

Određivanje moda prijevoza analizom spajanja korisnika na pojedine bazne stanice mobilne mreže

Rožman, Nikola

Undergraduate thesis / Završni rad

2021

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:119:248537>

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

Download date / Datum preuzimanja: **2024-09-22**



Repository / Repozitorij:

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



Sveučilište u Zagrebu
Fakultet prometnih znanosti

ZAVRŠNI RAD

**ODREĐIVANJE MODA PRIJEVOZA ANALIZOM SPAJANJA
KORISNIKA NA POJEDINE BAZNE STANICU MOBILNE MREŽE**

**DETERMINING MODE OF TRANSPORTATION BY ANALYSIS
OF USERS' CONNECTION TO SPECIFIC BASE STATIONS OF A
MOBILE NETWORK**

Mentor: izv. prof. dr. sc. Mario Muštra

Student: Nikola Rožman

JMBAG: 0067526941

Zagreb, rujan 2021.

SVEUČILIŠTE U ZAGREBU
FAKULTET PROMETNIH ZNANOSTI
ODBOR ZA ZAVRŠNI RAD

Zagreb, 5. rujna 2021.

Zavod: **Zavod za inteligentne transportne sustave**
Predmet: **Lokacijski i navigacijski sustavi**

ZAVRŠNI ZADATAK br. 6363

Pristupnik: **Nikola Rožman (0067526941)**
Studij: **Inteligentni transportni sustavi i logistika**
Smjer: **Inteligentni transportni sustavi**

Zadatak: **Određivanje moda prijevoza analizom spajanja korisnika na pojedine bazne stanice mobilne mreže**

Opis zadatka:

Opisati značajke čelijskih sustava koji su građevna osnova mobilnih mreža, uz poseban osvrt na generacije mobilnih mreža koje su trenutno u primjeni. Analizirati mogućnosti prikupljanja podataka o spojenosti korisnika na pojedinu baznu stanicu te značajke na temelju kojih se može odrediti kretanje korisnika kroz prometnu mrežu. Navesti i opisati postupke klasifikacije i strojnog učenja na podacima koji su različito strukturirani. Opisati podatke koji se koriste pri klasifikaciji u cilju određivanja moda prijevoza. Evaluirati klasifikacijske algoritme s obzirom na uspješnost klasifikacije prema dostupnim podacima o spajanju na bazne stanice.

Mentor:



izv. prof. dr. sc. Mario Muštra

Predsjednik povjerenstva za
završni ispit:

SAŽETAK

U ovome završnom radu opisano je strojno učenje, klasifikacija podataka kao i evaluacija klasifikatora za određivanje moda prijevoza. Predstavljena je osnovna definicija ćelija mobilnih mreža te pregled pet generacija mobilnih tehnologija. Također su opisani procesi ponovnog iskorištavanja frekvencija i preuzimanja poziva. Opisan je proces lociranja mobilnih uređaja koristeći tehnologije mobilne mreže. Uz to opisan je i proces izdvajanja značajki iz podataka. Navedene su podjele strojnog učenja te njihove značajke. Na kraju su opisane metode klasifikacije koje se promatraju u radu te njihova evaluacija.

Ključne riječi: bazna stanica, mobilna mreža, strojno učenje, klasifikator

SUMMARY

This paper describes machine learning, classification of data as well as evaluation of classifiers for determining mode of transportation. Basic definition of mobile cell networks is presented as well as overview of five generations of mobile technologies. Processes of frequency reuse and handoff are also mentioned. Process of locating mobile devices using mobile network technologies is described. Besides that, process of feature extraction is defined as well. Division of machine learning is listed as well as the features. Finally, methods of classifications being observed is presented and their evaluation.

Key words: base station, mobile network, machine learning, classifier

Sadržaj

| | |
|--|----|
| 1. UVOD | 1 |
| 2. Koncept ćelijskih sustava i mobilnih mreža | 1 |
| 2.1 Ćelijski sustav | 2 |
| 2.1.1 Ponovno iskorištavanje frekvencija | 3 |
| 2.1.2 Preuzimanje poziva | 4 |
| 2.2 Mobilna mreža | 4 |
| 2.2.1 Prva generacija mobilnih mreža – 1G | 5 |
| 2.2.2 Druga generacija mobilnih mreža – 2G | 5 |
| 2.2.3 Treća generacija mobilnih mreža – 3G | 5 |
| 2.2.4 Četvrta generacija mobilnih mreža – 4G | 6 |
| 2.2.5 Peta generacija mobilnih mreža – 5G | 6 |
| 3. Izdvajanje značajki iz podataka o spojenosti korisnika na baznu stanicu | 7 |
| 3.1 Lociranje mobilnih uređaja koristeći mrežne tehnologije | 7 |
| 3.2 Izdvajanje značajki | 8 |
| 4. Metode klasifikacije i strojno učenje | 9 |
| 4.1 Strojno učenje | 9 |
| 4.2 Metode klasifikacije | 10 |
| 4.2.1 Metoda k-najbližih susjeda | 10 |
| 4.2.2 Metoda slučajnih šuma | 11 |
| 4.2.3 ExtraTrees algoritam | 12 |
| 5. Opis podataka koji su korišteni u klasifikaciji | 13 |
| 6. Evaluacija klasifikacijskih algoritama na temelju dostupnih podataka | 15 |
| 6.1 Matrica zabune | 15 |
| 6.2 Točnost | 17 |
| 6.3 Preciznost | 18 |
| 6.4 Odziv | 19 |
| 7. Zaključak | 20 |
| LITERATURA | 21 |
| POPIS SLIKA | 22 |
| POPIS TABLICA | 23 |

1. UVOD

Mobilne mreže su jedan od najvažnijih faktora mobilnosti u svijetu. Razvojem mobilnih tehnologija povećavala se i potražnja za mobilnim uslugama. Rastom broja baznih stanica koje opskrbljuju mobilne uređaje sa signalom mobilnih mreža se povećala i kvaliteta, odnosno brzina prijenosa podataka i kvaliteta poziva i veze. Čelijski sustav je osmišljen na način da pokrije veliku zemljinu površinu sa signalima mobilnih tehnologija pomoću odašiljača male snage. Čelijski sustavi su raspoređeni u ćelije u kojima su smještene bazne stanice. Bazne stanice primaju i šalju podatke sa mobilnim uređajima. Prijelaz između baznih stanica je omogućen pomoću preuzimanja poziva što omogućava mobilnost mobilnim korisnicima. Lociranje mobilnih uređaja je proces identifikacije lokacije mobilnog uređaja. Lociranje mobilnih uređaja se obavlja svakodnevno uz pomoć GPS-a, ali uređaj se može locirati i pomoću infrastrukture mobilnih usluga. Podatci koji se prikupе pomoću baznih stanica se mogu svesti na skupine podataka koje su lakše za usporedbu. Izdvajanje značajki (engl. *feature extraction*) je naziv za metodu koja odabire varijable u značajke, smanjujući broj podataka za obradu, a u isto vrijeme ne umanjujući vjernost prikazivanja skupa podataka. Strojno učenje je programiranje računala da sama pronalaze algoritme za rješavanje problema. Klasifikacija je proces prepoznavanja i grupiranja podataka u kategorije koje su predodređene.

Ovaj rad je podijeljen u 7 poglavlja:

1. Uvod
2. Koncept ćelijskih sustava i mobilnih mreža
3. Izdvajanje značajki iz podataka o spojenosti korisnika na baznu stanicu
4. Metode klasifikacije i strojno učenje
5. Opis podataka koji su korišteni u klasifikaciji
6. Evaluacija klasifikacijskih algoritama na temelju dostupnih podataka
7. Zaključak.

Poslije uvodnog poglavlja, drugo opisuje koncept ćelijski sustava zajedno sa mobilnom mrežom. Navedeno je od čega se sastoji ćelijski sustav, procesi mobilne mreže te pregled generacija mobilnih tehnologija.

