

# Klasifikacija modova prijevoza višerazinskom metodom zasnovanom na matricama prijelaznih stanja i stablima odluke u stvarnom vremenu

---

Erdelić, Martina

Doctoral thesis / Disertacija

2022

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:119:316540>

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

Download date / Datum preuzimanja: **2024-07-19**



Repository / Repozitorij:

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





Sveučilište u Zagrebu  
Fakultet prometnih znanosti

Martina Erdelić

**KLASIFIKACIJA MODOVA PRIJEVOZA  
VIŠERAZINSKOM METODOM ZASNOVANOM NA  
MATICAMA PRIJELAZNIH STANJA I STABLIMA  
ODLUKE U STVARNOM VREMENU**

DOKTORSKI RAD

Zagreb, 2022.





Sveučilište u Zagrebu  
Fakultet prometnih znanosti

Martina Erdelić

**KLASIFIKACIJA MODOVA PRIJEVOZA  
VIŠERAZINSKOM METODOM ZASNOVANOM NA  
MATRICAMA PRIJELAZNIH STANJA I STABLIMA  
ODLUKE U STVARNOM VREMENU**

DOKTORSKI RAD

Mentor: prof. dr. sc. Tonči Carić

Zagreb, 2022.





University of Zagreb  
Faculty of Transport and Traffic Sciences

Martina Erdelić

**CLASSIFICATION OF TRANSPORT MODES  
USING MULTILEVEL METHOD BASED ON  
TRANSITION STATE MATRICES AND DECISION  
TREES IN REAL TIME**

DOCTORAL DISSERTATION

Supervisor: Professor Tonči Carić, Ph.D.

Zagreb, 2022



Doktorska disertacija izrađena je na Sveučilištu u Zagrebu, Fakultet prometnih znanosti, Zavod za inteligentne transportne sustave, Katedra za primijenjeno računalstvo.

Mentor: prof. dr. sc. Tonči Carić

Doktorska disertacija sadrži: 170 stranica

Povjerenstvo za ocjenu doktorske disertacije:

1. izv. prof. dr. sc. Edouard Ivanjko (predsjednik povjerenstva)  
Fakultet prometnih znanosti, Sveučilište u Zagrebu
2. prof. dr. sc. Tonči Carić  
Fakultet prometnih znanosti, Sveučilište u Zagrebu
3. prof. dr. sc. Ivana Šemanjski  
Faculty of Engineering and Architecture, Ghent University
4. izv. prof. dr. sc. Mario Muštra (zamjena)  
Fakultet prometnih znanosti, Sveučilište u Zagrebu

Datum:

29. kolovoza, 2022.





---

## **O mentoru**

---

**Tonči Carić** diplomirao je 1993. i magistrirao 2000. na Fakultetu elektrotehnike i računarstva, Sveučilišta u Zagrebu. Obranio je doktorsku disertaciju 2004. na Fakultetu prometnih znanosti, Sveučilišta u Zagrebu. Trenutno radi na Zavodu za Inteligentne transportne sustave Fakulteta prometnih znanosti kao redovni profesor i voditelj je Katedre za primijenjeno računalstvo. Sudjelovao je kao istraživač na znanstveno razvojnim projektima financiranim od Europske komisije, nacionalnim znanstvenim projektima i stručnim projektima u gospodarstvu. Njegovi glavni istraživački interesi vezani su za kombinatorne optimizacije, baze podataka pokretnih objekata i inteligentne transportne sustave. Član je udruženja INFORMS i IEEE.

---

## **About the mentor:**

---

**Tonči Carić** received his B.Sc. and the M.Sc. degree in Computer Science from the Faculty of Electrical Engineering and Computing, University of Zagreb, in 1993 and 2000, respectively. He received his Ph.D. degree in 2004 from the Faculty of Transport and Traffic Engineering, University of Zagreb. He is currently working at the Department of Intelligent Transportation Systems, Faculty of Transport and Traffic Sciences, University of Zagreb as Professor and Head of the Chair for Applied Computing. He participated in research and development projects funded by the European Commission, national research projects, and professional projects in cooperation with industry. His main research interests are combinatorial optimization, moving object databases, and intelligent transportation systems. He is a member of INFORMS and IEEE.



---

*Mojj obitelji*

---



---

## Zahvala

---

Zahvala je zadnji tekst koji pišem u svom doktorskom radu i zapravo je meni osobno savršen zaključak i najveći doprinos jer pokazuje onaj najljepši dio koji osjetiš i naučiš u ovakvo posebnom periodu života. Stoga koristim ovu priliku da se zahvalim svim, meni dragim ljudima, koji su na bilo koji način sudjelovali u mojoj životnoj priči tijekom izrade doktorskoga rada.

Najveću zahvalu dugujem svom suprugu Tomislavu. Hvala ti što si uvijek tu, što razumiješ i što si moj podsjetnik kada zaboravim na vrijeme. Ja znam da je teško biti moj slušatelj i hvala ti što si to tako dobro podnio. Osim toga, hvala ti na svakoj sugestiji, komentaru i osobito na odvojenom vremenu za pomoć u izradi aplikacije. Hvala ti što mi pokazuješ koliko smo snažni zajedno, s tobom i uz tebe je ovaj proces bio znatno lakši. Hvala našem Janu što je tako divno dijete uz koje dva doktorata nisu bila tako veliki problem i što u svaki trenutak unosi veselje onako kako samo on zna.

Veliku zahvalu dugujem svojim roditeljima i bratu, vi ste tu tijekom cjelokupnog procesa mog školovanja i moja vječna podrška. Hvala vam na tome. Mama, hvala ti što si moj vjerni slušatelj i što si tako pažljivo sa strane pratila moj život i time dopustila da se spontano i bez nametanja oblikujem u osobu kakva ja jesam i želim biti. Tata, ti preuzimaš zasluge za ovu moju temperamentniju stranu, a moj mali brat me puno toga naučio iako je mislio da samo on uči od mene. Hvala mojoj teti i tetku (sigurna sam da ćeš nekako ovo čuti) koji su oduvijek podupirali moje školovanje i odrastanje. Hvala teti Margret koja je prva osoba koja je govorila da ću doktorirati, čak prije nego li sam ja to sama znala.

Obožavam činjenicu što vas imam uz sebe, moja mala *Family*. Dijelimo cijeli život zajedno, od najranijeg djetinjstva, i sigurna sam da ćemo još puno toga podijeliti zajedno. Hvala vam za svaki razgovor, prikupljeni podatak, podršku, zagrljaj i bodrenje. S vama svaki težak trenutak postaje jednostavniji.

Osoba s kojom gradite odnos tijekom procesa izrade doktorata je mentor. Ja sam imala sreću da je moj mentor osoba koja me poticala na bolje i pustila da radim ono što želim. Tonči, hvala vam na tome. Moram priznati da me u početku frustriralo što nisam dobila jasnu uputu u kojem smjeru trebam ići, ali gledajući unazad neizmjereno sam vam zahvalna upravo na toj činjenici jer usmjeriti nekoga da pronađe svoj interes je znatno teže od davanja jasnih uputa. Izuzetno mi je drago da ste upravo vi bili moj mentor i naučili me mnogočemu tijekom ovog putovanja.

Veliko hvala mojoj uvijek mentorici Štefici i Marku. Vi ste bili moj prvi doticaj s istraživanjem i da nije bilo vas pitanje je da li bi ja tako zavoljela istraživanje. Hvala svim kolegama na Zavodu za ITS na svim komentarima, sugestijama i razgovorima. Lijepo je surađivati s tako pametnim i otvorenim ljudima. Hvala članovima povjerenstva, Ivani i prof. Ivanjku na izuzetno brzom čitanju doktorata i vrijednim sugestijama.

I zadnje, ali izuzetno bitno, hvala svim mojim *Collecty* obožavateljima koji su vjerno uklju-

čivali aplikaciju i prikupljali podatke. Vi ste dio ovog doktorata i bez vas on ne bi bio potpun. Moram se posebno zahvaliti svojoj Josipi i Mislavu koji su neumorno tijekom više mjeseci prikupljali podatke i prikupili daleko najveći broj podataka. Također, hvala Maji i Dadi koji su prikupili veliku količinu podataka i ko mravići ostavili trag do Siska. I posljednji vjerni sakupljač je moj Tomislav, hvala ti na svim prikupljenim podacima, a osobito na svim zimskim vožnjama biciklom i električnim romobilom.

---

## Sažetak

---

Poznavanje mobilnosti korisnika prometne mreže osnova je za određivanje prometne potražnje i planiranje transporta. Podatci s mobilnih uređaja postali su vrijedan izvor prometnih podataka o mobilnosti korisnika u prometnoj mreži. Za prikupljanje dodatnih informacija tijekom kretanja korisnika prometnom mrežom, poput modova prijevoza, koriste se razne metode kojima se iz prikupljenih podataka određuju informacije koje se ne mogu automatski prikupljati. Za detekciju moda prijevoza koriste se metode klasifikacije trajektorije, pri čemu se klasificira dio trajektorije u kojem je korišten jedan mod prijevoza. Trajektoriju je prije klasifikacije potrebno podijeliti na segmente koji sadrže jedan mod prijevoza. U literaturi su zastupljene mnogobrojne metode za klasifikaciju modova prijevoza koje se najčešće zasnivaju na metodama strojnog učenja dok su za segmentaciju trajektorije zastupljene metode koje se zasnivaju na nizu iskustveno postavljenih pravila. Nedostatak tako generiranih pravila je što su često prilagođena skupu podataka nad kojim su razvijena i nisu jedinstveno primjenjiva s jednakom točnošću na druge skupove podataka.

U ovoj doktorskoj disertaciji je razvijena višerazinska metoda za segmentaciju trajektorije i klasifikaciju modova prijevoza u stvarnom vremenu. Za odlučivanje se koriste matrice prijelaznih stanja i algoritam slučajnih šuma. Metoda je razvijena i testirana koristeći objavljene skupove podataka, a za validaciju je korišten novi skup podataka prikupljen u sklopu ovog istraživanja. U prvom dijelu doktorske disertacije detaljno je opisana implementacija višerazinske metode za segmentaciju i klasifikaciju trajektorije. Također, opisan je postupak obrade podataka koji uključuje podjelu sirovih podataka u vremenske prozore te preslikavanje podataka iz vremenske u frekvencijsku domenu. Na kraju prvog dijela doktorske disertacije analizirani su rezultati klasifikacije trajektorije razvijenom metodom i uspoređeni su s rezultatima postignutim na testnom skupu podataka dostupnom u literaturi. U drugom dijelu doktorske disertacije opisan je novi skup podataka prikupljen u sklopu ovog istraživanja s naglaskom na mobilnu aplikaciju i način prikupljanja podataka te distribuciju modova prijevoza u skupu podataka. Prikupljeni skup podataka korišten je za validaciju metode čiji su rezultati prikazani na kraju ovog dijela.

**Ključne riječi:** klasifikacija modova prijevoza, segmentacija trajektorije, urbana mobilnost, algoritam slučajnih šuma, matrice prijelaznih stanja, strojno učenje, podatci s osjetila mobilnih uređaja, mobilna aplikacija





---

## Extended abstract

---

### **Classification of Transport Modes Using Multilevel Method Based on Transition State Matrices and Decision Trees in Real Time**

---

#### **Introduction**

A transport network is a complex system whose analysis requires data from transport network users. With the growing degree of urbanization and population, transport is becoming one of the main factors affecting the quality of life in large cities. Therefore, research fields such as urban mobility, energy consumption, pollution and security are often the subject of scientific research. All these research fields are directly or indirectly related to the mobility of transport network users. The knowledge that can be acquired by observing the model of human mobility through data collected in different transport environments can significantly contribute to solving transport problems. These data can be collected from several sources and are divided into two basic types: sensors placed at fixed locations on the transport network and mobile sensors.

Data from mobile sensors require a different processing approach because they do not contain information about the static location on the transport network but contain a series of sequential locations of the transport entity depending on time. In that case, an in-depth analysis of the trajectory, which can be supported by clustering or classification methods, can describe the mobility of users through the transport network. Recognition of everyday human activities during various mobilities and uses of transport modes has several applications, such as educating users to change their pattern of behavior or promoting a healthy lifestyle. In addition, applications, which track user mobility, can help to public urban transport planning, smart parking and vehicle traffic monitoring. Also, some information about trajectories are not collected from the mobile sensors but can be identified afterwards from collected data. One additional information is the transport mode the user uses on the transport network. Information about the used transport mode gives new value to the data about the mobility of a group of users. For example, instead of total traffic demand on the part of the transport network, traffic demand can be described for each transport mode or demand for some particular bus line can be detected.

Trajectory classification methods are used to detect transport modes from the user trajectory. There are two basic types of methods, those that classify data during or after data collection. In both method groups, it is essential that the developed classifier is as accurate as possible and that it can classify as many transport modes as possible. In addition, a single user trajectory may include multiple connected transport modes that the user uses to complete the trip from the origin to the destination. Hence, such trajectories need to be segmented. Data from the mobile device sensors are often used to develop methods for classifying transport modes because mo-

mobile devices already have built-in sensors such as magnetometers, accelerometers, gyroscopes, Global Positioning System (GPS) and others. At the same time, it is possible to collect large datasets with mobile devices, given a large number of users.

Two challenges are still present in the applied methods. The first relates to the transport mode classification in real-time. Namely, solutions related to the classification of the entire trajectory, i.e. classification of historical data, are more common in the literature. As a consequence of the use of mobile phones for collecting traffic data, there is a need for real-time user feedback. Thus, there are requirements that the method for classifying transport modes must be adapted to work in real-time. The usefulness of real-time classification is manifested in the possibility of user confirmation of the accuracy of the collected data. On the other hand, users can be given real-time information about their trajectory. Another challenge relates to trajectory segmentation. Often methods for classifying transport modes include the assumption that one trajectory consists of only one transport mode, which is usually not true because user trajectories can contain several consecutive transport modes. For this reason, there is a need for trajectory segmentation methods to ensure that the part of the classified trajectory has only one transport mode.

## **Motivation**

The main motivation of this dissertation is to achieve a scientific contribution in the field of transport mode classification in parts that are not yet sufficiently researched in the literature. Accordingly, the main goal of the dissertation can be divided into two parts: (i) development of a method for segmentation and classification of the trajectory according to the used transport mode in real-time, and (ii) collecting a new dataset from the mobile phone sensors for validation of the proposed method.

So far, many methods have been developed for transport mode classification that use simple classifiers, ensemble classifiers, multilevel classification, or post-classification processing of results. However, the applied multilevel methods do not consider the classification of transport modes as a segmentation and classification process but implement the classification of transport modes at all their levels. Therefore, there is a need to develop a method that will recognize the point at which the user changes transport mode, meaning the point when one transport mode ends and the use of the following transport mode starts.

In addition, the datasets used to train and test methods for transport mode classification are usually not publicly available, as the authors test methods on their own datasets. The literature lists three datasets that have been made public and are suitable for trajectory segmentation and classification according to the transport mode. Two datasets contain data from mobile devices. Data from mobile devices may significantly differ depending on the user's behaviour while collecting data and the location where the mobile device was carried. For this reason, the

diversity of users and locations in the collected data is essential in the process of training the model. Both published datasets are relatively small in scope, at least those parts that have been made public, so there is a need in the field for new datasets that will complement existing ones and enable the validation of the newly developed methods.

## **Outline of the thesis**

This doctoral dissertation is organized into five chapters, and the first three provide the introduction and motivation for developing a multilevel method for real-time trajectory segmentation and classification. The fourth chapter describes the newly developed multilevel method for trajectory classification and segmentation, and presents the results achieved on the publicly available test dataset. The fifth chapter describes the developed mobile application for collecting data and newly collected dataset. In the same chapter, validation results for the multilevel method on the newly collected dataset are presented. Finally, the conclusion of the paper and topics for future work are given in the last chapter.

The first introductory chapter gives an overview of the area of urban mobility to which the problem of trajectory segmentation and classification according to transport mode belongs. It provides the motivation for this research and an overview of the main scientific contributions in this research.

The second chapter describes the problem of transport mode classification, emphasizing the classification of segmented and non-segmented trajectories. The most frequently used data are described, and an overview of public and private datasets used in the literature is given. Also, relevant features used in the literature to classify transport modes are presented.

The emphasis of the third chapter is on methods for solving the problem of transport mode classification. Basic trajectory segmentation methods from the supervised, unsupervised and semi-supervised methods are described. Also, an overview of data-driven methods for transport mode classification is given. At the end of the chapter, the structure of the test problem is described based on the Sussex-Huawei Locomotion (SHL) dataset used in this research for method training and testing. Accuracy is measured by the recall and precision ratio and the overall accuracy of the method.

The fourth chapter provides a detailed description of the multilevel method for real-time trajectory segmentation and classification. The method includes four basic components: (i) data preprocessing, (ii) method for trajectory segmentation, (iii) transport mode classification method, and (iv) rules in the multilevel method. During data preprocessing, the data are divided into time windows to simulate a real-time environment, and the features that are the input data for the multilevel method are calculated. In order to discover information from raw data, the discrete Fourier transform procedure was used to transform data from the time domain to the frequency domain. Transition state matrices are used to segment the trajectory. Processes of de-

velopment, testing, and reducing the required number of features in transition state matrices are described in detail in this chapter. Data discretization for transition state matrices development can be done in several ways. Three techniques were applied in this research: discretization by the method of equal spacing, equal frequency and  $k$  means method. The method for classifying transport modes is described through a random forest algorithm and an algorithm for hyperparameters optimization. Finally, the multilevel method for trajectory segmentation and classification is presented. At the end of the fourth chapter, the test results of the multilevel method on the test problem are presented.

Mobile application "Collecty" for collecting data from the mobile device's sensors, the basic structure and distribution of transport modes in the collected dataset are presented in chapter five. The mobile application contains introductory activities for login and reading the terms of use, the central part for transport mode selection, and the last part for trajectory validation by the user. The collected dataset contains 8 different transport modes and was collected by 15 users. One of the transport modes contained in the dataset is the electric scooter, for which no publicly available data are available in the literature. At the end of the fifth chapter, the validation results of the multilevel method for trajectory segmentation and classification on the collected dataset are presented.

In the final chapter, the basic conclusions of the doctoral dissertation are given, and the original scientific contributions achieved through the dissertation are highlighted. At the end of the conclusion chapter, a review of topics for future research is given.

## **Conclusion**

The main scientific contributions of this dissertation can be summarized in the following three statements:

1. Identification of features from the tracking user data in the time and frequency domain for trajectory segmentation according to the transport mode;
2. Development of a multilevel method for trajectory segmentation and classification according to transport mode in real-time;
3. Structuring a research dataset of mobile phone sensor data marked with multimodal characteristics suitable for recognizing users' activities and transport modes.

Through the achieved results of the dissertation, it can be concluded that the proposed multilevel method for trajectory segmentation and classification achieves better results in the most scenarios in the test problem. The validation results show that the method is spatially independent, and at the same time, this spatially independency is a new component in the test problem structure. The collected dataset is comparable in size to existing public datasets. In addition, the dataset contains data for the electric scooter, which are not represented in the published datasets so far.

**Keywords:** transport mode classification, trajectory segmentation, urban mobility, random forest, transition state matrices, machine learning, mobile phone sensors data, mobile application



# Sadržaj

<b>1. Uvod</b>	1
1.1. Motivacija istraživanja	3
1.2. Znanstveni doprinosi doktorske disertacije	4
1.3. Struktura doktorske disertacije	6
<b>2. Klasifikacija modova prijevoza u trajektoriji korisnika prometne mreže</b>	9
2.1. Klasifikacija trajektorije koja sadrži jedan mod prijevoza	11
2.1.1. Podatci o kretanju korisnika prometnom mrežom	11
2.1.2. Odabir relevantnih značajki	13
2.2. Klasifikacija trajektorije koja sadrži višestruke modove prijevoza	15
2.2.1. Segmentacija trajektorije	16
2.2.2. Klasifikacija modova prijevoza iz segmentirane trajektorije	17
<b>3. Metode za rješavanje problema klasifikacije modova prijevoza</b>	19
3.1. Metode za segmentaciju trajektorije	19
3.1.1. Nadzirane metode za segmentaciju trajektorije	20
3.1.2. Nenadzirane metode za segmentaciju trajektorije	21
3.1.3. Polunadzirane metode za segmentaciju trajektorije	24
3.2. Metode za klasifikaciju trajektorije	25
3.2.1. K najbližih susjeda	26
3.2.2. Stabla odluke	26
3.2.3. Neuronska mreža	29
3.2.4. Skriveni Markovljevi modeli	30
3.2.5. Višeslojna klasifikacija i združeni klasifikatori	31
3.3. Testni skup podataka	33
<b>4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu</b>	37
4.1. Obrada podataka	39
4.1.1. Podjela podataka u vremenske prozore	39
4.1.2. Preslikavanje podataka iz vremenske u frekvencijsku domenu	40



4.1.3.	Izračun relevantnih značajki . . . . .	42
4.2.	Segmentacija . . . . .	45
4.2.1.	Izrada matrica prijelaznih stanja . . . . .	46
4.2.2.	Diskretizacija podataka . . . . .	48
4.2.3.	Smanjenje broja relevantnih značajki . . . . .	50
4.2.4.	Testiranje metode . . . . .	52
4.3.	Klasifikacija . . . . .	53
4.3.1.	Odabir relevantnih značajki . . . . .	53
4.3.2.	Metoda slučajnih šuma . . . . .	55
4.3.3.	Ujednačavanje broja uzoraka . . . . .	56
4.3.4.	Optimizacija hiperparametara modela . . . . .	57
4.3.5.	Testiranje metode slučajne šume i podmodeli . . . . .	59
4.4.	Višerazinska metoda za segmentaciju i klasifikaciju trajektorije s obzirom na korišteni mod prijevoza . . . . .	59
4.5.	Rezultati . . . . .	62
4.5.1.	Rezultati obrade podataka . . . . .	62
4.5.2.	Rezultati metode za segmentaciju trajektorije . . . . .	67
4.5.3.	Rezultati metode za klasifikaciju trajektorije . . . . .	72
<b>5.</b>	<b>Validacija metode . . . . .</b>	<b>81</b>
5.1.	Mobilna aplikacija Collecty . . . . .	82
5.1.1.	Programski okvir <i>Sensors</i> . . . . .	82
5.1.2.	Klasa <i>LocationManager</i> . . . . .	84
5.1.3.	Struktura mobilne aplikacije . . . . .	85
5.1.4.	Pohrana podataka . . . . .	92
5.2.	Analiza istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama . . . . .	95
5.3.	Rezultati validacije . . . . .	103
5.3.1.	Rezultati validacije metode za segmentaciju i klasifikaciju trajektorije . . . . .	104
5.3.2.	Klasifikacija modova prijevoza u prikupljenom skupu podatka . . . . .	106
<b>6.</b>	<b>Zaključak . . . . .</b>	<b>109</b>
6.1.	Hipoteze i ostvareni znanstveni doprinosi istraživanja . . . . .	110
6.1.1.	Identifikacija značajki iz skupa podataka o kretanju korisnika prometnom mrežom u vremenskoj i frekvencijskoj domeni u svrhu segmentacije trajektorije s obzirom na mod prijevoza . . . . .	111
6.1.2.	Razvoj višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu . . . . .	111

6.1.3. Izrada istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama prikladnim za prepoznavanje aktivnosti i moda prijevoza korisnika . . . . .	112
6.2. Buduća istraživanja . . . . .	112
<b>Literatura . . . . .</b>	<b>115</b>
<b>Popis slika . . . . .</b>	<b>131</b>
<b>Popis tablica . . . . .</b>	<b>135</b>
<b>Popis algoritama . . . . .</b>	<b>137</b>
<b>Popis kratica . . . . .</b>	<b>139</b>
<b>Životopis . . . . .</b>	<b>143</b>



# Poglavlje 1

---

## Uvod

---

Prometna mreža predstavlja složen sustav za čiju su analizu potrebni podatci korisnika prometne mreže. Rastom stupnja urbanizacije i broja stanovnika, prijevoz postaje jedan od glavnih čimbenika koji utječe na kvalitetu života u velikim gradovima. Stoga se intenzivno istražuju područja poput urbane mobilnosti, potrošnje energije, zagađenja i sigurnosti, kao u istraživanju [1]. Sva ova područja su izravno ili neizravno povezana s mobilnošću korisnika prometne mreže. Znanje koje se može steći promatranjem modela ljudske mobilnosti kroz podatke prikupljene u različitim prometnim okruženjima može značajno doprinijeti rješavanju prometnih problema. Ti podatci mogu se prikupljati iz više izvora, a dijele se na dvije osnovne vrste: osjetila postavljena na fiksnim lokacijama na prometnoj mreži te pokretna osjetila. Ulogu pokretnih osjetila mogu imati satelitski sustavi za određivanje položaja, navođenje i određivanje vremena, npr. globalni sustav za satelitsku navigaciju (engl. *Global Navigation Satellite System*, GNSS), dugoročna evolucija mobilnih sustava četvrte generacije (engl. *Long Term Evolution forth Generation*, LTE 4G) ili raznovrsna druga osjetila s mobilnih uređaja. Također, mogu se koristiti podatci iz drugih izvora za praćenje kretanja korisnika, primjerice podatci s društvenih mreža.

Podatci koje generiraju korisnici pokretnih osjetila tijekom kretanja prometnim sustavom mogu se koristiti za prometne analize umjesto podataka dobivenih s fiksnih osjetila postavljenih na prometnoj mreži. Pritom je potreban različit pristup obradi podataka jer podatci prikupljeni s pokretnih osjetila ne sadrže informaciju o statičnoj lokaciji na prometnoj mreži već prikazuju niz slijednih lokacija prometnog entiteta u vremenu tzv. trajektoriju kako definiraju autori u [2]. U tom slučaju dubinskom analizom trajektorije putovanja, koja može biti podržana metodama grupiranja ili metodama klasifikacije, može se opisati mobilnost korisnika prometnom mrežom. Prepoznavanje svakodnevnih ljudskih aktivnosti tijekom različitih kretanja i korištenja prijevoznih sredstva ima brojne primjene poput podučavanja korisnika sa ciljem promjene obrasca

ponašanja ili poticanja zdravog načina života. Osim toga, aplikacije, koje prate mobilnost korisnika, mogu pomoći u planiranju javnog gradskog prijevoza, pametnom parkiranju i praćenju prometa vozila. Također, trajektorije unutar sebe sadrže informacije koje nisu izravno prikupljene s pokretnih osjetila, ali se mogu identificirati na temelju prikupljenih podataka. Jedna od dodatnih informacija je mod prijevoza kojim se korisnik služi na prometnoj mreži. Podatak o korištenom modu prijevoza daje novu vrijednost prikazu mobilnosti neke skupine korisnika jer, primjerice, umjesto prometne potražnje na nekom dijelu prometne mreže, može se opisati prometna potražnja za svaki mod prijevoza ili kako bi se detektirala potražnja korisnika za autobusnim prijevozom na nekoj relaciji jedna od mogućih analiza je analiza izbora moda putovanja, poput istraživanja [3].

Za detekciju moda prijevoza iz korisničke trajektorije koriste se metode za klasifikaciju trajektorije. Klasifikacija modova prijevoza je područje unutar polja prepoznavanja ljudskih aktivnosti (engl. *Human Activity Recognition*, HAR), pa se za rješenje ovog problema primjenjuju razne metode iz područja strojnog učenja koje se primjenjuju i za prepoznavanje ljudskih aktivnosti navode autori u [4]. Dvije su osnovne vrste metoda, one koje razvrstavaju podatke o korištenom modu prijevoza tijekom ili nakon prikupljanja podataka. Kod obje skupine metoda bitno je da je razvijeni klasifikator čim točniji te da ima mogućnost razvrstavanja što većeg broja modova prijevoza. Osim toga, jedna trajektorija korisnika može uključivati više povezanih modova prijevoza koje korisnik koristi kako bi završio putovanje od polazišta prema odredištu, pa je takve trajektorije potrebno segmentirati. U posljednje vrijeme često se koriste podatci s osjetila mobilnih uređaja za razvoj metoda za klasifikaciju modova prijevoza jer mobilni uređaji već imaju ugrađena osjetila poput magnetometra, akcelerometra, žiroskopa, GNSS-a i drugih, koji pokazuju dobre rezultate klasifikacije. Istovremeno je s mobilnim uređajima moguće prikupiti velike skupove podataka s obzirom na velik broj korisnika istih.

Dva su osnovna izazova koja su još uvijek prisutna u primijenjenim metodama. Prvi se odnosi na klasifikaciju modova prijevoza u stvarnom vremenu. Naime, u literaturi su zastupljenija rješenja koja se odnose na klasifikaciju cijele trajektorije, tj. klasifikaciju na povijesnim podacima. Međutim, prihvaćanjem mobilnih uređaja kao načina prikupljanja prometnih podataka razvila se potreba za stvarno vremenskom povratnom informacijom korisniku. Time se javljaju i zahtjevi da metoda za klasifikaciju modova prijevoza mora biti prilagođena za rad u stvarnom vremenu. Korisnost stvarno vremenske klasifikacije očituje se u mogućnosti korisničke potvrde o ispravnosti prikupljenih podataka, a sa druge strane korisnicima se mogu davati stvarnovremenske informacije o njihovoj trajektoriji. Drugi izazov odnosi se na segmentaciju korisničke trajektorije. Često metode za klasifikaciju modova prijevoza uključuju pretpostavku da se jedna trajektorija sastoji od isključivo jednog moda prijevoza što najčešće nije točno jer korisničke trajektorije mogu sadržavati više uzastopnih modova prijevoza. Iz tog razloga postoji potreba za metodama segmentacije trajektorije kojima bi se osiguralo da dio trajektorije koji se

klasificira pripada isključivo jednom modu prijevoza.

## 1.1 Motivacija istraživanja

Osnovna motivacija ove disertacije je postizanje znanstvenog doprinosa u području klasifikacije modova prijevoza u dijelovima koji nisu još uvijek dovoljno istraženi u literaturi. U skladu s time osnovni cilj disertacije može se podijeliti u dva dijela: (i) razvoj metode za klasifikaciju i segmentaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu, i (ii) prikupljanje skupa podataka s osjetila mobilnih uređaja koji će se koristiti za validaciju predložene metode.

Do sada su primijenjene različite metode za klasifikaciju modova prijevoza. Stabla odluke (engl. *Decision Tree*, DT) u istraživanju [5], algoritam k najbližih susjeda (engl. *k Nearest Neighbor*, kNN) u istraživanju [6], naivan Bayesov klasifikator (engl. *Naive Bayes*, NB) u istraživanju [7] te algoritam slučajnih šuma (engl. *Random Forest*, RF) u istraživanju [8] su najčešće korištene metode. Koriste se i metoda neizrazite logike (engl. *Fuzzy Logic*, FL) u istraživanju [9], metoda potpornih vektora (engl. *Support Vector Machine*, SVM) u istraživanju [10] te metode zasnovane na samostalnom određivanju značajki učenjem, kao što su neuronska mreža (engl. *Neural Network*, NN) u istraživanju [11], duboka neuronska mreža (engl. *Deep Neural Network*, DNN) u istraživanju [12] i konvolucijska neuronska mreža (engl. *Convolutional Neural Network*, CNN) kao vrsta primijenjene DNN u istraživanjima [13, 14]. Razvijeno je nekoliko pristupa koji koriste združene klasifikatore, višerazinske klasifikatore ili obradu rezultata nakon klasifikacije. Kod združenih klasifikatora najčešće se koriste AdaBoost, poput istraživanja [5] i RF u istraživanju [15] te se korištenjem takve kombinacije postižu bolji rezultati klasifikacije. Višerazinska klasifikacija se odnosi na dvije razdvojene faze odlučivanja kako bi se došlo do konačne odluke. Prvo se provodi odlučivanje kreće li se korisnik motoriziranim ili nemotoriziranim modom prijevoza. Ako je u prvoj fazi detektiran motorizirani mod prijevoza, u drugoj fazi se odlučuje o kojem se motoriziranom modu prijevoza radi poput istraživanja [16]. Obrada nakon klasifikacije odnosi se na proučavanje korelacije između susjednih vremenskih prozora kako navode autori u [16] ili primjenu skrivenih Markovljevih lanaca (engl. *Hidden Markov Model*, HMM) za dobivanje vjerojatnosti tranzicije iz jednog moda prijevoza u drugi kao u istraživanju [17]. Metode dubokog učenja koriste se za klasifikaciju modova prijevoza u kontekstu polunadziranog i nenadziranog učenja. Autori u [18] testirali su 3 polunadzirana algoritma i 6 nadziranih algoritama, a rezultati su pokazali da je jedan od polunadziranih algoritama postigao bolje rezultate od ostalih nadziranih algoritama. Duboka neuronska mreža je također korištena kao polunadzirana metoda učenja s minimalnim brojem označenih podataka u istraživanju [19]. Nenadzirana metoda koja uključuje kombinaciju grupiranja i dubokog konvolucijskog autoenkodera primijenili su autori u [20]. Metode koje su višerazinske ne razmatraju problem klasifikacije modova prijevoza kao segmentaciju i klasifikaciju trajektorije već

provode klasifikaciju modova prijevoza na svim svojim razinama. Stoga, postoji potreba za razvojem metode koja će osim klasifikacije modova prijevoza u korisničkoj trajektoriji prepoznati točke prestanka korištenja jednog moda prijevoza i početka korištenja sljedećeg moda prijevoza.

