koje eliminirati.

Proces funkcionira tako da RFE trenira logističku regresiju na skupu značajki i zatim koristi koeficijente koje je model izračunao kako bi rangirao značajke. Značajke s najnižim apsolutnim vrijednostima koeficijenata prve se eliminiraju jer imaju najmanji utjecaj na predikciju. Ovaj postupak se ponavlja sve dok ne ostane željeni broj značajki.

Logistička regresija, kada se koristi unutar RFE-a, omogućuje da se odaberu najvažnije značajke koje najbolje predviđaju ishod, dok se istovremeno eliminiraju one koje povećavaju složenost modela bez značajnog doprinosa [16].

### 6.3. Primjena metode RFE na odabir najboljih meteoroloških značajki i značajki o stanju na kolniku

Na već postojeći kod putem kojeg su odabrane najbolje značajke iz podatkovnog skupa, implementiran je dodatak kodu koji sadrži RFE metodu s logističkom regresijom. Na slici 26, prikazan je dio koda koji je dodatno ubačen kako bi se mogle dobiti 3 značajke koje najviše utječu na ishod. Dodane su dvije metode u sklopu *sklearn* biblioteke za svrhu primjene metode RFE s logističkom regresijom. Kod je dodan na već postojeći kod za odabir 3 najbolje značajke, u svrhu uspoređivanja između filtracijske metode te metode RFE. Isto je napravljeno i za izvršavanje metoda za odabir četiri i pet značajki.

```
# Uvoz dodatnih biblioteka za rad s RFE
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression

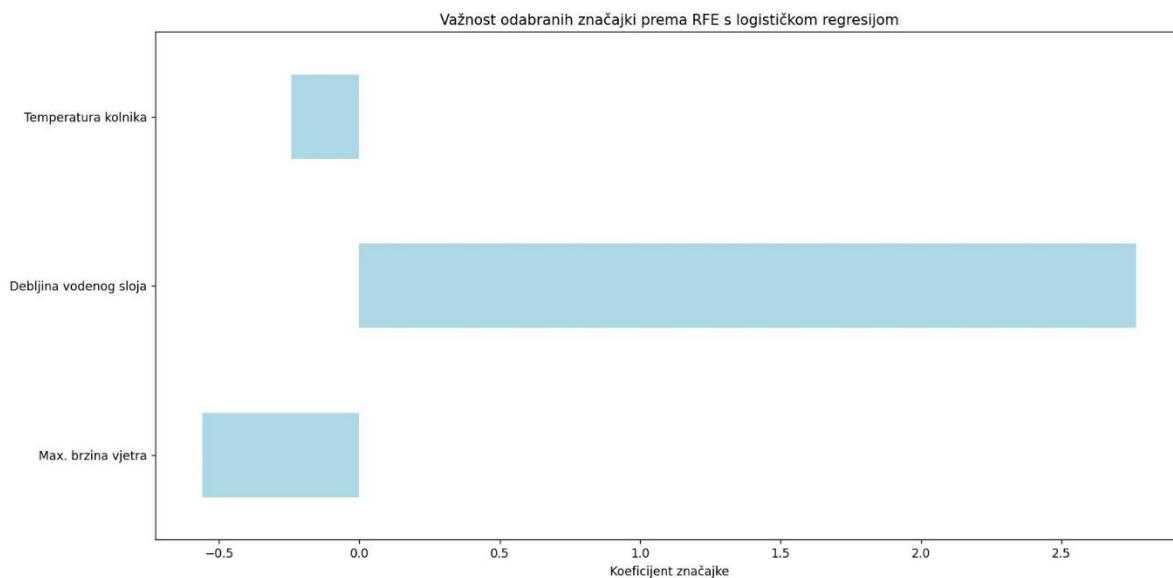
# Implementacija RFE koristeći logističku regresiju
model_log_reg = LogisticRegression()
rfe = RFE(estimator=model_log_reg, n_features_to_select=3)
rfe.fit(X_train, y_train)

# Prikaz rezultata RFE
selected_features = X.columns[rfe.support_]
print("Odabrane značajke:", selected_features)
print("Rangiranje značajki:", rfe.ranking_)

# Vizualizacija rezultata RFE (koeficijenti logističke regresije)
coefs = pd.Series(rfe.estimator_.coef_[0], index=selected_features)
plt.figure(figsize=(10, 6))
coefs.plot(kind='barh', color='lightblue')
plt.title('Važnost odabranih značajki prema RFE s logističkom regresijom')
plt.xlabel('Koeficijent značajke')
plt.show()
```

Slika 26: Dodatak kodu radi primjene RFE metode

Nakon izvršavanja koda, napravljena je vizualizacija koja prikazuje koja su to 3 najbolja atributa koja najviše utječu na ishod, koristeći metodu Rekurzivne eliminacije značajki s logističkom regresijom (Slika 27.).