Treće poglavlje opisuje načine lociranja mobilnih uređaja. Također opisuje i izdvajanje značajki iz podataka koji se dobiju pomoću baznih stanica na koju su korisnici spojeni.

Četvrto poglavlje prikazuje i opisuje metode klasifikacije koje se evaluiraju u ovom radu. Opisano je strojno učenje i podjela.

U petom poglavlju predstavljene su podatci koji se obrađuju te značajke navedenih podataka.

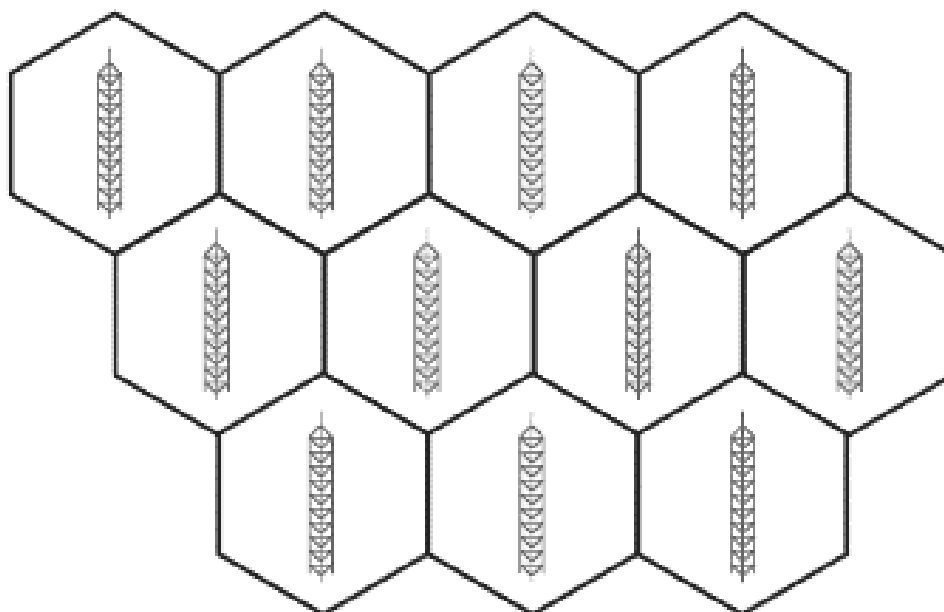
Šesto poglavlje prikazuje evaluaciju klasifikacijskih algoritama.

Sedmo poglavlje donosi zaključak koji se odnosi na cjelokupan rad, kao i konačno mišljenje autora o temi završnog rada.

2. Koncept ćelijskih sustava i mobilnih mreža

Mobilna mreža je komunikacijska mreža u kojoj se prijenos podataka između čvorova odvija bežično. Mobilna mreža je podijeljena na ćelije. Ćelije su područja koja unutar svoje površine imaju najmanje jedan primopredajnik koji je postavljen na trajnoj lokaciji. Slika 1, [2], shematski prikazuje ćelije. Bazne stanice pružaju ćeliji dovoljnu mrežnu pokrivenost za prijenos podataka koji može biti glasovni, podatkovni ili drugi.

Bazne stanice imaju drugačiji skup valnih duljina od susjednih ćelija kako bi se smanjila mogućnost šuma i interferencija u prijenosu podataka. Skupa spojene ćelije pokrivaju veliku površinu što omogućuje međusobnu komunikaciju između velikog broja prijenosnih terminalnih uređaja, kao što su mobilni uređaji, tableti, prijenosna računala, kao i komunikaciju sa fiksnim primopredajnicima i telefonskim uređajima. Spojene ćelije također omogućavaju komunikaciju čak i kada terminalni uređaj prelazi iz jedne ćelije u drugu tijekom prijenosa podataka [1].



Slika 1 Shematski prikaz ćelija

Prijenosni medij kojim se razmjenjuju podatci bežično je otvoren prostor, a podatci se prenose putem elektromagnetskih valova. Elektromagnetsko zračenje je medij za svakodnevnu interakciju sa okolinom: svjetlost koje nam omogućava da vidimo, rendgenske zrake u medicinske svrhe, radio valovi koji omogućuju slušanje radijskih programa i gledanje televizijskog sadržaja i slično.

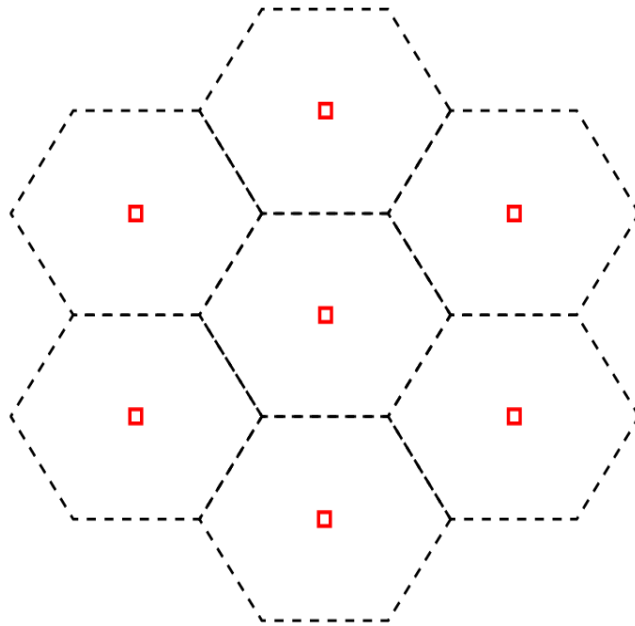
Tablica 1 Elektromagnetski spektar

| Elektromagnetski spektar | | |
|--------------------------|-----------------|---|
| Naziv | Valna duljina | Frekvencija (Hz) |
| Gama-čestica | < 0,02 nm | > $15 \cdot 10^{18}$ Hz |
| Rendgenske zrake | 0,01 nm – 10 nm | $30 \cdot 10^{18}$ Hz – $30 \cdot 10^{15}$ Hz |
| Ultraljubičasto zračenje | 10 nm – 400 nm | 30 PHz – 750 THz |
| Vidljiva svjetlost | 390 nm – 750 nm | 770 THz – 400 THz |
| Infracrveno zračenje | 750 nm – 1 mm | 400 THz – 300 GHz |
| Mikrovalovi | 1 mm – 1 m | 300 GHz – 300 MHz |
| Radio valovi | 1 mm – 100 km | 300 GHz – 3 kHz |

U tablici 1, [3], prikazan je elektromagnetski spektar. Elektromagnetski spektar je raspon frekvencija elektromagnetskog zračenja prema valnim duljinama (frekvencijama). Radio spektar je podijeljen u 8 frekvencijskih područja (pojasa), rangirani od vrlo niskih frekvencija (*very low frequency* – VLF) koje započinju na 3 kHz i proteže se do ekstremno visokih frekvencija (*extremely high frequency* – EHF) koje završavaju na 300 GHz [3].

2.1 Čelijski sustav

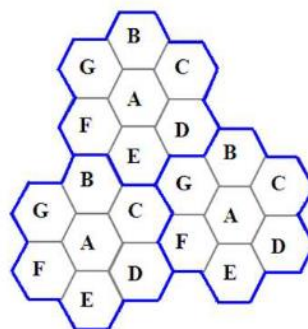
U mobilnim mrežama čelijski sustavi su organizirani na način da se koristi više odašiljača male snage, najčešće 100W i manje, koji su raspodijeljeni na određenom području. Područje koje pokriva snop primopredajnika je raspoređen u ćelije. Bazna stanica sastoji se od odašiljača, prijamnika i kontrolne jedinice. Model čelijskog sustava je osmišljen tako da su odašiljači susjednih ćelija međusobno jednako udaljeni pa je oblik samih ćelija šesterokut [4]. Primopredajnici baznih stanica su smješteni u sami centar ćelija što je prikazano slikom 2 [5].