Podatci koji se koriste za učenje i testiranje metode za klasifikaciju modova prijevoza najčešće nisu javno objavljeni već autori testiraju metode na vlastitim skupovima podataka. U literaturi se navode tri testna skupa podataka koja su javno objavljena te su prikladna za klasifikaciju i segmentaciju trajektorije s obzirom na mod prijevoza. GeoLife je skup podataka koji sadrži zapise prikupljene s globalnog položajnog sustava (engl. *Global Positioning System*, GPS) koji su prikupljeni od 182 korisnika u razdoblju od više od tri godine. Skup podataka opisan je u istraživanju [21]. Preostala dva skupa podataka, Sussex-Huawei Locomotion (SHL), opisan u istraživanju [22] i Transport Mode Detection (TMD) skup podataka opisan u istraživanju [23] sadrže mjerenja s osjetila mobilnih uređaja, ali nisu u potpunosti označeni GPS lokacijama. Podatci s mobilnih uređaja mogu pokazivati bitno različito ponašanje u ovisnosti o korisniku koji podatke prikuplja i o lokaciji na kojoj je mobilni uređaj nošen. Zbog toga, bitna je raznolikost korisnika i lokacija u prikupljenim podacima. Oba objavljena skupa podataka, koja se odnose na podatke mobilnih uređaja, su relativno malog opsega, barem oni dijelovi koji su javno objavljeni, pa se javlja potreba u ovom istraživačkom području za novim skupovima podataka koji će dopuniti postojeće te omogućiti validaciju razvijenih metoda.

## 1.2 Znanstveni doprinosi doktorske disertacije

Prvi znanstveni doprinos doktorske disertacije odnosi se na identifikaciju značajki iz skupa podataka o kretanju korisnika prometnom mrežom u svrhu segmentacije trajektorije s obzirom na mod prijevoza. Cilj je razviti metodu za segmentaciju trajektorije koja će pomoću automatski određenih značajki detektirati vremenski trenutak u trajektoriji kada je korisnik prestao koristiti jedan mod prijevoza i počeo koristiti sljedeći mod prijevoza. Najbitniji dio ove metode je automatsko određivanje značajki koje će se koristiti za segmentaciju trajektorije te primjena metode u stvarnom vremenu. Naime, metode koje se trenutno koriste u literaturi za segmentaciju trajektorije uglavnom se odnose na segmentaciju povijesnih podataka trajektorije te se zasnivaju na nizu pravila za segmentaciju koja su iskustveno određena. Glavni cilj identifikacije značajki za segmentaciju trajektorije je priprema trajektorije za klasifikaciju s obzirom na mod prijevoza. Za ostvarenje ovog doprinosa iskorišten je SHL testni skup podataka.

Drugi znanstveni doprinos odnosi se na razvoj višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu. Ostvarenjem ovog znanstvenog doprinosa provjerava se prva hipoteza istraživanja kojom se pretpostavlja da se višerazinska metoda za klasifikaciju i segmentiranje trajektorije s obzirom na mod prijevoza može primijeniti za rješavanje problema klasifikacije šest modova prijevoza u stvarnom vre-

menu. Višerazinska metoda za klasifikaciju zasniva se na matricama prijelaznih stanja (engl. *Transition State Matrices*, TSM) i metodi RF. Pritom se TSM koriste za segmentaciju, odnosno za automatsku detekciju točke promjene moda prijevoza, a sama klasifikacija se provodi združenim klasifikatorom RF kako bi se uz više naučenih stabala odluke postigla što veća točnost. Za ostvarivanje stvarnovremenske komponente metode podatci prikupljeni od korisnika će se segmentirati u vremenske prozore veličine 30 s na kojima se metoda primjenjuje. Usporedba točnosti razvijene metode provest će se koristeći strukturirani testni problem na SHL skupu podataka. Svrha višerazinske metode nakon razvoja je korištenje metode tijekom prikupljanja podataka s mobilne aplikacije kako bi se podacima s osjetila automatski pridružio mod prijevoza.

Treći znanstveni doprinos obuhvaća izradu istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama. Ostvarenjem ovog znanstvenog doprinosa provjerava se druga hipoteza istraživanja kojom se pretpostavlja da rezultati testiranja i validacije predložene metode na javno objavljenom i prikupljenom skupu podataka pokazuju usporedivu točnost s rezultatima dostupnim u literaturi. Za stvaranje novog, označenog i uređenog podatkovnog skupa za provođenje istraživanja u području segmentacije i klasifikacije trajektorija koristiti će se aplikacija za mobilne uređaje Collecty koja je razvijena u sklopu ovog istraživanja. Svrha aplikacije je praćenje i pohrana podataka s osjetila mobilnog uređaja tijekom kretanja korisnika prometnom mrežom. Praćeni korisnici unose pomoću aplikacije precizno i točno podatke o svom modu prijevoza, a aplikacija automatski prikuplja podatke s osjetila mobilnog uređaja. Podatci se pohranjuju lokalno na mobilnom uređaju ili se automatski šalju na poslužitelj. Skup podataka je oblikovan jedinstvenim značajkama te pohranjen na dva načina: u obliku relacijske baze podataka i tekstualne datoteke.

Za ostvarenje znanstvenih doprinosa koristit će se metode eksperimentalnog mjerenja, statističke analize, računalne analize te sinteze.

Temeljem prethodnog objašnjenja, osnovni znanstveni doprinosi ove disertacije mogu se sažeti u sljedeća tri navoda:

1. Identifikacija značajki iz skupa podataka o kretanju korisnika prometnom mrežom u vremenskoj i frekvencijskoj domeni u svrhu segmentacije trajektorije s obzirom na mod prijevoza;
2. Razvoj višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu;
3. Izrada istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama prikladnim za prepoznavanje aktivnosti i moda prijevoza korisnika.



## 1.3 Struktura doktorske disertacije

Doktorska disertacija organizirana je u pet poglavlja, od kojih prva tri daju uvod i motivaciju za razvoj metode za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu. U četvrtom poglavlju opisana je višerazinska metoda za klasifikaciju i segmentaciju trajektorije te su prikazani ostvareni rezultati na testnom skupu podataka. Peto poglavlje daje opis podataka i načina prikupljanja istih te su prikazani rezultati validacije metode za segmentaciju i klasifikaciju trajektorije. Posljednje poglavlje donosi zaključak o provedenom istraživanju i smjernice za buduća istraživanja.

U poglavlju 1 dan je osvrt na područje urbane mobilnosti kojem pripada problem segmentacije i klasifikacije trajektorije s obzirom na mod prijevoza. U uvodnom poglavlju je navedena motivacija ovog istraživanja kao i pregled osnovnih znanstvenih doprinosa ostvarenih u istraživanju.

Poglavlje 2 opisuje područje klasifikacije modova prijevoza s posebnim naglaskom na klasifikaciju segmentiranih i nesegmentiranih trajektorija. Opisani su najčešće korišteni podatci te je dan pregled javnih i privatnih skupova podataka korištenih u literaturi. Prikazane su relevantne značajke koje su korištene u literaturi za klasifikaciju modova prijevoza. Zatim je prikazan detaljan opis klasifikacije trajektorije koja uključuje višestruke modove prijevoza tako definirajući segmentaciju trajektorije s obzirom na mod prijevoza.

Naglasak poglavlja 3 je na metodama za rješavanje problema klasifikacije modova prijevoza. Opisane su osnovne metode za segmentaciju trajektorije iz skupine nadziranih, nenadziranih te polunadziranih metoda. Također dan je pregled podatkovno vođenih metoda za klasifikaciju modova prijevoza te zastupljenost pojedinih metoda u literaturi. Na kraju poglavlja opisana je struktura testnog problema koji je zasnovan na SHL skupu podataka na kojem je ispitana točnost razvijene metode. Točnost se mjeri omjerom odziva i preciznosti te ukupnom točnošću metode.

U poglavlju 4 dan je detaljan opis višerazinske metode za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu. Metoda uključuje četiri osnovne komponente: (i) obradu podataka, (ii) metodu za segmentaciju trajektorije, (iii) metodu za klasifikaciju modova prijevoza, te (iv) višerazinsku metodu za segmentaciju i klasifikaciju modova prijevoza. Tijekom obrade podataka sami podatci su podijeljeni u vremenske prozore kako bi se ostvarila stvarnovremenska komponenta metode te su izračunate značajke koje su ulazni podatci višerazinske metode. U svrhu otkrivanja novih informacija iz sirovih podataka korišten je postupak diskretne Fourierove transformacije (engl. *Discrete Fourier Transform*, DTF) kako bi se prikazala frekvencijska domena podataka. Za segmentaciju trajektorije koriste se TSM-ovi čija su izrada, testiranje te postupak smanjenja potrebnog broja značajki detaljno opisani u ovom poglavlju. Diskretizacija podataka u svrhu izrade TSM-a može se provesti na više načina. U radu su primijenjene tri teh-

nike: diskretizacija metodom jednakog razmaka, metodom jednake frekvencije te metodom  $k$  sredine. Metoda za klasifikaciju modova prijevoza opisana je kroz algoritam za izradu slučajne šume te algoritam za optimizaciju hiperparametara metode. Naposljetku je prikazan algoritam višerazinske metode za segmentaciju i klasifikaciju trajektorije koji obuhvaća metodu za segmentaciju i klasifikaciju trajektorije, a njegov izlaz je vremenski prozor s pridruženim modom prijevoza. Na kraju četvrtog poglavlja uspoređeni su rezultati testiranja ostvarene metode na testnom problemu gdje se promatra točnost metode uzimajući u obzir korisničku, lokacijsku i vremensku nezavisnost metode.

Poglavlje 5 opisuje mobilnu aplikaciju Collecty za prikupljanje podataka s osjetila mobilnog uređaja te osnovnu strukturu i distribuciju modova prijevoza u prikupljenom skupu podataka. Mobilna aplikacija sastoji se od uvodnih aktivnosti koje su namijenjene prijavi i upoznavanju korisnika s aplikacijom, središnjeg dijela pomoću kojeg korisnik odabire mod prijevoza te zadnjeg dijela koji se koristi za potvrdu ispravnosti rute od strane korisnika. Prikupljeni podatci sadrže 8 različitih modova prijevoza koje je prikupljalo 15 korisnika. Jedan od modova prijevoza sadržanih u skupu podataka je električni romobil za koji nema dostupnih javno objavljenih podataka u literaturi. Na kraju petog poglavlja prikazani su validacijski rezultati višerazinske metode za segmentaciju i klasifikaciju trajektorije na prikupljenom skupu podataka.

U završnom poglavlju 6, dani su osnovni zaključci doktorske disertacije te su istaknuti izvorni znanstveni doprinosi koji su ostvareni kroz rad. Na kraju poglavlja, dan je osvrt na teme za buduće istraživanje u području klasifikacije modova prijevoza u stvarnom vremenu.



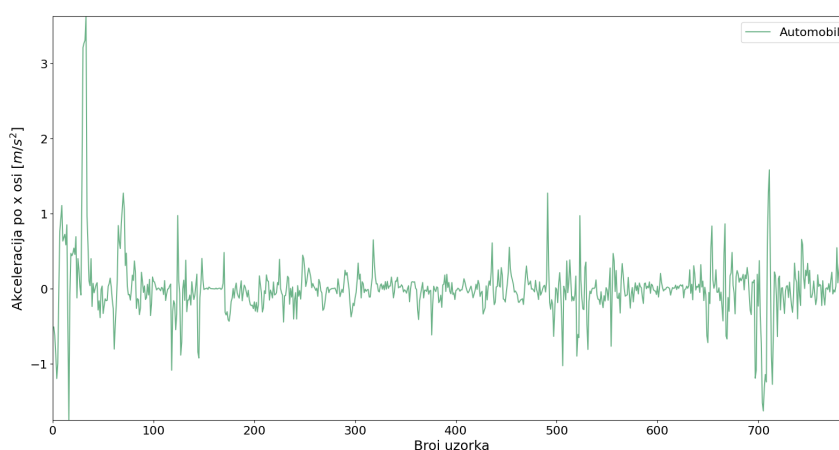
# Klasifikacija modova prijevoza u trajektoriji korisnika prometne mreže

Podatci o mobilnosti korisnika prometnom mrežom prikupljaju se raznovrsnim osjetilima. Neka osjetila, poput brojila prometa, postavljena su fiksno na prometnoj mreži, dok osjetila poput mobilnih uređaja nemaju primarnu ulogu prikupljanja prometnih podataka, ali se mogu iskoristiti u tu svrhu. Često se za prometne analize koristi podatak o korištenom modu prijevoza na nekom dijelu prometne mreže, a ta informacija nije sadržana u sirovim podacima. Dugo se u praksi primjenjuje metoda prikupljanja podataka o putovanju korisnika prometne mreže ispitivanjem korisnika kroz anketne obrasce o dnevnim migracijama. Ovaj pristup prikupljanja podataka uključuje i informaciju o korištenom modu prijevoza. Kroz godine je ovaj pristup pokazao niz nedostataka, poput netočnih informacija zbog lošeg sjećanja korisnika ili nedovoljni odaziv korisnika koji žele dati informacije o svojim dnevnim migracijama. Iz tog razloga se počinju razvijati metode za automatizaciju prikupljanja dnevnika putovanja.

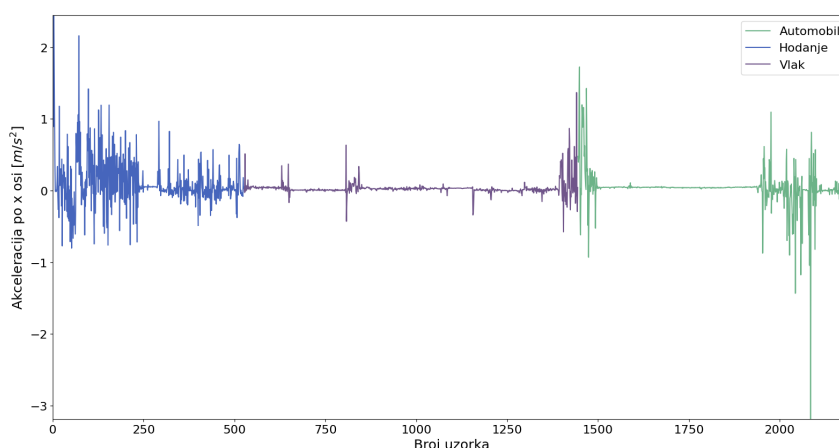
Prva istraživanja koja proučavaju automatsko prikupljanje dnevnika putovanja su opisana u istraživanjima [24, 25, 26]. Podatci su prikupljeni GPS prijammikom ugrađenim u vozilo, što isključuje velik broj modova prijevoza. Stoga se istraživanja usmjeravaju prema prikupljanju podataka uređajima izvan vozila. Jedno od prvih istraživanja tog tipa provode autori u [27] pri čemu koriste GPS prijammike nošene u ruci. Naglasak istraživanja je, obzirom na težinu tada dostupnih uređaja, na teretu nošenja uređaja te pogrešci uređaja. Kasnije se razvijaju prilagođene platforme za prikupljanje podataka kretanja korisnika prometnom mrežom, ali se ne razmatraju metode klasifikacije poput istraživanja [28, 29, 30, 31, 32, 33]. Nakon toga su se istraživanja usmjerila ka razvoju metoda klasifikacije trajektorije.

Klasifikacija modova prijevoza može se podijeliti na dva osnovna zadatka: (i) određivanje

postoji li kretanje korisnika, te (ii) u slučaju da se korisnik kreće, detekcija prijevoznog sredstva koje korisnik koristi za kretanje. U svrhu razlikovanja dvije osnovne vrste trajektorija u kontekstu korištenih modova prijevoza na Slici 2.1 prikazana su mjerenja akceleracije po  $x$  osi gibanja korisnika ( $y$  os grafa) u vremensko poredanom broju uzorka ( $x$  os grafa). Na Slici 2.1a prikazane su očitane vrijednosti s akcelerometra tijekom vožnje automobilom, dok su na Slici 2.1b prikazane očitane vrijednosti tijekom kretanja trima uzastopnim modovima prijevoza: hodanje, vlak i automobil. Prema tome, trajektorija korisnika može sadržavati jedan (Slika 2.1a) ili više uzastopnih modova prijevoza (Slika 2.1b) te su potrebni različiti postupci obrade podataka trajektorije za svaku vrstu trajektorija. Stoga će se u ovom radu razmatrati dva odvojena problema: (i) klasifikacija trajektorije koja sadrži jedan mod prijevoza, i (ii) klasifikacija trajektorije koja uključuje višestruke modove prijevoza.



(a) Trajektorija korisnika s jednim modom prijevoza



(b) Trajektorija korisnika s više modova prijevoza

**Slika 2.1:** Vrijednosti mjerenja akceleracije u smjeru  $x$  osi kretanja korisnika a) jednim ili b) više modova prijevoza

## 2.1 Klasifikacija trajektorije koja sadrži jedan mod prijevoza

Na trajektoriju koja sadrži jedan mod prijevoza primjenjuju se dva osnovna procesa: (i) odabir relevantnih značajki, te (ii) primjena odabrane metode klasifikacije.

Osnovni cilj odabira relevantnih značajki je što realističniji opis ponašanja uzorka, koji se klasificira, kako bi klasifikator strukturom što točnije opisivao klase obuhvaćene u klasifikaciji. U mnogim istraživanjima je prihvaćeno da su značajke zasnovane na brzini vrlo bitne za klasifikaciju modova prijevoza kako navode autori u [34, 35]. Osim brzine, vrlo često se uzima u obzir i akceleracija kako je opisano u istraživanju [36]. Naravno, značajke koje se uzimaju u obzir za klasifikaciju ovise i o tehnologiji koja se koristi za prikupljanje podataka.

### 2.1.1 Podatci o kretanju korisnika prometnom mrežom

U Tablici 2.1 prikazane su tri skupine najčešće korištenih podataka za klasifikaciju: lokacijski podatci u obliku GNSS točaka tijekom kretanja korisnika (oznaka L), podatci s osjetila mobilnih uređaja (oznaka OMU) te kombinacija prva dva skupa podataka (oznaka H). GNSS podatci mogu također biti prikupljeni mobilnim uređajem ili bilo kojim drugim uređajem koji se koristi za prikupljanje korisničkih GNSS točaka. U istraživanjima je najčešće korišten vlastiti skup podataka. Ipak, neka istraživanja koriste javno objavljene skupove podataka od kojih su najpoznatiji:

- GeoLife skup podataka opisan je u članku [21], a sastoji se od 9043 trajektorija pri čemu su prikupljeni GPS podatci,
- SHL skup podataka kojeg čine 2800 sati prikupljenih podataka kretanja korisnika sa 16 osjetila mobilnog uređaja, a opisuju ga autori u [22],
- TMD skup podataka kojeg čini 31 sat podataka sa 23 osjetila mobilnog uređaja tijekom kretanja korisnika prometnom mrežom, opisan je u članku [23].

Istraživanja koja uključuju lokacijske informacije uglavnom koriste GeoLife skup podataka ili privatne skupove podataka koji nisu objavljeni istraživačkoj zajednici. Podatci koji se koriste kod te skupine su GPS podatci koji uključuju niz zapisa o geografskoj širini i dužini kretanja korisnika. Iz tog tipa podataka najčešće se izračunavaju veličine poput prijeđene udaljenosti i brzine. Dio istraživanja GPS podatke dopunjuje geoprostornim podacima kao npr. podatci geografskog informacijskog sustava (engl. *Geographic Information System*, GIS) i obrađuje u programskom okruženju geoinformacijskog sustava kako bi se povećala točnost klasifikacije modova prijevoza koje karakteriziraju fiksne lokacije na mreži, kako je opisano u istraživanju [57]. Geoprostorni podatci koji se koriste u klasifikaciji modova prijevoza najčešće uključuju značajke poput udaljenosti do stanice autobusa i stanice vlaka, kako primje-

## 2. Klasifikacija modova prijevoza u trajektoriji korisnika prometne mreže

**Tablica 2.1:** Istraživanja u području klasifikacije modova prijevoza na povijesnim podatcima

Vrsta podataka	Skup podataka				Modovi prijevoza	Referenca	
	Dostupnost	Veličina uzorka	Osjetila	Broj korisnika			Uređaj
L	Javni skup podataka GeoLife	7112 trajektorija	GPS	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus i automobil	[37]
		17621 trajektorija	GPS	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus, automobil, taksi, vlak i podzemna željeznica	[5]
		23062 trajektorije	GPS	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus, automobil, taksi, vlak i podzemna željeznica	[38]
		17621 trajektorija	GPS	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus, automobil, vlak	[13]
		17621 trajektorija	GPS	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus, automobil, vlak	[20]
		25955 segmenata	GPS i akceleracija	182	Mobilni uređaj i GPS zapisnik podataka	Hodanje, bicikl, autobus, automobil, vlak	[14]
	Privatni skup podataka	4685 trajektorija	GPS	30	Mobilni uređaj	Hodanje, bicikl, e-bicikl, autobus i automobil	[39]
		6,2 sati	GPS, GIS	6	Mobilni uređaj	Stajanje, hodanje, bicikl, autobus, automobil i vlak	[40]
		30000 trajektorija	GPS, GIS	8000	Mobilni uređaj i digitalna karta	Hodanje, bicikl, autobus, automobil i vlak	[41]
		900 sati	GPS	81	Mobilni uređaj	Hodanje, bicikl, autobus, automobil, vlak i podzemna željeznica	[33]
		340 trajektorija	GPS, GIS	63	GPS zapisnik podataka	Hodanje, autobus, automobil, vlak i podzemna željeznica	[42]
114 sati		GPS	4	Mobilni uređaj	Hodanje, autobus i automobil	[11]	
OMU	Javni skup podataka SHL	498 sati	GPS, WiFi	3	Mobilni uređaj i kamera	Hodanje, trčanje, bicikl, autobus, automobil, vlak i podzemna željeznica	[43]
		470 sati	Akc, Mag, Žir	3	Mobilni uređaj i kamera	Hodanje, trčanje, bicikl, autobus, automobil, vlak i podzemna željeznica	[44]
	Javni skup podataka TMD	31 sat	Akc, Mag, Žir, Zvuk, Svjetlina	13	Mobilni uređaj	Stajanje, hodanje, autobus, vlak i automobil	[23]
	Privatni skup podataka	1249 sati	Akc, Mag, Žir	45	Mobilni uređaj	Stajanje, hodanje, trčanje, bicikl, motor, autobus, automobil, vlak i metro	[16, 45, 46]
		12 sati	Akc, Mag, Žir	7	Mobilni uređaj	Hodanje, bicikl, autobus, automobil i vlak	[47]
		25 sati	Akc, Mag, Žir	10	Mobilni uređaj	Hodanje, bicikl, autobus, automobil, vlak i podzemna željeznica	[48]
		150 sati	Akc	16	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[49]
		3 sata	Akc, Mag, Žir, Bar	5	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[50]
		4 sata	Akc	6	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[51]
		30 sati	Akc	15	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[52]
		890 sati	Akc	46	GPS i akcelerometar zapisnik podataka	Hodanje, autobus, automobil, vlak i podzemna željeznica	[15]
		2,5 sata	Akc	4	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[53]
		9 sati	Akc	5	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[36]
3 sata	Akc, Mag, Žir	/	Mobilni uređaj	Hodanje, autobus, automobil, vlak i podzemna željeznica	[54]		
H	Privatni skup podataka	120 sati	GPS, Akc	16	Mobilni uređaj	Hodanje, trčanje, bicikl, vozilo	[34]
		355 sati	GPS, Akc	15	Mobilni uređaj	Hodanje, bicikl, motor, autobus, automobil, vlak, tramvaj i podzemna željeznica	[55]
		266 sati	GPS, Akc	15	Mobilni uređaj	Stajanje, hodanje, bicikl, motor, autobus, automobil, vlak, tramvaj i podzemna željeznica	[56]

njuju autori u [40] ili dopunu podataka za podzemnu željeznicu gdje nema GPS signala, poput istraživanja [42]. Druga skupina podataka odnosi se na mjerenja s osjetila mobilnih uređaja

tijekom kretanja korisnika prometnom mrežom. Ova skupina podataka koristi se u HAR području kojem pripada problem prepoznavanja moda prijevoza u korisničkoj trajektoriji. Najčešće se koriste mjerenja s osjetila akcelerometra, žiroskopa i magnetometra. I u ovoj skupini su najčešće zastupljeni privatni skupovi podataka, međutim dostupni su javni SHL i TMD skupovi podataka. Zadnja skupina podataka uključuje kombinaciju lokacijskih podataka i podataka s osjetila mobilnih uređaja. Akcelerometar i GPS podatci su kombinacija koja ima značajniju pojavnost u ovoj skupini. Broj modova prijevoza uključenih u klasifikaciju, razmatran u dostupnoj literaturi, nije uniformiran i varira između dva i devet. Klasifikacija dva moda prijevoza najčešće uključuje klasifikaciju na motorizirane i ne motorizirane modove prijevoza, poput istraživanja [16, 49, 50], dok klasifikacija više modova prijevoza uključuje razlikovanje između motoriziranih modova prijevoza.

### 2.1.2 Odabir relevantnih značajki

Odabir relevantnih značajki jedan je od glavnih i bitnijih dijelova pretprocesiranja podataka. Tehnike odabira značajki koriste se za smanjenje broja ulaznih varijabli eliminiranjem suvišnih značajki na one koje su najrelevantnije za model strojnog učenja. Time se poboljšavaju performanse prediktivnog modela te se smanjuje vrijeme potrebno za učenje modela.

U Tablici 2.2 prikazane su značajke koje se najčešće odabiru u području klasifikacije modova prijevoza. Podatci s mobilnih uređaja mogu se interpretirati kao signal u vremenskoj domeni, pa se nad tim podacima često primjenjuju metode preslikavanja podataka kako bi se podatci preslikali iz vremenske u frekvencijsku domenu, opisuju autori u [58, 59]. Podatci koji se prikupljaju s mobilnog uređaja sadrže podatke s osjetila poput akcelerometra, magnetometra i žiroskopa. U definiranim vremenskim trenutcima bilježi se promjena mjerenja na osjetilu u tri smjera:  $x$ ,  $y$  i  $z$  osi. S obzirom na to da je orijentacija mobilnog uređaja promjenjiva (uređaj ne mora nužno biti usmjeren prema smjeru kretanja) onda su očitavanja po smjerovima osjetljiva na promjene vrijednosti, pa se često koristi magnituda nekog osjetila umjesto podataka po pojedinom smjeru, kao u istraživanjima [60, 61, 62]. Za takav oblik podataka u vremenskoj domeni često se izračunavaju uobičajene statističke mjere poput srednje vrijednosti, standardne devijacije, moda, medijana, maksimalne i minimalne vrijednosti kako navode autori u [5, 63]. Također često se koriste mjere koje su povezane s vrijednostima ispod ili iznad odabrane kvartalne vrijednosti ili međukvartalni raspon te mjere zaobljenosti i simetrije koje opisuju distribuciju podataka u promatranom vremenskom intervalu.

Prednost GPS podataka je što neposredno sadrže podatke o brzini i prijeđenoj udaljenosti, pa se skup značajki može nadopuniti podacima poput ukupne prijeđene udaljenosti, srednje brzine u nekom intervalu ili stopi promjene smjera kretanja. Time se može utjecati na sličnosti ponašanja modova prijevoza. Primjerice, zbog promjenljivog ponašanja brzine prometnog toka tijekom dana, kako navode autori u [64], vozilo može imati jednaku maksimalnu brzinu kao



pješak ili biciklist tijekom kraćeg vremenskog intervala. Jedno rješenje problema je uvođenje novih značajki poput prvih pet najvećih vrijednosti brzine umjesto maksimalne brzine u nekom vremenskom intervalu. U domeni GPS podataka autori u [65] prvi uvode osnovne značajke poput prijedene udaljenosti, brzine i akceleracije. Kasnije, autori u [66] uvode dodatne značajke: stupanj promjene brzine, stupanj promjene smjera te stupanj stajanja u ruti.

**Tablica 2.2:** Značajke u području klasifikacije modova prijevoza

Domena podataka	Vrsta podataka	Značajka
Vremenska	Magnituda akceleracije, žiroskopa, magnetometra i barometra, 3 dimenzije akcelerometra	Srednja vrijednost
		Standardna devijacija
		Stopa prelaska nule
		Energija
		Korelacija
		Mjere zaobljenosti i simetrije podataka
		Minimalna i maksimalna vrijednost
		Međukvartilni raspon
		Treći kvartil
		Kvartil (5, 25, 50, 75, 90)
	GPS podatci (brzina, akceleracija, udaljenost, kut skretanja)	Srednja vrijednost
		Standardna devijacija
		Maksimalna i minimalna vrijednost
		$N$ najmanjih/najvećih vrijednosti
		Interkvartilni raspon
		Stopa promjene brzine
		Stopa stajanja
		Stopa promjene smjera
		Frekvencijska
Standardna devijacija		
Medijan energije		
Frekvencija s najvećom amplitudom		
Stopa između dvije najveće amplitude		
Vektor amplituda		
Zbroj amplitude između 0-2 Hz		
Stopa energije između frekvencijskih pojasa različite širine		
Amplituda za frekvenciju u rasponu od 1-10 Hz		

Frekvencijska domena podataka sadrži dvije osnovne komponente: raspon frekvencije u nekom vremenskom intervalu te pripadajuće energije (amplitude) za svaku vrijednost frekvencije. Stoga su i značajke, koje se povezuju s frekvencijskom domenom, orijentirane na frekvenciju signala ili na energiju signala. Jednako kao u vremenskoj domeni, izračunavaju se osnovne statističke veličine poput srednje vrijednosti, standardne devijacije i medijana. U istraživanjima

se navodi da specifična širina frekvencijskog pojasa daje bolji uvid u kretanje korisnika u usporedbi s cijelom širinom frekvencijskog pojasa. Tako se značajke stope energije i ukupne energije često referiraju upravo na specifičnu širinu frekvencijskog pojasa. Najčešće se koristi raspon od 1-10 Hz.

U literaturi postoji i niz drugih značajki i podataka koji uglavnom uključuju dodatne podatke. Primjerice autori u [63] razmatraju dodatne značajke poput prosječne temperature i temperature zraka na početku i kraju putovanja koji zahtijevaju informaciju o meteorološkom stanju. Autori uključuju značajke tog tipa jer razmatraju i svrhu putovanja. Dio istraživanja uključuje demografske podatke (spol, godine, vozačka dozvola i slično) kao i geoprostorne podatke kao autori u [42]. Osim uobičajenih statističkih mjera autori u [17] koriste i značajke poput vlasnik bicikla, automobila ili motora (binarna varijabla s vrijednostima *Da* ili *Ne*), broj dostupnih satelita te pogrešku uzorkovanja po prijeđenoj udaljenosti. Autori u [67] koriste značajke promjene kuta u odnosu na sjever, pri čemu promatraju promjenu kuta u vremenu te učestalost opisanih promjena. Autori u [68] promatraju distribuciju značajki za detekciju značajnijih promjena u uzorcima (spektralna entropija). Isti autori u radu [69] promatraju četiri skupa atributa s različitih osjetila: akcelerometra, žiroskopa, GPS-a te vektor rotacije, s izračunatim osnovnim statističkim mjerama u sve tri osi. Stupanj promjene akceleracije prvi put koriste autori u [13]. U istraživanjima se koriste dodatne značajke poput temperature tijela, otkucaja srca te intenziteta svjetlosti [70].

Kako bi se izbjegao gubitak podataka izostavljanjem bitnih značajki, dio istraživanja koristi sirove podatke za klasifikaciju pri čemu se metodama dubokog učenja automatski izračunavaju značajke iz sirovih podataka trajektorije navode autori u [38].

Većina istraživanja kod razvoja klasifikatora dijeli prikupljeni skup podataka na dva dijela: skup podataka za učenje te skup podataka za testiranje klasifikatora. Formiranje skupa podataka za učenje i testiranje utječe na rezultate klasifikatora za detekciju moda prijevoza. Podjela podataka svih praćenih korisnika na slučajan način daje znatno veću točnost nego korištenje skupa podataka od jednog dijela korisnika za učenje, a drugog dijela korisnika za testiranje kako navode autori u [71]. Metode klasifikacije primijenjene u literaturi su raznolike te dolaze iz sve tri grupe metoda strojnog učenja: nadzirano, nenadzirano te polunadzirano učenje.