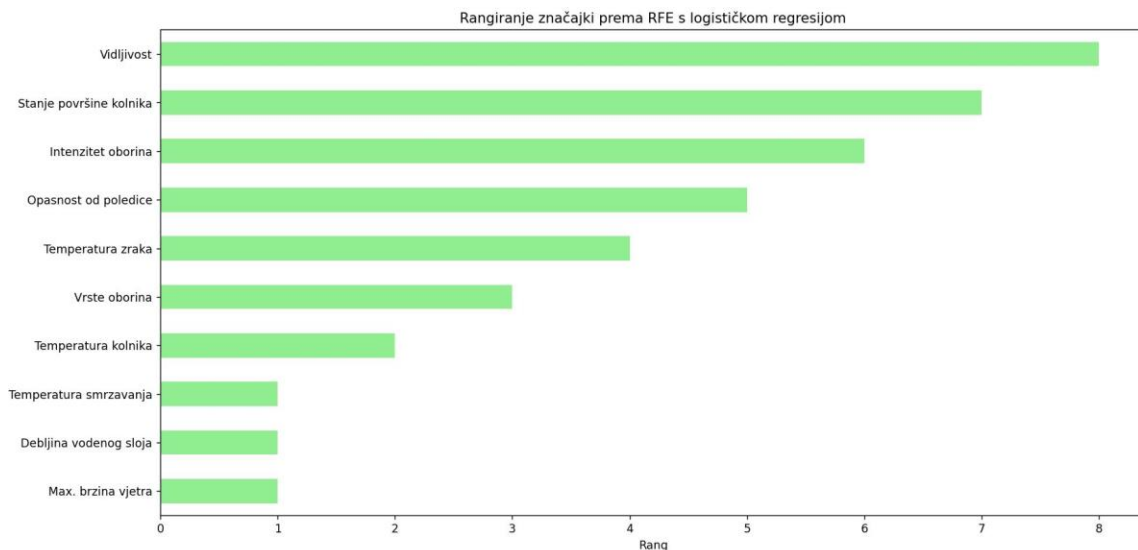
Slika 27: Vizualizacija 3 odabrane značajke po važnosti prema RFE

Nakon pokretanja koda s metodom RFE, tri najvažnija atributa su: Maksimalna brzina vjetra, debljina vodenog sloja i temperature kolnika. Njihovi koeficijenti su iskazani u bar grafu. Što je koeficijent značajke veći (pozitivan ili negativan), to značajka više utječe na ishod.

- Pozitivan koeficijent – značajka povećava vjerojatnost pozitivnog ishoda
- Negativan koeficijent – značajka smanjuje vjerojatnost pozitivnog ishoda

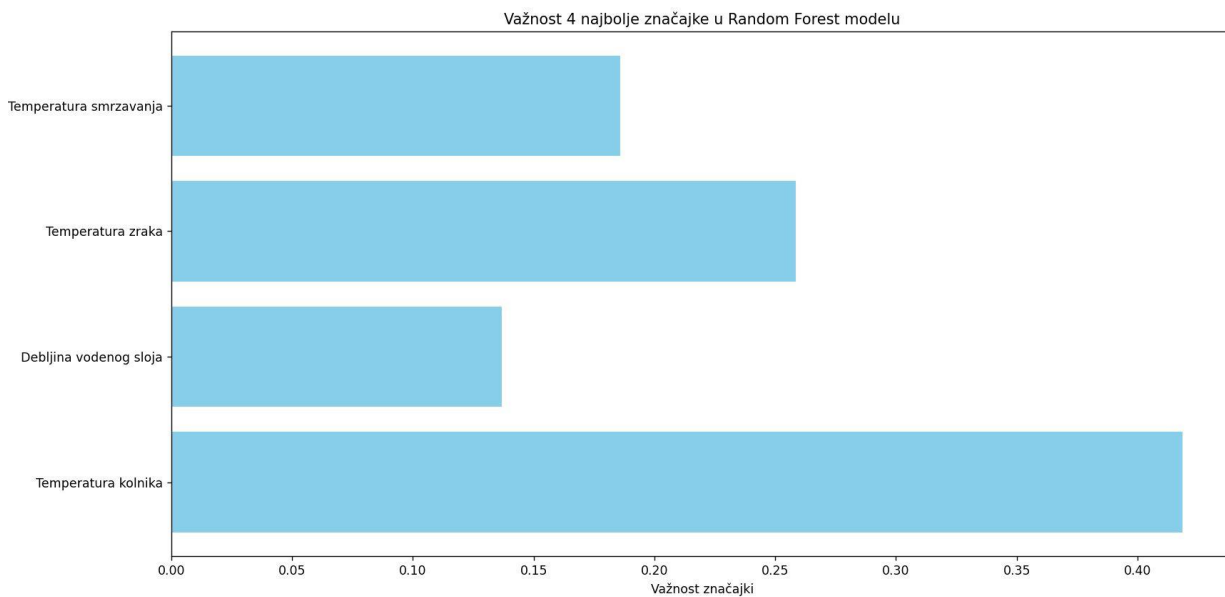
Atribut s najvećim koeficijentom jest debljina vodenog sloja, njegov koeficijent značajno povećava vjerojatnost pozitivnog ishoda. Dok, maksimalna brzina vjetra i temperature kolnika smanjuju vjerojatnost pozitivnog ishoda. Također, nakon izvršavanja koda, značajke su rangirane po važnosti (Slika 28.), što niži rang to je veća važnost.