Slika 2 Pozicija položaja baznih stanica unutar ćelija

2.1.1 Ponovno iskorištavanje frekvencija

Svakoj ćeliji je dodijeljena skupina frekvencija. Susjednim ćelijama je dodijeljena različita skupina frekvencija kako bi se smanjila interferencija i šum u izmjeni podataka. Antene baznih stanica su osmišljene na način da ostvare željenu pokrivenost unutar granica ćelije. Ograničavajući područje pokrivanja signalom unutar granica ćelije, ista skupina frekvencija može se koristiti za pokrivanje geografskog područja druge ćelije, ali samo ukoliko su te dvije ćelije dovoljno geografski udaljene jedna od druge kako bi se smanjio stupanj interferencije unutar prihvatljivih okvira [6].



Slika 3 Koncept ponovnog iskorištavanja frekvencija

Slika 3, [6], prikazuje ponovno iskorištavanje frekvencija (engl. *frequency reuse*) gdje je svakoj ćeliji dodijeljeno slovo. Svako slovo sadrži jedinstveni skup frekvencija u određenoj skupini ćelija. Skupina ćelija je ocrtana plavom bojom.

Elementi koji određuju ponovno iskorištavanje frekvencija su udaljenost (engl. *reuse distance*) i faktor ponovnog iskorištanja (engl. *reuse factor*). Udaljenost, D , se izračunava po formuli:

$$D = R * \sqrt{3N}$$

gdje je R radijus ćelije, a N je broj ćelija u skupini ćelija. Radijus ćelija varira od jednog kilometra do trideset kilometara. Faktor ponovnog iskorištavanja je stopa po kojoj se frekvencije mogu iskoristiti u mreži [6].

2.1.2 Preuzimanje poziva

Jedan od najbitnijih zahtjeva mobilne mreže je omogućavanje neprekinute i nesmetane veze između dva ili više korisnika. Preuzimanje poziva (engl. *handover*) je proces ostvarivanja neprekidne usluge prijenosom poziva ili veze u tijeku sa jedne frekvencije na drugu frekvenciju koje mobilna mreža koristi [4].

Unutar ćelijskih sustava postoje dvije vrste preuzimanja poziva, a to su horizontalno preuzimanje poziva (engl. *intra-system handover*) i vertikalno preuzimanje poziva (engl. *inter-system handover*). Horizontalno preuzimanje poziva odvija se između različitih ćelija unutar iste mreže. U takvoj mreži preuzimanje poziva je potrebno kada jedna bazna stanica više nije u mogućnosti posluživati mobilnu mrežu [7].

Vertikalno preuzimanje poziva se odvija u situacijama kada:

- korisnik prelazi između mreža različitih pristupnih tehnologija
- kada je potrebno ukupno opterećenje mreže raspodijeliti između više različitih sustava [8].

2.2 Mobilna mreža

Mobilna mreža se može podijeliti u tri glavna dijela: mobilne stanice, bazne stanice i mrežnog podsistema. Mobilna stanica se sastoji od mobilnog terminalnog uređaja i identifikacijske SIM (*Subscriber Identity Module*) kartice. SIM kartica omogućuje mobilnost pa korisnik nije vezan isključivo za jedan, određeni mobilni uređaj već SIM karticu može ubaciti u bilo koji uređaj. Svaka SIM kartica sadrži tajni kod koji služi za njegovu identifikaciju. Bazna stanica se sastoji od komunikacijskog i kontrolnog dijela. Komunikacijski dio definira ćelije i rukuje protokolima za komunikaciju s mobilnim stanicama. Kontrolni dio podešava radio-kanale, izmjenjuje frekvencije i obavlja preuzimanje poziva. Kontrolni dio predstavlja vezu između mobilne stanice i mrežnog podsistema koji sadrži mobilnu stanicu MSC (*Mobile Switching Center*) koji uspostavlja vezu između korisnika [4].

2.2.1 Prva generacija mobilnih mreža – 1G

Prva generacija mobilnih mreža koristila je analogni signal za prijenos govora. Tehnologija prve generacije mobilnih mreža koristila je metodu multipleksiranja s podjelom frekvencije FDMA (*Frequency Division Multiple Access*). FDMA je metoda pomoću koje se prenose podatci kod koje se frekvencije iz šireg pojasa frekvencija dijele na više užih pojaseva frekvencija koji se zovu kanali. Tehnologija je postala komercijalno dostupna 1984. godine. Tehnologija prve generacije imala je brojne nedostatke. Pokrivenost je bila loša, kao i kvaliteta zvuka. Nije bilo podrške između operatera različitih država, kao niti kompatibilnosti između sustava s obzirom da su različiti sustavi radili na različitim frekvencijama [4].

2.2.2 Druga generacija mobilnih mreža – 2G

Druga generacija mobilnih mreža pokrenuta je u Finskoj 1991. pod standardom GSM (engl. *Global System for Mobile Communications*). Druga generacija prenosila je govor digitalnim signalom. 2G je također uvela usluge kao što su SMS poruke, roaming te mogućnost korištenja multimedijских sadržaja preko mobilnih uređaja. Za razliku od poziva prve generacije, pozivi obavljani drugom generacijom mobilnih mreža su bili enkriptirani. 2G je također omogućavala veći broj korisnika po frekvencijskom spektru s obzirom na veću efikasnost korištenja radijskog frekvencijskog spektra. Većina sustava druge generacije mobilnih mreža koristi metodu multipleksiranja s podjelom vremena TDMA (*Time Division Multiple Access*). TDMA je metoda pomoću koje se prenose podatci kod koje se frekvencijski pojas dijeli na vremenske intervale tijekom kojih se prenose podatci. Konstantnim tehnološkim napretkom 2G je zamijenjen sa 2.5G, 2.75G, 3G, 4G i naposljetku sa 5G. Teoretski maksimalna brzina prijenosa podataka 2G mreže je bila 14,4 kbit/s [4].

2.2.3 Treća generacija mobilnih mreža – 3G

3G je nadogradnja na 2G kako bi se standardizirali protokoli prijenosa podataka. Za razliku od prošlih generacija mobilnih mreža koje su koristile vremenske i frekvencijske metode multipleksiranja, treća generacija mobilnih mreža koristi WCDMA (*Wideband Code Division Multiple Access*) metodu. Koristeći tehnologije treće generacije korisnici nisu odvojeni terminima ili frekvencijama, već im je dodijeljen jedinstvena šifra. Propusnost je

također povećana što je u početku omogućavalo pojedinom korisniku teoretski maksimalnu brzinu prijenosa podataka od 384 kbit/s, a kasnije se povećalo do 2 Mbit/s [4].

2.2.4 Četvrta generacija mobilnih mreža – 4G

Četvrta generacija mobilnih mreža je prvo razvijena u Stockholmu i Oslu 2009. godine kao LTE (engl. *Long Term Evolution*) standard. Cilj četvrte generacije je povećanje kapaciteta i brzine bežičnih mobilnih mreža kao i smanjenje latencije tijekom prijenosa podataka. Maksimalna brzina prijenosa podataka je 300 Mbit/s [4].

2.2.5 Peta generacija mobilnih mreža – 5G

Peta generacija mobilnih mreža je predstavljena 2016. godine i trenutno nije još rasprostranjena kao 4G. Tri glavna kriterija za razvoj i komercijalnu upotrebu pete generacije mobilnih mreža su:

- brzine prijenosa podataka 1 Gbit/s i više
- latencija ispod 1 milisekunde
- energetski učinkovitija tehnologija od prethodnika [9]
- do 100 puta veći kapacitet prijenosa podataka [10].

Za korištenje pete generacije mobilnih mreža potreban je uređaj dizajniran za korištenje 5G mreže.