## 2.2 Klasifikacija trajektorije koja sadrži višestruke modove prijevoza

U prethodnim poglavljima objašnjen je postupak obrade trajektorije kada ona sadrži samo jedan mod prijevoza. No, u stvarnosti jedna trajektorija korisnika može uključivati više povezanih modova prijevoza koje korisnik koristi kako bi završio putovanje od polazišta prema odredištu. U tom slučaju nije moguće klasificirati trajektorije na uobičajen način jer je izlaz klasifikatora

uvijek jedan mod prijevoza. Štoviše, ako se trajektorije s višestrukim modovima prijevoza pret hodno ne obrađuju, klasifikator ne može generirati pouzdana predviđanja s visokom točnošću. Stoga se proces klasifikacije modova prijevoza dijeli na dva osnovna zadatka:

- **Segmentacija trajektorije** - grupiranje homogenih dijelova trajektorije zasnovano na odabranom kriteriju. Ako je kriterij segmentacije trajektorije mod prijevoza, onda se segmentacija trajektorije definira kao podjela trajektorije na segmente koji sadrže samo jedan mod prijevoza;
- **Klasifikacija moda prijevoza u segmentiranoj trajektoriji** - prepoznavanje korištenog moda prijevoza u jednom segmentu trajektorije.

### 2.2.1 Segmentacija trajektorije

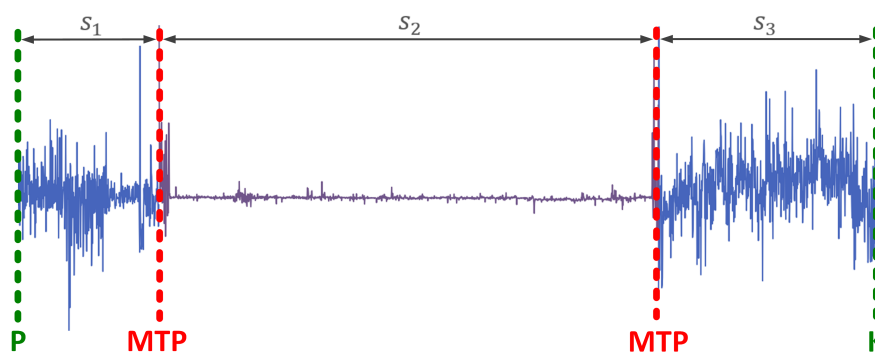
Jedan od temeljnih zadataka obrade trajektorija je segmentacija, tj. podjela neobrađene putanje na segmente. Naziva ju se temeljnim zadatkom budući da točnost segmentacije utječe na kvalitetnije značajke koje bolje predstavljaju ponašanje pokretnog objekta. Dakle zadatak segmentacije je razlikovati homogene ili slične dijelove putanje na osnovu zadanih kriterija.

Segmentacija trajektorije primjenjuje se u mnogim područjima te su raznovrsni razlozi za rješavanje ovog problema. Često se primjenjuju metode segmentacije trajektorije radi pripreme ulaznih podataka za algoritme koji zahtijevaju jedinstvenu strukturu podataka, kao što je neuronska mreža u istraživanju [13]. Razlog segmentacije također može biti maksimiziranje homogenosti podataka koji pripadaju jednom segmentu, kao u istraživanju [72], pronalaženje kritičnih točaka u trajektoriji u istraživanju [73] ili pronalaženje uzoraka u trajektoriji u istraživanju [74].

U kontekstu moda prijevoza, svrha segmentacije trajektorije je da se na trajektoriji  $T$  detektira točka promjene moda prijevoza (engl. *Mode Transfer Point*, MTP) i segment  $S$  u kojem je korišten jedan mod prijevoza, uz sljedeće pretpostavke problema:

- (i) Trajektorija  $T$  je skup uzastopnih zapisa s osjetila uređaja za praćenje kretanja korisnika od početka kretanja korisnika (točka  $P$ ) do završetka kretanja korisnika (točka  $K$ ). Svaki zapis u trajektoriji je skup od više podataka koji opisuju mjerenja s dostupnih osjetila u datom trenutku.
- (ii) MTP je lokacija u skupu zapisa na kojoj je korisnik promijenio mod prijevoza.
- (iii) Segment  $S$  predstavlja jedan dio trajektorije  $T$  u kojem je korišten isključivo jedan mod prijevoza.

Na Slici 2.2 je prikazan primjer jedne trajektorije korisnika koja se sastoji od tri uzastopna moda prijevoza (različita boja krivulje predstavlja različite modove prijevoza). Zelena isprekidana linija prikazuje početak ( $P$ ) i završetak ( $K$ ) trajektorije. Zadatak segmentacije je pronaći MTP-ove (crvene isprekidane linije) kako bi se trajektorija korisnika podijelila na segmente ( $s_1$ ,  $s_2$  i  $s_3$ ) koji sadrže dio trajektorije u kojoj je korišten samo jedan mod prijevoza.



**Slika 2.2:** Mjerenja s osjetila za jednu trajektoriju korisnika s više uzastopnih modova prijevoza

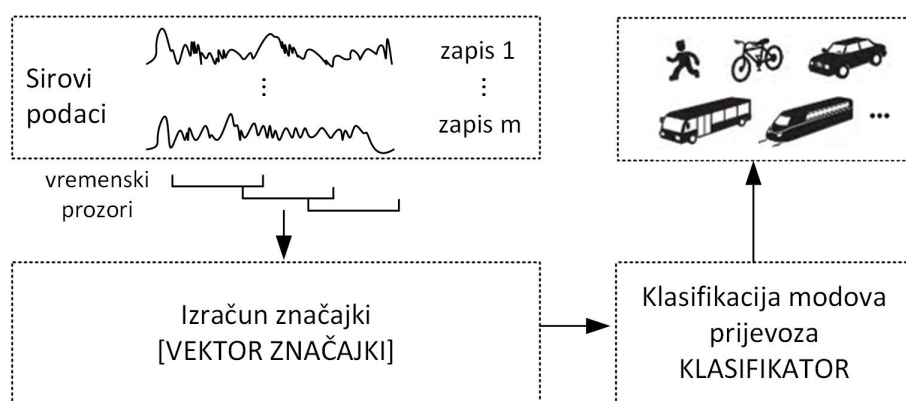
Općenito se segmentacija trajektorije može provoditi na dva načina: (i) segmentacija koja je zasnovana na vremenskom trenutku promjene moda prijevoza, tj. detekcija točke (engl. *Point-Based Classification*, PBC) ili (ii) segmentacija koja detektira dio trajektorije u kojoj se promijenio mod prijevoza, tj. detekcija segmenta (engl. *Segment-Based Classification*, SBC). PBC je skupina metoda koja se povezuje s podjelom vremenske serije podataka u vremenske prozore pri čemu se uspoređuju podatci unutar jednog vremenskog prozora sa susjednim vremenskim prozorom. Uglavnom se izračunava sličnost između dva vremenska prozora te se tako donosi odluka da li je u promatranom trenutku došlo do promjene moda prijevoza ili nije. Skupina metoda SBC, obično proučava cijelu trajektoriju te ju segmentira na temelju sličnosti njezinih dijelova. Prema tome, PBC se obično koristi za klasifikaciju koja se provodi u stvarnom vremenu ili približno stvarnom vremenu, dok su SBC metode prilagođenije za segmentaciju povijesnih podataka. U ovom istraživanju naglasak je na PBC metodama jer se klasifikacija trajektorije provodi u stvarnom vremenu.

### 2.2.2 Klasifikacija modova prijevoza iz segmentirane trajektorije

Ako se razmatraju trajektorije koje sadrže više uzastopnih modova prijevoza onda se segmentacijom trajektorije osigurava da se klasificira segment trajektorije koji sadrži jedan mod prijevoza. Time je omogućeno da se klasificira trajektorija prema modu prijevoza u stvarnom vremenu. Ako se klasifikacija provodi na povijesnim podacima, postupak je jednak onom opisanom u Poglavlju 2.1.

Potreba za klasifikacijom modova prijevoza u stvarnom vremenu javlja se kao tendencija automatizacije prikupljanja dnevnika putovanja. U tu svrhu mogu se izraditi aplikacije na mobilnim uređajima koje u pozadini prikupljaju podatke s GPS-a ili akcelerometra ili oboje, a na grafičkom sučelju korisnik može odabrati trenutni mod prijevoza. Tako su razvijene aplikacije Future Mobility Survey u istraživanju [75], MEILI u istraživanju [76], Move Smarter u istra-

živanju [77] i Peacox u istraživanju [78]. Kako aplikacije tog tipa ne bi ovisile o korisničkoj potvrdi moda prijevoza potrebno je automatskim metodama detektirati korišteni mod prijevoza u trajektoriji. Proces klasifikacije modova prijevoza u stvarnom vremenu prikazan je na Slici 2.3. Sirovi podaci prikupljaju se s različitih osjetila tijekom trajanja jednog vremenskog prozora. Vektor značajki izračunava se za svaki vremenski prozor te predstavlja ulazne podatke klasifikatora. Značajke i vrste podataka koje se koriste u području jednake su onima opisanim u poglavlju 2.1.2. Izlaz klasifikatora je jedan od razmatranih modova prijevoza. Osnovna razlika ovog pristupa klasifikaciji u odnosu na klasifikaciju povijesnog skupa podataka je primjena vremenskih prozora. Značajke koje opisuju trajektoriju izračunavaju se nakon isteka definiranoga vremenskog intervala (vremenski prozor), a ne za cijelu trajektoriju jednog korisnika. Vremenski prozori mogu biti fiksne ili promjenjive veličine bez ili s preklapanjem kako navode autori u [79]. Veličina vremenskog prozora je tema mnogih istraživanja u kojima postoje različiti zaključci o optimalnoj veličini istog. Međutim, uglavnom se razmatraju veličine vremenskog prozora od 2 s do 1 min. Manji vremenski prozori do 30 s primjenjuje se na podacima s mobilnih uređaja, dok se veći vremenski prozori primjenjuju na GPS podacima.



Slika 2.3: Proces klasifikacije modova prijevoza u stvarnom vremenu

Kako se najčešće klasifikacija modova prijevoza povezuje s podacima s mobilnih uređaja, otvara se novi problem, a to je potrošnja energije na mobilnim uređajima uzrokovana raznovrsnim osjetilima razmatrana u istraživanjima [6, 11, 34, 49, 50, 54, 80]. Pregledom literature može se vidjeti da najmanje energije troši akcelerometar. Žiroskop troši značajno više energije od akcelerometra (za red veličine), a magnetometar troši dvostruko više energije od žiroskopa. GPS je najveći potrošač energije, a troši za red veličine više od magnetometra kako navode autori u [81]. Unatoč tome, podaci prikupljeni s mobilnih uređaja u svrhu detekcije modova prijevoza najčešće uključuju GPS podatke i podatke s akcelerometra.

U ovom radu naglasak je na klasifikaciji modova prijevoza u stvarnom vremenu pri čemu se klasificira trajektorija koja sadrži više uzastopnih modova prijevoza.

# Metode za rješavanje problema klasifikacije modova prijevoza

Nad velikim skupovima podataka najčešće se primjenjuju tehnike strojnog učenja za automatsko pronalaženje obrazaca i novog znanja unutar složenih podataka. Skriveni obrasci i znanje o problemu mogu se koristiti za predviđanje budućih događaja i izvođenje složenih odluka. Metode za segmentaciju i klasifikaciju modova prijevoza uglavnom dolaze iz skupine metoda strojnog učenja. U istraživanjima se razlikuju one metode koje segmentaciju provode kao zasebnu metodu koja prethodi klasifikaciji te metode koje dijele trajektoriju u vremenske prozore i pretpostavljaju da se u jednom vremenskom prozoru nije dogodila promjena moda prijevoza. Prva skupina metoda primjenjuje se na povijesnim podacima, dok se druga skupina odnosi na metode razvijane za stvarnovremensko okruženje. U sljedećim potpoglavljima objašnjene su metode koje se najčešće koriste za segmentaciju trajektorije (potpoglavlje 3.1) i klasifikaciju modova prijevoza (potpoglavlje 3.2). U potpoglavlju 3.3 opisana je struktura testnog problema i podatci na kojima je problem zasnovan.

### 3.1 Metode za segmentaciju trajektorije

Većina istraživanja u području segmentacije modova prijevoza zasnovana je na subjektivnim metodama u kojima autori postavljaju iskustvene zaključke na osnovu kojih se kreiraju pravila za detekciju MTP-a. Metode su uglavnom zasnovane na neizrastoj logici u istraživanjima [9, 82, 83, 84, 85] ili na nizu heurističkih pravila u istraživanjima [29, 86].

Najzastupljenija metoda segmentacije trajektorije je neizrastita logika. Neizrastita logika zasniva se na iskazima o stanju i nizu pravila zaključivanja zasnovanima na viševrijednosnoj

mjeri istinitosti, a čiji je cilj, u određivanju moda prijevoza, detektirati neuobičajen zapis tijekom prikupljanja podataka. Ako se jedno od pravila aktivira, točka koja je uzrokovala aktivaciju smatra se mjestom tranzicije sa jednog moda prijevoza u drugi. Vrlo često su pravila zasnovana na gubitku signala tijekom kojeg GPS daje neuobičajene zapise, tj. pokazuje značajna odstupanja u bilježenju. Drugo često pravilo detektira višestruke GPS zapise na istoj lokaciji što vrlo često signalizira stajanje. Gubitak signala se događa kod ulaska u autobus, tramvaj ili vlak radi metalne konstrukcije prijevoznog sredstva koja ometa GPS signal [77].

Među prvim istraživanjima, koja su promatrala segmentaciju trajektorije, je istraživanje [29]. Autori detektiraju MTP koristeći dva pravila. Prvim pravilom detektiraju se brze tranzicije iz jednog moda prijevoza u drugi (pješač-automobil), a drugim pravilom sporije tranzicije (pješač-autobus). U prvom pravilu, ako je povećanje brzine iznad 10 km/h i razlika u vremenu između dva zapisa veća od 5 s točka se smatra MTP. Brzina od 10 km/h se smatra maksimalnom brzinom pješaka. Drugim pravilom se prate mjesta gubitka signala, primjerice ulazak u autobus. Ako je gubitak signala duži od 150 metara detektirana je MTP. Pravila ovog tipa su učestala kod segmentacije trajektorije, te se pravila uglavnom generiraju iskustveno kako je navedeno u [87].

Autori u [85] definiraju tri tipa MTP-a: (i) početak pješaćenja (engl. *Start of Walk*, SoW), točka u kojoj korisnik prelazi iz bilo kojeg moda prijevoza na pješaćenje; (ii) kraj pješaćenja (engl. *End of Walk*, EoW), točka u kojoj korisnik prelazi s pješaćenja na bilo koji drugi mod prijevoza, te (iii) kraj perioda bez zapisa o kretanju korisnika (engl. *End of Gap*, EoG), točka u kojoj prestaje duži period u kojem se nije pojavio niti jedan zapis o kretanju korisnika. Sličnu metodologiju koriste autori u [37] na podacima prikupljenim u sklopu GeoLife projekta u kojima su trajektorije segmentirane koristeći pravilo pješaćkog segmenta koji signalizira tranziciju s jednog moda prijevoza na drugi. Pješaćke segmente detektirali su pomoću brzine i akceleracije, tj. ako je brzina manja od 9 km/h, a akceleracija manja od  $1,5 \text{ m/s}^2$ .

Osim metoda zasnovanih na pravilima, u području segmentacije trajektorije zastupljene su mnogobrojne metode koje pripadaju skupini nadziranih, nenadziranih i polunadziranih metoda strojnog učenja. Kao i u drugim područjima, metode iz skupine nadziranih i polunadziranih metoda zahtijevaju da je skup podataka označen klasama dok metode iz skupine nenadziranih metoda ne zahtijevaju označeni skup podataka. Klasa u području segmentacije modova prijevoza ukazuje na događaj promjene moda prijevoza.

#### 3.1.1 Nadzirane metode za segmentaciju trajektorije

Subjektivne metode zahtijevaju poznavanje podataka i moda prijevoza u trajektoriji. Na temelju podataka izračunavaju se vrijednosti parametara koje se koriste u odlučivanju. Najpoznatija nadzirana metoda za segmentaciju trajektorije iz skupine subjektivnih metoda je metoda za detekciju stajanja (engl. *Staying Point Detection*, SPD) opisana u istraživanju [88]. U ovoj metodi prisutna su dva ograničenja: udaljenosti i vremena, čije su vrijednosti izračunavaju iz

skupa podataka za učenje. U Algoritmu 3.1 je prikazan postupak detekcije stajanja. Ulazni podatci su GPS trajektorija te unaprijed zadano ograničenje udaljenosti i vremena. Tijekom izvršavanja algoritma iterativno se ispituje prijedena udaljenost i proteklo vrijeme između svake dvije točke u trajektoriji (linija 4-22). Ako je udaljenost između dvije uzastopne GPS točke veća od zadanog ograničenja udaljenosti provjerava se vrijeme provedeno između promatranih GPS točaka  $\Delta T$ . Ako je  $\Delta T$  veća od zadanog ograničenja, GPS koordinata koja se nalazi na polovištu dužine između dvaju promatranih GPS točaka dodaje se u skup točaka stajanja  $S$  (linija 8-17).

---

**Algoritam 3.1** Metoda za detekciju stajanja, [88]
 

---

```

1: Ulazni parametri: GPS trajektorija T, GPS točke P, ograničenje udaljenosti  $\theta_d$  i ograničenje
   vremena  $\theta_t$ 
2: Izlazni parametri: skup točaka stajanja  $S=\{s\}$ 
3:  $i = 0, brojZapisa = |T|$ 
4: dok je  $i < brojZapisa$  ponavlja
5:    $j := i + 1$ 
6:   dok je  $j < brojZapisa$  ponavlja
7:     udaljenost = Udaljenost( $P_i, P_j$ )
8:     ako udaljenost  $> \theta_d$  onda
9:        $\Delta T = p_j.T - p_i.T$ 
10:      ako  $\Delta T > \theta_t$  onda
11:        s.Gsirina= SrednjaVrijednost( $\{p_k.Gsirina | i \leq k \leq j\}$ )
12:        s.Gduljina= SrednjaVrijednost( $\{p_k.Gduljina | i \leq k \leq j\}$ )
13:        s.pocVrijeme =  $p_i.T$ 
14:        s.krajVrijeme =  $p_j.T$ 
15:        S.dodaj(s)
16:      kraj ako
17:    kraj ako
18:     $i := j$ 
19:    break
20:  kraj dok je
21:   $j := j + 1$ 
22: kraj dok je
23: vрати S

```

---

### 3.1.2 Nenadzirane metode za segmentaciju trajektorije

Metode nenadziranog učenja mogu se podijeliti u tri skupine metoda: (i) metode grupiranja, (ii) metode zasnovane na interpolaciji i (iii) metode zasnovane na funkciji cilja. Za grupiranje se najčešće koristi algoritam zasnovan na gustoći (engl. *Density-Based Spatial Clustering of Applications with Noise*, DBSCAN) [89], tj. dvije inačice tog algoritma koje su prilagođene za segmentaciju trajektorije, CB-SMoT (engl. *Clustering-Based Stops and Moves of Trajectories*) [90] i T-DBSCAN (engl. *Trajectory DBSCAN*) [91]. Algoritam 3.2 prikazuje osnovni postupak DBSCAN algoritma. Algoritam zahtjeva tri ulazna parametra: skup točaka koje se grupiraju  $D = \{t_1, t_2, \dots, t_N\}$  te parametre  $Eps$  i  $MinT$ . Parametrom  $Eps$  se definira udaljenost



između dvije točke da bi pripadale nekoj grupi. Dakle, ako je udaljenost između dvije točke manja ili jednaka vrijednosti  $Eps$  one se smatraju susjednim točkama.  $MinT$  je potreban minimalan broj susjednih točaka da bi one formirale grupu. Primjerice, ako je  $MinT = 10$  onda je potrebno minimalno 10 susjednih točaka da bi se formirala grupa. Nakon zadanih vrijednosti ulaznih parametara, algoritam za svaku točku koja nije dodijeljena niti jednoj grupi pronalazi skup susjednih točaka funkcijom *dohvatiSusjedneTocke()*. Funkcija *dohvatiSusjedneTocke* dohvaća sve točke koje su udaljenje manje od vrijednosti parametra  $Eps$  do promatrane točke. Ako je dohvaćen broj susjednih točaka veći od minimalnog broja točaka u grupi  $MinT$  onda se susjedne točke i promatrana točka dodaju u novu grupu  $C_K$  (linije 4-12). Točke koje na kraju izvođenja algoritma ne pripadaju niti jednoj grupi smatraju se vrijednostima izvan granica. Algoritam T-DBSCAN slijedi isti postupak kao izvorni algoritam, ali uvodi novi parametar  $Ceps$  kojim se zaustavlja pretraga u trenutku kada je prva izračunata udaljenost veća od vrijednosti  $Ceps$ . Tako se osigurava da se prati vremenski slijed trajektorije i sprječava grupiranje sličnih, ali vremenski udaljenih točaka u trajektoriji. Primjerice ako je udaljenost između grupa i promatrane točke manja od vrijednosti parametra  $Eps$ , promatrana točka se smatra susjednom, u protivnom ako je udaljenost veća od  $Ceps$  završava se proces pretraživanja za trenutnu grupu. CB-SMoT algoritam mijenja način izračuna parametra  $Eps$  tako da se ne računa izravno udaljenost između dvije točke već apsolutna udaljenost točaka u trajektoriji. Vrijednost parametra se ne zadaje iskustveno već se izračunava na osnovi distribucije udaljenosti za svaku trajektoriju.

---

**Algoritam 3.2** Algoritam DBSCAN, [89]

---

```
1: Ulazni parametri:  $D = \{t_1, t_2, \dots, t_N\}$ ,  $MinT$ ,  $Eps$ 
2: Izlazni parametar:  $C = \{c_1, c_2, \dots, c_K\}$ 
3:  $K = 0$ 
4: for  $i = 1$  do  $N$  do
5:   if  $t_i$  nije u grupi then
6:      $SusjedneTocke = dohvatiSusjedneTocke(t_i, Eps)$ 
7:     if broj elemenata  $SusjedneTocke > MinT$  then
8:        $K = K + 1$ 
9:        $C_K.Dodaj(SusjedneTocke, t_i)$ 
10:    end if
11:  end if
12: end for
```

---

Inačica algoritma  $k$  sredine (engl. *Warped-kMeans*, Wk-Means), koji opisuju autori u [92], je algoritam grupiranja koji se koristi za segmentaciju trajektorije. Algoritam 3.3 prikazuje pseudokod algoritma  $k$ -sredina (engl. *k-Means*) opisan u istraživanju [93]. Prvi zadatak algoritma je inicijalizirati  $k$  centara na slučajan način. Zatim se svaka točka u trajektoriji dodjeljuje najbližem centru. Zadnji korak je izračunavanje nove lokacije centara s obzirom na lokacije točaka koje su dodijeljene centru (linija 7). Postupak se ponavlja sve dok ima promjena u lokacijama centara (linija 2-9). Osnovna razlika između izvornog algoritma  $k$  sredine i Wk-Means

algoritma je tijekom dodjeljivanja točaka centrima, tj. premještanja točke iz trenutnog centra u centar do kojeg je izračunata manja udaljenost. Naime kod Wk-Means algoritma je postavljeno ograničenje da se prva polovica točaka iz trenutne grupe  $j$  smije prebaciti u grupu koja prethodi  $j - 1$ , a druga polovica se smije prebaciti u grupu koja slijedi  $j + 1$ .

---

**Algoritam 3.3** Algoritam  $k$ -sredina, [93]
 

---

```

1: Inicijaliziraj centre grupa  $c_1, c_2, \dots, c_k$  na slučajan način
2: while ima promjena do
3:   for svaku točku  $i$  do
4:      $C^{(i)} := \arg \min_j \|x^{(i)} - c_j\|$ 
5:   end for
6:   for svaki centar  $c$  do
7:     Izračunaj novu poziciju centra
8:   end for
9: end while

```

---

Metode koje se zasnivaju na interpolaciji uglavnom koriste linearnu ili kinematičku interpolaciju kako bi predvidjele sljedeću geografsku lokaciju u trajektoriji. Predviđena i stvarna lokacija se uspoređuju te se izračunava pogreška. Evaluacijom distribucije dobivenih pogrešaka zaključuje se da li je došlo do promjene u kretanju u trajektoriji te se trajektorija dijeli na segmente. U području segmentacije trajektorije često se koriste metode zasnovane na traženju pogreške (engl. *Octal Window Segmentation*, OWS) poput istraživanja u [94] ili vremenskoj segmentaciji trajektorije (engl. *Sliding Window Segmentation*, SWS) poput istraživanja u [95].

Funkcija cilja koristi se za modifikaciju broja i lokacija točaka na kojima će se segmentacija provesti u pohlepnoj heuristici (engl. *Greedy Randomized Adaptive Search Procedure for Unsupervised Trajectory Segmentation*, GRASP-UTS) autora u [72]. Algoritam 3.4 prikazuje osnovne korake izvođenja GRASP-UTS algoritma pri čemu se inicijalizira početno rješenje na slučajan način, a zatim se vrši lokalna pretraga početnog rješenja. Ako se lokalnom pretragom pronađe rješenje bolje od trenutnog ono se zapisuje kao trenutno najbolje rješenje (linije 4-8 u Algoritmu 3.4). Ulazni parametri u algoritam su koeficijent sličnosti  $coef$  kojim se definira razina sličnosti između dva segmenta koja je ujedno i granica za spajanje dva segmenta te faktor particije  $p_f$  kojim se definira omjer za particioniranje točaka između dva centra.

Osnovni cilj lokalne pretrage je minimizirati distorziju te maksimizirati kompresiju trajektorije. Minimalna distorzija postiže se povećanjem sličnosti unutar segmenata trajektorije dok se maksimalna kompresija postiže smanjenjem broja segmenata. Ova dva kriterija su suprotstavljena jer su obrnuto proporcionalni, pa se funkcijom cilja postiže približno optimalan omjer između dva kriterija. Funkcija cilja zasnovana je na principu minimalne duljine opisa (engl. *Minimum Description Length*, MDL), Izraz 3.1, pri čemu je funkcija cilja zasnovana na dvije komponente  $L(H)$  i  $L(D|H)$  u kojima  $H$  označava hipotezu, a  $D$  podatke. Općenito,  $L(H)$  je duljina hipoteze u bitovima, a  $L(D|H)$  duljina opisa podataka u bitovima ako su podatci struk-

---

**Algoritam 3.4** GRASP-UTS, [72]

---

- 1: Ulazni parametri: skup tocaka u trajektoriji poredanih po vremenu  $T_n = (p_1, p_2, \dots, p_n)$ , minimalno vrijeme ( $\text{minVrijeme}$ ), koeficijent sličnosti  $\text{coef}$ , faktor particije  $p_f$ , ograničenje pohlepnosti algoritma  $\alpha$ , maksimalan broj iteracija  $\text{max}_{iter}$
  - 2: Izlazni parametri: skup segmenata trajektorije  $\theta_T = (s_1, s_2, \dots, s_T)$
  - 3:  $k=0$
  - 4: **for**  $k < \text{max}_{iter}$  **do**
  - 5:    $\theta_T \leftarrow$  početno rješenje na pohlepan slučajan način( $T, \text{minVrijeme}, p_f, \alpha$ )
  - 6:    $\theta_T \leftarrow$  Lokalna pretraga( $\theta_T, \text{coef}, \text{minVrijeme}, p_f$ )
  - 7:   Ažuriraj trenutno najbolje rješenje ( $\theta_T, \text{Best}\theta_T$ )
  - 8: **end for**
  - 9: **vрати**  $\text{Best}\theta_T$
- 

turirani hipotezom  $H$ . U GRASP-UTS algoritmu hipoteza se sastoji od odabira optimalnog broja centara segmenata, a podatci, koji se za to koriste, su značajke koje opisuju svaku točku u trajektoriji. Najbolje rješenje je ono koje daje minimalnu vrijednost MDL pravila.

U Izrazu 3.2 prikazana je komponenta  $L(H)$  u GRASP-UTS algoritmu pri čemu su  $\phi_T = (lm_1, lm_2, \dots, lm_T)$  skup centara segmenata trajektorije,  $\text{max}_{udaljenost}$  predstavlja maksimalnu moguću udaljenost između dvije značajke trajektorije, a funkcija  $\text{udaljenost}()$  koristi se za izračun udaljenost između dva slijedna centra  $lm_j$  i  $lm_{j+1}$ . Na ovaj način izračunava se funkcija cilja ako se trajektorija prikaže odabranim skupom centara  $\phi_T$ . S druge strane, Izrazom 3.3, izračunava se homogenost rješenja, pri čemu je  $\theta_T = (s_1, s_2, \dots, s_T)$  skup segmenata trajektorije, a  $[p_u \dots p_v]_k$  predstavlja skup točaka u  $k$ -tom segmentu, kako je to opisano u istraživanju [72].

$$MDL_{princip} = L(H) + L(D|H) \quad (3.1)$$

$$L(H) = f(\phi_T) = \log_2 \left( 1 + \sum_{j=1}^{T-1} \text{max}_{udaljenost} - \text{udaljenost}(lm_j, lm_{j+1}) \right) \quad (3.2)$$

$$L(D|H) = g(\theta_T) = \log_2 \left( 1 + \sum_{k=1}^T \text{Homogenost}(lm_k, [p_u \dots p_v]_k) \right) \quad (3.3)$$

### 3.1.3 Polunadzirane metode za segmentaciju trajektorije

Polunadzirane metode u klasifikacijskim problemima koriste kombinaciju podataka čija klasa je poznata i onih podataka čija klasa nije poznata. Inačica GRASP algoritma (engl. *Reactive GRASP for semantic Semi-supervised Trajectory Segmentation*, RGRASP-SemTS) opisana je u istraživanju [96] kao polunadzirani algoritam koji je nadogradnja Algoritma 3.4. Osnovna razlika između izvornog i nadograđenog algoritma je što algoritam RGRASP-SemTS koristi skup podataka koji nije u potpunosti označen izlaznim klasama.

## 3.2 Metode za klasifikaciju trajektorije

Metode za klasifikaciju trajektorije pripadaju skupini nadziranih metoda te je do sada primijenjen velik broj metoda na različitim skupovima podataka za klasifikaciju modova prijevoza. Dio istraživanja s primijenjenim metodama, postignutom točnošću te veličinom vremenskog prozora (klasifikacija modova prijevoza u stvarnom vremenu ili klasifikacija cijele trajektorije) prikazan je u Tablici 3.1. Razlikuju se istraživanja u kojima je primijenjena klasifikacija trajektorije podijeljene u vremenske prozore te klasifikacija cijele trajektorije. Klasifikacija cijele trajektorije primjenjuje se na povijesnim podacima (nakon prikupljanja podataka) dok se metode koje dijele trajektoriju u vremenske prozore mogu primijeniti u stvarnom vremenu kako navode autori u [97]. Veličina vremenskog prozora u kojem se prikupljaju podatci uglavnom je u rasponu od 1 - 180 s. U nastavku poglavlja su opisane metode koje su najzastupljenije u istraživanjima ili one pomoću kojih su postignuti najbolji rezultati. Navedene metode pripadaju metodama nadziranog učenja te se mogu primijeniti na probleme regresije i klasifikacije. Za potrebe ovog istraživanja sve metode interpretirane su u kontekstu klasifikacije.