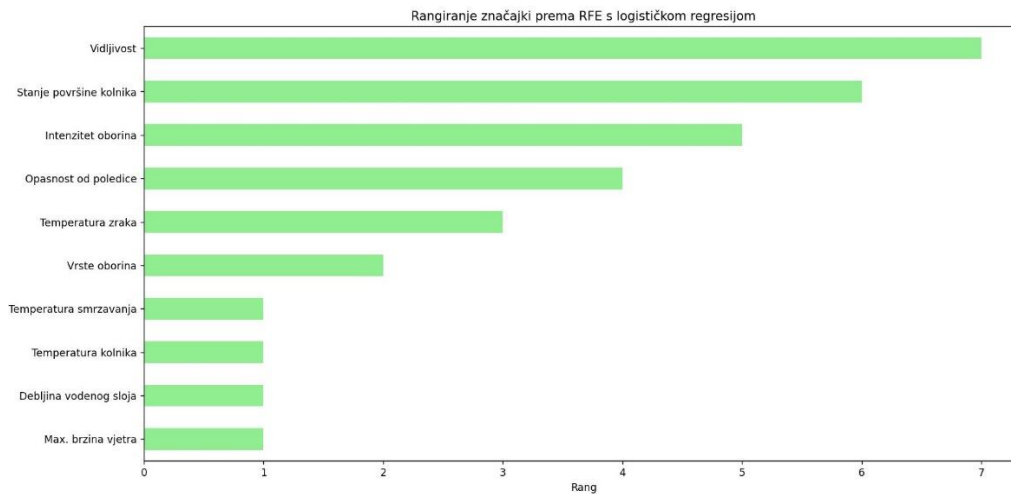
Slika 28: Rangiranje značajki nakon odabira tri značajke prema RFE metodi

Isto je napravljeno i za odabir najbolja četiri atributa koristeći RFE metodu s logističkom regresijom. Atributi su opisani na slici 29.



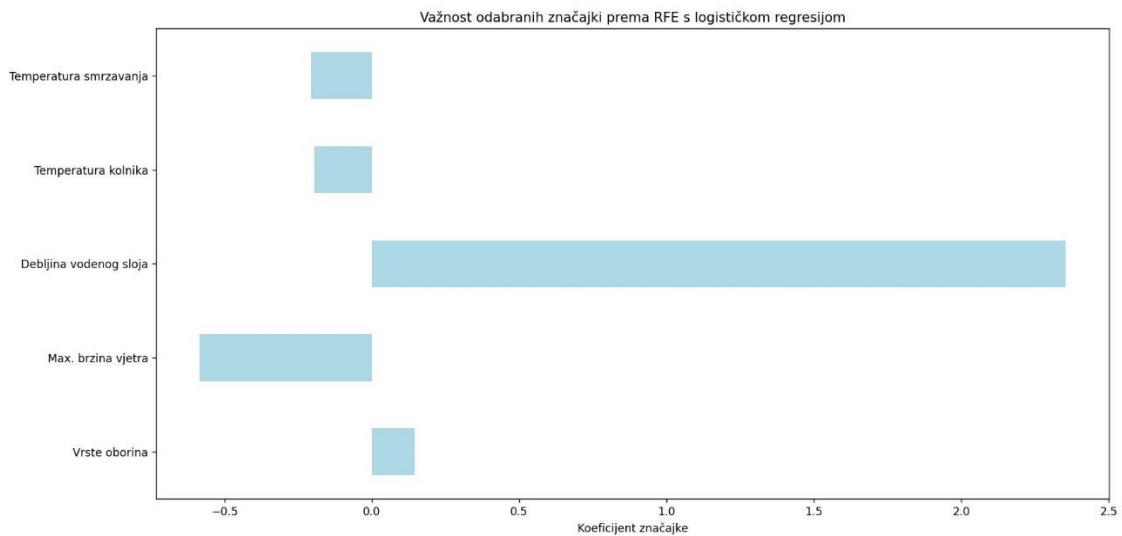
Slika 29: Vizualizacija 4 odabrane značajke prema RFE

Četvrti atribut koji je pridodan prva tri, jest “Intenzitet oborina” čiji koeficijent ne predstavlja velik značaj, unatoč tome, koeficijent je negativan i smanjuje vjerojatnost pozitivnog ishoda. Na slici 30, očituje se novo rangiranje značajki nakon izvršavanja koda za 4 atributa.