3. Izdvajanje značajki iz podataka o spojenosti korisnika na baznu stanicu

U današnje vrijeme većina ljudi koristi mobilni uređaj koji se spaja na bazne stanice. Komunikacija između mobilnog terminalnog uređaja i bazne stanice se odvija konstantno kako bi mobilni uređaj bio spojen na ispravnu baznu stanicu, tj. koristio signal ćelije u kojoj ima najkvalitetniju uslugu. Podatci koji se izmjenjuju u procesu handovera pružaju informaciju o približnoj lokaciji korisnika.

3.1 Lociranje mobilnih uređaja koristeći mrežne tehnologije

Lociranje mobilnih uređaja je proces identifikacije lokacije mobilnog uređaja. Lociranje mobilnih uređaja je danas svakodnevica uz servise kao što su Google Karte koje koriste GPS (engl. *Global Positioning System*) i mobilnu mrežu za lociranje korisnika. Lokacija mobilnog uređaja može se odrediti i samo pomoću infrastrukture mobilne mreže.

Lokacija mobilnog uređaja može se odrediti pomoću infrastrukture pružatelja mobilnih usluga. Lociranje mobilnih uređaja koristeći mrežne tehnologije je razvijeno prije pojave GPS-a u komercijalno dostupnim uređajima. Tehnologija lociranja se bazira na mjerenju razina snage signala i lokacijama antena koristeći pretpostavku da uključeni mobilni uređaj uvijek komunicira bežično sa najbližom baznom stanicom. Određeni napredni sustavi lociranja mogu odrediti sektor ćelije u kojoj je mobilna mreža locirana i ugrubo procijeniti udaljenost od bazne stanice.

Preciznost takvih sustava lociranja je do 50 metara u urbanim sredinama gdje je promet mobilnim podacima dovoljno velik i gustoća baznih stanica zadovoljavajuća [11]. Ruralne sredine i nenaseljena mjesta često imaju bazne stanice međusobne udaljenosti od više kilometara pa je preciznost manja.

Preciznost lociranja korisnika koristeći mrežne tehnologije ovisi o gustoći baznih stanica. Dvije najčešće metode koje se koriste za lokaciju korisnika koristeći mrežne tehnologije su višestranost (engl. *multilateration*) i triangulacija (engl. *triangulation*). Višestranost je tehnika za utvrđivanje lokacije korisnika na osnovu metodologije vremenske razlike dolaska (engl. *Time Difference of Arrival - TDOA*) koja se može upotrijebiti na dva načina:

- signal mobilnog uređaja se mjeri na određenom broju poznatih, fiksnih lokacija
- signal određenog broja fiksnih lokacija se mjeri na mobilnom uređaju.

Vremenska razlika dolaska signala odašiljača omogućava lociranje mobilnog entiteta [12].

Interpolacija je metoda pronalaska novih podatkovnih točaka na osnovu opsega diskretnog skupa poznatih podatkovnih točaka [13].

3.2 Izdvajanje značajki

Izdvajanje značajki (engl. *feature extraction*) je postupak u kojem se neobrađeni podatci svode na skupine podataka koje su lakše za usporedbu.

Glavna karakteristika velikih skupova podataka je značajan broj varijabli koje zahtijevaju popriličnu količinu računalnih resursa za obradu. Izdvajanje značajki je naziv metoda koja odabiru i/ili kombiniraju varijable u značajke, značajno smanjujući broj podataka koji se moraju obraditi, a u isto vrijeme i dalje vjerno prikazujući skup podataka. Izdvajanje značajki također smanjuje broj suvišnih podataka za analizu [14].

U slučaju da su ulazni podatci koje algoritam mora obraditi preveliki za obradu u razumnom vremenu i sumnja se da postoje redundantni podatci, onda se takvi podatci mogu reducirati u skup značajki. Određivanje podskupa inicijalnih značajki zove se odabir značajki (engl. *feature selection*).

Očekuje se da odabrane značajke sadrže relevantne informacije iz inicijalnih podataka tako da se traženi zadatak može odraditi sa novim skupom podataka iz odabranih značajki umjesto sa inicijalnim skupom podataka. Analiza s velikim brojem varijabli zahtjeva veliku količinu računalne memorije i računalne procesorske snage [14].

4. Metode klasifikacije i strojno učenje

Strojno učenje je ogranak umjetne inteligencije baziran na ideji da računala mogu učiti iz podataka, prepoznavati uzorke u podacima na temelju podatkovnih primjera ili prethodnog iskustva.

4.1 Strojno učenje

U strojnom učenju računala otkrivaju načine rješavanja problema bez da su eksplicitno programirana da riješe određeni problem. Računala uče načine rješavanja problema na temelju podataka koji su pruženi računalu. Za jednostavne zadatke moguće je isprogramirati algoritam koji će računalu točno reći način na koji treba izvršiti sve korake potrebne za rješavanje problema. U ovom slučaju računalo nije potrebno naučiti način rješavanja problema s obzirom da je egzaktan algoritam rješavanja isprogramiran [15].

Ovisno o ulaznim podacima, strojno učenje dijelimo u sljedeće kategorije: nadzirano učenje (engl. *supervised learning*), nenadzirano učenje (engl. *unsupervised learning*) i podržano učenje (engl. *reinforcement learning*) [15].

Nadzirano učenje karakteriziraju ulazni podatci i odgovarajuće izlazne kategorije. Cilj nadziranog učenja je naučiti općenito pravilo koje bi povezal ulazne podatke i izlazne kategorije. Počevši s analizom ulaznih podataka, algoritam učenja proizvede funkciju pomoću koje se rade predviđanja o izlaznim kategorijama. Algoritam učenja je u mogućnosti usporediti izlazne, predviđene, podatke sa ispravnim izlaznim kategorijama i pronaći greške te tako optimizirati algoritam učenja [15].

Nenadzirano učenje je vrsta strojnog učenja koja koristi skup ulaznih neoznačenih podataka i pušta algoritam uči iz podataka bez navođenja. Cilj nenadziranog učenja je pronalazak pravilnosti, trendova, uzoraka i sličnih karakteristika. Algoritam koristi sljedeće metode za obradu ulaznih podataka: metoda grupiranja (engl. *clustering*), otkrivanje stršćih vrijednosti (engl. *outlier detection*) i smanjenje dimenzionalnosti (engl. *dimensionality reduction*) [15].

Podržano učenje je vrsta strojnog učenja u kojem algoritam kreira matematički model kako bi maksimizirao sumu nagrada. Za razliku od nadziranog učenja, gdje ulazni podatci sadrže određenu kategoriju, kod podržanog učenja model nema točan odgovor kojoj kategoriji pripada ulazni podataka. Sukladno tome, algoritam podržanog učenja mora učiti iz svog iskustva [15].

Osim po ulaznim podacima, strojno učenje se može podijeliti i po vrsti izlazne vrijednosti: na regresiju (engl. *regression*) i klasifikaciju (engl. *classification*). Algoritam je

regresijski ukoliko su izlazne vrijednosti kontinuirane vrijednosti, a algoritam je klasifikacijski ukoliko se koristi za predviđanje unaprijed određenih kategorija [15].

Također, svaki algoritam se sastoji od tri skupa podataka: skup za treniranje, skup za validaciju i skup za testiranje. Model uči isključivo iz skupa za treniranje, dok ostala dva skupa ne vidi do same validacije odnosno testiranja [15].

4.2 Metode klasifikacije

Klasifikacija je proces prepoznavanja, razumijevanja i grupiranja ulaznih podataka u izlazne kategorije koje su predodređene. Klasifikacijski algoritmi koji se koriste u strojnom učenju iskorištavaju skup za treniranje u svrhu predviđanja u kojoj predodređenoj kategoriji će budući, dosad neviđeni, podatci pripadati.

Metode klasifikacije koje će se promatrati i evaluirati u ovom radu su: metoda k -najbližih susjeda (engl. *k-nearest neighbors*), algoritam slučajnih šuma (engl. *random forrests*) i ExtraTrees algoritam.