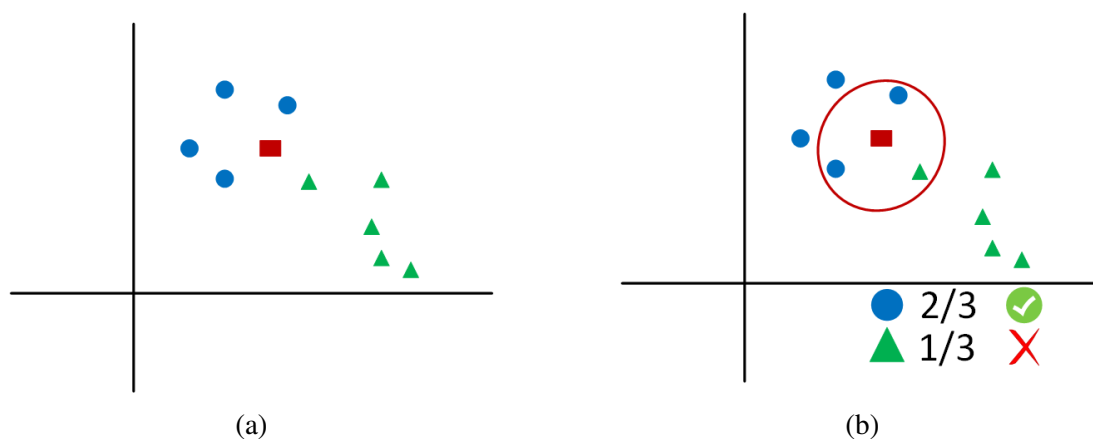
**Tablica 3.1:** Primijenjene metode na problemu klasifikacije modova prijevoza

Istraživanje	Metoda klasifikacije														Veličina prozora	Točnost
	DT	RF	SVM	KNN	BN	NN	DNN	CNN	AdaBoost	HMM	CRF	XGBoost	NB	H		
[45, 46]	✓		✓	✓			✓								17,2 s	86,94%
[16]	✓		✓					✓							17,2 s	90,66%
[47]	✓		✓	✓											8 s	70,73%
[48]	✓	✓	✓	✓											1 s	95,01%
[49]								✓	✓						1,2 s	84,20%
[50]			✓												12,8 s	98,55%
[51]				✓											7,5 s	95,80%
[52]	✓														8 s	88,00%
[15]	✓	✓	✓					✓							7,8 s	99,86%
[53]	✓		✓	✓									✓		Nepoznato	90,40%
[36]			✓										✓		4 s	97,00%
[54]	✓	✓		✓											5 s	97,00%
[66]	✓		✓		✓										Cijela trajektorija	76,20%
[5]	✓	✓	✓	✓								✓			Cijela trajektorija	90,77%
[38]							✓								Cijela trajektorija	67,90%
[39]					✓										Cijela trajektorija	92,74%
[40]	✓	✓			✓								✓		30 s	93,50%
[41]			✓												Cijela trajektorija	91,58%
[33]			✓												180 s	88,00%
[42]													✓		Cijela trajektorija	82,60%
[11]						✓									Cijela trajektorija	91,23%
[17]					✓										60 s	91,70%
[34]	✓								✓						1 s	82,14%
[55]									✓						10 s	75,80%

### 3.2.1 K najbližih susjeda

Algoritam kNN je nadzirani algoritam čija je osnovna pretpostavka da su slični uzorci uvijek blizu u prostoru svojih značajki, a izvorno je prikazan u istraživanjima [98, 99]. Uzorci za učenje s poznatom klasom, koja se predviđa, se pohranjuju i predstavljaju model ovog algoritma.

Testiranje modela se vrši s podacima kojima nije poznata klasa te se ona predviđa. Na Slici 3.1 je slikoviti prikaz odlučivanja u kNN modelu. Na Slici 3.1a su prikazani pohranjeni uzorci iz skupa podataka za učenje modela plavim krugovima i zelenim trokutima. Trokut i krug su dvije klase modela, tj. dva moguća rezultata predviđanja. Crvenim pravokutnikom je prikazan uzorak iz skupa podataka za testiranje kojemu se treba odrediti klasa. Na Slici 3.1b je prikazan proces donošenja odluke. Na osnovi izračunatih udaljenosti od crvenog pravokutnika do svih drugih točaka u modelu odabire se  $k$  uzoraka do kojih je udaljenost najmanja. U ovom primjeru odabrana je vrijednost tri za parametar  $k$ . U skupu tri najbliža susjeda nalaze se dva kruga i jedan trokut. Prema tome većinski broj uzoraka pripada klasi kruga, pa se novi uzorak klasificira kao plavi krug.



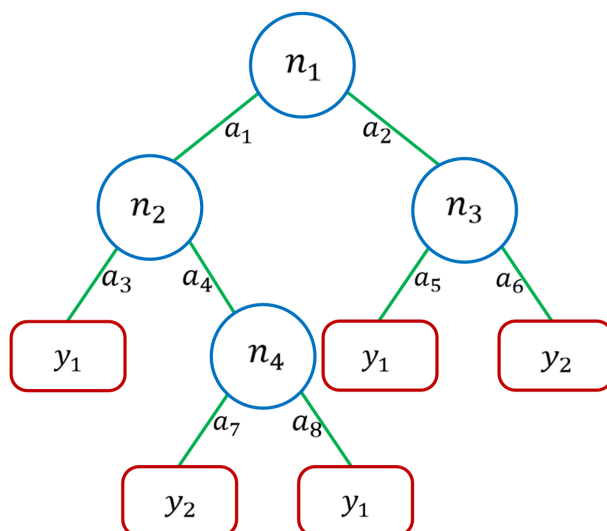
Slika 3.1: Proces donošenja odluke tijekom klasifikacije kNN algoritmom

### 3.2.2 Stabla odluke

Modeli zasnovani na DT-u dijele skup podataka prema određenim graničnim vrijednostima njihovih značajki više puta, kako je opisano u istraživanju [100]. Podjelom se stvara više podskupova podataka, a svaki uzorak iz skupa podataka za učenje pripada jednom podskupu. Podskupovi podataka se zatim koriste za grananje stabla.

Osnovni resursi DT-a su čvorovi (unutarnji i korijenski), grane i listovi. Čvorovi su zapravo značajke, grane ispod svakog čvora odgovaraju vrijednostima značajke, a listovi su klasifikacijske odluke. Prvi čvor u stablu je korijenski čvor (krenuvši od vrha), a svi drugi čvorovi u stablu su unutarnji čvorovi. Na Slici 3.2 su prikazani osnovni dijelovi DT. Plavim krugovima ( $n_1, \dots, n_4$ ) označeni su čvorovi, crvenim pravokutnicima ( $y_1, y_2$ ) listovi, a zelenim linijama

$(a_1, \dots, a_8)$  vrijednosti grane DT. Novi uzorak se klasificira slijednim ispitivanjem vrijednosti značajki, krenuvši od korijenskoga čvora prema listovima stabla. Kada se prolaskom kroz stablo dođe do lista, novi uzorak se klasificira klasom koja je određena određnim listom. Dakle, jedno stablo može se interpretirati kao sažeti skup zapisa pravila dobivenih na osnovu nekog skupa podataka.



Slika 3.2: Osnovni dijelovi DT

Interpretacija postojećeg DT je jednostavna, ali izgradnja samog stabla je složena. Algoritama za izgradnju stabla ima mnogo, a najpoznatiji je algoritam za izgradnju klasifikacijskog i regresijskog stabla (engl. *Classification And Regression Trees*, CART) opisan u istraživanju [100]. Osnovni koncept kod svih algoritama je jednak, a to su dva osnovna zadatka:

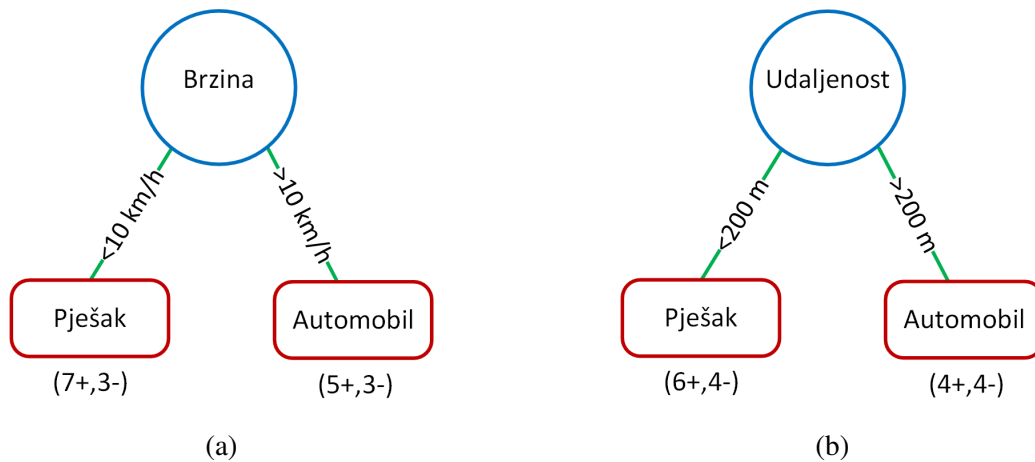
1. strukturirati problem kao stablo tako da se kreiraju čvorovi i grane stabla,
2. dodijeliti vjerojatnost svakom ishodu.

Prvi korak u izgradnji DT je odabir značajke koja se ispituje za svaki od čvorova. Postoji više numeričkih kriterija na osnovu kojih se zaključuje pomoću koje značajke se bolje razlikuju klase, primjerice entropija i gini indeks navedene su u istraživanju [101]. Entropija je mjera homogenosti nekog skupa podataka, a izračunava se prema Izrazu 3.4 gdje je  $p_{(+)}$  vjerojatnost pozitivne klase,  $p_{(-)}$  vjerojatnost negativne klase, a  $S$  skup podataka za učenje kako se opisuje u istraživanju [102]. Što je entropija veća teže je iz podataka dobiti neke zaključke.

$$E(S) = -p_{(+)} \log p_{(+)} - p_{(-)} \log p_{(-)} \quad (3.4)$$

Kod izgradnje stabla entropija zapravo mjeri stupanj slučajnosti nekog čvora. Primjerice na Slici 3.3 je prikazan primjer dvije značajke, brzina i udaljenost, iz skupa podataka za koji su navedene značajke odabrane kao čvor stabla. Skup podataka ima dva moguća izlaza: pješak i automobil. Ako je brzina odabrani čvor stabla i dvije su grane čvora ( $brzina < 10$  km/h i  $brzina > 10$  km/h) onda izlaz pješak rezultira sa 7 točnih i 3 pogrešna predviđanja, a izlaz auto-

mobil rezultira sa 5 točnih i 3 pogrešna predviđanja. Isto se određuje i za značajku udaljenost koja za izlaz pješak ima 6 točnih i 4 pogrešna predviđanja, a za izlaz automobil 4 točna i 4 pogrešna predviđanja.



**Slika 3.3:** Primjer donošenja odluke na osnovi jedne značajke skupa podataka

Uzimajući u obzir broj ispravno i pogrešno klasificiranih uzoraka za svaku značajku izračunata je entropija na osnovi Izraza 3.5 i 3.6 pri čemu je entropija za značajku brzina 0,284, a za udaljenost 0,296. Pomoću entropije za svaku značajku se izražava informacijska dobit i odlučuje se koja značajka ima prednost kod izgradnje stabla. U ovom primjeru je to značajka brzina.

$$E(S, Brzina) = \frac{10}{18} \left( -\frac{7}{10} \log\left(\frac{7}{10}\right) - \frac{3}{10} \log\left(\frac{3}{10}\right) \right) + \frac{8}{18} \left( -\frac{3}{8} \log\left(\frac{3}{8}\right) - \frac{5}{8} \log\left(\frac{5}{8}\right) \right) = 0,284 \quad (3.5)$$

$$E(S, Udalj) = \frac{10}{18} \left( -\frac{6}{10} \log\left(\frac{6}{10}\right) - \frac{4}{10} \log\left(\frac{4}{10}\right) \right) + \frac{8}{18} \left( -\frac{4}{8} \log\left(\frac{4}{8}\right) - \frac{4}{8} \log\left(\frac{4}{8}\right) \right) = 0,296 \quad (3.6)$$

Ista odluka može se donijeti pomoću mjere gini indeks koja je dana Izrazom 3.7 pri čemu je  $P(x = k)$  vjerojatnost da uzorak  $x$  pripada klasi  $k$ .

$$gini = 1 - \sum_{k=1}^n (P(x = k))^2 \quad (3.7)$$

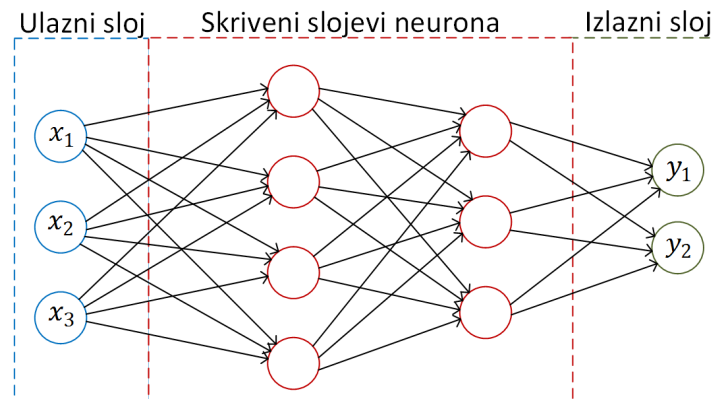
Proces dodjele značajke promatranom čvoru ponavlja se ovisno o tijeku algoritma koji se koristi za grananje stabla. Najpoznatiji su algoritmi:

1. CART algoritam - koristi binarno grananje tijekom podijele uzoraka u čvoru;
2. ID3 (engl. *Iterative Dichotomiser 3*) - može koristiti višestruke podijele tijekom podjele uzoraka u čvoru.

Metoda DT je sklona prenaučivosti jer se grananjem postiže kvalitetan skup pravila za razdvajanje uzoraka u skupu podataka za učenje, ali isto stablo neće jednako dobro opisivati drugi skup podataka iz jednakog problema. Prenaučivost se općenito odnosi na modele koji jako dobro modeliraju podatke na kojima su naučeni, a pokazuju znatno lošije rezultate na novom skupu podataka. Za smanjenje vjerojatnosti prenaučivosti koriste se hiperparametri metode kojima se može spriječiti prenaučivost metode. Osnovni hiperparametri su maksimalna dubina stabla (broj razina unutarnjih čvorova u stablu) i minimalan broj uzoraka (broj uzoraka u podskupu podataka koji se koristi za izračun entropije ili gini indeksa neke značajke).

### 3.2.3 Neuronska mreža

Neuronska mreža je metoda iz područja strojnog učenja koja je inspirirana načinom komunikacije neurona u ljudskom mozgu navode autori u [103]. Osnovna struktura svake neuronske mreže prikazana je na Slici 3.4, a sastoji se od ulaznog sloja, skrivenih slojeva neurona te izlaznog sloja. Ulazni sloj je prvi sloj neuronske mreže i on prihvaća ulazni skup podataka ( $x_1, x_2, x_3$ ). Izlazni sloj je posljednji sloj NN i on se sastoji od mogućih predviđanja NN ( $y_1$  ili  $y_2$ ). U skrivenim slojevima se nalaze neuroni koji obrađuju podatke. U svim neuronima nalazi se aktivacijska funkcija te težine na sinapsama između neurona u kojima je pohranjeno znanje mreže.



Slika 3.4: Osnovna struktura neuronske mreže

Za slučaj unaprijedne neuronske mreže neuroni u skrivenom sloju neurona imaju strukturu poput neurona prikazanog na Slici 3.5. Ako su ulazni podatci  $x_1, x_2, x_3$  onda se izlaz neurona može prikazati kako je prikazano u Izrazu 3.8 pri čemu je  $w_{ji}$  težina na sinapsi između ulaznog podatka i neurona ili između dva neurona, a  $x_i$  ulazni podatak u neuron.

$$n_j = \sum_{i=1}^3 w_{ji} x_i \quad (3.8)$$

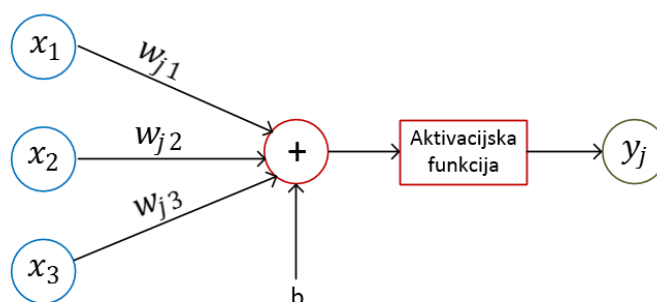
Izlazna vrijednost iz neurona je ulazni parametar u aktivacijsku funkciju na osnovu koje



se određuje predviđanje  $y_j$  ili ulazna vrijednost za drugi povezani neuron prema Izrazu 3.9 pri čemu je  $n_j$  izlaz neurona, a  $b$  prag koji se određuje kao ulazni parametar NN.

$$y_j = \text{Aktivacijska funkcija}(n_j - b) \quad (3.9)$$

Aktivacijske funkcije su važan dio NN i osnovni im je zadatak donošenje odluke hoće li neuron biti aktiviran ili ne. Aktivacijskih funkcija ima više vrsta poput skokomične, linearne ili sigmoidalne funkcije kako se navodi u istraživanju [104].



Slika 3.5: Struktura neurona

Proces učenja NN je iterativni proces podešavanja težina na sinapsama neurona i pragova. Tako mreža uči iz svoje okoline, tj. ulaznog skupa podataka. Metoda za učenje NN ima više, a najčešće su zasnovane na učenju korekcijom pogreške pri čemu se pogreška propagira unazad te se tako korigiraju težine sinapsi i postiže bolji rezultat NN.

### 3.2.4 Skriveni Markovljevi modeli

Skriveni Markovljev model je vjerojatnosni model koji se koristi kako bi se opisala vjerojatnosna karakteristika nekog procesa zasnovana na Markovljevim lancima. Osnovna pretpostavka u Markovljevim lancima je da ako se želi predvidjeti buduće stanje nekog slijednog događaja ono ovisi isključivo o trenutnom stanju. Dakle, povijesni događaji nemaju utjecaja na predviđanje kako opisuju autori u [105].

HMM model primijenjen na odabrani problem je jedinstven, ali osnovni dijelovi svakog modela mogu se opisati kroz sljedećih pet parametara:

1. skup od  $N$  stanja:

$$S = \{s_1, s_2, \dots, s_N\}, \quad (3.10)$$

2. matrica prijelaznih stanja pri čemu je  $t_{ij}$  vjerojatnost prelaska iz stanja  $i$  u stanje  $j$ :

$$T = [t_{11}, t_{12}, \dots, t_{n1}, \dots, t_{nm}], \quad (3.11)$$

3. distribucija vjerojatnosti za početna stanja pri čemu je  $\alpha_i$  vjerojatnost da će prvo stanje Markovljevog lanca biti  $i$ :

$$A = \{\alpha_1, \alpha_2, \dots, \alpha_N\}, \quad (3.12)$$

4. niz opažanja:

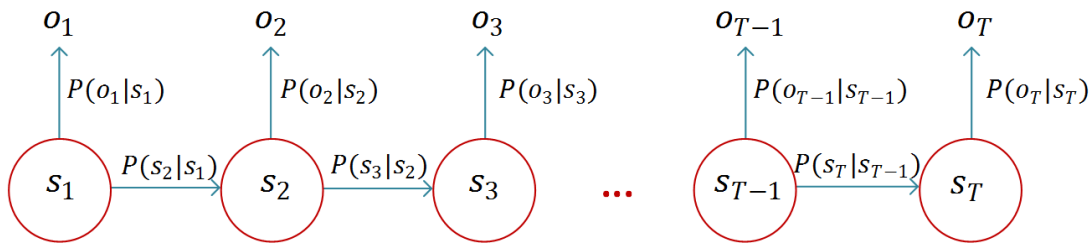
$$O = \{o_1, o_2, \dots, o_T\}, \quad (3.13)$$

5. matrica vjerojatnosti emitiranja opažanja  $o_t$  iz stanja  $i$ :

$$B = b_i(o_t). \quad (3.14)$$

Na ovaj način može se opisati proces prelaska iz jednog u drugo skriveno stanje koje je dano skupom  $S$  te prelazak iz jednog opažanja u sljedeće koje je dano skupom  $O$ . Za predviđanje prelaska iz jednog stanja u sljedeće koriste se matrice vjerojatnosti prijelaznih stanja  $T$  i  $B$  prema istraživanju u [105].

Na Slici 3.6 je prikazana osnovna struktura HMM koja pokazuje proces odlučivanja nakon što se slučajan proces interpretira kroz varijable HMM. Dakle, skrivena stanja prikazana su crvenim krugovima te vjerojatnost pojave nekog stanja na mjestu primjerice stanja  $s_2$  ovisi o vrijednosti vjerojatnosti pojave stanja  $s_2$  uz uvjet da je sustav prethodno bio u stanju  $s_1$ . Dok vjerojatnost pojave nekog opažanja ovisi o vjerojatnosti promatranog opažanja uz uvjet da je sustav prethodno bio u nekom stanju. Tipična pretpostavka u HMM-u je da su opažanja poznata vrijednost, a skrivena stanja nisu poznata.



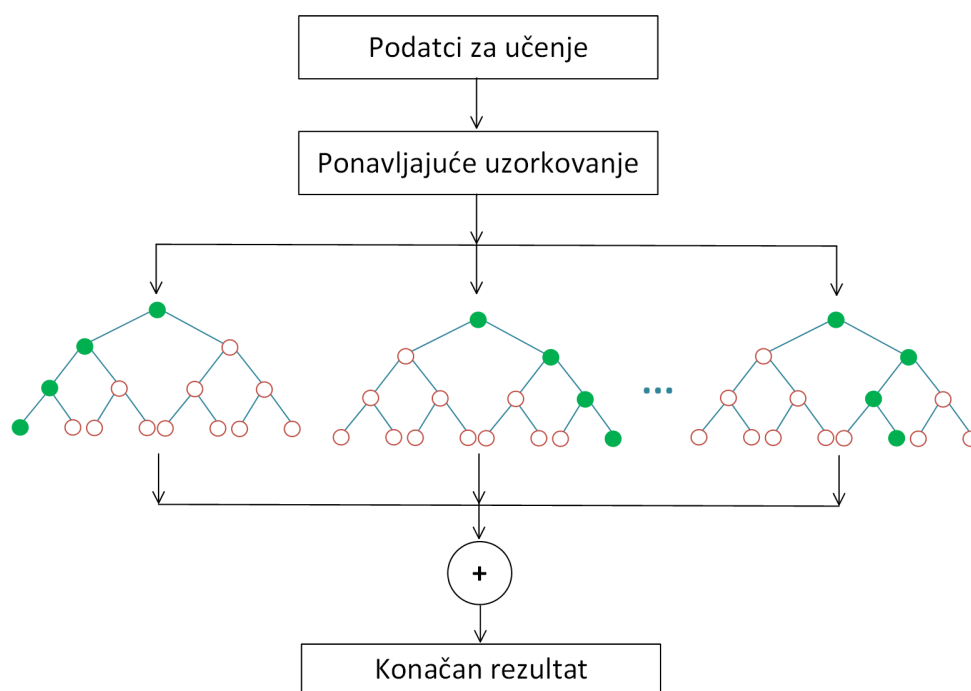
Slika 3.6: Struktura HMM

### 3.2.5 Višeslojna klasifikacija i združeni klasifikatori

U skupinu višeslojnih i združenih klasifikatora pripadaju metode koje unutar sebe sadrže više različitih ili jednakih vrsta metoda. Osnovna razlika između združenih klasifikatora i višeslojne klasifikacije je što združeni klasifikatori nemaju hijerarhiju donošenja odluke dok je ona kod višeslojnih klasifikatora prisutna. Združeni klasifikatori su skupina metoda koje unutar sebe sadrže više jednostavnih klasifikatora te svaki od njih donosi odluku o klasifikaciji istog uzorka navode autori u [106]. Dok je kod višeslojne klasifikacije proces klasifikacije podijeljen u više

slojeva, najčešće dva, pri čemu se u prvom koraku odabranim klasifikatorom donosi odluka o grupiranim klasama, a detaljnija klasifikacija se provodi u sljedećem sloju, tj. koraku.

U području klasifikacije modova prijevoza najčešći združeni klasifikatori su RF i metoda adaptivnog povećanja točnosti predikcije (engl. *Adaptive Boosting*, AdaBoost). Metoda RF je zasnovana na metodi DT. Na Slici 3.7 je prikazana osnovna ideja metode RF. Ulazni skup podataka za učenje se uzorkuje i svaki podskup podataka služi za izgradnju jednog stabla. Proces izgradnje jednog stabla jednak je postupku opisanom u poglavlju 3.2.2. U procesu odlučivanja, svako stablo donosi odluku o pripadnosti nekog uzorka klasi (zeleno obojeni krugovi). Konačni rezultat je klasa koja je rezultat odluka najvećeg broja stabala prema istraživanju u [107].

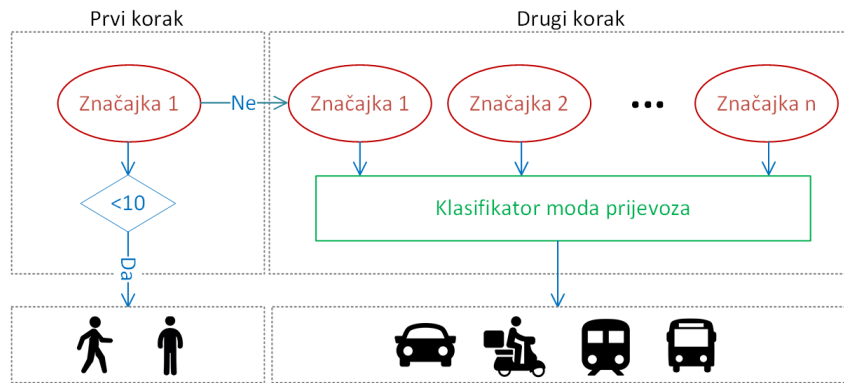


**Slika 3.7:** Osnovna struktura metode RF

AdaBoost metoda se također zasniva na više jednostavnih klasifikatora od kojih svaki donosi svoju odluku o klasi. Međutim kod procesa učenja se iterativno ponavlja učenje tako da se pogrešno klasificiranim uzorcima povećava težina kako bi se maksimizirala točnost modela poput istraživanja u [108].

Višeslojne metode u području klasifikacije modova prijevoza uglavnom sadrže dva sloja, pri čemu se u prvom sloju odlučuje kreće li se korisnik motoriziranim ili ne motoriziranim modom prijevoza, a u idućem koraku se odlučuje o kojem se točno modu prijevoza radi navode autori u [16]. Na Slici 3.8 je prikazan osnovni koncept višeslojnih metoda.

U prvom koraku se na osnovi jedne značajke i praga vrijednosti značajke odlučuje hoda li korisnik ili miruje. Ako vrijednost značajke ne zadovoljava uvjet zadanog praga onda se prelazi na drugu razinu odlučivanja u kojoj se na osnovi više značajki i drugog modela klasifikatora odlučuje o kojem se motoriziranom modu prijevoza radi.



Slika 3.8: Osnovna struktura višeslojnih metoda

### 3.3 Testni skup podataka

U području klasifikacije modova prijevoza osnovni problem usporedivosti rezultata je nedovoljno korištenje testnog skupa podataka na kojem autori mogu usporediti svoje rezultate s postojećim. Kao što je prikazano u Tablici 2.1 veliki broj istraživanja koristi vlastite skupove podataka te nije uniformiran broj modova prijevoza koji se klasificira. Nije jednaka složenost metode koja klasificira modove prijevoza na tri klase, primjerice hodanje, bicikl i vozilo, i metode koja klasificira sedam modova prijevoza. Ovaj problem prepoznali su i vlasnici javnog skupa podataka SHL te su strukturirali testni skup podataka tako da otklone navedene probleme usporedivosti u članku [81]. Za potrebe ove disertacije koristit će se upravo SHL testni skup podataka, pa će u nastavku biti opisana struktura tog javnog skupa podataka.

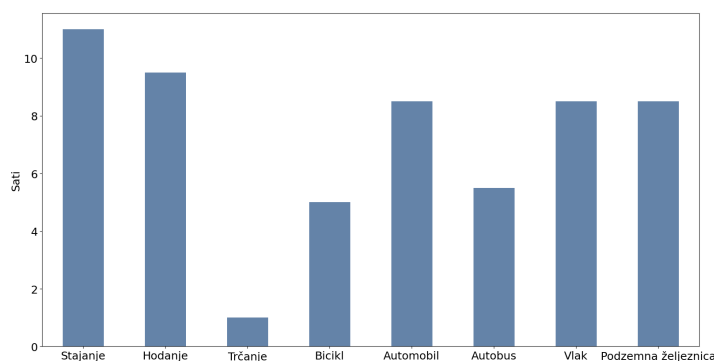
SHL skup podataka su prikupljala tri korisnika pomoću mobilnih uređaja. Osnovne karakteristike SHL skupa podataka prikazane su u Tablici 3.2. Tijekom prikupljanja podataka korisnici prikupljaju istovremeno podatke s četiri mobilna uređaja pozicionirana na četiri različite lokacije na tijelu: torba, ruka, torzo i kuk. Osim toga korisnici su na prsima nosili kameru iz čijih se snimaka naknadno podacima pridružio korišteni mod prijevoza. S mobilnih uređaja prikupljali su se podatci s osjetila akcelerometra, žiroskopa, magnetometra, barometra i gravitacije s

**Tablica 3.2:** Osnovne karakteristike objavljenog skupa podataka SHL

Obilježje skupa podataka	
Broj korisnika	3
Pozicije prikupljanja podataka	ruka, torzo, torba i kuk
Osjetila	Akceleracija, žiroskop, magnetometar, linearna akceleracija, gravitacija, barometar
Modovi prijevoza i aktivnosti	Hodanje, trčanje, bicikl, automobil, autobus, vlak, podzemna željeznica i stajanje
Broj podataka [h]	$4 \times 703 = 2.812$ , objavljeno $4 \times 59 = 236$
Frekvencija uzorkovanja [Hz]	100

frekvencijom uzorkovanja 100 Hz. Svakom zapisu pridružen je jedan mod prijevoza iz skupa: hodanje, trčanje, bicikl, automobil, autobus, vlak i podzemna željeznica ili aktivnost: stajanje. Ako za zapis nije bilo moguće procijeniti mod prijevoza, zapis je označen klasom nula.

SHL skup podataka sadrži 2.812 sati zapisa o kretanju korisnika, međutim nije cijeli skup podataka objavljen već je objavljen samo dio podataka koji čini sveukupno 236 sati kretanja tri korisnika. Na Slici 3.9 je prikazana distribucija podataka pojedinog moda prijevoza u objavljenom skupu podataka. Najmanje su zastupljeni podatci o trčanju, biciklu i autobusu. Podatci o ostalim modovima prijevoza su podjednako distribuirani.



**Slika 3.9:** Zastupljenost modova prijevoza u SHL skupu podataka

Testni skup podataka organiziran je tako da ispituje točnost metode s obzirom na mod prijevoza koji se predviđa, poziciju mobilnog uređaja na tijelu korisnika, vremenski period kada su podatci prikupljeni, utjecaj ponašanja pojedinog korisnika te osjetila s kojih se podatci prikupljaju. Kako u dosadašnjim istraživanjima nije uniformno određeno koji modovi prijevoza moraju biti uključeni u klasifikaciju korisničkih trajektorija s obzirom na mod prijevoza onda može se pronaći različit broj i vrsta klasa za navedeno. Stoga je testiranje u testnom problemu strukturirano kroz 12 različitih scenarija kako bi se pokrile sve kombinacije grupiranja modova prijevoza dostupne u literaturi. Grupiranje 8 modova prijevoza i aktivnosti dostupnih u SHL skupu podataka (Mirovanje (M1), Hodanje (H2), Trčanje (T3), Bicikl (B4), Automobil (A5), Autobus (A6), Vlak (V7), Podzemna željeznica (P8)) provedeno je na sljedeći način:

- Scenarij 1 - fizičke aktivnosti (H2, T3, B4) i ne fizičke aktivnosti (M1, A5, A6, V7, P8);
- Scenarij 2 - M1, pješak (H2, T3, B4) i vozilo (A5, A6, V7, P8);
- Scenarij 3 - M1, pješak (H2, T3), B4 i vozilo (A5, A6, V7, P8);
- Scenarij 4 - M1, H2, T3, B4 i vozilo (A5, A6, V7, P8);
- Scenarij 5 - M1, pješak (H2, T3), B4, cestovno vozilo (A5, A6) i tračnički sustavi (V7, P8);
- Scenarij 6 - M1, H2, T3, B4, cestovno vozilo (A5, A6) i tračnički sustavi (V7, P8);
- Scenarij 7 - M1, pješak (H2, T3), B4, privatno prijevozno sredstvo (A5) i javno prijevozno

- sredstvo (A6, V7, P8);
- Scenarij 8 - M1, H2, T3, B4, privatno prijevozno sredstvo (A5) i javno prijevozno sredstvo (A6, V7, P8);
  - Scenarij 9 - M1, pješak (H2, T3), B4, privatno prijevozno sredstvo (A5), javno cestovno prijevozno sredstvo (A6) i javni tračnički sustavi (V7, P8);
  - Scenarij 10 - M1, H2, T3, B4, privatno prijevozno sredstvo (A5), javno cestovno prijevozno sredstvo (A6) i javni tračnički sustavi (V7, P8);
  - Scenarij 11 - M1, pješak (H2, T3), B4, A5, A6, V7 i P8;
  - Scenarij 12 - M1, H2, T3, B4, A5, A6, V7 i P8.