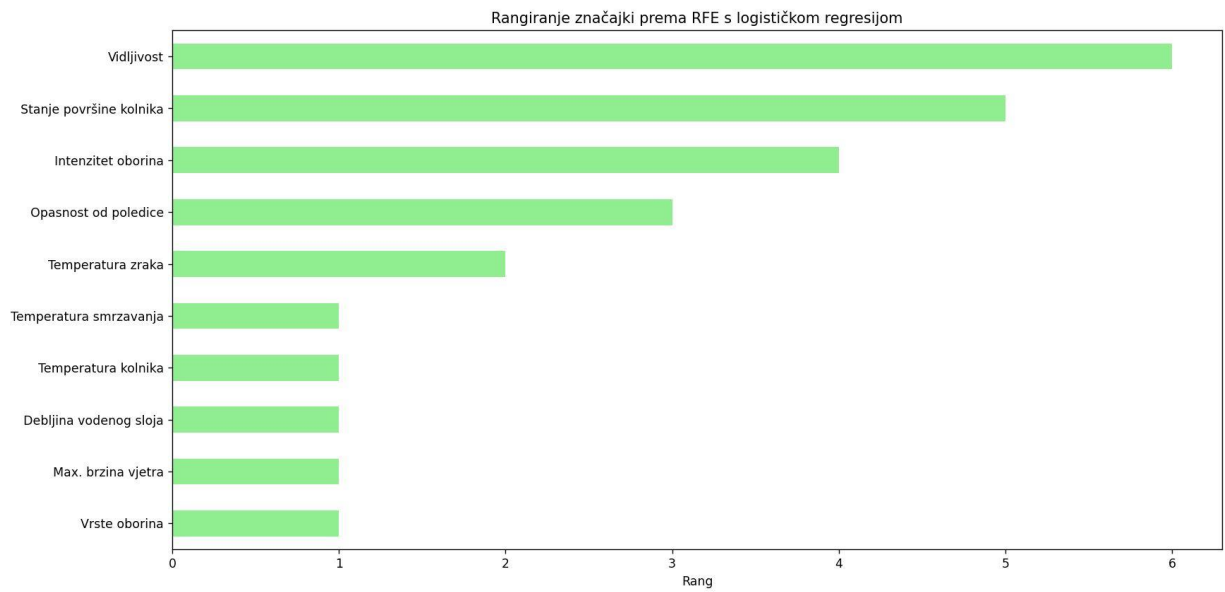


Slika 30: Rangiranje značajki nakon odabira 4 značajke prema RFE

Implementacija koda s RFE metodom izvršena je i za odabir pet značajki koje najviše utječu na promjenu brzine na autocesti, njegov rezultat vizualiziran je na slici 31. Peti atribut je “Vrste oborina”, njegov je koeficijent pozitivan, što bi značilo da pozitivno utječe na vjerojatnost ishoda. Ovo potvrđuje da bi vozači poštovali ograničenje na cesti više ukoliko su pod utjecajem neke oborine. Na slici 32, prikazano je rangiranje podataka nakon egzekucije koda koji je tražio 5 najutjecajnijih atributa.



Slika 31: Vizualizacija pet odabranih značajki prema RFE



Slika 32: Rangiranje značajki nakon odabira 5 značajki prema RFE

## 7. Analiza poštivanja ograničenja s obzirom na odabrane značajke

Kada se koristi skup podataka s velikim brojem značajki, svi podaci ne doprinose jednako modelu. Neke značajke mogu biti redundantne, a druge možda ne pružaju nikakve korisne informacije. Postoje dva glavna cilja odabira značajki:

1. Poboljšanje performansi modela – uklanjanjem nepotrebnih ili redundantnih značajki možemo poboljšati performance modela, kako u smislu točnosti, tako i u smislu vremena potrebnog za treniranje modela.
2. Smanjenje složenosti modela – manji broj značajki smanjuje složenost modela, što vodi jednostavnijoj interpretaciji i manjoj mogućnosti prekomjernog prilagođavanja