4.2.1 Metoda k -najbližih susjeda

Algoritam k -najbližih susjeda je jedan od najjednostavnijih algoritama strojnog učenja. Algoritam k -najbližih susjeda pohranjuje uzorke skupa za učenje i klasificira nove uzorke usporedbom njihove sličnosti sa postojećim uzorcima koristeći funkcije udaljenosti. k -najbližih susjeda može se koristiti i za regresiju, ali se najčešće koristi za klasifikacijske probleme. Algoritam k -najbližih susjeda je algoritam koji ne uči na temelju seta podataka za treniranje, već sprema skup podataka i u trenutku klasifikacije izvršava radnju nad podacima.

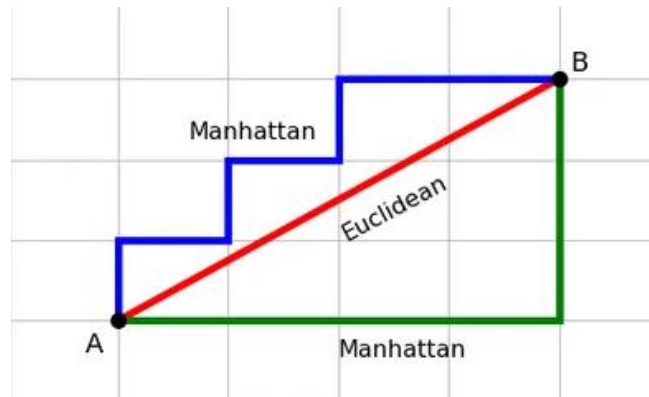
Algoritam k -najbližih susjeda se može promatrati kroz 5 koraka:

- korak 1: odrediti k broj susjeda
- korak 2: izračunati udaljenost između k broja susjeda, kao na slici 3
- korak 3: uzima se k -ti najbliži susjed prema izračunatoj udaljenosti iz koraka 2
- korak 4: između ovih k susjeda, prebrojati podatkovne točke u svakoj kategoriji
- korak 5: dodijeliti nove podatkovne točke kategoriji za koju je najveći broj k susjeda [17].

Udaljenost između k broja susjeda može se izračunati pomoću:

- Euklidska udaljenosti (engl. *Euclidean distance*)
- Manhattanske udaljenosti (engl. *Manhattan distance*),

prikazano na slici 4 [17].



Slika 4 Prikaz Euklidske i Manhattanske udaljenosti

Euklidska udaljenost računa se prema sljedećoj formuli:

$$d(A, B) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

gdje je d udaljenost između točke A , sa koordinatama x_1 i y_1 i točke B , sa koordinatama x_2 i y_2 .

Manhattanska udaljenost računa se prema sljedećoj formuli:

$$d(A, B) = |x_1 - x_2| + |y_1 - y_2|$$

gdje je d udaljenost između točke A , sa koordinatama x_1 i y_1 i točke B , sa koordinatama x_2 i y_2 .

4.2.2 Metoda slučajnih šuma

Metoda slučajnih šuma je popularan algoritam strojnog učenja. Kao i k -najbližih susjeda, može se koristiti i za regresijske i za klasifikacijske probleme. Bazira se na konceptu ansambla, odnosno procesa kombiniranja više klasifikatora za rješenje kompleksnih problema i za unaprijeđenje performansi modela. Metoda slučajnih šuma je klasifikator koji sadrži broj stabala odluka (engl. *decision trees*) nad određenim skupom podataka i uzima prosjek radi povećanja preciznosti predviđanja skupa podataka.

Algoritam slučajnih šuma radi u dvije faze. U prvoj fazi kreira nasumičnu šumu spajajući N broj stabala odluka. U drugoj fazi radi predviđanje za svako stablo kreirano u prvoj.

Algoritam slučajnih šuma promatra se kroz sljedećih 5 koraka:

- korak 1: odabrati k nasumičnih podataka iz skupa za treniranje,
- korak 2: stvoriti stabla odluka povezanih sa odabranim podacima iz prvog koraka,
- korak 3: kreirati N broj stabala odluka,
- korak 4: koraci jedan i dva se ponavljaju,
- korak 5: za nove podatke pronaći predviđanje svakog stabla odluke i dodijeliti taj podatak kategoriji koja ima najviše glasova [18].

4.2.3 ExtraTrees algoritam

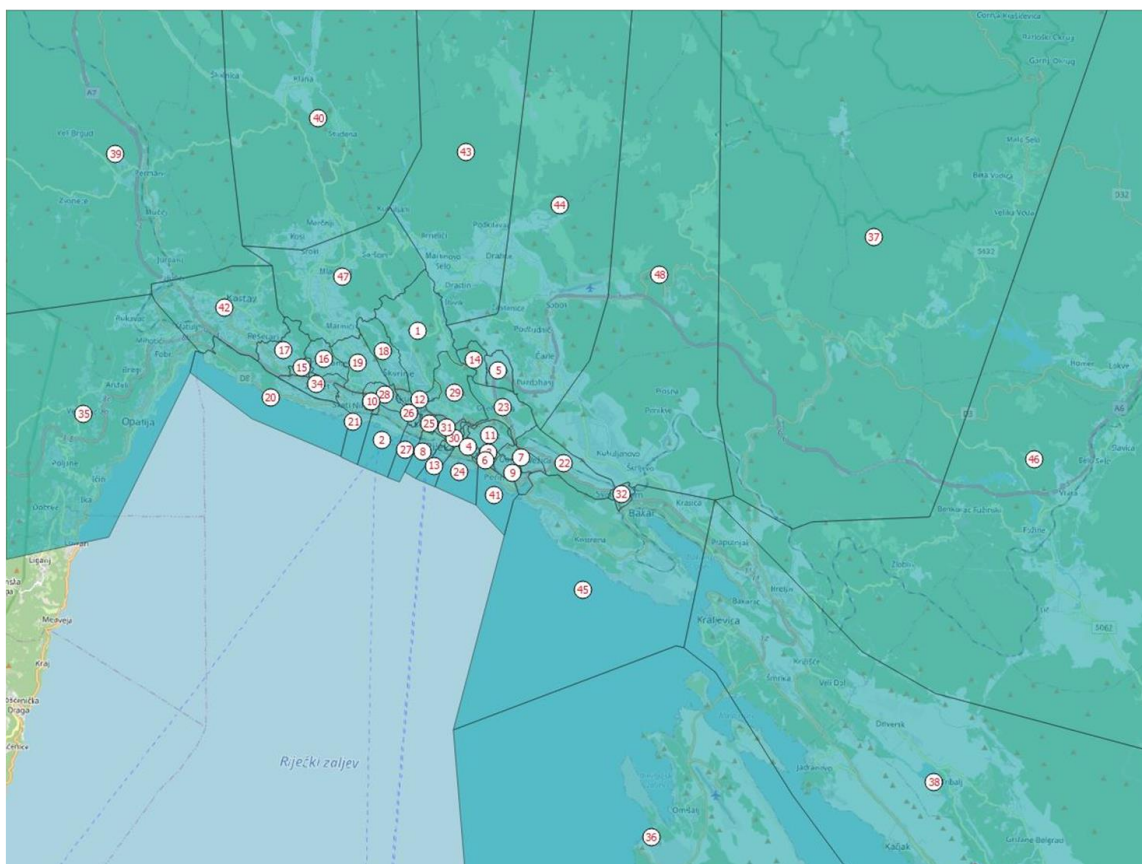
ExtraTrees algoritam je sličan metodu slučajnih šuma uz dvije glavne razlike:

- metoda slučajnih šuma koristi "*bootstrap replica*" kao set za treniranje, "*bootstrap replica*" je postupak u kojem se uzima podatak iz originalnog skupa za treniranje i prepisuje u novi skup podataka za treniranje s tim da se duplicirani podatci mogu pojaviti, dok ExtraTrees algoritam koristi izvorni skup podataka za treniranje
- metoda slučajnih šuma odabire optimalan rez podataka na grane dok ExtraTrees odabire taj rez nasumično.