Grupiranjem modova prijevoza u scenarije omogućena je usporedba metoda koje razvrstavaju modove prijevoza u veći ili manji broj klasa. Osim toga, većina dosadašnjih istraživanja može se svrstati u jedan od navedenih scenarija. Kako bi se testirala korisnička, vremenska i lokacijska neovisnost metode unakrsno se izbacuje jedno od svojstva svake komponente iz skupa podataka za učenje i koristi se u skupu podataka za testiranje kao što je prikazano u Tablici 3.3. Primjerice, za testiranje korisničke neovisnosti modela iz skupa podataka za učenje uklanjaju se podatci jednog korisnika te se uklonjeni podatci koriste za testiranje metode. Postupak se ponavlja tri puta za svakog korisnika u skupu podataka. Tako se testira koliko dobro metoda generalizira podatke o modovima prijevoza, tj. koliko je prenaučena na ponašanje pojedinog korisnika. Uzrok prenaučivosti može biti primjerice velik broj potencijalnih značajki, pa će se kao posljedica manifestirati znatno lošiji rezultati predviđanja na podacima novog korisnika.

**Tablica 3.3:** Svojstva testnog skupa podataka

Izostavljanje jednog elementa svakog svojstva		Scenarij	Osjetila
Korisnički zavisno	Korisnik 1	1 – 12	Akceleracija Akceleracija + Žiroskop Akceleracija + Žiroskop + Magnetometar
	Korisnik 2		
	Korisnik 3		
Lokacijski zavisno	Ruka		
	Kuk		
	Torzo		
	Torba		
Vremenski zavisno	Period 1		
	Period 2		
	Period 3		
	Period 4		

Isto se ponavlja kod vremenske neovisnosti za različite periode prikupljanja podataka te za testiranje lokacijske neovisnosti za različite lokacije nošenja uređaja. Osim toga, testira se

utjecaj uključenih osjetila na točnost metode. Na ovaj način svrha metode postaje klasifikacija moda prijevoza promatranog zapis, a bez obzira kada je zapis prikupljen, od kojeg korisnika ili na način kako korisnik nosi mobilni uređaj tijekom prikupljanja podataka.

Kako SHL skup podataka nije objavljen u cijelosti onda je za potrebe testiranja u doktorskoj disertaciji vremenska komponenta prilagođena. Naime, cjelokupni skup podataka sastoji se od ukupno 152 dana, dok je javno objavljen skup podataka znatno manji i sastoji se od svega 9 dana. Izuzetak su podatci korisnika 1, za kojeg je objavljeno svih 82 dana, ali samo s pozicije mobilnog uređaja koji je nošen u džepu (kuk). Smanjenje broja dostupnih podataka zahtjeva prilagodbu podataka za testiranje vremenske neovisnosti metode. Za simulaciju vremenskog odmak prikupljanja podataka korišteno je uzorkovanje podataka korisnika 1, s razmakom od otprilike 30 dana. Podjela podataka prikazana je u Tablici 3.4. Korisnik 2 i 3 ne unose vremensku komponentu u skup podataka već se nalaze u raspodijeli kako se testiranje ne bi provodilo na samo jednom korisniku i jednoj lokaciji nošenja uređaja.

**Tablica 3.4:** Vremenska podjela SHL skupa podataka

Ukupan broj dana	Vremenski period 1	Vremenski period 2	Vremenski period 3	Vremenski period 4
Korisnik 1	1 – 3	30 – 33	60 – 63	79 – 82
Korisnik 2	1 – 3	1 – 3	1 – 3	1 – 3
Korisnik 3	1 – 3	1 – 3	1 – 3	1 – 3

Za ocjenu metode potrebno je poznavati preciznost i odziv sustava koji se izračunavaju koristeći vrijednosti sadržane u matrici zabune. Neka je  $x_{ij}$   $i, j$ -ti element matrice zabune koji predstavlja koliki je broj uzoraka koji izvorno pripadaju klasi  $i$  klasificirano kao klasa  $j$  onda se mjere odziva i preciznosti mogu zapisati Izrazima 3.15 i 3.16. U oba izraza  $C$  je broj klasa.

$$odziv_i = \frac{x_{ii}}{\sum_{j=1}^C x_{ij}} \quad (3.15)$$

$$preciznost_j = \frac{x_{jj}}{\sum_{i=1}^C x_{ij}} \quad (3.16)$$

Za ocjenu točnosti metode predložene su  $F1$  mjera i ukupna točnost.  $F1$  mjera izračunava se kao harmonijska sredina odziva i preciznosti 3.17. Dok je točnost  $T$  definirana kao omjer ispravno klasificiranih uzoraka i ukupnog broja uzoraka 3.18.

$$F1 = \frac{1}{C} \sum_{i=1}^C \frac{2 \cdot odziv_i \cdot preciznost_i}{odziv_i + preciznost_i} \quad (3.17)$$

$$T = \frac{\sum_{i=1}^C x_{ii}}{\sum_{i=1}^C \sum_{j=1}^C x_{ij}} \quad (3.18)$$

# Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

U ovom poglavlju opisana je novo razvijena višerazinska metoda za segmentaciju i klasifikaciju trajektorije. Osnovni zadatak metode je segmentacija trajektorije s obzirom na promjenu moda prijevoza te identifikacija moda prijevoza za svaki segment trajektorije. Predložena metoda je zasnovana na Markovljevim lancima i združenoj RF metodi te uključuje stvarnovremensku komponentu predviđanja. Osnovni koncept Markovljevih lanaca je prilagođen kako bi se iz podataka ekstrahirale matrice prijelaznih stanja kojima će se opisati uobičajeno ponašanje korisnika tijekom prelaska iz jednog u drugi mod prijevoza. Također hiperparametri metode RF su optimizirani te je za predviđanja korištena vjerojatnost pripadanja uzorka svakom modu prijevoza.

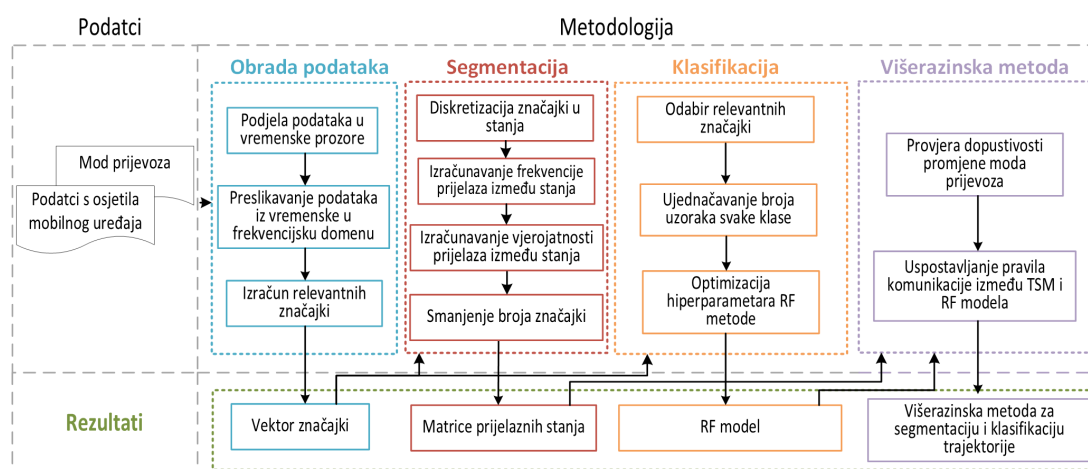
Poglavlje je strukturirano u pet osnovnih dijelova. U potpoglavlju 4.1 prikazani su postupci obrade podataka koji uključuju podjelu podataka u vremenske prozore kako bi se simuliralo stvarnovremensko okruženje, preslikavanje podataka iz vremenske u frekvencijsku domenu te izračun relevantnih značajki. Postupak diskretizacije podataka i izrade TSM-a, koje se koriste za segmentaciju trajektorije, opisan je u potpoglavlju 4.2. Klasifikacija modova prijevoza u segmentiranim dijelovima trajektorije koristeći metodu RF, optimizacija hiperparametara metode te metoda smanjenja broja značajki opisana je u potpoglavlju 4.3. U potpoglavlju 4.4 prikazan je algoritam višerazinske metode za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu koji obuhvaća metode opisane u potpoglavljima 4.2 i 4.3. Usporedba dobivenih rezultata opisane metode s rezultatima postignutim na testnom problemu dana je u potpoglavlju 4.5.

Izrada višerazinske metode prikazana je dijagramom na Slici 4.1 kroz četiri osnovna procesa: (i) obrada podataka, (ii) segmentacija trajektorije, (iii) klasifikacija modova prijevoza



i (iv) pravila komunikacije u višerazinskoj metodi. Ulazni podatci s osjetila mobilnog uređaja prvo prolaze kroz proces obrade podataka. Obrada podataka uključuje podjelu podataka u vremenske prozore, preslikavanje podataka iz vremenske u frekvenciju domenu te izračun relevantnih značajki. Rezultat obrade podataka je vektor značajki za sve podatke podijeljene u vremenske prozore. Dakle, broj vektora značajki je jednak broju vremenskih prozora podataka. Skup vektora značajki je ulazni podatak za segmentaciju trajektorije. Kako se za segmentaciju trajektorije koriste TSM u čijim su redcima i stupcima diskretne vrijednosti, potrebno je sve značajke diskretizirati. Tako se dobiva niz diskretnih vrijednosti za svaku razmatranu značajku (dalje u tekstu stanje). Provedene su tri različite metode diskretizacije te su uspoređeni rezultati točnosti metode za svaku vrstu TSM pojedinačno. Nakon formiranja vrijednosti redaka i stupaca TSM-a, izračunata je frekvencija prijelaza između svakog od stanja na temelju uzoraka u skupu podataka za učenje. Koristeći frekvencije svakog prijelaza i ukupnog broja prijelaza između stanja izračunata je vjerojatnost prijelaza između svih stanja u TSM. U konačnici je svaka razmatrana značajka prikazana dvjema TSM-ovima, pa je proveden postupak smanjenja broja značajki kako bi konačan broj TSM-ova u modelu bio čim manji.

Treći proces odnosi se na izradu metode za klasifikaciju modova prijevoza. Ulazni podatak ovog procesa je također vektor značajki koji je rezultat obrade podataka. Konačni RF model za klasifikaciju trajektorije s obzirom na mod prijevoza rezultat je tri procesa koja se odnose na odabir relevantnih značajki, ujednačavanje uzoraka svake klase te optimizaciju hiperparametara RF metode. Ujednačavanje uzoraka svake klase provedeno je zbog grupiranja klasa prema pravilima scenarija u testnom problemu. RF model te TSM su ulazni podatci višerazinske metode. U višerazinskoj metodi uspostavljaju se pravila komunikacije između dva ulazna modela te se postavljaju ograničenja koja moraju biti ispunjena da bi promjena moda prijevoza bila izvediva. Detaljniji opis svakog od potprocesa, koji vodi ka izradi višerazinske metode za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu, nalazi se u nastavku poglavlja.



Slika 4.1: Dijagram procesa izrade višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza

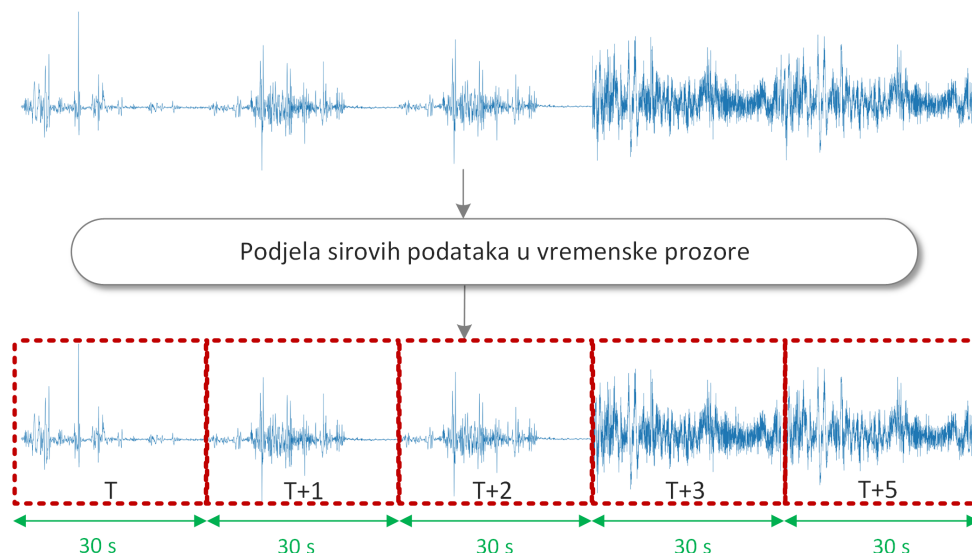
## 4.1 Obrada podataka

Obrada podataka smatra se jednim od najbitnijih postupaka tijekom podatkovnog učenja modela [109]. SHL skup podataka je skup podataka prikupljen s mobilnih uređaja za koje su karakteristični utjecaj šuma i nedostajućih vrijednosti zbog degradacije signala. Iz tog razloga sirovi podatci trebaju biti očišćeni. U prvom koraku iz cijelog skupa podataka uklonjene su nedostajuće vrijednosti i oni podatci za koje nije poznata klasa. Podatci su brisani horizontalno, dakle cijeli redak u kojem nije poznata klasa ili je prisutna nedostajuća vrijednost mjerenja na nekom osjetilu. Zatim su na podatcima provedena sljedeća tri postupka: (i) podjela podataka u vremenske prozore, (ii) preslikavanje podataka iz vremenske u frekvencijsku domenu, te (iii) izračun relevantnih značajki. Ovi postupci su jedinstveni za segmentaciju i klasifikaciju trajektorije. Dodatni postupci prije učenja modela koji se odnose na pojedinu metodu bit će objašnjeni u pojedinačnim poglavljima.

### 4.1.1 Podjela podataka u vremenske prozore

SHL skup podataka pohranjen je u formi trajektorija u kojima su pojedinačni zapisi vremenski razmaknuti 10 ms. Kako bi se provela simulacija stvarno vremenskog predviđanja cijeli skup podataka podijeljen je u ne preklapajuće vremenske prozore od 30 s. Dakle podatci sa svih osjetila prikupljeni tijekom perioda od 30 s pripadaju istom vremenskom prozoru, čijim istekom vremena započinje sljedeći vremenski prozor. Nakon brisanja podataka može se dogoditi da svi vremenski prozori ne sadrže isti broj zapisa, ali ne može se dogoditi da u isti vremenski prozor pripadaju dva podatka između kojih je proteklo više od 30 s, iako su potencijalno ti zapisi slijedni. Dakle, vremenski kriterij je osnovni kriterij tijekom podjele podataka u vremenske prozore. Na Slici 4.2 je prikazan primjer podjele sirovih podataka s jednog osjetila u vremenske prozore ( $T, T+1, \dots, T+5$ ). Sirovi podatci su najprije posloženi u vremenski slijed kako bi se simuliralo okruženje stvarno vremenskog pristizanja podataka. Zatim su tako posloženi podatci podijeljeni u vremenske prozore. Svaki vremenski prozor je veličine 30 s što znači da svaki vremenski prozor sadrži točno 3.000 mjerenja s jednog osjetila.

Veličina prozora nije jednoobrazno dogovorena u području klasifikacije trajektorije kretanja korisnika. Uvijek se radi o odmjeravanju između dužih vremenskih prozora koji mogu uključivati opširnije informacije o kretanju korisnika, ali mogu povećati pogrešnu klasifikaciju kada podatci dvaju različitih modova prijevoza budu uključeni u jedan vremenski prozor. Sa druge strane, kraći vremenski prozori poboljšavaju praćenje brzih prijelaza, ali neki bitni obrasci možda neće biti prepoznati [110]. U različitim istraživanjima se koriste različite veličine vremenskih prozora kao što je prikazano u Tablici 3.1 te se veličina vremenskog prozora uglavnom proizvoljno određuje bez prethodne analize. Sumiranjem takvih odabira može se primijetiti da je učestalo odabrana veličina vremenskog prozora od 5 s u HAR području. Odabir tako krat-



Slika 4.2: Proces podijele sirovih podataka u vremenske prozore

kih vremenskih prozora može se obrazložiti činjenicom da se u HAR području ne klasificiraju samo kretanja prometnom mrežom već i druge aktivnosti poput otvaranja vrata, hodanja po stepenicama i sl., čije je trajanje znatno kraće od primjerice vožnje u autobusu. Autori u [97] su istraživali utjecaj veličine vremenskih prozora na klasifikaciju aktivnosti u SHL skupu podataka s posebnim naglaskom na vrijeme prijelaza s jednog na drugi mod prijevoza. Razmatrane su veličine vremenskog prozora od 15 s, 20 s, 30 s i 1 min te su uspoređeni postignuti rezultati klasifikatora. Najbolji rezultati postignuti su s vremenskim prozorom od 30 s. U ovom istraživanju odabrana je upravo veličina vremenskog prozora od 30 s.

### 4.1.2 Preslikavanje podataka iz vremenske u frekvencijsku domenu

Podatci prikupljeni s osjetila mobilnih uređaja su zapravo niz numeričkih vrijednosti prikupljenih u nekom vremenskom periodu. Tako promatrani podatci mogu se smatrati nizom diskretnih podataka (diskretni signal) u vremenskoj domeni. Na diskretnim signalima u vremenskoj domeni može se primijeniti postupak diskretne Fourierove transformacije (engl. *Discrete Fourier Transform*, DFT) kako bi se prikazala frekvencijska domena signala, tj. diskretnih podataka [111]. Ovo je često korišten postupak u HAR području kako bi se bolje opisali podatci vremenskog prozora [112, 113]. U ovom istraživanju za preslikavanje podataka iz vremenske u frekvencijsku domenu korištena je metoda DFT-a, tj. algoritam brzih Fourierovih transformacija (engl. *Fast Fourier Transform*, FFT) [114]. Fourierova analiza je skupina matematičkih tehnika koja svoje osnovne postupke zasniva na dekompoziciji signala na zbroj sinusoida. Cilj dekompozicije signala je izdvojiti oblik signala koji je jednostavniji od izvornog. DFT je jedna od tehnika koja se koristi u digitaliziranim signalima. Naime, izvorno je Fourierova transformacija (engl. *Fourier Transform*, FT) namijenjena signalima koji su kontinuirani i periodični. Niti

jedno od ovih svojstava nije izravno zadovoljeno kod podataka poput podataka s mobilnih uređaja. U ovom istraživanju koristi se inačica DFT-a koja se primjenjuje na izvorno aperiodične i diskretne signale. U Izrazu 4.1 prikazan je osnovni izraz za preslikavanje niza diskretnih zapisa u vremenskoj domeni u frekvencijsku domenu [111]. Prema tome DFT se može definirati i kao metoda za preslikavanje konačnog diskretnog niza od  $N$  uzoraka  $\{x_0, x_1, \dots, x_{N-1}\}$  u niz od  $N$  kompleksnih brojeva za sve  $0 \leq k \leq N - 1$ .  $x_n$  je vrijednost funkcije u jednako razmaknutim vremenskim trenucima  $t = 0, 1, \dots, N - 1$ .

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi kn/N} \quad (4.1)$$

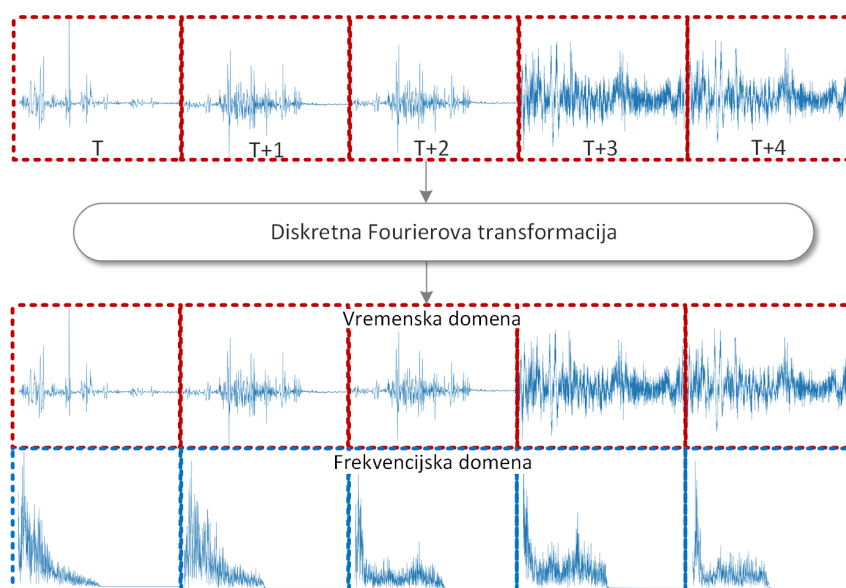
Izlazna vrijednost  $X_k$  je kompleksan broj koji predstavlja amplitudu i fazu sinusnog signala s frekvencijom  $k/N$  perioda po vremenskoj jedinici. Ovo se dobiva iz izraza za eksponencijalni (Eulerov) oblik kompleksnog broja, Izraz 4.2 [115].

$$e^{-i2\pi kn/N} = \cos(2\pi kn/N) + i\sin(2\pi kn/N) \quad (4.2)$$

Za provedbu DFT-a korišten je Cooley-Tukey FFT algoritam [116]. Cilj ovog algoritma je smanjenje složenosti postupka preslikavanja podataka iz vremenske u frekvencijsku domenu. Izračun DFT-a ima složenost  $\mathcal{O}(n^2)$  jer je za izračun svakog  $X_k$  potrebno  $n$  operacija. Osnovna ideja Cooley-Tukey FFT algoritam je "podijeli, pa vladaj" kako bi se smanjenjem veličine niza smanjila i složenost algoritma koja u konačnici iznosi  $\mathcal{O}(n \log n)$ .

Iako podatci prikupljeni s mobilnih uređaja u jednom vremenskom prozoru nisu periodični, uobičajeno se primjenjuje interpretacija cijelog signala kao beskonačne periodične duplikacije trenutnog vremenskog prozora [111]. Kako je veličina vremenskog prozora 30 s, a brzina uzorkovanja je 100 Hz, razmatran period signala je 3.000 uzoraka. Kao što je prikazano na Slici 4.3, nakon primjene FFT-a, svaki vremenski prozor je predstavljen podatcima u dvije domene: vremenskoj i frekvencijskoj domeni. Fundamentalne karakteristike valnog oblika zadanog signala mogu se izdvojiti iz vremenske domene, dok se iz frekvencijske domene može izdvojiti struktura signala. Točnije, magnituda predstavlja amplitudu kosinusnih valova na određenoj frekvenciji. Primjerice ako se promatra signal akceleracije u jednom vremenskom prozoru onda vremenska domena signala pokazuje amplitudu signala, na primjer, pokazujući ubrzanje li osoba naglo (na početku trčanja) ili sporijim tempom (početak kretanja vlaka). Frekvencijska domena prikazuje spektar signala. Primjerice, ako osoba hoda ujednačenim tempom od 1 koraka u 1 s, budući da signal sadrži 3.000 uzoraka, trebalo bi doći do skoka u frekvencijskoj domeni na 1 Hz ( $30/3.000 = 0,01$  stope uzorkovanja). Drugi primjer, za automobil koji ponavlja svoje ubrzanje svakih 3 s, dolazi do skoka u frekvencijskoj domeni na 0,33 Hz (udio od  $10/3.000 = 0,0033$  stope uzorkovanja). Dakle, ako frekvencijski spektar pokazuje više propusnosti visokih frekvencija, to znači da se ubrzanje često mijenjalo, a ako pokazuje propusnost niskih frekvencija,

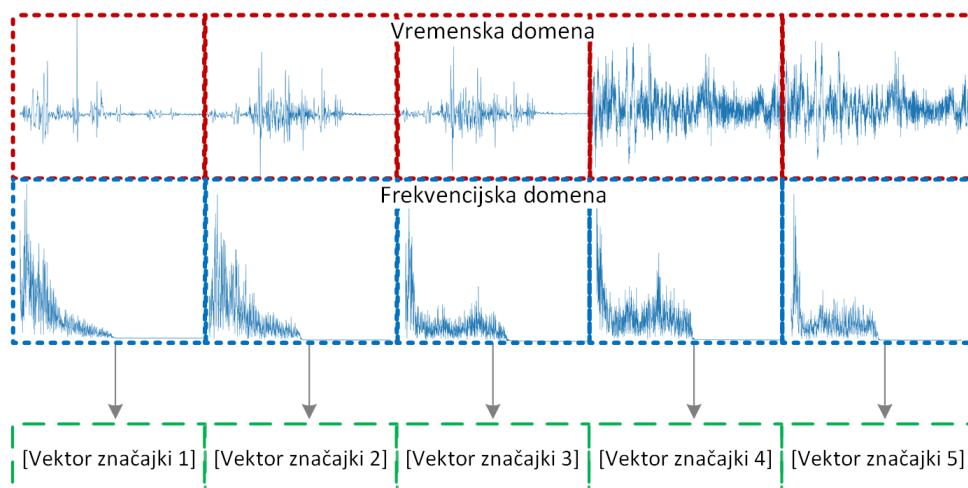
vrijednosti ubrzanja nisu se tako često mijenjale.



Slika 4.3: Preslikavanje podataka iz vremenske u frekvencijsku domenu

### 4.1.3 Izračun relevantnih značajki

Mnoga istraživanja pokazala su da izračun relevantnih značajki za problem koji se promatra uvelike utječe na konačne rezultate predloženih metoda [28, 117, 118]. Stoga je u ovom istraživanju izračunato 873 značajki kako bi se mogla provesti analiza utjecaja pojedine značajke na točnost metode za segmentaciju i klasifikaciju trajektorije. Sirovi podatci za svaki vremenski prozor se pretvaraju u vektor značajki kako je to prikazano na Slici 4.4. Dakle, svi podatci vremenske i frekvencijske domene svakog vremenskog prozora koriste se za izračunavanje značajki koje se onda pohranjuju u vektor značajki. Vektor značajki je ulazni podatak metode za segmentaciju i klasifikaciju trajektorije.



Slika 4.4: Proces pretvorbe sirovih podataka u vektor značajki

U Tablici 4.1 navedene su sve značajke koje su izračunate za svaki vremenski prozor s pripadajućim izrazom za izračun značajke te oznakom na koju domenu podataka se značajka odnosi.

Ne koriste se iste značajke za segmentaciju i za klasifikaciju. U potpoglavljima 4.2 i 4.3 su referirane pojedinačne značajke za svaku metodu. Metoda za segmentaciju trajektorije predviđa jedan od dva moguća ishoda (promjena ili bez promjene moda prijevoza), dok metoda za klasifikaciju predviđa jedan od osam mogućih ishoda (modova prijevoza), pa je u skladu s time broj primijenjenih značajki za klasifikaciju veći. Izračunate značajke se odnose na osnovne statističke veličine poput aritmetičke sredine, medijana, standardne devijacije, varijance, srednje apsolutne vrijednosti, međukvartalnog raspona, maksimalne i minimalne vrijednosti te korijena srednje kvadratne vrijednosti. Mjere poput mjere asimetrije i spljoštenosti podataka te razlike vršnih vrijednosti opisuju distribuciju podataka unutar jednog vremenskog prozora. Značajka  $z$  vrijednost pokazuje udaljenost podataka u trenutnom vremenskom prozoru od srednje vrijednosti pri čemu se udaljenost mjeri u standardnim devijacijama.

Faktori asimetrije, spljoštenosti i oblikovanja daju uvid u ponašanje podataka u jednom vremenskom prozoru, dok su faktori impulsa, porasta i razmaka usredotočeni na opis vršnih vrijednosti unutar vremenskog prozora.

Za frekvencijsku domenu svakog vremenskog prozora osim značajki navedenih u Tablici 4.1 izračunate su i značajke koje se zasnivaju na centru frekvencije i širini frekvencijskog pojasa. Značajke su predložene u istraživanju [81] kao odgovor na često izostavljanje dijela frekvencijskog pojasa za izračun značajki korištenjem nisko ili visoko propusnog filtra [119, 120, 121]. Umjesto da se iskustveno odabere širina frekvencijskog pojasa koja će se koristiti za izračun značajki, na ovaj način se ispituje utjecaj svih mogućih vrijednosti centra frekvencije,  $f_c$ , i širine frekvencijskog pojasa,  $\omega_b$ .

Kako je frekvencija uzorkovanja u SHL skupu podataka 100 Hz onda je prema Shannon-Nyquistovom teoremu najveća dozvoljena frekvencija u podacima 50 Hz. Širina frekvencijskog pojasa može poprimiti jednu od vrijednosti iz skupa  $\omega_b = \{1, 2, 3, 4, 5, 10, 15, 20, 25\}$  Hz. Za svaku širinu frekvencijskog pojasa ispituje se centar frekvencije iz raspona vrijednosti od  $\omega_b/2$  do  $50 - \omega_b/2$  s korakom od 1 Hz. Kod širine vremenskog pojasa od 1 Hz primjenjuje se korak od 0,5 Hz. Na ovaj način izračunato je niz potpojaseva  $\omega \in [f_c - \frac{\omega_b}{2}, f_c + \frac{\omega_b}{2}]$  od kojih se za svaki može izračunati koliko informacije nosi za razlikovanje između dva ili više moda prijevoza. Za svaki potpojas izračunate su dvije značajke: energija i omjer energije potpojasa i cijele širine frekvencijskog pojasa  $f_c$ . U Tablici 4.2 su prikazani svi potpojasevi s pripadajućim brojem izračunatih značajki. Ukupno je za svaki vremenski prozor i za svako osjetilo koji se prati izračunato 846 značajki koje se odnose na frekvencijske potpojaseve. Dakle, vektor značajki za svako osjetilo sadrži ukupno 873 značajki od kojih se 20 značajki odnosi na vremensku domenu, a 853 na frekvencijsku domenu.

Tablica 4.1: Značajke u vremenskoj i frekvencijskoj domeni

Naziv	Izraz *	T	F
Aritmetička sredina	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	✓	✓
Medijan	$X = \begin{cases} \frac{(N+1)^{ti}}{2} el, & \text{ako je N neparan} \\ \frac{N^{ti}}{2} el + \frac{(N+1)^{ti}}{2} el, & \text{ako je N paran} \end{cases}$	✓	✓
Standardna devijacija	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$	✓	✓
Varijanca	$\sigma = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$	✓	✓
Z-vrijednost	$z_i = \frac{x_i - \bar{x}}{s}$	✓	
Mjera asimetrije podataka	$g = \frac{\sum_{i=1}^N (x_i - \bar{x})^3 / N}{s^3}$	✓	
Mjera spljoštenosti	$k = \frac{\sum_{i=1}^N (x_i - \bar{x})^4 / N}{s^4}$	✓	
Maksimalna vrijednost	$x_{max} = MAX()$	✓	
Minimalna vrijednost	$x_{min} = MIN()$	✓	
Međukvartalni raspon	$IQR = Q_3 - Q_1$	✓	
Energija	$E = \sum_{i=1}^n  x_i ^2$	✓	
Razlika vršnih vrijednosti	$PTP = x_{max} - x_{min}$	✓	
Srednja apsolutna vrijednost	$\bar{x}_{abs} = \frac{\sum_{i=1}^n  x_i }{n}$	✓	
Korijen srednje kvadratne vrijednosti	$RMS = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$	✓	
Korijen varijance frekvencije	$RVS = \sqrt{\frac{\sum_{i=1}^n (f_i - \mu_{freq})^2 a_i}{\sum_{i=1}^n a_i}}$		✓
Centar frekvencije	$F_c = \frac{\sum_{i=1}^n f_i a_i}{\sum_{i=1}^n a_i}$		✓
Faktor impulsa	$F_i = \frac{x_{max}}{\bar{x}_{abs}}$	✓	
Faktor oblikovanja	$F_s = \frac{RMS}{\bar{x}_{abs}}$	✓	✓
Faktor porasta	$F_p = \frac{x_{max}}{RMS}$	✓	
Faktor razmaka	$F_r = \frac{x_{max}}{\frac{\sum_{i=1}^n \sqrt{ x_i }}{n}}$	✓	
Faktor asimetrije podataka	$F_g = \frac{g}{RMS^3}$	✓	
Faktor spljoštenosti	$F_k = \frac{k}{RMS^4}$	✓	

\*U svim izrazima sirovi podatci prikupljeni u jednom vremenskom prozoru označeni su vektorom  $x = \{x_1, x_2, \dots, x_n\}$ , amplituda u frekvencijskoj domeni vektorom  $a = \{a_1, a_2, \dots, a_n\}$ , frekvencija u frekvencijskoj domeni vektorom  $f = \{f_1, f_2, \dots, f_n\}$ ,  $Q_3$  označava vrijednost trećeg kvartila, a  $Q_1$  vrijednost prvog kvartila vektora x.