*SelectKBest* je jedna od najjednostavnijih i najčešće korištenih metoda za odabir značajki. Ova metoda koristi različite statističke testove kako bi ocijenila koliko svaka značajka doprinosi cilju modela. U ovom radu koristi se *ANOVA f-test* za evaluaciju značajki, koji mjeri koliko svaka značajka varira unutar različitih klasa ciljane varijable. Značajke s najvećim ocjenama (*score*) smatraju se najvažnijima te se odabiru kao ključne varijable za izgradnju modela. Međutim, nedostatak ove metode je taj što ne uzima u obzir interakcije između značajki. Svaka značajka se ocjenjuje samostalno, što znači da značajke koje same po sebi nisu važne, ali u kombinaciji s drugim značajkama doprinose modelu, mogu biti zanemarene.

*Recursive Feature Elimination (RFE)* je naprednija metoda odabira značajki koja koristi model strojnog učenja za iterativnu procjenu značajki. Ova metoda radi tako da prvo trenira model koristeći sve dostupne značajke. Zatim uklanja značajku koja najmanje doprinosi točnosti modela, te ponavlja postupak sve dok ne ostane željeni broj značajki. U ovom radu korištena je logistička regresija kao model za RFE.

Cilj usporedbe *SelectKBest* i *RFE* metoda u ovom radu jest razumjeti kako svaka metoda procjenjuje značajke te u kojoj mjeri doprinose točnosti modela. *SelectKBest* se oslanja na čiste statističke testove, ne uzimajući u obzir međusobne interakcije značajki, dok *RFE* koristi model za procjenu važnosti značajki unutar samog konteksta modela. Ove dvije metode mogu dati slične

rezultate, ali u nekim slučajevima će se značajke razlikovati, osobito kad su interakcije među značajkama važne.

U svrhu analize poštivanja ograničenja s obzirom na odabrane značajke putem dvije različite metode, kreirana je tablica (Tablica 3.) koja sadrži sve odabrane značajke, bilo u iteraciji s tri, četiri ili pet značajki. Kod za kreiranje tablice prikazan je na slikama 33 i 34. Tablica prikazuje odnos i ocjene tih značajki iz obje metode, uključujući:

- Ocjenu značajke (*Score*) – dobivenu putem *SelectKBest metode*
- Koeficijent značajke – dobiven iz logističke regresije korištene u *RFE metodi*

```
# Uvoz potrebnih biblioteka
import pandas as pd
from sklearn.feature_selection import SelectKBest, f_classif
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.feature_selection import RFE
import matplotlib.pyplot as plt

# Učitavanje podataka
file_path = r'C:\Users\zvoni\OneDrive\Desktop\Faks\diplomski\Master_tablica_v6.xlsx'
df = pd.read_excel(file_path)

# Izbacivanje nepotrebnih stupaca
df_cleaned = df.drop(columns=["Broj vozila <=100", "Broj vozila >100"])
df_cleaned.rename(columns={"Razlika između stupaca = outcome": "outcome"}, inplace=True)

# Definiranje značajki (X) i ciljane varijable (y)
X = df_cleaned.drop(columns=["Datum", "outcome"])
y = df_cleaned["outcome"]

# Podjela podataka na trening i testni skup
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# PRVI DIO: SelectKBest (ANOVA test)
select_kbest = SelectKBest(f_classif, k='all')
select_kbest.fit(X_train, y_train)
kbest_scores = pd.DataFrame({'Feature': X.columns, 'Score': select_kbest.scores_})
kbest_scores = kbest_scores.sort_values(by='Score', ascending=False)
```

Slika 33: Prvi dio koda za kreiranje tablice

```

# DRUGI DIO: RFE s logističkom regresijom
model_log_reg = LogisticRegression()
rfe = RFE(estimator=model_log_reg, n_features_to_select=5)
rfe.fit(X_train, y_train)
rfe_selected_features = pd.Series(rfe.support_, index=X.columns)
rfe_coefficients = pd.Series(rfe.estimator_.coef_[0], index=X.columns[rfe.support_])

# Kreiranje usporedne tablice
comparison_df = pd.DataFrame(columns=['Feature', 'Method', 'SelectKBest Score', 'RFE Coefficient'])

# Popunjavanje podataka za SelectKBest
for feature in kbest_scores['Feature'].head(5): # Prvih 5 značajki iz SelectKBest
    comparison_df = comparison_df.append({
        'Feature': feature,
        'Method': 'SelectKBest',
        'SelectKBest Score': kbest_scores[kbest_scores['Feature'] == feature]['Score'].values[0],
        'RFE Coefficient': '-'
    }, ignore_index=True)

# Popunjavanje podataka za RFE
for feature in rfe_selected_features[rfe_selected_features == True].index:
    if feature in comparison_df['Feature'].values:
        comparison_df.loc[comparison_df['Feature'] == feature, 'Method'] = 'Obje'
        comparison_df.loc[comparison_df['Feature'] == feature, 'RFE Coefficient'] = rfe_coefficients[feature]
    else:
        comparison_df = comparison_df.append({
            'Feature': feature,
            'Method': 'RFE',
            'SelectKBest Score': '-',
            'RFE Coefficient': rfe_coefficients[feature]
        }, ignore_index=True)

# Ispis tablice