5. Opis podataka koji su korišteni u klasifikaciji

Korišteni skup podataka u ovom radu sastoji se od 505 992 kategoriziranih podataka izvorišno-odredišne matrice, koje predstavljaju podatke dobivenih pomoću mobilne mreže o putovanju u gradu Rijeci. Podatci su ustupljeni od strane tvrtke Ericsson Nikola Tesla.

Svaki put se opisuje pomoću 9 atributa prikazanih u tablici 2. Atribut 'Interval' predstavlja vrijeme intervala u kojem se put zapisao. Moguće vrijednosti su 00:00-06:00, 06:00-09:00, 09:00-14:00, 14:00-18:00 ili 18:00-00:00 sati. 'Trajanje' predstavlja ukupno vrijeme putovanja izraženo u sekundama. 'Izvorišni sektor' predstavlja sektor početka putovanja pod odgovarajućim ID-em prikazanih na slici 5. 'Odredišni sektor' predstavlja sektor završetka putovanja pod odgovarajućim ID-em, kako je prikazano na slici 5, [19]. 'Zračna udaljenost' i 'zračna brzina' predstavljaju izračunatu euklidsku udaljenost i brzinu izračunatu pomoću euklidske udaljenosti i razlike u vremenu početka i završetka putovanja. 'Cestovna udaljenost' i 'cestovna brzina' predstavljaju stvarne udaljenosti i brzine korisnika koji su putovali cestovnom infrastrukturom. 'Mod' predstavlja mod prijevoza koji se koristio za završetak putovanja. Mogući modovi prijevoza u podacima su 'automobil', 'javni prijevoz', 'hodanje' i 'bicikl'.

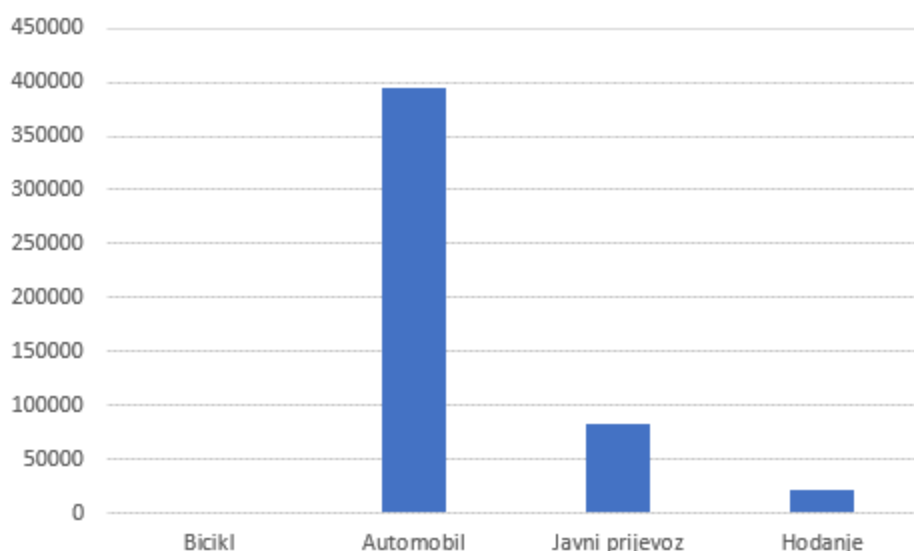


Slika 5 Prostorne zone grada Rijeke

Tablica 2 Atributi skupa podataka

| Atribut | Opis |
|---------------------|---|
| Interval | Interval u kojem je put započeo |
| Trajanje | Vrijeme trajanja puta [s] |
| Izvorišni sektor | ID sektora početka putovanja |
| Odredišni sektor | ID sektora završetka putovanja |
| Zračna udaljenost | Euklidska udaljenost između početka i kraja putovanja [m] |
| Zračna brzina | Brzina dobivena iz euklidske udaljenosti [m/s] |
| Cestovna udaljenost | Prava udaljenost putovanja prometnicom između početku i kraja putovanja [m] |
| Cestovna brzina | Brzina dobivena iz cestovne udaljenosti |
| Mod prijevoza | Mod prijevoza koji je korišten za put |

S obzirom na to da se mod prijevoza 'bicikl' ponavlja samo 220 puta u cijelom skupu podataka, za samu evaluaciju su uklonjeni podaci čiji je mod prijevoza 'bicikl'. Mod prijevoza 'automobil' pojavljuje se 394 290 puta, mod prijevoza 'javni prijevoz' pojavljuje se 82 166 puta, a 'hodanje' 22 368 puta, što se vidi i na slici 6.



Slika 6 Broj podataka po modu prijevoza

6. Evaluacija klasifikacijskih algoritama na temelju dostupnih podataka

Za evaluaciju klasifikacijskih algoritama na temelju dostupnih podataka korišten je programski jezik Python te biblioteka SciKit-learn. Svaki klasifikacijski algoritam evaluiran je pomoću matrica zabune (engl. *confusion matrix*), točnosti (engl. *accuracy*), osjetljivosti (engl. *recall*) i preciznosti (engl. *precision*).

Skup za testiranje je postavljen na 30% od ukupnog skupa podataka dok je skup za treniranje postavljen na 70%. Cilj testiranja je predvidjeti mod prijevoza na osnovi atributa koji su sljedeći: 'trajanje', 'Izvorišni sektor', 'Odredišni sektor', 'Zračna udaljenost', 'Zračna brzina', 'Cestovna udaljenost' i 'Cestovna brzina'. Klase su modovi prijevoza 'automobil', 'javni prijevoz' i 'hodanje'.

6.1 Matrica zabune

Matrica zabune je $N \times N$ matrica gdje je N klasa te služi za evaluaciju klasifikacijskog algoritma. U ovom radu su tri klase s obzirom da se provjeravaju klasifikacijski algoritmi koji predviđaju mod prijevoza. Matrica uspoređuje stvarne ciljne vrijednosti, odnosno modove, sa onima koje je algoritam predvidio.

U tablici 3 se nalaze objašnjenja polja matrica. Vrijednost je pozitivna ukoliko se stvarna i predviđena vrijednost preklapaju, odnosno ukoliko je algoritam ispravno predvidio klasu. Vrijednost je negativna ukoliko je algoritam neispravno predvidio klasu. Broj koji se nalazi u polju matrice označava postotak ispravno ili neispravno predviđenih klasa.

Tablica 3 Matrica zabune veličine 3 x 3

| Stvarne vrijednosti | | | | |
|------------------------|---------|----------------|----------------|----------------|
| Predviđene vrijednosti | | Klasa 1 | Klasa 2 | Klasa 3 |
| | Klasa 1 | Pozitivna 1 | Negativna 2 | Negativna 3 |
| | Klasa 2 | Negativna 4 | Pozitivna 5 | Negativna 6 |
| | Klasa 3 | Negativna 7 | Negativna 8 | Pozitivna 9 |

Matrica zabune se sastoji od 4 kategorije:

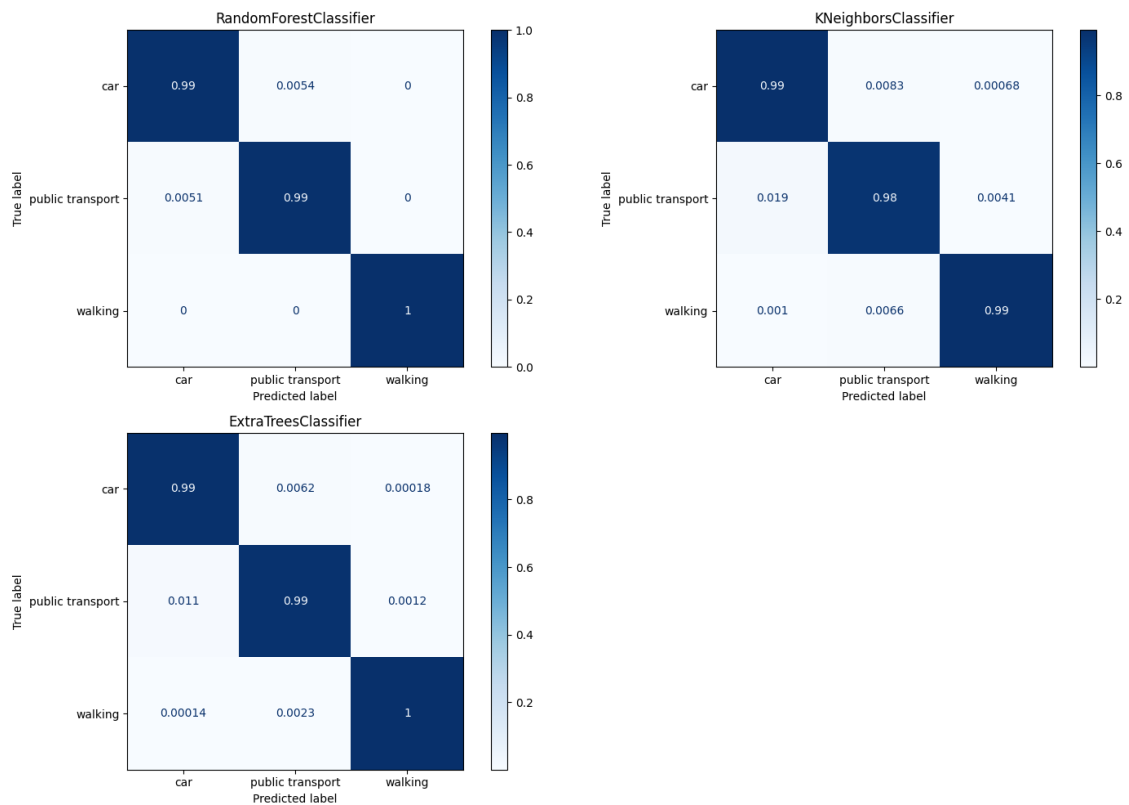
- TP – true positive, predviđena vrijednost odgovara stvarnoj vrijednosti
- FP – false positive, predviđena vrijednost ne odgovara stvarnoj vrijednosti, stvarna vrijednost je negativna, a algoritam je predvidio pozitivnu vrijednost
- FN – false negative, predviđena vrijednost ne odgovara stvarnoj vrijednosti, stvarna vrijednost je pozitivna, a algoritam je predvidio negativnu vrijednost
- TN – true negative, predviđena vrijednost odgovara stvarnoj vrijednosti, stvarna vrijednost je bila negativna, a algoritam je predvidio pozitivnu vrijednost.

Gore navedene kategorije računaju se po sljedećem principu za matricu veličine 3 x 3 koji je vidljiv u tablici 4 [20].

Tablica 4 Izračunavanje TP, FP, FN i TN vrijednosti

| | Klasa 1 | Klasa 2 | Klasa 3 |
|----|---|---|---|
| TP | Polje 1 | Polje 5 | Polje 9 |
| FP | Polje 2 + Polje 3 | Polje 4 + Polje 6 | Polje 7 + Polje 8 |
| FN | Polje 4 + Polje 7 | Polje 2 + Polje 8 | Polje 3 + Polje 6 |
| TN | Polje 5 + Polje 6 + Polje 8 + Polje 9 | Polje 1 + Polje 2 + Polje 7 + Polje 9 | Polje 1 + Polje 2 + Polje 4 + Polje 5 |

Na slici 7 vidljive su matrice zabune za tri klasifikatora. Sva tri klasifikatora su točno predvidjela klasu automobila za 99% podataka. Javni prijevoz je točno predviđen za 98% podataka u klasifikatoru k-najbližih susjeda, a za 99% podataka u klasifikatorima slučajne šume i ExtraTrees. Hodanje je predviđeno za 100% podataka sa klasifikatorima slučajne šume i ExtraTrees dok je k-najbližih susjeda ispravno predvidio 99% podataka.



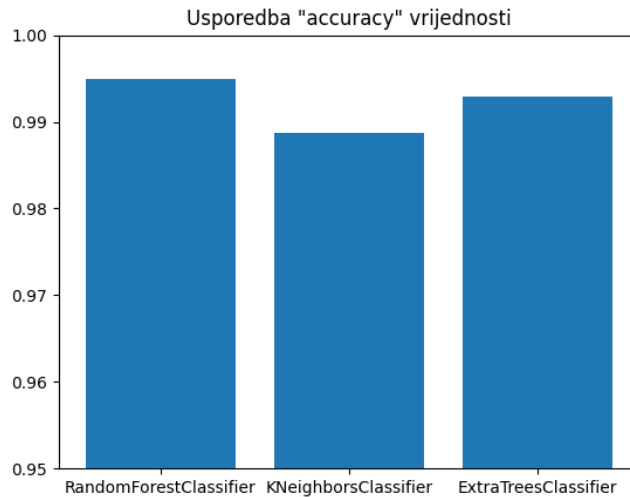
Slika 7 Matrice zabune za klasifikatore k-najbližih susjeda, slučajne šume i ExtraTrees

6.2 Točnost

Klasifikator ispravno klasificira na dijagonali matrice zabune. Točnost se definira kao omjer ispravno klasificiranih primjera i ukupnog N broja primjera. Formula za točnost, [21], je sljedeća:

$$Acc = \frac{TP + TN}{N}$$

Na slici 8 je usporedba točnosti za klasifikatore. Klasifikator slučajne šume ima točnost 99,49 %, k-najbližih susjeda ima točnost 98,87%, a ExtraTrees 99,29%.



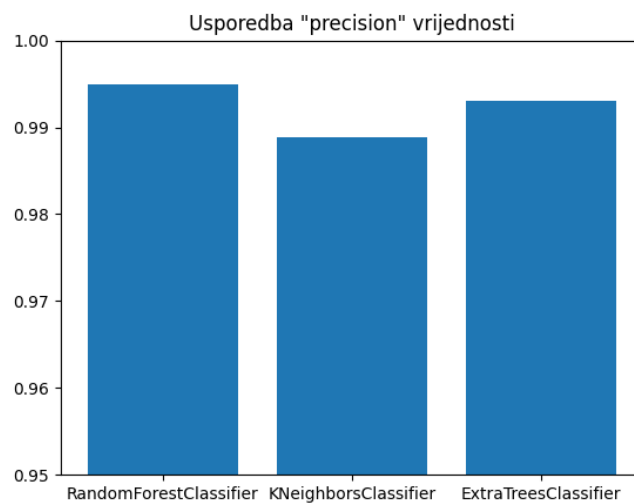
Slika 8 Usporedba točnosti klasifikatora

6.3 Preciznost

Preciznost pokazuje omjer uzoraka neke klase koji uistinu pripadaju toj klasi u odnosu na broj uzoraka kojima je predviđeno za pripadaju toj klasi. Formula za preciznost, [21], je sljedeća:

$$P = \frac{TP}{TP + FP}$$

Na slici 9 je usporedba točnosti za klasifikatore. Klasifikator slučajne šume ima točnost 99,49%, k-najbližih susjeda ima točnost 98,88%, a ExtraTrees 99,3%.