**Tablica 4.2:** Značajke frekvencijskih potpojava

Rd.broj	Naziv	Broj značajki
1	$E$ i $f_e$ sa širinom pojasa 1 Hz i korakom od 0,5 Hz	198
2	$E$ i $f_e$ sa širinom pojasa 2 Hz i korakom od 1 Hz	98
3	$E$ i $f_e$ sa širinom pojasa 3 Hz i korakom od 1 Hz	96
4	$E$ i $f_e$ sa širinom pojasa 4 Hz i korakom od 1 Hz	94
5	$E$ i $f_e$ sa širinom pojasa 5 Hz i korakom od 1 Hz	92
6	$E$ i $f_e$ sa širinom pojasa 10 Hz i korakom od 1 Hz	82
7	$E$ i $f_e$ sa širinom pojasa 15 Hz i korakom od 1 Hz	72
8	$E$ i $f_e$ sa širinom pojasa 20 Hz i korakom od 1 Hz	62
9	$E$ i $f_e$ sa širinom pojasa 25 Hz i korakom od 1 Hz	52

Podatci s osjetila mobilnog uređaja sadrže za svako osjetilo vrijednost mjerenja po  $x$ ,  $y$  i  $z$ -osi. Kako mobilni uređaj nije nošen u fiksnoj orijentaciji onda na vrijednosti mjerenja može utjecati promjena orijentacije mobilnog uređaja. Što znači da iste vrijednosti osjetila u različitim slučajevima zapravo mogu predstavljati kretanje u različitim smjerovima. Orijehtacija osjetila automatski mijenja distribuciju podataka, pa je teško razlučiti je li došlo do promjene vrijednosti mjerenja ili promjene u orijentaciji. Za rješavanje ovog problema koriste se razne metode koje mogu uključivati heurističku transformaciju i dekompoziciju singularnih vrijednosti za transformaciju signala u vremensku domenu koja nije osjetljiva na promjenu orijentacije [122]. Međutim, zastupljeniji pristup u literaturi je izračunavanje magnitude 3D signala kako bi se postigla orijentacijska neovisnost mjerenja s osjetila [60, 61, 62]. Opisani pristup korišten je i u ovom istraživanju, a magnituda svih osjetila izračunata je prema Izrazu 4.3 gdje je  $M$  magnituda, a  $S_x$ ,  $S_y$  i  $S_z$  vrijednosti osjetila duž  $x$ ,  $y$ , odnosno  $z$  osi.

$$M = \sqrt{S_x^2 + S_y^2 + S_z^2} \quad (4.3)$$

## 4.2 Segmentacija

Svrha segmentacije trajektorije u kontekstu klasifikacije modova prijevoza je priprema podataka za klasifikaciju tako da se razdvoje dijelovi trajektorije u kojima je korišten različit mod prijevoza. Drugim riječima, svrha ovog procesa je detektirati MTP i početak korištenja drugog moda prijevoza.

U ovom istraživanju predložena je nadzirana metoda za detekciju MTP-a pri čemu je korišten SHL skup podataka za učenje i testiranje metode. U svrhu osiguravanja testiranja točnosti



metode u stvarnom vremenu trajektorija je segmentirana kako je to opisano u poglavlju 4.1.1. Osim toga, izvorno su podaci označeni klasama koje odgovaraju korištenim modovima prijevoza, a za segmentaciju je potrebna isključivo informacija o prisutnosti događaja promjene moda prijevoza. Dakle, cijeli skup podataka je označen binarnim oznakama koje označavaju vremenski prozor s promjenom (klasa: 1) ili bez promjene (klasa: 0). Tako označen skup podataka ima značajno veći broj vremenskih prozora bez promjene nego s promjenom. Primjerice putovanje koje traje 80 minuta tijekom kojeg korisnik promijeni tri moda prijevoza imat će samo dva vremenska prozora u kojima se dogodila MTP, a 158 vremenskih prozora u kojim se nije dogodila promjena moda prijevoza. Iz tog razloga, rezultati točnosti metode ne mogu se ocijeniti isključivo ukupnom točnošću već se moraju uvesti i druge mjere: matrica zabune i odziv, a kako bi se dobio relevantan podatak o uspješnosti detekcije MTP-a.

### 4.2.1 Izrada matrica prijelaznih stanja

Osnovna ideja matrica prijelaznih stanja proizlazi iz modela HMM. HMM je jedan od najpopularnijih modela u literaturi za obradu slijednih podataka zbog ne zahtjevnih izračuna i sposobnosti modeliranja dinamičkog ponašanja vremenskih serija s jednostavnim modelom varijabli [123, 124, 125]. Osim toga, vjerojatnosni modeli su otporniji na šum koji je prisutan kod podataka s osjetila mobilnih uređaja.

TSM je matrica koja opisuje prijelaze između dva uzastopna vremenska prozora,  $t$  i  $t + 1$ , koristeći vektor značajki svakog vremenskog prozora. Za svaku značajku koja se koristi za segmentaciju trajektorije izrađuju se dvije matrice: matrica koja opisuje prijelaze između uzastopnih vremenskih prozora bez MTP-a i matrica koja opisuje prijelaze s MTP-om. Kao što je prikazano na Slici 4.5, svaka matrica se sastoji od  $k$  redaka i  $k$  stupaca. Vrijednosti indeksa stupaca i redaka matrice su stanja dobivena diskretizacijom vrijednosti značajki u skupu podataka za učenje (više o diskretizaciji u poglavlju 4.2.2). Dakle, svaka vrijednost vektora značajki trenutnog vremenskog prozora pripada točno jednom stanju značajke, a prijelaz definiraju dvije vrijednosti iste značajke koje pripadaju dva uzastopna vremenska prozora. Vrijednosti u matricama su dobivene prebrojavanjem događaja MTP-a tijekom prijelaza između dva susjedna vremenska prozora. Svaki prijelaz između stanja tijekom kojeg se dogodila MTP se dodjeljuje jednoj kombinaciji redak-stupac u matrici, tj. odgovarajućim vrijednostima promatrane značajke u prozoru  $t$  i prozoru  $t + 1$  u TSM koja prikazuje događaj MTP-a. Na primjer, za prijelaz  $Y_{0 \rightarrow 1}$ , bilježi se koliko puta,  $S_{Y_{0 \rightarrow 1}}$ , vremenski prozor  $t$  prelazi iz stanja 0 u stanje 1 u vremenskom prozoru  $t + 1$  za promatranu značajku, a da se pritom dogodio MTP za matricu na Slici 4.5a. Isti postupak se provodi za prijelaze u kojima nije došlo do MTP-a za matricu na Slici 4.5b.

Rezultat ovog postupka su dvije TSM matrice (promjena i bez promjene) za svaku značajku koje predstavljaju učestalost pojave prijelaza između pojedinih stanja. Za izradu opisanih TSM

$$\begin{array}{c}
\text{Stanje 0} \\
\text{Stanje 1} \\
\text{Stanje 2} \\
\vdots \\
\text{Stanje } k
\end{array}
\begin{array}{c}
\text{Stanje 0} \quad \text{Stanje 1} \quad \text{Stanje 2} \quad \dots \quad \text{Stanje } k \\
\left( \begin{array}{cccc}
S_{Y_{0 \rightarrow 0}} & S_{Y_{0 \rightarrow 1}} & S_{Y_{0 \rightarrow 2}} & \dots & S_{Y_{0 \rightarrow k}} \\
S_{Y_{1 \rightarrow 0}} & S_{Y_{1 \rightarrow 1}} & S_{Y_{1 \rightarrow 2}} & \dots & S_{Y_{1 \rightarrow k}} \\
S_{Y_{2 \rightarrow 0}} & S_{Y_{2 \rightarrow 1}} & S_{Y_{2 \rightarrow 2}} & \dots & S_{Y_{2 \rightarrow k}} \\
\vdots & \vdots & \vdots & & \vdots \\
S_{Y_{k \rightarrow 0}} & S_{Y_{k \rightarrow 1}} & S_{Y_{k \rightarrow 2}} & \dots & S_{Y_{k \rightarrow k}}
\end{array} \right)
\end{array}
\quad
\begin{array}{c}
\text{Stanje 0} \\
\text{Stanje 1} \\
\text{Stanje 2} \\
\vdots \\
\text{Stanje } k
\end{array}
\begin{array}{c}
\text{Stanje 0} \quad \text{Stanje 1} \quad \text{Stanje 2} \quad \dots \quad \text{Stanje } k \\
\left( \begin{array}{cccc}
S_{Y_{0 \rightarrow 0}} & S_{Y_{0 \rightarrow 1}} & S_{Y_{0 \rightarrow 2}} & \dots & S_{Y_{0 \rightarrow k}} \\
S_{Y_{1 \rightarrow 0}} & S_{Y_{1 \rightarrow 1}} & S_{Y_{1 \rightarrow 2}} & \dots & S_{Y_{1 \rightarrow k}} \\
S_{Y_{2 \rightarrow 0}} & S_{Y_{2 \rightarrow 1}} & S_{Y_{2 \rightarrow 2}} & \dots & S_{Y_{2 \rightarrow k}} \\
\vdots & \vdots & \vdots & & \vdots \\
S_{Y_{k \rightarrow 0}} & S_{Y_{k \rightarrow 1}} & S_{Y_{k \rightarrow 2}} & \dots & S_{Y_{k \rightarrow k}}
\end{array} \right)
\end{array}$$

(a) Matrica vremenskih prozora s MTP                      (b) Matrica vremenskih prozora bez MTP

**Slika 4.5:** Formiranje TSM za jednu značajku u vektoru značajki

korištene su samo trajektorije iz skupa podataka za učenje.

Kako bi se provela diskretizacija svake značajke potrebno je odrediti broj diskretnih stanja koji će se koristiti, tj. veličina  $k$  u TSM-u. Za određivanje broja  $k$  korištena su dva pravila: Freedman–Diaconisovo [126] i Sturgovo pravilo [127], koja se učestalo koriste u statistici za određivanje širine razreda (broj diskretnih stanja) u histogramu. Prema Sturgovom pravilu broj razreda ovisi o ukupnom broju zapisa ( $n$ ) u skupu podataka (Izraz 4.4). Problem ovog pravila je da može imati tendenciju zaglađivanja podataka kod velikog broja zapisa u skupu podataka. Iz tog razloga, broj razreda analiziran je i prema Freedman–Diaconisovom pravilu, Izraz 4.5. Kod ovog pravila za određivanje širine razreda koristi se međukvartilni raspon  $IQR$ , vektora svih zapisa  $x$  u skupu podataka te ukupan broj zapisa  $n$ . Prema Sturgovom pravilu, broj razreda u korištenom skupu podataka za učenje je 13 dok je prema Freedman–Diaconisovom pravilu najmanji broj razreda 17. Za konačan broj razreda u ovom istraživanju odabrana je srednja vrijednost ova dva pravila što iznosi 15. Kao rezultat toga, svaka TSM ima  $15 \times 15 = 225$  prijelaza između stanja svake značajke.

$$k_S = \lceil \log_2 n \rceil + 1 \quad (4.4)$$

$$k_{FD} = 2 \frac{IQR(x)}{\sqrt[3]{n}} \quad (4.5)$$

Nakon provedbe prebrojavanja prijelaza između stanja u svakom elementu TSM-a je pohranjena frekvencija pojave pojedinih prijelaznih stanja u skupu podataka za učenje. Kako bi se takva matrica mogla koristiti za predviđanje budućih prijelaza potrebno je izračunati vjerojatnost pojave MTP-a za svako prijelazno stanje. Vjerojatnost je izračunata za svako prijelazno stanje koristeći Izraz 4.6 pri čemu je  $S_{Y_{i \rightarrow j}}$  broj pojavljivanja događaja MTP-a za svaki prijelaz između stanja  $i$  i stanja  $j$ . U nazivniku je zbroj svih elemenata matrice, tj. ukupan broj događaja MTP-a.

$$P_{ij} = \frac{S_{Y_{i \rightarrow j}}}{\sum_{i=0}^k \sum_{j=0}^k S_{Y_{i \rightarrow j}}} \quad (4.6)$$

U TSM matrici za prijelaze kada nije bilo MTP-a,  $S_{Y_{i \rightarrow j}}$  predstavlja broj pojavljivanja prijelaza bez MTP-a. Koristeći TSM matrice u fazi testiranja, za svaka dva uzastopna vremenska prozora predviđa se da li se dogodila promjena moda prijevoza ili nije. Postupak glasovanja opisan je u poglavlju 4.2.4.

### 4.2.2 Diskretizacija podataka

Diskretizacija podataka je proces pretvorbe kontinuirane varijable u diskretni oblik tako što se kontinuirane vrijednosti dijele u odgovarajuće razrede kojih ima konačan broj. Određivanje razreda je potrebno kako bi se odredila stanja, tj. vrijednosti indeksa stupaca i redaka TSM-a. Diskretizacija podataka je provedena na tri različita načina te su dobivene tri različite vrste TSM-a. Sve vrste su testirane zasebno na skupu podataka za testiranje te je odabrana vrsta matrica kojom je postignuta najveća točnost predviđanja MTP-a.

#### Diskretizacija metodom jednakog razmaka

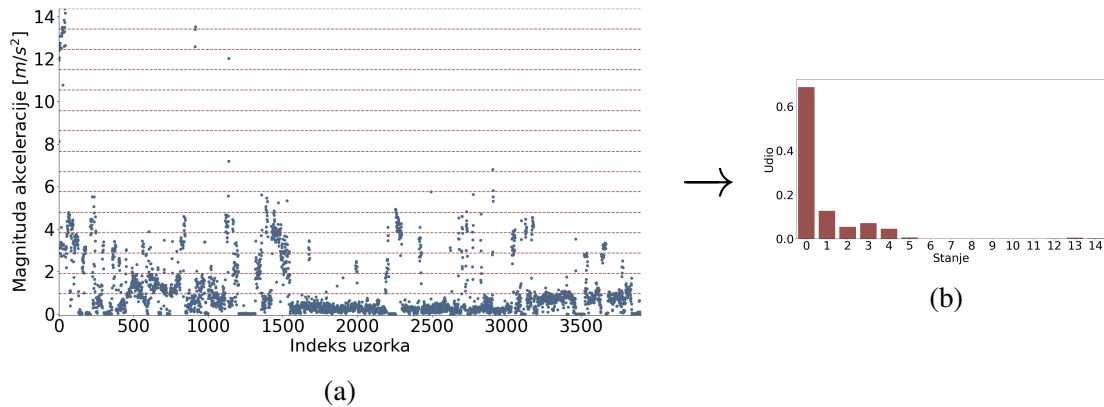
Diskretizacija metodom jednakog razmaka je najjednostavnija metoda za diskretizaciju podataka u kojoj se raspon vrijednosti neke značajke dijeli s unaprijed određenim brojem razreda prema Izrazu 4.7 gdje su  $X_{max}$  i  $X_{min}$  maksimalna i minimalna vrijednost svih zapisa neke značajke, a  $N$  ukupan broj zapisa [128].

$$w = \frac{X_{max} - X_{min}}{N} \quad (4.7)$$

Osnovna prednost ove metode diskretizacije podataka je što se zadržava stvarna distribucija podataka. Ipak, ako podatci nisu ravnomjerno raspoređeni, različite su gustoće u nekim dijelovima ili imaju stršće vrijednosti, ova metoda će generirati mnoge prazne razrede ili razrede s malim udjelom podataka. Na Slici 4.6 je prikazan primjer diskretizacije podataka srednje vrijednosti magnitude akceleracije. Na  $x$ -osi Slike 4.6a su prikazani indeksi zapisa, a na  $y$ -osi srednja vrijednost magnitude akceleracije. Raspon podataka, tj. razlika između minimalne i maksimalne vrijednosti akceleracije, podijeljen je u 15 razreda jednake širine (crvene isprekidane linije). Svaki razred predstavlja jedno stanje u TSM matrici. Rezultat diskretizacije prikazan je na Slici 4.6b, gdje je prikazana distribucija podataka između razreda. Broj podatkovnih točaka u svakom razredu podijeljen je s ukupnim brojem podatkovnih točaka. Jednakim razmakom razreda je uzrokovan mali udio podataka za neke razrede, primjerice razrede 5-14.

#### Diskretizacija metodom jednake frekvencije

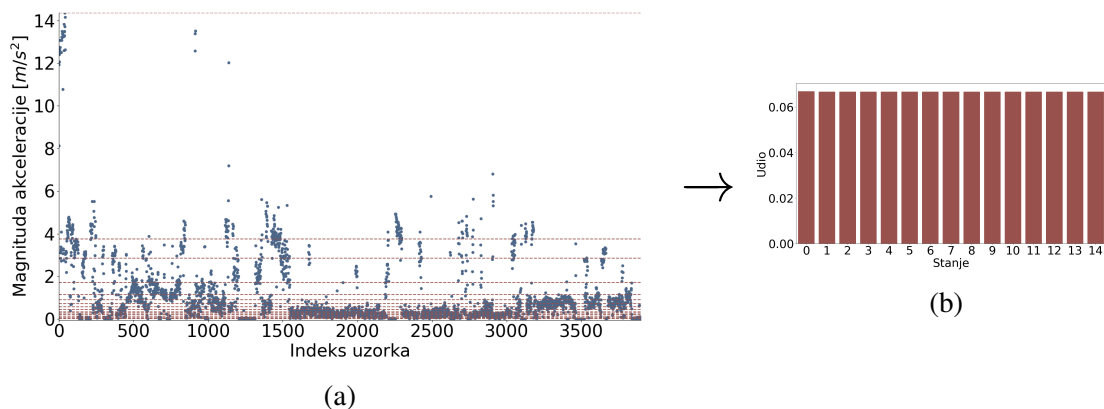
Metodom diskretizacije podataka uz pravilo jednake frekvencije razdvajaju se podatci u unaprijed definiran broj razreda pri čemu svaki razred sadrži jednak broj zapisa [128]. Na ovaj način



**Slika 4.6:** Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda jednakog razmaka

generira se umjetna distribucija podataka, ali su eliminirani prazni razredi i razredi s malim udjelom podataka.

Na Slici 4.7 prikazan je primjer diskretizacije podataka srednje vrijednosti magnitude akceleracije. Na  $x$ -osi Slike 4.7a su prikazani indeksi zapisa, a na  $y$ -osi srednja vrijednost magnitude akceleracije. Širine pojedinih razreda značajno odstupaju, primjerice razred 14, no broj zapisa svakog razreda je približno jednak. Približno jednaka distribucija podataka po razredima prikazana je na Slici 4.7b.



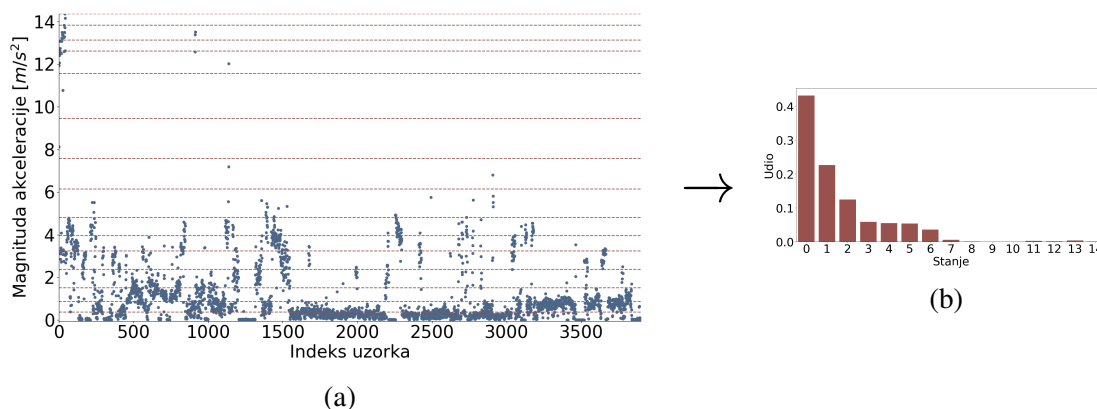
**Slika 4.7:** Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda jednake frekvencije

### Diskretizacija metodom $k$ sredine

Grupiranje je jedan od učestalih postupaka u analizi podataka kako bi se shvatila struktura promatranih podataka. Može se definirati kao zadatak razdvajanja podskupina u podacima tako da su zapisi u istoj podskupini vrlo slični dok su zapisi u različitim podskupinama vrlo različiti. Jedan od algoritama koji se koristi za grupiranje je algoritam  $k$  sredina, Algoritam 3.3. I u ovom kontekstu algoritam  $k$  sredina je iterativni postupak koji pokušava podijeliti skup

podataka u  $k$  različitih podskupina koje se ne preklapaju i svaki zapis pripada točno jednoj podskupini [129]. Rezultat algoritma  $k$  sredina je zapravo centar podskupine i točke koje pripadaju svakoj podskupini. Za potrebe diskretizacije rezultat ovog algoritma je gornja i donja granica svake podskupine.

Na Slici 4.8 prikazan je rezultat diskretizacije srednje vrijednosti magnitude akceleracije algoritmom  $k$  sredina. Dobiveni razredi su različitih širina, Slika 4.8a, tako su razredi koji obuhvaćaju vrijednost akceleracije  $0 - 5 \text{ m/s}^2$  manjih širina u odnosu na razrede većih vrijednosti akceleracije. Ovo je uzrokovano i većom frekvencijom zapisa u području niže akceleracije. Ako se veličina razreda uspoređi s prethodnim metodama diskretizacije, rezultat je kombinacija rezultata obje metode. Naime, pri većim frekvencijama zapisa je učestalija dioba razreda dok je kod manjih frekvencija dijeljenje na razrede manje učestalo. Ipak u usporedbi s razredima nastalim metodom diskretizacije jednakom frekvencijom, odstupanja širina razreda su manja.



**Slika 4.8:** Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda grupiranih algoritmom  $k$  sredine

Distribucija podataka unutar razreda prikazana je na Slici 4.8b gdje je prisutna veća zastupljenost podataka u stanjima nižih akceleracija dok razredi viših akceleracija imaju značajno manje podataka.

### 4.2.3 Smanjenje broja relevantnih značajki

Za segmentaciju trajektorije razmatrane su sve značajke navedene u Tablici 4.1. Dakle, isključene su značajke koje se odnose na promatranje različitih širina frekvencijskog pojasa značajki. Sve značajke izračunate su za podatke s akcelerometra, žiroskopa, gravitacije i magnetometra što rezultira s 432 značajki osjetila ( $27 \text{ značajki} \times (3 \text{ osi} + 1 \text{ magnituda}) \times 4 \text{ osjetila} = 432$ ). Ako se uzme u obzir da je svaka značajka opisana s dvije TSM (jedna za prijelaze kada se dogodila promjena moda prijevoza, a druga kada se promjena nije dogodila) onda je jasno da je bitno provesti analizu značajki kojom bi se broj značajki, a time i matrica, smanjio samo na one koje doprinose novim informacijama. Drugi razlog analize značajki je detektirati koje značajke

donose više informacije u identifikaciji trenutka kada se događa promjena moda prijevoza osobito zbog disbalansa između broja vremenskih prozora u kojima se dogodila promjena moda prijevoza i onih koji predstavljaju vremenske prozore u kojima nema promjene moda prijevoza.

Proces smanjenja broja značajki podijeljen je u dva dijela: (i) značajke su sortirane prema testu analizi varijance (engl. *ANalysis Of VAriance*, ANOVA) te su, (ii) iz skupa značajki uklonjene značajke sa značajnim utjecajem na detekciju lažno pozitivnih primjeraka (klasifikacija vremenskih prozora u kojima se nije dogodila promjena kao one u kojima se dogodila promjena).

ANOVA test je statistička metoda koja se koristi za provjeru srednjih vrijednosti značajki za dvije ili više klasa koje se međusobno značajno razlikuju [130]. Drugim riječima, vrijednost dobivena ANOVA testom kazuje koliko dobro neka značajka razlikuje dvije klase. Za izračun ANOVA testa koriste se dvije veličine: varijabilnost unutar grupa i varijabilnost između grupa. Obje veličine koriste zbroj kvadrata za određivanje disperzije u podacima. Varijabilnost unutar grupa  $SS_u$  pokazuje disperziju podataka u cijelom skupu podataka pri čemu se zasebno razmatraju podatci koji pripadaju različitim klasama, Izraz 4.8, te varijabilnost između grupa  $SS_i$  koja pokazuje varijabilnost podskupa podataka koji pripadaju istoj klasi u odnosu na cijeli skup podataka, Izraz 4.9. U oba izraza, Izraz 4.8 i Izraz 4.9,  $k$  predstavlja broj klasa,  $x_{ij}$  je  $i$ -ta vrijednost promatrane značajke koja pripada klasi  $j$  u skupu podataka veličine  $N$ .  $\bar{x}$  je srednja vrijednost svih zapisa za promatranu značajku, a  $\bar{x}_j$  srednja vrijednost svih zapisa koji pripadaju klasi  $j$ . Ukupan broj zapisa neke klase označen je s  $n_j$ . Postavljanjem dobivenih vrijednosti za  $SS_u$  i  $SS_i$  u omjer dobiva se ANOVA  $F$  vrijednost, Izraz 4.10, te na ovaj način ANOVA  $F$  vrijednost pokazuje omjer varijanci između i unutar podataka koji pripadaju nekoj klasi [130].

$$SS_u = \frac{\sum_{i=1}^N \sum_{j=1}^k (x_{ij} - \bar{x}_j)^2}{N - k} \quad (4.8)$$

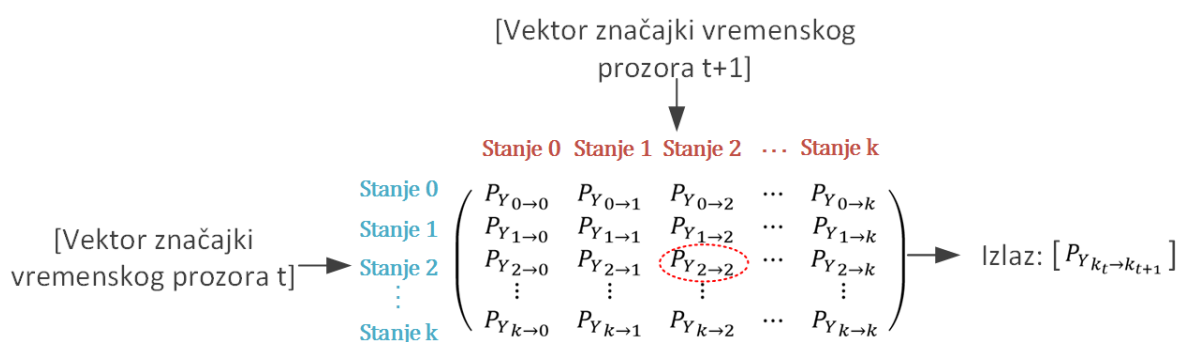
$$SS_i = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2}{k - 1} \quad (4.9)$$

$$F = \frac{SS_u}{SS_i} \quad (4.10)$$

Nakon primjene ANOVA testa odabrane značajke će tijekom predviđanja davati prednost klasi promjene stoga su iz skupa značajki uklonjene značajke koje uzrokuju detekciju lažno pozitivnih primjeraka. Za testiranje ovisnosti nezavisnih varijabli (značajki) i zavisne varijable (klasa) korišten je test stope lažno pozitivnih uzoraka.

#### 4.2.4 Testiranje metode

Testiranje točnosti metode za predviđanje promjena moda prijevoza koristeći TSM provedeno je na testnom skupu podataka. Testni skup podataka sadrži 25 % cjelokupnog skupa podataka. Za svaki prijelaz između vremenskih prozora ispituje se onoliko parova vrijednosti vjerojatnosti koliko je odabranih značajki koje se koriste za predviđanje. Kao što je prikazano na Slici 4.9, vjerojatnost svakog prijelaza određena je s dva vektora značajki, vektorom značajki vremenskog prozora  $t$  i susjednim vremenskim prozorom  $t + 1$ . Vrijednost u vektoru značajki za vremenski prozor  $t$  predstavlja odgovarajući redak matrice, dok vrijednost u vektoru značajke za vremenski prozor  $t + 1$  predstavlja odgovarajući stupac matrice. Jedan element matrice (crvena isprekidana elipsa na Slici 4.9) predstavlja vjerojatnost da se promjena moda prijevoza dogodila ili nije (ovisno o promatranoj matrici) između vremenskog prozora  $t$  i vremenskog prozora  $t + 1$ .



Slika 4.9: Testiranje TSM-a

Ako je vjerojatnost promjene moda prijevoza u TSM-u, koja predstavlja vremenske prozore u kojima se dogodila promjena moda prijevoza, veća od vjerojatnosti da nema promjene u TSM-u koji predstavlja vremenske prozore u kojima se nije dogodila promjena moda prijevoza, tada promatrana značajka daje glas za događaj promjene moda prijevoza. U suprotnom daje glas za događaj bez promjene moda prijevoza. Postupak se ponavlja za sve vrijednosti značajki u vektoru  $t$  i vektoru  $t + 1$ . Izlazna vrijednost (promjena moda prijevoza ili bez promjene) se odabire na temelju većeg broja glasova svih TSM-a.

U ovoj fazi izračunata je točnost modela na temelju usporedbe stvarne i predviđene vrijednosti. Zbog već opisanog disbalansa u broj uzoraka svake klase za ocjenu točnosti metode korištene su mjere ukupne točnosti, Izraz 3.18, i odziva, Izraz 3.15.

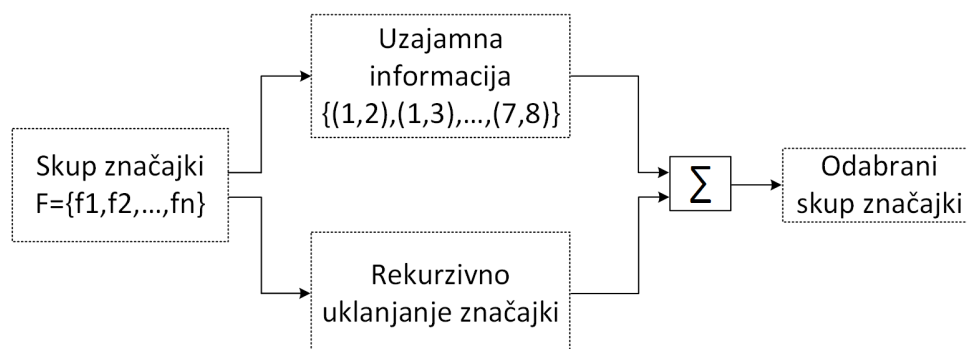
## 4.3 Klasifikacija

Osnovni zadatak klasifikacije modova prijevoza je raspoznavanje modova prijevoza u segmentiranom dijelu trajektorije. Modovi prijevoza su prilagođeni scenarijima testnog skupa podataka opisanih u poglavlju 3.3 za potrebe testiranja metode za klasifikaciju. Skup podataka je obrađen na način kako je to opisano u poglavlju 4.1.

Korištenjem skupa podataka za učenje razvijen je niz pravila RF modela kojim se nastoji čim bolje opisati razmatran problem razvrstavanja. Međutim povećanjem broja značajki može se dogoditi da model počinje dobro opisivati skup podataka na kojem je provedeno učenje, ali ne opisuje dovoljno dobro problem čijem rješavanju je namijenjen. Takvi modeli pokazuju dobre rezultate na skupu podataka za učenje, ponekad i za testiranje, ali se u procesu validacije značajno smanjuje točnosti modela. Kako bi se utjecalo na prenaučenos modela provode se metode odabira relevantnih značajki te se prilagođavaju parametri modela. Tako se postiže generalizacija pravila tijekom procesa učenja. Oba procesa zajedno s algoritmom za izgradnju RF-a opisana su u sljedećim poglavljima.

### 4.3.1 Odabir relevantnih značajki

Za klasifikaciju modova prijevoza razmatrane su sve značajke opisane u poglavlju 4.1.3. Kako bi se uklonile značajke koje unose šum u podatke i istaknule one značajke koje donose najviše informacija za razlikovanje modova prijevoza, korištene su dvije metode odabira značajki: izračun uzajamne informacije te metoda rekurzivnog uklanjanja značajki. Metode se provode paralelno, Slika 4.10. Ulazni parametar obje metode je skup svih značajki  $F = \{f_1, f_2, \dots, f_m\}$ , a konačni rezultat je unija rezultata obje metode.



Slika 4.10: Proces smanjenja broja značajki

Uzajamna informacija (engl. *Mutual Information*, MI) je mjera kojom se kvantificira količina informacije koju sadrži jedna varijabla  $X$  o varijabli  $Y$ . Izraz 4.11 prikazuje izračun MI vrijednosti pri čemu  $x_j$  predstavlja značajku, a  $y_i$  klasu problema [131].



$$MI(X, Y) = \sum_{i=1}^{N_y} \sum_{j=1}^{N_x} P(x_j, y_i) \log \frac{P(x_j, y_i)}{P(x_j)P(y_i)} \quad (4.11)$$

Kako bi konačni podskup značajki sadržavao značajke koje uspješno opisuju svaki od modova prijevoza, postupak izračuna MI vrijednosti provodi se za svaki par klasa [81]. Rezultat formiranja parova klasa iz 8 modova prijevoza je 28 različitih kombinacija modova prijevoza. Svaka kombinacija sadrži 2 moda prijevoza, tzv. par modova prijevoza ( $\binom{8}{2} = 28$ ). Dakle, MI vrijednost izračunava se za sve značajke za primjerice par klasa Hodanje-Trčanje, Hodanje-Bicikl, Hodanje-Automobil itd. što rezultira s ukupno 28 podskupova značajki za svaku kombinaciju modova prijevoza. Za svaki par modova prijevoza odabrano je 20 značajki čija je MI vrijednost najveća. Ovim postupkom neke značajke mogu biti odabrane više puta, pa je ukupan broj značajki akceleracije 137, magnetometra 126 te žiroskopa 138.