```

Slika 34: Drugi dio koda za kreiranje tablice

Obje metode imaju svoje prednosti i nedostatke, a njihova primjena ovisi o specifičnom zadatku i skupu podataka. *SelectKBest* pruža jednostavnu, brzu i efikasnu metodu za odabir značajki, dok *RFE* nudi dublju analizu, uzimajući u obzir interakcije među značajkama. U konačnici, odabir metode treba ovisiti o složenosti problema i vremenu dostupnom za analizu. U ovom radu, usporedba između ove dvije metode pruža bolji uvid u karakteristike podataka te omogućuje odabir najvažnijih značajki koje najviše doprinose točnosti modela.

Tablica 3: *Usporedba provedenih metoda za odabir najznačajnijih atributa*

Značajka	Metoda	SelectKBest Rezultat	RFE Koeficijent
Temperatura smrzavanja	Obje	138.8182688	-0.207120807
Temperatura zraka	SelectKBest	138.1314781	-
Debljina vodenog sloja	Obje	137.9343084	2.354461049
Temperatura kolnika	Obje	125.4246465	-0.194922144
Max. brzina vjetra	Obje	43.91047099	-0.585477505
Vrste oborina	RFE	-	0.146360713

Na osnovu tablice iznad, moguće je usporediti kako se značajke rangiraju u svakoj metodi i koje su značajke zajedničke, a koje jedinstvene za svaku metodu. Značajke koje su odabrane u obje metode su temperatura smrzavanja, debljina vodenog sloja, temperatura kolnika te maksimalna brzina vjetra.

Temperatura smrzavanja ima ocjenu 138.81 prema *SelectKBest* metodi, što je čini relativno važnom u kontekstu statističke razlike među klasama, dok *RFE* metoda pokazuje rezultat negativnog doprinosa s koeficijentom -0.207.

Temperatura zraka sa sličnom ocjenom kao i temperatura smrzavanja (138.13), ali nije odabrana metodom *RFE*. To može sugerirati da ova značajka, iako statistički značajna, možda nije ključna za model predikcije kada se uzmu u obzir sve značajke zajedno.

Debljina vodenog sloja je odabrana u obje metode, što je čini vrlo važnom za model. *SelectKBest* joj dodjeljuje ocjenu 137.93, dok *RFE* procjenjuje da je njen doprinos pozitivan, s koeficijentom 2.57. Kombinirajući obje metode, može se zaključiti da debljina vodenog sloja ima važnu ulogu u modelu jer se statistički razlikuje među klasama, a *RFE* također prepoznaje njen značajan doprinos predikcijskoj snazi modela.

Temperatura kolnika odabrana u obje metode, iako s nešto manjom ocjenom prema *SelectKBest* metodi (125.42). *RFE* metoda dodjeljuje negativan koeficijent (-0.227), što sugerira da povećanje temperature kolnika smanjuje vjerojatnost pozitivnog ishoda. Ova značajka je važna jer oba pristupa potvrđuju njen doprinos modelu.

Vrste oborina je značajka koja nije odabrana metodom *SelectKBest*, ali *RFE* je prepoznaje kao značajnu, s pozitivnim koeficijentom od 0.19, što sugerira da ova značajka pozitivno doprinosi modelu. To može značiti da vrste oborina igraju važnu ulogu u kombinaciji s drugim značajkama.

Maksimalna brzina vjetra također je prepoznata u *SelectKBest* metodi, ali i u *RFE* gdje joj dodjeljuje negativan koeficijent od -0.59. Negativan koeficijent sugerira da, kako brzina vjetra raste, smanjuje se vjerojatnost pozitivnog ishoda. Značajka očito ima utjecaj na model, ali možda ne statistički značajan prema *ANOVA* testu korištenom u *SelectKBest*.

Razlike u rezultatima između *SelectKBest* i *RFE* metoda ističu važnost razumijevanja interakcija među značajkama. Značajke koje su same po sebi statistički značajne (prema *SelectKBest*) možda neće imati velik utjecaj na predikciju kada se uzimaju u obzir druge značajke (prema *RFE*).