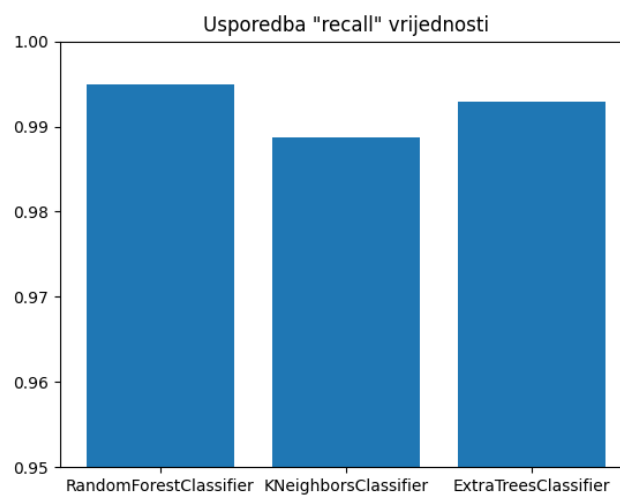
Slika 9 Usporedba preciznosti klasifikatora

6.4 Odziv

Odziv je definiran omjerom broja primjera koje je klasifikator označio kao pozitivne i ukupnog broja pozitivnih primjera. Formula za odziv, [21], je sljedeća:

$$R = \frac{TP}{TP + FN}$$

Na slici 10 je usporedba točnosti za klasifikatore. Klasifikator slučajne šume ima točnost 99,49%, k-najbližih susjeda ima točnost 98,87%, a ExtraTrees 99,29%.



Slika 10 Usporedba odziva klasifikatora

7. Zaključak

Strojno učenje je iznimno koristan proces u današnje vrijeme s obzirom na to da smanjuje obujam posla ljudima i donosi veću preciznost rada. Prije pojave strojnog učenja ljudi su za svaki problem morali kreirati algoritam rješavanja problema korak po korak, dok se pomoću strojnog učenja može kreirati model koji efikasno i efektivno rješava više vrsta problema. Korištenje strojnog učenja je zahtjevan posao s obzirom na to da, ako se napravi greška u procesu programiranja algoritma strojnog učenja, može doći do rješenja problema koje nije zadovoljavajuće.

Mogućnosti strojnog učenja uz podatke iz mobilnih mreža su ogromne. Svaki trenutak se šalji veliki broj podataka, između mobilnih uređaja i baznih stanica. Veoma je bitno imati način interpretacije tolike količine podataka što je moguće uz ispravno korištenje strojnog učenja. Strojno učenje se izvodi u tri vrste i najbolji rezultati se vide kada podatci prođu kroz sve tri vrste strojnog učenja, što najčešće nije slučaj u praksi.

U ovom radu su evaluirane tri klasifikacijske metode na velikom skupu podataka koji sadrži nekoliko bitnih informacija za algoritme klasifikacije kao što su trajanje putovanja, brzina, udaljenost između početne i krajnje točke putovanja. Evaluacija se izvršavala programskim jezikom Python uz biblioteku SciKit-learn.

Sve tri klasifikacijske metode su imale izvrsne rezultate sa preciznošću, točnošću i odzivom preko 98%. Međutim, čak i među takvim metodama postoji razlika u navedenim metrikama. Metoda slučajnih šuma je imala preciznost, točnost i odziv oko 99,5%, metoda k-najbližih susjeda oko 98,8%, a ExtraTrees oko 99,3%.

Navedenim mjerenjima je pokazano da sve tri metode imaju zadovoljavajuće predviđanje zadanih 30% testnog skupa podataka. U budućem radu fokus istraživanja bi bio proširiti broj klasifikatora kako bi se istražilo koje metode nisu prihvatljive za determiniranje moda prijevoza pomoću podataka koji su se koristili.

LITERATURA

- [1] Miao G, Zander J, Sung K W, Ben Slimane S. Principles of cellular systems. Fundamentals of Mobile Data Networks. Cambridge University Press; 2016.
- [2] Miller M. Wireless Networking Absolute Beginner's Guide. Pearson Education, Inc.; 2013.
- [3] Kimura W D. Electromagnetic Waves and Lasers. Morgan & Claypool Publishers; 2017.
- [4] Sauter M. From GSM to LTE-Advanced Pro and 5G. John Wiley & Sons Ltd; 2017.
- [5] Truong K T, Heath R W. Effects of channel aging in massive MIMO systems. Journal of Communications and Networks; 2013.
- [6] Rappaport T S. Wireless Communications: Principles and Practice Subsequent Edition. Prentice Hall; 2001.
- [7] Dhiman, A, Sandha Singh, K. Vertical and horizontal handover in heterogeneous wireless networks using opnet, international journal of engineering research & technology; 2013.
- [8] Pramod Kumar, P, Thirupathi, V, Monica, D: enhancements in mobility management for future wireless networks, international journal of advanced research in computer and communication engineering; 2013.
- [9] <https://www.techrepublic.com/article/does-the-world-really-need-5g/> (pristupljeno 1.9.2021.)
- [10] <https://www.qualcomm.com/5g/what-is-5g> (pristupljeno 1.9.2021.)
- [11] Laitinen H, Lahteenmaki J, Nordstrom T. Database correlation method for GSM location. 53rd Vehicular Technology Conference; May 06-09 2001.; Rhodes, Greece
- [12] Miller F P, Vandome A F, McBrewster J. Multilateration. VDM Publishing; 2010.
- [13] Steffensen, J. F. , INTERPOLATION, 2006.
- [14] <https://deepai.org/machine-learning-glossary-and-terms/feature-extraction> (pristupljeno 28.8.2021.)
- [15] Alpaydin E. Introduction to machine learning, The MIT Press; 2014.
- [16] <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning> (pristupljeno 30.8.2021.)
- [17] <https://www.omnicalculator.com/math/manhattan-distance> (pristupljeno 5.9.2021.)
- [18] <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (pristupljeno 27.8.2021.)
- [19] Tišljarić L, Cvetek D, Vareškić V, Gregurić M. Classification of Travel Modes from Cellular Network Data Using Machine Learning Algorithms. 2021.
- [20] <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning> (pristupljeno 3.9.2021.)
- [21] https://www.fer.unizg.hr/download/repository/SU-2015-Vrednovanje_modela.pdf (pristupljeno 2.8.2021.)

POPIS SLIKA

| | |
|---|----|
| Slika 1 Shematski prikaz ćelija..... | 1 |
| Slika 2 Pozicija baznih stanica unutar ćelija | 3 |
| Slika 3 Koncept ponovnog iskorištavanja frekvencija..... | 3 |
| Slika 4 Prikaz Euklidske i Manhattanske udaljenosti | 11 |
| Slika 5 Prostorne zone grada Rijeke..... | 13 |
| Slika 6 Broj podataka po modu prijevoza | 14 |
| Slika 7 Matrice zabune za klasifikatore k-najbližih susjeda, slučajne šume i ExtraTrees | 17 |
| Slika 8 Usporedba točnosti klasifikatora..... | 18 |
| Slika 9 Usporedba preciznosti klasifikatora | 18 |
| Slika 10 Usporedba odziva klasifikatora | 19 |

POPIS TABLICA

| | |
|---|----|
| Tablica 1 Elektromagnetski spektar | 2 |
| Tablica 2 Atributi skupa podataka | 14 |
| Tablica 3 Matrica zabune veličine 3 x 3 | 15 |
| Tablica 4 Izračunavanje TP, FP, FN i TN vrijednosti | 16 |



Sveučilište u Zagrebu
Fakultet prometnih znanosti
10000 Zagreb
Vukelićeva 4

IZJAVA O AKADEMSKOJ ČESTITOSTI I SUGLASNOST

Izjavljujem i svojim potpisom potvrđujem kako je ovaj _____ završni rad
isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na
objavljenu literaturu što pokazuju korištene bilješke i bibliografija.
Izjavljujem kako nijedan dio rada nije napisan na nedozvoljen način, niti je prepisan iz
necitiranog rada, te nijedan dio rada ne krši bilo čija autorska prava.
Izjavljujem također, kako 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 _____ završnog rada
pod naslovom ODREĐIVANJE MODA PRIJEVOZA ANALIZOM SPAJANJA
KORISNIKA NA POJEDINE BAZNE STANICU MOBILNE MREŽE
na internetskim stranicama i repozitoriju Fakulteta prometnih znanosti, Digitalnom akademskom
repozitoriju (DAR) pri Nacionalnoj i sveučilišnoj knjižnici u Zagrebu.

U Zagrebu, 06/09/2021

Student/ica:

Milica Kozina
(potpis)