Paralelno se provodi metoda rekurzivnog uklanjanja značajki (engl. *Recursive Feature Elimination*, RFE). Ova metoda pripada skupini metoda omotača koje zapravo provode postupak uklanjanja značajki koji se zasniva na rezultatima unutarnje metode [132]. Algoritam 4.1 prikazuje osnovni postupak metode rekurzivnog uklanjanja značajki.

---

**Algoritam 4.1** Metoda rekurzivnog uklanjanja značajki, [132]

---

- 1: Ulazni parametri: skup podataka za učenje  $x = \{x_1, x_2, \dots, x_n\}$ , skup značajki  $F = \{f_1, f_2, \dots, f_m\}$ , broj značajki za odabir  $Z = \{m, m-1, m-2, \dots, 10\}$
  - 2: Izlazni parametri: skup značajki koji rezultira najvećom točnošću
  - 3:  $f_{best} = F$
  - 4:  $t_{best} = \text{minVrijednost}()$
  - 5: **for** svaki  $z$  u  $Z$  **do**
  - 6:   Odaberi  $z$  značajki  $f$
  - 7:   Nauči RF model s  $f$  značajki
  - 8:   Izračunaj točnost  $T_z$  modela na testnom skupu podataka
  - 9:   **if**  $T_z > t_{best}$  **then**
  - 10:      $f_{best} = f$
  - 11:      $t_{best} = T_z$
  - 12:   **end if**
  - 13:   Sortiraj značajke prema važnosti značajke
  - 14:   Izbaci značajku najmanje važnosti
  - 15: **end for**
- 

Ulazni parametri u algoritam su skup podataka za učenje  $x = \{x_1, x_2, \dots, x_n\}$  koji uključuje sve značajke iz skupa podataka pohranjene u skupu  $F = \{f_1, f_2, \dots, f_m\}$ . Broj značajki koji će biti ispitan u svakoj iteraciji algoritma definiran je parametrom  $z$ . Izlazni parametar je podskup značajki koji daje najveću točnost unutarnje metode  $f_{best}$ , u ovom slučaju odabrana je metoda RF. U svakoj iteraciji algoritma na temelju  $z$  značajki uči se model RF te se ispituje njegova točnost. Ako je dobivena točnost  $T_z$  veća od do sada najbolje postignute točnosti  $t_{best}$  onda se trenutni podskup značajki zapisuje u vektor  $f_{best}$ . Nakon toga se značajke sortiraju prema njho-

voj važnosti te se uklanja jedna značajka čija je važnost najmanja (linija 5- 15 u Algoritmu 4.1).

Odabrani skup značajki je rezultat unije podskupa značajki nakon provedbe RFE metode i podskupa značajki s najvećom MI vrijednošću. Neke značajke pojavljuju se u oba podskupa, pa je konačan broj značajki za klasifikaciju za akcelerometar 147, magnetometar 146 te žiroskop 144 značajki. Na ovaj način zadržane su značajke koje sadrže najviše informacija za razlikovanje pojedinog moda prijevoza (MI vrijednost) te su uklonjene značajke koje u međusobnoj interakciji stvaraju redundantnost.

### 4.3.2 Metoda slučajnih šuma

Za klasifikaciju modova prijevoza u stvarnom vremenu korištena je metoda slučajne šume. RF je metoda koja pripada združenim klasifikatorima koji koriste više jednostavnih klasifikatora za donošenje odluke o klasi kojoj uzorak pripada. Kod RF metode jednostavni klasifikatori su stabla odluke. Za izgradnju svakog stabla odluke koristi se procedura odabira podskupa podataka iz skupa podataka za učenje prikazana u Algoritmu 4.2. Ulazni podatak algoritmu RF je skup podataka za učenje  $x = \{x_1, x_2, \dots, x_n\}$  i broj stabala  $N$ . Skup podataka dijeli se  $N$  puta na slučajan način formirajući  $N$  podskupova podataka (linija 2- 5 Algoritma 4.2).

---

#### Algoritam 4.2 Algoritam slučajne šume, [107]

---

- 1: Ulazni parametri: skup podataka za učenje  $x = \{x_1, x_2, \dots, x_n\}$ , broj stabala u šumi  $N$
  - 2: **for** svaki  $n$  in  $N$  **do**
  - 3:   Na slučajan način odaberi podskup  $D_n$  skupa podataka za učenje  $x$
  - 4:     $CART(D_n)$
  - 5: **end for**
- 

Za svaki podskup podataka poziva se algoritam CART kako bi se izgradilo jedno stablo u šumi. Podjela skupa podataka za učenje provodi se sa zamjenom što znači da se isti uzorak može pronaći u više podskupova podataka. Na ovaj način postiže se raznolikost stabala u šumi, od kojih neka stabla bolje generaliziraju jednu skupinu modova prijevoza, dok druga daju prednost preostalim modovima prijevoza. Uvođenje slučajne komponente odabira podskupova podataka najviše pridonosi sprječavanju prenaučivosti modela na skup podataka za učenje.

Postupak izgradnje svakog stabla CART algoritmom prikazan je u Algoritmu 4.3. Ulazni podatci algoritma su podskup podataka  $D_n$ , skup značajki  $F$  i uvjet prekida grananja. U svakoj iteraciji algoritma odabire se značajka koja donosi najveću informacijsku dobit te se ista pohranjuje u varijablu *podijeli<sub>f</sub>* (linija 7- 13 Algoritma 4.3). Informacijska dobit izračunava se na temelju entropije, Izraz 3.4. Značajka, čija je informacijska dobit najveća, koristi se za dijeljenje podskupa podataka na dva dijela (binarno grananje) koristeći metodu Podijeli() koja prima dva parametra: podskup skupa podataka za učenje  $D_n$  i naziv značajke *podijeli<sub>f</sub>* na temelju koje se podskup dijeli. Binarno grananje je postupak kojim se određuje vrijednost značajke kojom će se uzorci podijeliti u dvije skupine (linija 14 Algoritma 4.3).

---

**Algoritam 4.3** CART, [100]

---

```
1: Ulazni parametri: podskup podataka  $D_n$ , skup značajki  $F$ , uvjetPrekida
2: if  $D_n \neq \emptyset$  and  $F \neq \emptyset$  then
3:   Na slučajan način odaberi  $n\_znacajki$  iz  $F$ 
4:   while uvjetPrekida do
5:      $Max\_dobitakInformacije = min\_vrijednost$ 
6:      $podijeli_f \leftarrow \emptyset$ 
7:     for svaku  $f$  in  $n\_znacajki$  do
8:       Izračunaj informacijsku dobit  $DI_f$ 
9:       if  $DI_f > max\_dobitakInformacije$  then
10:         $max\_dobitakInformacije = DI_f$ 
11:         $podijeli_f = f$ 
12:      end if
13:    end for
14:    Podijeli ( $D_n, podijeli_f$ )
15:  end while
16: end if
```

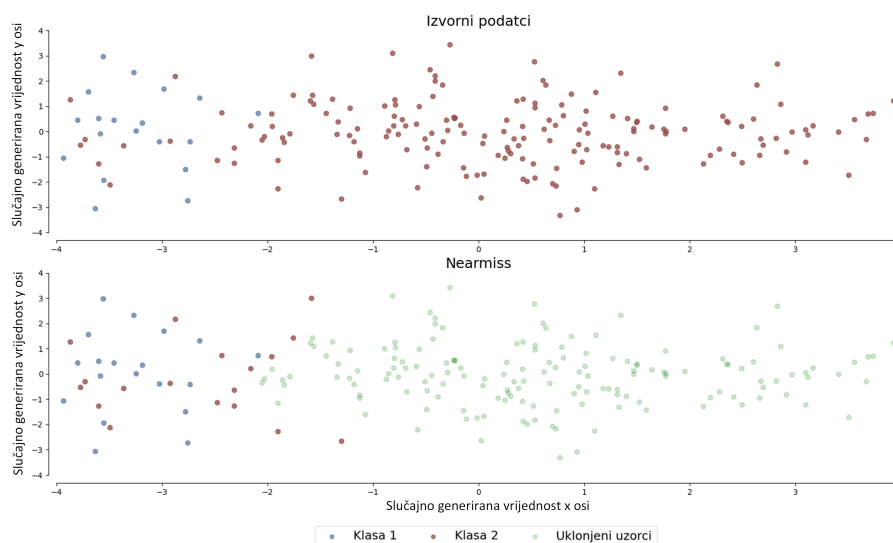
---

Postupak se ponavlja sve dok nije zadovoljen uvjet prekida. Uvjeta prekida može biti više, neki od najčešće korištenih su maksimalna dubina stabla, minimalan broj uzoraka za podjelu čvora ili minimalan broj uzoraka u čvoru odluke (list stabla). Ako je postavljeno više uvjeta prekida grananja, onda se grananje izvršava sve dok jedan od uvjeta ne dosegne svoju graničnu vrijednost.

### 4.3.3 Ujednačavanje broja uzoraka

Iako metoda RF pripada skupini metoda koje nisu osjetljive na neujednačenost broja podataka različitih klasa u skupu podataka, eksperimentalno je zaključeno da se postiže bolji rezultat klasifikacije modova prijevoza ako su uklonjena značajna odstupanja u broju klasa osobito zbog grupiranja modova prijevoza tijekom ispitivanja različitih scenarija testnog problema. Stoga je provedena metoda grupiranja podataka kako bi se broj podataka za sve modove prijevoza koji imaju znatno veći broj podataka smanjio.

Za ujednačavanje broja klasa korišten je algoritam *NearMiss* [133]. Osnovna ideja algoritma je promatranje zapisa iz skupa podataka kao problem grupiranja, u kojem zapisi koji pripadaju jednoj klasi čine jednu grupu. Na Slici 4.11 je prikazan primjer podataka u izvornom obliku (gornja slika) pri čemu podatci pripadaju dvjema klasama, plavi kružići pripadaju klasi 1, a crveni klasi 2 koja je ujedno i brojnija klasa. Nakon smanjenja broja uzoraka *NearMiss* algoritmom (donja slika) zadržani su uzorci iz klase 2 koji su najbliži uzorcima klase 1. Zelenim krugovima su prikazani uzorci koji nisu odabrani. Skup podataka izvorno sadrži 20 umjetno generiranih značajki na kojima je primijenjen algoritam *NearMiss*, ali su zbog preglednosti vizualizacije odabrane samo 2 značajke iz skupa podataka.



**Slika 4.11:** Postupak ujednačavanja broja uzoraka

Iz svake grupe odabire se unaprijed određen broj zapisa prema sljedećim koracima:

1. Izračunavaju se udaljenosti svakog zapisa brojnije klase do svih zapisa manje brojne klase;
2. Za svaki zapis manje brojne klase odabire se  $M$  najbližih zapisa brojnije klase;
3. Između odabranih zapisa brojnije klase odabiru se oni čija je prosječna udaljenost do tri najbliža susjeda najveća.

Broj zapisa za svaki mod prijevoza koji se zadržava ne mora biti jednak. U ovom istraživanju broj uzoraka svih klasa postavljen je na broj uzoraka druge najmanje klase prema broju uzoraka. Time je broj korištenih podataka za učenje dodatno smanjen.

#### 4.3.4 Optimizacija hiperparametara modela

Za svaki model potrebni su parametri kojima se postiže unaprjeđenje istog. Dio parametara može se naučiti iz podataka, a dio njih mora biti unaprijed definiran tijekom kreiranja modela. Parametri koji moraju biti unaprijed definirani nazivaju se hiperparametrima modela. Pomnim odabirom hiperparametara RF modela može se utjecati na više komponenti RF metode. Primjerice jedan od parametara RF modela je najmanji broj uzoraka za dijeljenje čvora čijim se povećanjem smanjuje vjerojatnost prenaučivosti stabala. S tim ciljem provode se postupci optimizacije hiperparametara modela.

U ovom istraživanju korišten je algoritam zasnovan na slijednom stablastom modelu optimizacije (engl. *Tree based Parzen Estimator*, TPE) za provedbu optimizacije hiperparametara. Ovaj algoritam provodi iterativno pretraživanje koje je zasnovano na Bayesovoj optimizaciji, a strategija pretraživanja je podijeljena u dvije faze prikazane u Algoritmu 4.4. Tijekom početne faze nasumično se odabiru kombinacije vrijednosti različitih hiperparametara. Odabrane kom-

binacije vrijednosti hiperparametara koriste se za treniranje modela te se za svaku kombinaciju bilježi postignuta točnost modela.

Prva faza traje do  $n$  iteracija, u ovom istraživanju vrijednost parametra  $n$  je postavljena na 20 (linija 3- 8 u Algoritmu 4.4). Rezultat ove faze je funkcija koja je temeljena na Bayesovom pravilu kako je to prikazano u Izrazu 4.12 pri čemu je  $P(t|hiperp)$  vjerojatnost da će se neka točnost modela postići određenim skupom vrijednosti hiperparametara. Na temelju točnosti, kombinacije vrijednosti hiperparametara dijele se na dobre i loše, linija 9 Algoritma 4.4. U ovom istraživanju prvih 25% kombinacija hiperparametara kojima se postiže najveća točnost pripadaju skupini dobrih rješenja, a preostalih 75% pripadaju skupini loših rješenja. Za skupinu loših i skupinu dobrih rješenja izračunava se funkcija gustoće iz kojih se u daljnjim koracima algoritma odabiru vrijednosti hiperparametara.

---

**Algoritam 4.4** TPE, [134]

---

```

1: hiperp=[]
2: t=[]
3: for i=1 do n do
4:   Na slučajan način odaberi vrijednosti  $hiperp_i$ 
5:   Izračunaj točnost  $t_i$ 
6:   hiperp.Dodaj( $hiperp_i$ )
7:   t.Dodaj( $t_i$ )
8: end for
9: Podijeli hiperp u  $p^{dobri}$  i  $p^{losi}$ 
10: for j=1 do  $n_{test}$  do
11:    $hiper_{OP} = []$ 
12:    $OP = []$ 
13:   for z=1 do  $n_{OP}$  do
14:      $hiper_{OP}$ .Dodaj(Uzmi parametre iz grupe  $p^{dobri}$ )
15:      $OP$ .Dodaj(Izračunaj OP za  $hiper_{OP_z}$ )
16:   end for
17:    $hiperp_{bestOP}$ =Odaberi iz  $hiper_{OP}$  kombinaciju s najvećim OP
18:   Izračunaj  $t_j$  modela s  $hiperp_{bestOP}$ 
19:   hiperp.Dodaj( $hiperp_{bestOP}$ )
20:   t.Dodaj( $t_j$ )
21:   Podijeli hiperp u  $p^{dobri}$  i  $p^{losi}$ 
22: end for
23: Odaberi  $hiperp$  s najvećom vrijednosti  $t$ 

```

---

$$P(t|hiperp) = \frac{P(hiperp|t) \times P(hiperp)}{P(t)} \quad (4.12)$$

Cilj druge faze algoritma je maksimizirati omjer Očekivanog Poboljšanja ( $OP$ ) prikazanog Izrazom 4.13. Maksimiziranje  $OP$  vrijednosti može se postići odabirom vrijednosti hiperparametara s visokom vjerojatnošću pod  $p^{dobri}(hiperp)$  i malom vjerojatnošću pod  $p^{losi}(hiperp)$ . Ova faza provodi se u  $n_{OP}$  iteracija pri čemu se u svakoj iteraciji ispituje vrijednost očekivanog

poboljšanja s trenutnim vrijednostima hiperparametara te se pohranjuje u listu  $OP$ . U ovom istraživanju vrijednost parametra  $n_{OP}$  je 24 (linija 13- 16 u Algoritmu 4.4). U sljedećem koraku algoritma odabire se rješenje koje je rezultiralo najvećom vrijednošću očekivanog poboljšanja  $OP$  (linija 17) te se za odabranu kombinaciju hiperparametara izračunava točnost modela. Dobbiveno rješenje dodaje se u vektor svih kombinacija parametara  $hiperp$  te se ponavlja dijeljenje rješenja u dobre i loše, linija 19- 21 u Algoritmu 4.4. Postupak se ponavlja do unaprijed zadannog broja iteracija  $n_{test}$  čija je vrijednost u ovom istraživanju postavljena na 100 (linija 10- 22 u Algoritmu 4.4).

$$OP = \frac{p^{dobri}(hiperp)}{p^{loši}(hiperp)} \quad (4.13)$$

### 4.3.5 Testiranje metode slučajne šume i podmodeli

Testiranje RF metode provodi se na testnom skupu podataka, pri čemu se odluka kojoj klasi pripada neki uzorak ne donosi na temelju većinskog broja glasova već je rezultat RF metode distribucija vjerojatnosti koja pokazuje koliko je vjerojatno da neki uzorak pripada svakoj od klasa. Konačna odluka o klasi pripadanja nekog uzorka donosi se u dva koraka:

1. Određuju se dva moda prijevoza čije su vjerojatnosti najveće,
2. Ako modovi prijevoza pripadaju sličnim modovima prijevoza provodi se dodatno testiranje klase podmodelom.

Slični parovi modova prijevoza su: autobus i automobil te vlak i podzemna željeznica. Ako dobiveni rezultat ne pripada jednom od ova dva para, rezultat klasifikacije je mod prijevoza čija je vjerojatnost najveća.

Podmodeli su DT čiji izlaz može biti jedna od dvije klase. DT koja se koriste u svrhu podmodela naučena su na istom skupu podataka za učenje kao i osnovni model, a koriste podskup značajki čija je MI vrijednost najveća za promatrani par modova prijevoza. Ako je uzorak testiran podmodelom, tada je predviđanje podmodela konačni rezultat predviđanja.

## 4.4 Višerazinska metoda za segmentaciju i klasifikaciju trajektorije s obzirom na korišteni mod prijevoza

Višerazinska metoda za segmentaciju i klasifikaciju trajektorije objedinjuje metodu za segmentaciju (TSM) i metodu za klasifikaciju (RF). Ona predstavlja pravila komunikacije između ove dvije metode kako bi se donijela konačna odluka. Višerazinska metoda je kombinacija iskustvenog pristupa i pravila većinskog glasanja. Iskustveni pristup očituje se u dva postavljena pravila koja su opisana u nastavku, a pravilo većinskog glasanja koristi se tijekom odlučivanja o promjeni moda prijevoza koristeći TSM. Razvijena metoda je prilagođena za stvarnovremen-

#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

sku klasifikaciju kojim modom prijevoza se korisnik trenutno kreće kao i detekciju vremenskog prozora u kojem je korisnik prestao koristiti jedan mod prijevoza i počeo koristiti neki drugi mod prijevoza. Stoga je osnovna pretpostavka ove metode da su podatci koji stižu na poslužitelj slijedni. Veličina vremenskog prozora u kojem se prikupljaju sirovi podatci je 30 s nakon kojih se izračunavaju značajke koje su ulazni parametar metode za klasifikaciju.

Osim što se ovom metodom uspostavljaju pravila komunikacije između metoda RF i TSM, kreirana su dva pravila kojima se nastoji korigirati pogrešno predviđanje. Prvo pravilo odnosi se na pojavu različitog moda prijevoza od onog neposredno prije i neposredno nakon promatranog vremenskog prozora. Budući da nije izvedivo da korisnik u 30 s promijeni tri moda prijevoza, ove se ispravke mogu provoditi odmah tijekom klasifikacije. Na Slici 4.12 prikazan je niz vremenskih prozora kojima je pridružen mod prijevoza. U prvom retku prikazana su predviđanja RF metode, a u drugom retku predviđanja nakon primjene prvog pravila. Ispod svakog vremenskog prozora naznačen je indeks vremenskog prozora.

T3	T3	T3	T3	T3	A6	A6	A6	A6	V7	A6	A6	A6	A6	H2	H2	H2	H2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

T3	T3	T3	T3	T3	A6	A6	A6	A6	A6	A6	A6	A6	A6	H2	H2	H2	H2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

**Slika 4.12:** Rezultat predviđanja RF metode prikazan u prvom retku i rezultat nakon primjene prvog pravila u drugom retku

Prvih 5 vremenskih prozora klasificirano je kao trčanje te posljednjih 4 kao hodanje. Između se nalazi mod prijevoza autobus, pri čemu je deseti vremenski prozor klasificiran kao vlak. Nakon primjene prvog pravila (drugi redak) deseti vremenski prozor je korigiran u mod prijevoza autobus jer su vremenski prozori 9 i 11 klasificirani kao autobus.

Drugim pravilom nastoji se procijeniti točnost predviđanja moda prijevoza za vremenski prozor u kojem se dogodila promjena moda prijevoza, a ne može se primijeniti prvo pravilo. Za svaku promjenu moda prijevoza, promatra se sljedećih šest vremenskih prozora te se odabire najbrojniji mod prijevoza za predikciju moda prijevoza u trenutnom vremenskom prozoru. Na Slici 4.13 prikazan je niz vremenskih prozora s pridruženim modom prijevoza. U prvom retku prikazan je rezultat klasifikacije RF metode, a u drugom retku rezultat klasifikacije nakon primjene drugog pravila. Predviđen mod prijevoza za šesti vremenski prozor je vlak, pri čemu vremenskom prozoru prethodi mod prijevoza trčanje, a slijedi autobus. Prema tome, za predviđanje moda prijevoza u šestom vremenskom prozoru aktivira se drugo pravilo i promatraju se predviđanja za sljedećih šest vremenskih prozora (vremenski prozori 6-11). Najzastupljeniji mod prijevoza je autobus (2 predviđanja vlaka i 4 autobusa). Stoga se predviđanje šestog vremenskog prozora korigira u autobus. Ovime je ujedno i nastupila promjena moda prijevoza

#### 4.4. Višerazinska metoda za segmentaciju i klasifikaciju trajektorije s obzirom na korišteni mod prijevoza

iz trčanja u mod prijevoza autobus, pa se za konačnu odluku poziva metoda za segmentaciju trajektorije kako bi se donijela odluka o promjeni moda prijevoza.

T3	T3	T3	T3	T3	V7	A6	A6	A6	V7	A6	A6	A6	A6	H2	H2	H2	H2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

T3	T3	T3	T3	T3	A6	A6	A6	A6	V7	A6	A6	A6	A6	H2	H2	H2	H2
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

**Slika 4.13:** Rezultat predviđanja RF metode prikazan u prvom retku i rezultat nakon primjene drugog pravila u drugom retku

U Algoritmu 4.5 je prikazan slijed višerazinske metode za segmentaciju i klasifikaciju trajektorije. Ulazni podatci algoritma su vremenski slijed sirovih podataka koji se svakih 30 s transformiraju u značajke koje reprezentiraju podatke jednog vremenskog prozora. Izlazni podatak algoritma je lista *trajektorija* sa slijednim vrijednostima modova prijevoza pri čemu  $i$ -ti mod prijevoza liste *trajektorija* odgovara  $i$ -tom vremenskom prozoru.

---

#### Algoritam 4.5 Višerazinska metoda za segmentaciju i klasifikaciju trajektorije

---

```

1: Ulazni parametri: podatci s osjetila mobilnog uređaja za vremenski prozor  $tw$ 
2: Izlazni podatci: trajektorija lista modova prijevoza za svaki vremenski prozor
3: while kraj vremenskogniza do
4:    $mod_i = RF(tw_i)$ 
5:   if  $mod_i == mod_{i-1}$  then
6:     trajektorija.Dodaj( $mod_i$ )
7:   else
8:     if  $mod_{i-1} == mod_{i+1}$  then
9:       trajektorija.Dodaj( $mod_{i+1}$ )
10:    else
11:       $mod_i = MOD(mod_i, mod_{i+1}, mod_{i+2}, mod_{i+3}, mod_{i+4}, mod_{i+5})$ 
12:      if  $mod_i != mod_{i-1}$  then
13:        promjena = TSM( $tw_i$ )
14:        if promjena == True then
15:          trajektorija.Dodaj( $mod_i$ )
16:        else
17:          trajektorija.Dodaj( $mod_{i-1}$ )
18:        end if
19:      else
20:        trajektorija.Dodaj( $mod_i$ )
21:      end if
22:    end if
23:  end if
24: end while

```

---

Sve do kraja vremenskog niza ponavlja se postupak za svaki vremenski prozor  $tw_i$  pri čemu



se kroz niz pravila predviđa i ispituje mod prijevoza (linija 3- 24 u Algoritmu 4.5). U prvom koraku algoritma poziva se metoda  $RF(tw_i)$  čiji je rezultat predviđeni mod prijevoza u vremenskom prozoru  $tw_i$ . Zatim se u prvoj grani petlje (linija 5 Algoritma 4.5) provjerava jesu li prethodni i trenutni mod prijevoza jednaki, ako jesu u listu *trajektorija* se dodaje predviđeni mod prijevoza. U slučaju nejednakosti provjerava se uvjet prvog pravila. Ako je uvjet istinit postupak prvog pravila se provodi i u listu *trajektorija* se dodaje predviđeni mod prijevoza (linija 8- 9 u Algoritmu 4.5). U suprotnom se provodi drugo pravilo. Ako je drugim pravilom predviđena promjena moda prijevoza poziva se metoda za segmentaciju trajektorije TSM() koja daje glas o promjeni moda prijevoza. Ovisno o odluci metode TSM() dopušta se ili ne dopušta promjena moda prijevoza te se odgovarajući mod prijevoza dodaje u listu *trajektorija* (linija 11- 17 u Algoritmu 4.5).

Dakle, Algoritam 4.5 omogućuje komunikaciju između metode RF i TSM tako da svaku promjenu moda prijevoza, koju predikcijom moda prijevoza trenutnog vremenskog prozora izvrši RF, mora dopustiti metoda TSM. U slučajevima kada je mod prijevoza jednak prethodnom, funkcija TSM() nije uključena u predviđanje.

## 4.5 Rezultati

U ovom poglavlju prikazani su rezultati metode za segmentaciju i klasifikaciju trajektorije. Potpoglavljje je strukturirano u tri osnovna dijela. U prvom potpoglavljju 4.5.1 prikazani su rezultati obrade podataka koji su opisani kroz izračunate značajke u vremenskoj i frekvencijskoj domeni. U drugom potpoglavljju 4.5.2 prikazani su dobiveni rezultati metode za segmentaciju trajektorije dok treće potpoglavljje 4.5.3 prikazuje rezultate klasifikacije trajektorije uključujući i rezultate višerazinske metode za segmentaciju i klasifikaciju trajektorije.

Sve metode i postupci obrade podataka implementirani su u programskom jeziku *Python*. Za učenje i testiranje metoda korišteno je prijenosno računalo s Intel i7 procesorom te 8 GB radne memorije.

### 4.5.1 Rezultati obrade podataka

Nakon obrade podataka, skup podataka sadrži 27.319 vremenskih prozora u vremenskoj i frekvencijskoj domeni, što čini 196,2 h prikupljenih podataka. Navedeni vremenski prozori odnose se na sve korisnike i sve lokacije nošenja uređaja.

U Tablici 4.3 prikazani su rezultati obrade podataka zajedno s ukupnom količinom podataka prije obrade. Odvojeno su prikazane količine podataka za četiri podskupa podataka: skup podataka prikupljan uređajem na torzu i kuku te u ruci i torbi korisnika. Prije obrade podataka sva četiri podskupa podataka sadrže oko 30.000.000 zapisa, svaki zapis je jedno mjerenje s osjetila mobilnog uređaja u razmaku od 10 ms u odnosu na prethodni zapis. Broj zapisa s nedostajućim

vrijednostima je varijabilan u pojedinom podskupu podataka te je njegova vrijednost u rasponu 1.268-17.222. Broj zapisa koji nema pridruženu klasu je značajnije veći te je jednak za sve podskupove podataka. Kada se broj zapisa podijeli u vremenske prozore duljine 30 sek (svaki vremenski prozor sadrži oko 3000 pojedinačnih zapisa) onda je broj vremenskih prozora za sva četiri podskupa podataka oko 6.800 sa manjim odstupanjima.

**Tablica 4.3:** Broj zapisa nakon brisanja podataka

		Skup podataka torzo		Skup podataka ruka		Skup podataka torba		Skup podataka kuk		Ukupno	
Ukupno zapisa prije obrade		30.048.209		30.048.209		30.048.209		30.048.209		120.192.836	
Zapisi s nedostajućim vrijednostima		6.150		1.269		15.996		17.222		40.637	
Zapisi bez pridružene klase		9.563.541		9.563.541		9.563.541		9.563.541		38.254.164	
Ukupno zapisa nakon obrade		20.478.518		20.483.399		20.468.672		20.467.446		81.898.035	
Ukupan broj vremenskih prozora		6.833		6.828		6.830		6.828		27.319	
		Ukupno	Uklonjeno	Ukupno	Uklonjeno	Ukupno	Uklonjeno	Ukupno	Uklonjeno	Ukupno	Uklonjeno
Broj vremenskih prozora po modu prijevoza	Stajanje	1.284	718	1.289	722	1.283	715	1.283	715	5.139	2.870
	Hodanje	1.144	578	1.143	576	1.142	574	1.140	572	4.569	2.300
	Trčanje	136	0	136	0	136	0	136	0	544	0
	Bicikl	566	0	567	0	568	0	568	0	2.269	0
	Automobil	1.013	447	1.012	445	1.014	446	1.015	447	4.054	1.785
	Autobus	636	70	632	65	635	67	635	67	2.538	269
	Vlak	1.025	459	1.023	456	1.026	458	1.025	457	4.099	1.830
	Podzemna željeznica	1.029	463	1.026	459	1.026	458	1.026	458	4.107	1.838

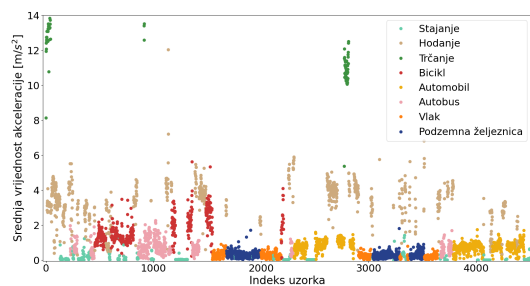
Raspodjela broja vremenskih prozora prema modovima prijevoza nije jednolika. Najveći broj vremenskih prozora odnosi se na aktivnost stajanja, dok se najmanji broj vremenskih prozora odnosi na trčanje. Nakon ujednačavanja broja modova prijevoza i dalje je najmanji broj vremenskih prozora za trčanje dok je broj vremenskih prozora za sve ostale modove prijevoza jednak. Ujednačavanje broja vremenskih prozora za različite modove prijevoza iznimno je bitno zbog grupiranja modova prijevoza u različitim scenarijima prema strukturi testnog problema. Tada značajna odstupanja u količini podataka za neke modove prijevoza dolaze do izražaja.

S obzirom na složenost problema i sličnost pojedinih modova prijevoza ne postoji jedna značajka koja može dobro rasporediti uzorke prema klasi kojoj pripadaju. Na Slici 4.14 prikazane su vrijednosti četiri značajke: srednja vrijednost i međukvartilni raspon akceleracije u vremenskoj domeni te centar frekvencije i standardno odstupanje akceleracije u frekvencijskoj domeni. Jedan krug označava vrijednost značajke u jednom vremenskom prozoru. Različitim bojama označeni su krugovi koji pripadaju različitim klasama, tj. modovima prijevoza. Prikazani su samo vremenski prozori dijela skupa podataka koji se odnose na podatke prikupljene s mobilnog uređaja nošenog u ruci.

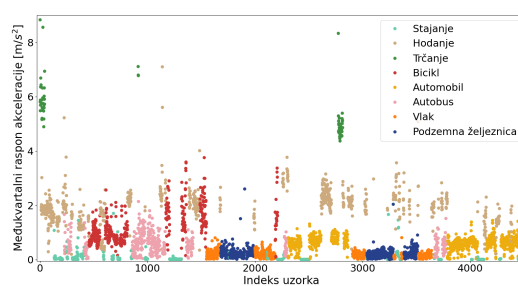
Niti jedna značajka u cijelom skupu podataka ne odjeljuje različite modove prijevoza samostalno. Primjerice značajke na Slikama 4.14a i 4.14b prikazuju jasnu razliku između trčanja i svih ostalih modova prijevoza dok se vrijednosti značajki za ostale modove prijevoza nalaze u

#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

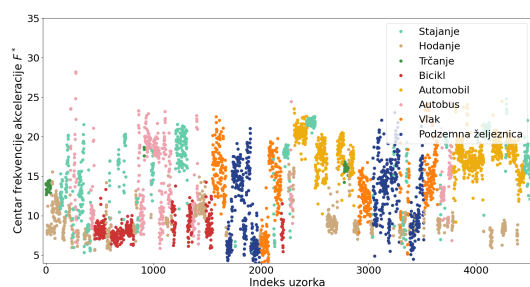
istom rasponu vrijednosti. Na Slici 4.14a se raspoznaje nešto bolje odvajanje klase hodanje od svih ostalih modova prijevoza, pa tako i bicikla. Za razliku od značajki u vremenskoj domeni, značajke u frekvencijskoj domeni na Slikama 4.14c i 4.14d pokazuju bolje odvajanje modova prijevoza koji pripadaju skupini motoriziranih modova prijevoza. Primjerice na Slici 4.14d se mogu razlikovati modovi prijevoza automobil i autobus od modova prijevoza vlak i podzemna željeznica. Također na Slici 4.14c u većem području razdvojni su modovi prijevoza hodanje i bicikl u odnosu na stajanje.