## 8. Zaključak

Koristeći ogroman set prometnih i meteoroloških podataka na autocesti Zagreb-Karlovac, u ovom diplomskom radu opisan je i analiziran utjecaj tih podataka na prometnu sigurnost vozila i vozača. S obzirom na količinu podataka, nužno je bilo prilagoditi te podatke kako bi bili spremni za obradu i upotrebu.

Svaki prometni parametar morao je biti obrađen i spojen u veliku glavnu tablicu na način da zadovoljava uvjete kreiranja i izvršavanja modela strojnog učenja. Tim parametrima su se vrijednosti morale konvertirati iz jednog tipa u drugi, sumirati, praviti prosjek od njih, popunjavati prazne vrijednosti te grupirati u vremenskom periodu od jednog reprezentativnog mjeseca po svakom satu svakoga dana.

Nakon početne pripreme, svi parametri su prošli kroz određenu vizualizaciju kako bi se potvrdila vjerodostojnost podataka prije nego što bi se radila bilo kakva statistička analiza ili implementacija modela tj. algoritma strojnog učenja. Izvršavanjem različitih metoda obrade podataka, mogli su se izvući zaključci o utjecaju svakog parametra na konačni ishod tog podatkovnog skupa. Konačni ishod odnosi se na prometnu sigurnost, na njega parametri djeluju negativno ili pozitivno, čime izravno utječu na prosječnu brzinu vozila na autocesti.

Daljnjim analiziranjem podataka, stvara se slika da postoji mnogo različitih metoda koje mogu dati različite rezultate, a koje se mogu mjeriti po sličnostima i podudaranjima između tih rezultata. Svaka sličnost može ukazivati na točnost podataka čime se ulijeva povjerenje da se mogu početi tražiti rješenja za negativne utjecaje na sigurnost u prometu ili podržati i proširiti pozitivne utjecaje na dijelove prometa gdje je potrebno.

## Literatura

- [1] Couto A., Ferreira S., Iglesias I., Lobo A. *Urban Road Crashes and Weather Conditions: Untangling the Effects*, Sustainability; 2019. Preuzeto s: <https://www.mdpi.com/2071-1050/11/11/3176> [Pristupljeno: 15. rujna 2024.]
- [2] Meena G., Sharma D., Mahrishi M. *3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, 2020. Preuzeto s: <https://ksra.eu/wp-content/uploads/2020/07/10.1109@ICETCE48199.2020.9091758.pdf> [Pristupljeno: 15. rujna 2024.]
- [3] Numpy the fundamental package for scientific computing with python. <https://numpy.org/>. [Pristupljeno: 15. rujna 2024.]
- [4] Pandas data analysis and manipulation tool. <https://pandas.pydata.org/>. [Pristupio: 15. rujna 2024.]
- [5] Sklearn machine learning in python. <https://scikit-learn.org/stable/>. [Pristupio: 15. rujna 2024.]
- [6] Matplotlib: Visualization with Python. <https://matplotlib.org/>. [Pristupljeno: 15. rujna 2024.]
- [7] Seaborn: statistical data visualization. <https://seaborn.pydata.org/> [Pristupljeno 15. Rujna 2024.]
- [8] Brijs S., Brijs T., Brijs T., Hermans E. *The Impact of Weather Conditions on Road Safety Investigated on an Hourly Basis*, TRB 85th Annual Meeting Compendium; 2007. [Pristupljeno: 12. rujna 2024.]
- [9] What is Correlation Analysis? A Definition and Explanation. <https://blog.flexmr.net/correlation-analysis-definition-exploration>. [Pristupljeno: 13. rujna 2024.]
- [10] Machine Learning Models. <https://www.javatpoint.com/>. [Pristupio:10. rujna 2024.]
- [11] Vaiti T. *Predikcija emisija štetnih plinova vozila korištenjem algoritama strojnog učenja s podacima prikupljenih OBD2 uređajem*. ,Diplomski rad, Sveučilište u Zagrebu, Fakultet prometnih znanosti, Zagreb, 2022. Dostupno na: <https://repositorij.fpz.unizg.hr/islandora/object/fpz%3A2877/datastream/PDF/view>

- [12] Goodfellow I., Bengio Y., Courville A. *Deep learning*, The MIT Press; 2016. [Pristupljeno: 16. rujna 2024.]
- [13] Guyon I., Elisseeff A., *An Introduction to Variable and Feature Selection*, Journal of Machine Learning Research; 2003. Preuzeto s: <https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>. [Pristupljeno: 08. rujna 2024.].
- [14] Python Feature Selection Tutorial: A Beginner's Guide. <https://www.datacamp.com/tutorial/feature-selection-python>. [Pristupljeno: 09. rujna 2024.]
- [15] Guyon I., Weston J., Barnhill S., Vapnik V. *Gene Selection for Cancer Classification using Support Vector Machines*, Springer Link; 2002. Preuzeto s: <https://link.springer.com/article/10.1023/A:1012487302797>. [Pristupljeno: 11. rujna 2024.].
- [16] Hosmer D.W., Lemeshow S., Sturdivant R. *Applied Logistic Regression*, John Wiley & Sons, Inc.; 2013. Preuzeto s: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387> [Pristupljeno: 12. rujna 2024.]

## Popis slika

Slika 1: Prikaz čvora Donja Zdenčina .....	3
Slika 2: Prikaz snimanja kamere na promatranom čvoru .....	4
Slika 3: <i>Ukupan broj vozila po mjesecima za 2018. godinu (smjer Karlovac)</i> .....	7
Slika 4: Prosječna brzina po mjesecima za 2018. godinu (smjer Karlovac).....	8
Slika 5: Usporedba prosječnih brzina po prometnim tracima.....	9
Slika 6: Usporedba prosječnog broja vozila po ograničenju .....	10
Slika 7: Prosječna razlika po satu za ožujak 2018. ....	10
Slika 8: Distribucija vrijednosti 0 i 1 po satima.....	11
Slika 9: Grupiranje podataka o temperaturi zraka .....	13
Slika 10: Predobrada tekstualnih podataka .....	14
Slika 11: Promjene temperature zraka i kolnika kroz vrijeme.....	15
Slika 12: Distribucija brzine vjetra .....	16
Slika 13: Promjena vidljivosti kroz mjesec .....	17
Slika 14: Korelacijska matrica svih atributa .....	18
Slika 15: Kategorije učenja u strojnim sustavima.....	19
Slika 16: Metoda filtriranja.....	21
Slika 17: Omotna metoda.....	21
Slika 18: Prvi dio koda za odabir najbolje 3 značajke .....	22
Slika 19: Drugi dio koda za odabir 3 najbolje značajke .....	23
Slika 20: Izvješće o klasifikaciji za 3 najbolje značajke.....	24
Slika 21: Vizualizacija 3 najbolje značajke u Random Forest modelu .....	24
Slika 22: Vizualizacija važnosti 4 najbolje značajke .....	25
Slika 23: Izvješće o klasifikaciji za 4 atributa .....	26
Slika 24: Vizualizacija važnosti 5 najboljih značajki.....	26
Slika 25: Izvješće o klasifikaciji za 5 najboljih atributa .....	27
Slika 26: Dodatak kodu radi primjene RFE metode .....	30
Slika 27: Vizualizacija 3 odabrane značajke po važnosti prema RFE .....	31
Slika 28: Rangiranje značajki nakon odabira tri značajke prema RFE metodi.....	32

Slika 29: Vizualizacija 4 odabrane značajke prema RFE .....	32
Slika 30: Rangiranje značajki nakon odabira 4 značajke prema RFE .....	33
Slika 31: Vizualizacija pet odabranih značajki prema RFE .....	33
Slika 32: Rangiranje značajki nakon odabira 5 značajki prema RFE .....	34
Slika 33: Prvi dio koda za kreiranje tablice .....	36
Slika 34: Drugi dio koda za kreiranje tablice.....	37

## **Popis tablica**

Tablica 1: primjer jednog seta podataka (Intenzitet oborina) .....	5
Tablica 2: Master tablica sa svim grupiranim atributima.....	12
Tablica 3: Usporedba provedenih metoda za odabir najznačajnijih atributa .....	38

Sveučilište u Zagrebu  
Fakultet prometnih znanosti  
Vukelićeva 4, 10000 Zagreb

## IZJAVA O AKADEMSKOJ ČESTITOSTI I SUGLASNOSTI

Izjavljujem i svojim potpisom potvrđujem da je \_\_\_\_\_ diplomski rad \_\_\_\_\_  
(vrsta rada)

isključivo rezultat mojega vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju upotrijebljene bilješke i bibliografija. Izjavljujem da nijedan dio rada nije napisan na nedopušten način, odnosno da je prepisan iz necitiranog rada te da nijedan dio rada ne krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Svojim potpisom potvrđujem i dajem suglasnost za javnu objavu diplomskog rada pod naslovom **ANALIZA UTJECAJA METEOROLOŠKIH PRILIKA I STANJA KOLNIKA NA POŠTIVANJE OGRANIČENJA BRZINE**, u Nacionalni repozitorij završnih i diplomskih radova ZIR.

Student/ica:

U Zagrebu, 20.09.2022.

*Zvonimir Batarilo*  
(ime i prezime, potpis)