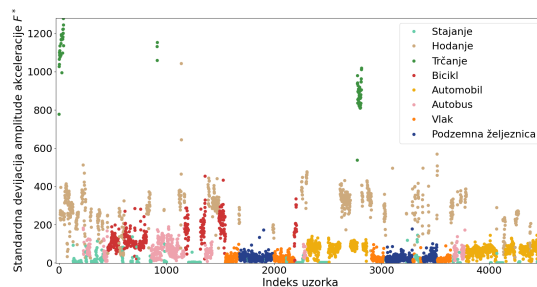
(a) Srednja vrijednost akceleracije



(b) Međukvartilni raspon akceleracije



(c) Centar frekvencije

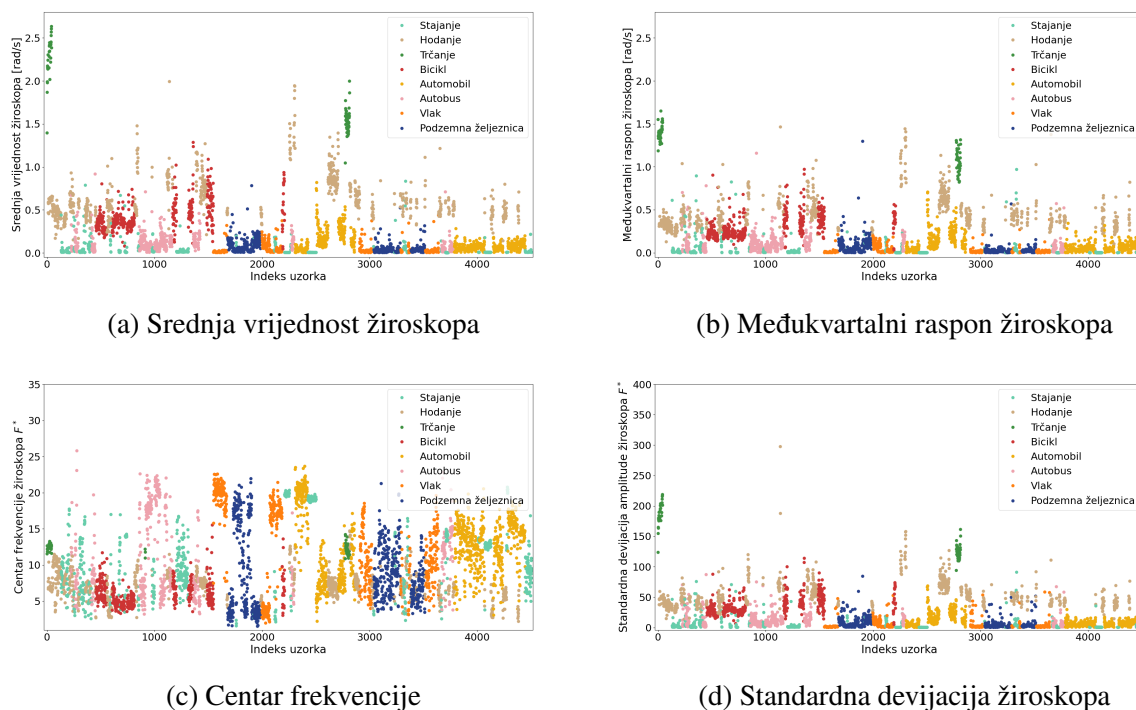


(d) Standardno odstupanje amplitude

**Slika 4.14:** Vrijednosti mjerenja akceleracije za vremenske prozore prikazano kroz više značajki

Za razliku od značajki akceleracije na Slici 4.15 prikazane su iste značajke, ali za vrijednosti magnitude žiroskopa. Boja kruga označava mod prijevoza kojem izmjerena vrijednost pripada. Prikazane značajke srednje vrijednosti i međukvartilnog raspona (Slika 4.15a i 4.15b) pokazuju bolje razdvajanje između bicikla i motoriziranih modova prijevoza. Također prikazane značajke iz frekvencijske domene na Slikama 4.15c i 4.15d puno lošije razdvajaju modove prijevoza unutar skupine motoriziranih modova prijevoza, ali pokazuju vrijednosti kojima se mogu razdvojiti skupina motoriziranih i skupina nemotoriziranih modova prijevoza. Dakle, jasno je da se kombinacijom različitih značajki mogu bolje odrediti granične vrijednosti kojima bi se mogli razlikovati različiti modovi prijevoza međusobno. Također različita osjetila doprinose razlikovanju pojedinih skupina modova prijevoza, pa je očekivano da će kombinacija sva tri osjetila najviše doprinijeti točnosti metode za klasifikaciju.

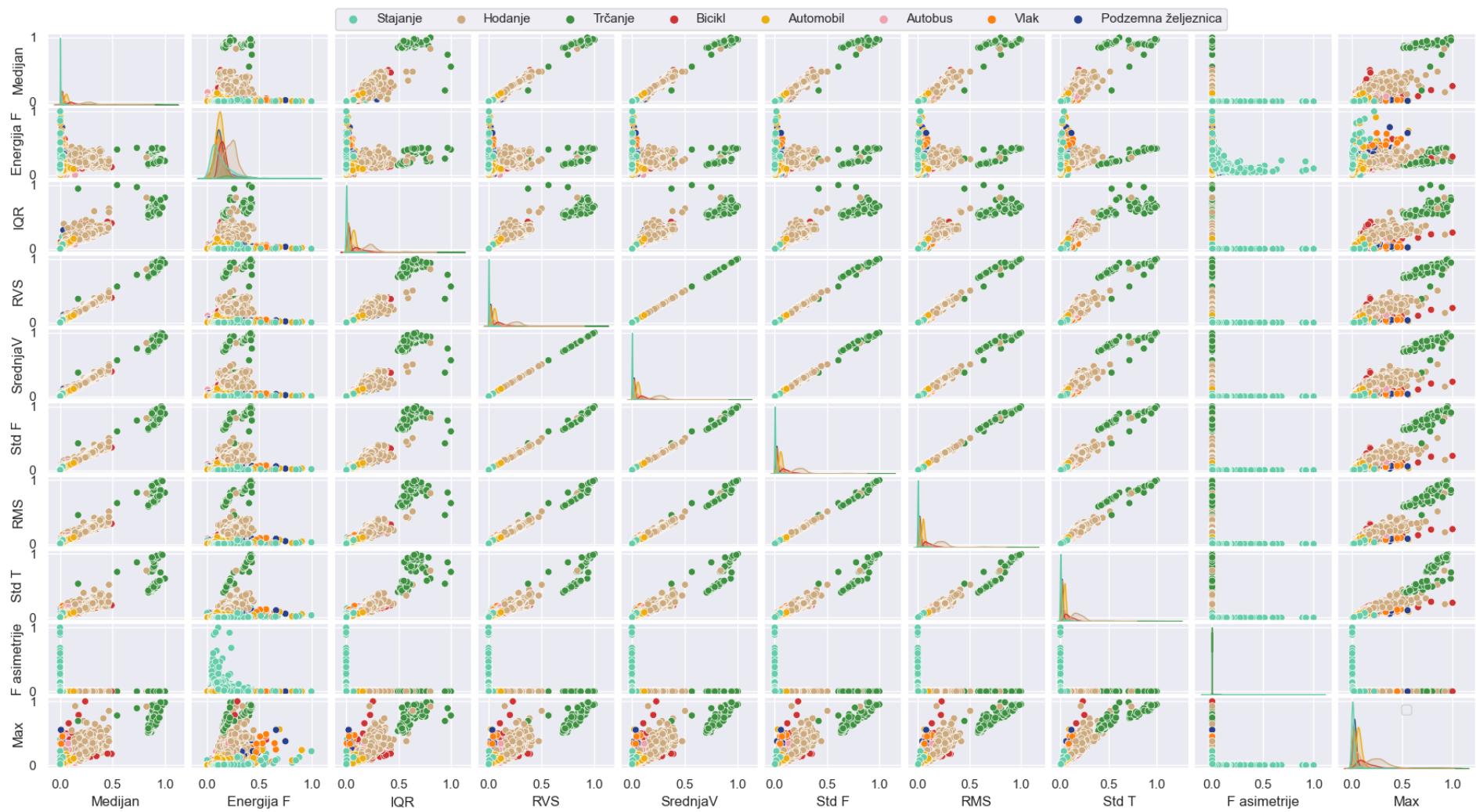
Na Slici 4.16 prikazana je matrica čiji elementi prikazuju odnos značajke koje se nalaze u pripadajućem retku i stupcu matrice. Različitim bojama krugova označena je njihova pri-



**Slika 4.15:** Vrijednosti mjerenja žiroskopa za vremenske prozore prikazano kroz više značajki

padnost određenom modu prijevoza. Odabrano je 10 izračunatih značajki akcelerometra koje pokazuju najveću MI vrijednost nakon primjene metode za odabir značajki opisane u potpoglavlju 4.3.1.

Element koji se nalazi u retku i stupcu iste značajke prikazuje distribuciju vrijednosti te značajke po modovima prijevoza. I ovim prikazom je potvrđeno da mali broj značajki ne može jednoznačno razlikovati navedene modove prijevoza. Međutim odnos između nekih značajki pokazuje sposobnost razlikovanja nekih grupa modova prijevoza. Većina prikaza zapravo pokazuje jasnu razliku između trčanja i svih ostalih modova prijevoza. Također faktor asimetrije u odnosu s bilo kojom drugom značajkom dobro razdvaja stajanje od svih drugih modova prijevoza. Preostali odnosi značajki uglavnom pokazuju sposobnost razlikovanja između tri skupine modova prijevoza: trčanje, stajanje i svi ostali modovi prijevoza, primjerice par značajki maksimalna vrijednost akceleracije i energija u frekvencijskoj domeni ili međukvartilni raspon i korijen varijance frekvencije. Iz ovog prikaza je jasno da je odnos između različitih motoriziranih modova prijevoza u složenijem odnosu od onog dvodimenzionalnog te su iz tog razloga primijenjene naprednije metode za odabir relevantnih značajki.



Slika 4.16: Prikaz ovisnosti značajki s najvećom MI vrijednosti

## 4.5.2 Rezultati metode za segmentaciju trajektorije

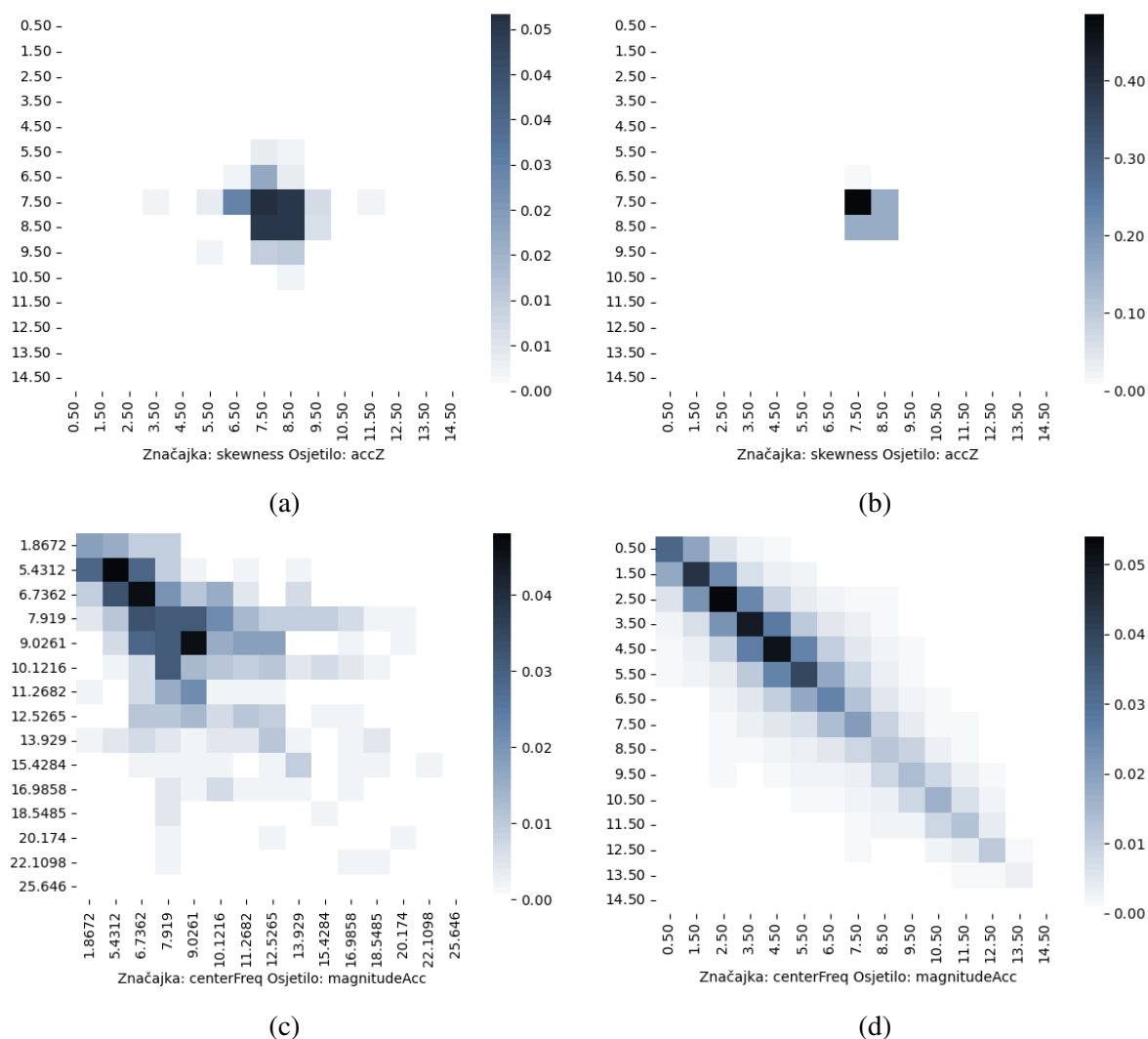
Skup podataka za učenje sadrži 17.659 zapisa, dok skup podataka za testiranje sadrži 5.885 zapisa za svako osjetilo. Skup podataka za učenje formiran je prema pravilima formiranja skupa podataka za učenje u testnom problemu opisanom u potpoglavlju 3.3. Za prikaz rezultata metode za segmentaciju odabrana je podjela prema pravilima za ispitivanje lokacijske neovisnosti metode. Primjerice dio skupa podataka prikupljen u torbi koristi se za testiranje metode, a preostala tri dijela skupa podataka s preostale tri lokacije za učenje metode, tj. izradu TSM-a. Postoji 408 prijelaza iz jednog moda prijevoza u drugi mod prijevoza u skupu podataka za učenje i 139 takvih prijelaza u skupu podataka za testiranje. Za testiranje višerazinske metode, skup podataka za testiranje i skup podataka za učenje mijenjaju se ovisno o neovisnosti koja se ispituje. U tom slučaju koristi se jednaka podjela skupa podataka, prema pravilima formiranja testnog problema, kod metode za segmentaciju trajektorije i metode za klasifikaciju modova prijevoza.

Broj TSM-a je dvostruko veći od broja značajki jer je svaki TSM kreiran za klasu promjenu i klasu bez promjene. Na  $x$  i  $y$ -osi TSM-ova na Slici 4.17 nalaze se vrijednosti odgovarajuće TSM značajke podijeljene u 15 stanja. Intenzitet boje od bijele do tamnoplave ukazuje na vjerojatnost prijelaza između stanja.

Slike 4.17a i 4.17b predstavljaju TSM-ove za značajku asimetrije i osjetilo akcelerometra duž  $z$ -osi, a Slike 4.17c i 4.17d su TSM-ovi za značajku centar frekvencije i magnitudu akceleracije. U lijevom stupcu prikazani su TSM-ovi za vremenske prozore u kojima je došlo do promjene moda prijevoza, a u desnom stupcu za vremenske prozore u kojima nije došlo do promjene moda prijevoza. Na Slikama 4.17a i 4.17b su najčešći prijelazi između stanja uglavnom isti. Međutim, postoji veće raspršenje različitih stanja u matrici koja prikazuje klasu s promjenama. Za razliku od ovog para matrica, matrice na Slikama 4.17c i 4.17d pokazuju pomak u najčešćim stanjima za klasu u kojoj se dogodila promjena moda prijevoza i onu u kojoj nema promjene moda prijevoza. Najčešća stanja kod vremenskih prozora u kojima se nije dogodila promjena moda prijevoza smještene su na dijagonali matrice što upućuje da najčešće korisnik zadržava postojeće vrijednosti akceleracije u vremenskim prozorima kada ne mijenja mod prijevoza. Suprotno tome, za vremenske prozore u kojima se događa promjena moda prijevoza, pojavljuju se i stanja koja ukazuju na značajnije promjene vrijednosti akceleracije.

Za evaluaciju metode za segmentaciju trajektorije korištena je matrica zabune s pripadajućim mjerama. Ako je metodom predviđena promjena moda prijevoza provjerava se u vektoru, koji sadrži stvarne vrijednosti klase, postoji li promjena moda prijevoza u susjednih šest vremenskih prozora. Usporedba vrijednosti klasa u susjednim vremenskim prozorima uvedena je jer je uočeno da model detektira promjenu moda prijevoza u vremenskom prozoru u kojem je došlo do promjene moda prijevoza i u susjednim vremenskim prozorima. To se može objasniti kratkim trajanjem vremenskog prozora dok je najčešće trajanje promjene moda prijevoza duže

#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

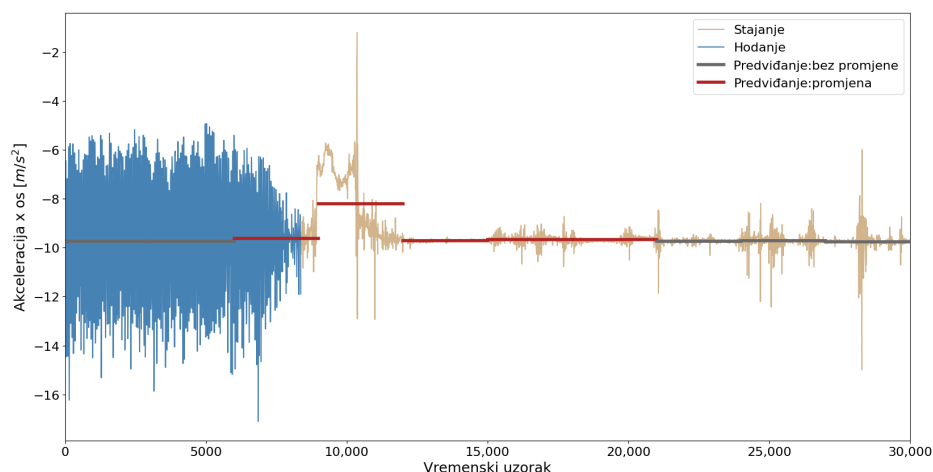


**Slika 4.17:** TSM za (a) značajku asimetrije podataka i akceleraciju po  $z$ -osi za vremenske prozore s promjenom moda prijevoza; (b) značajku asimetrije podataka i akceleraciju po  $z$ -osi za vremenske prozore bez promjene moda prijevoza, (c) značajku centar frekvencije i magnitudu akceleracije za vremenske prozore s promjenom moda prijevoza; i (d) značajku centar frekvencije i magnitudu akceleracije za vremenske prozore bez promjene moda prijevoza

od 30 s. Stoga se promjena moda prijevoza odražava u susjednim vremenskim prozorima.

Primjerice, na  $y$ -osi Slike 4.18 prikazani su sirovi podatci po  $x$ -osi akcelerometra i vremenski uzorci na  $x$ -osi, pri čemu se u 8.372-om vremenskom uzorku događa promjena moda prijevoza (promjena boje krivulje na slici). Središnja crta predstavlja srednju vrijednost akceleracije duž  $x$ -osi u promatranim vremenskim prozorima. Boja središnje crte označava je li vremenski prozor klasificiran kao vremenski prozor u kojem je došlo do promjene moda prijevoza (crvena) ili ne (siva). U ovom slučaju promjena moda prijevoza predviđena je u točnom vremenskom prozoru, ali se predviđa i promjena moda prijevoza u sljedećih nekoliko vremenskih prozora. Time je promjena koja se dogodila u jednom trenutku utjecala na vrijednosti značajki u susjednim prozorima, što uzrokuje lažno pozitivna predviđanja za te vremenske prozore. Međutim, takvi se slučajevi ne mogu smatrati pogreškama predviđanja jer se odnose na promjenu moda

prijevoza koja se dogodila u protekle tri minute, a ovo vrijeme se ponekad naziva i vremenom koje je potrebno korisniku da se fizički prebaci iz jednog moda prijevoza u drugi [135]. Stoga se u ovom radu predikcija smatra točnom ako se promjena moda prijevoza detektira unutar tri minute od trenutka stvarne promjene načina transporta.



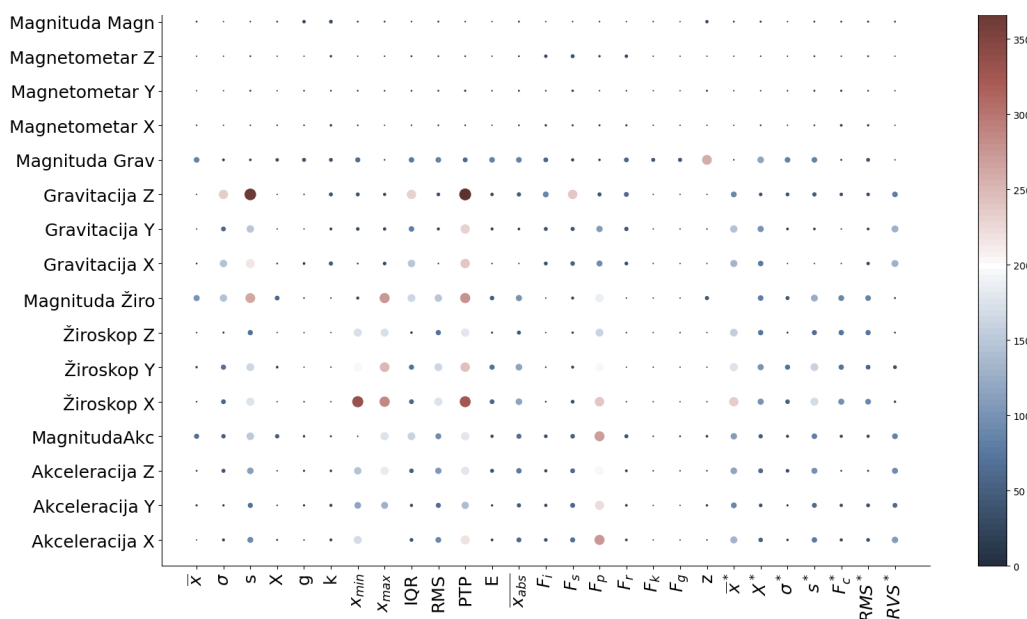
**Slika 4.18:** Vrijednost sirovih podataka u trajektoriji po x osi akcelerometra i srednja vrijednost akceleracije

Analiziran je utjecaj pojedine značajke i osjetila na rezultat detekcije promjene moda prijevoza kako bi se pronašle značajke koje bolje opisuju izvorne podatke. Značajkama su pridružene numeričke vrijednosti koje ukazuju na relevantnost istih te su izdvojene značajke koje daju najviše informacija tijekom prijelaza iz jednog vremenskog prozora u sljedeći kako je to opisano u potpoglavlju 4.2.3. Slika 4.19 prikazuje važnost pojedine značajke koja je prikazana ispunjenim krugovima s dva atributa: (i) veličina kruga predstavlja važnost značajke, što je veća važnost, veći je krug, te (ii) boja kruga, od tamnoplave do tamnocrvene, predstavlja ANOVA F vrijednost. Os y predstavlja osjetila, a x-os značajke. Značajke navedene na x osi označene su jednakim oznakama koje su im pridružene u Tablici 4.1. Najbolji ANOVA F rezultat postignut je za standardnu devijaciju i osjetilo gravitacija duž z-osi. Od svih osjetila magnetometar pokazuje najmanje ANOVA F vrijednosti dok su značajke s najmanjim utjecajem medijan te mjera i faktori asimetrije i spljoštenosti.

U sljedećem koraku potrebno je odrediti potreban broj značajki koje će se koristiti za predviđanje promjene moda prijevoza. S tim ciljem testiran je utjecaj broja značajki na točnost predviđanja vremenskog prozora u kojem se dogodila promjena moda prijevoza. Za ovo testiranje odabrano je 200 značajki čija je  $F$  vrijednost najveća. Točnost predviđanja promatrana je kroz mjere ukupne točnosti, odziva i  $F$  vrijednosti modela. Na Slikama 4.20, 4.21 i 4.22 prikazani su rezultati točnosti TSM-a koji su izrađeni opisanim trima metodama diskretizacije. Na svim slikama na x-osi prikazan je uključen broj značajki, a na y-osi postotak točnosti metode.

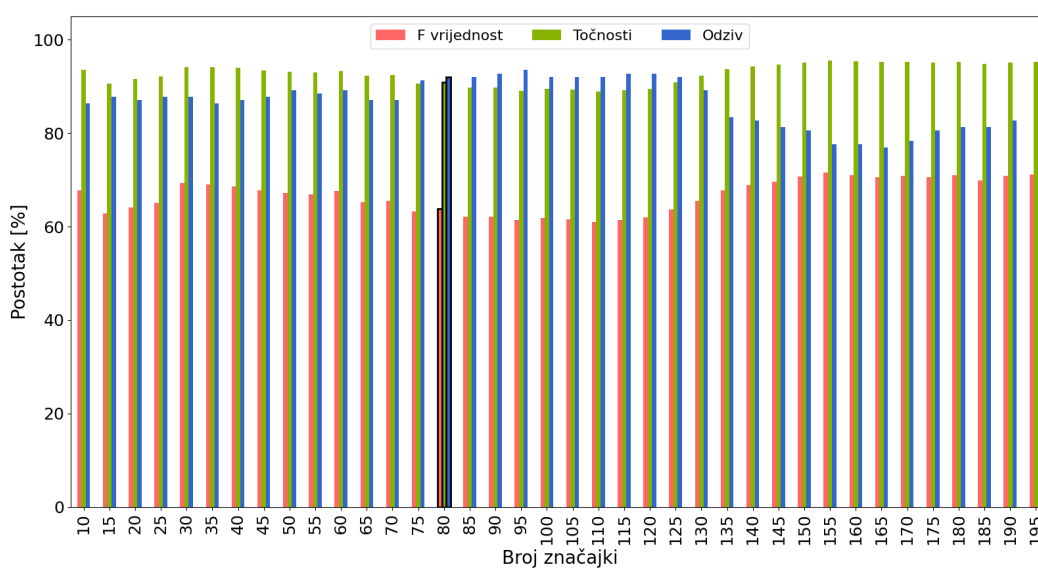


#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu



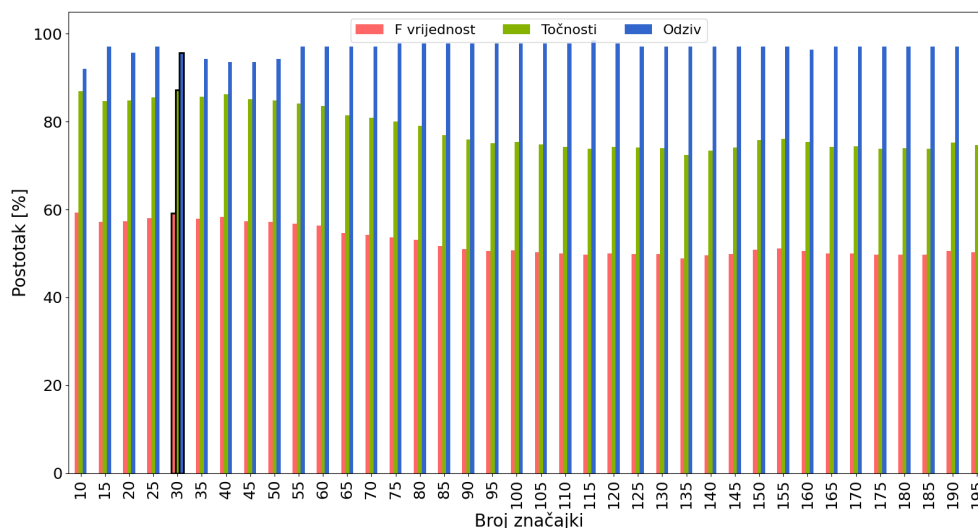
Slika 4.19: Važnost značajki prema F vrijednosti

Slika 4.20 prikazuje rezultat testiranja TSM-a izrađenih metodom diskretizacije jednake duljine. Povećanjem broja značajki povećava se ukupna točnost međutim smanjuje se odziv. Odziv predstavlja točnost predviđanja vremenskih prozora u kojima se dogodila promjena moda prijevoza.  $F$  vrijednost prikazuje omjer pogrešnih predviđanja promjena i ne promjena. Ipak, najveću težinu za ocjenu metode za segmentaciju trajektorije ima odziv jer je bitno čim više smanjiti propuštene detekcije promjene moda prijevoza sa što manjim narušavanjem ukupne točnosti. Stoga, bi kod ove vrste diskretizacije najbolji rezultat predstavljao broj značajki 80 (stupac je podebljan okvirom crne boje) kod kojeg odziv poprima jednu od najvećih vrijednosti, a smanjenje ukupne točnosti je najmanje u odnosu na prikazane primjere.



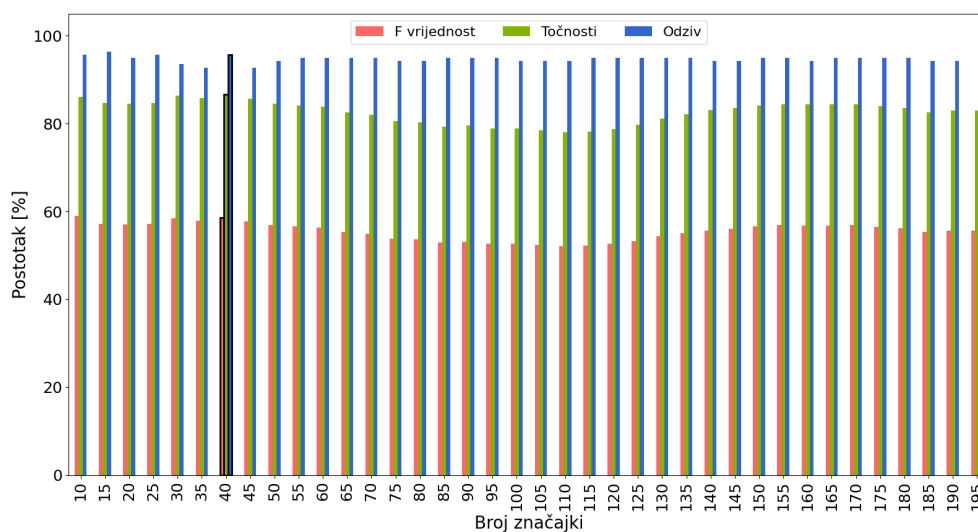
Slika 4.20: Rezultat testiranja TSM-a izrađenih metodom diskretizacije jednake duljine za različit broj značajki

Na Slici 4.21 prikazani su rezultati testiranja različitog broja značajki za matrice izrađene metodom diskretizacije jednake frekvencije. Najbolji omjer ukupne točnosti, odziva i  $F$  vrijednosti postiže se koristeći 30 značajki (stupac je podebljan okvirom crne boje). U usporedbi s metodom diskretizacije jednake duljine smanjen je potreban broj značajki, ali je smanjen i broj detektiranih vremenskih prozora u kojima se događa promjena moda prijevoza.



**Slika 4.21:** Rezultat testiranja TSM-a izrađenih metodom diskretizacije jednake frekvencije za različit broj značajki

Rezultati TSM-a izrađenih metodom diskretizacije algoritmom  $k$  sredina prikazani su na Slici 4.22. Najbolji rezultat postignut je koristeći 40 značajki (stupac podebljan okvirom crne boje). Ponovno je postignut najbolji rezultat manjim brojem značajki, ali je detektiran manji broj vremenskih prozora u kojima se događa promjena moda prijevoza u usporedbi s metodom diskretizacije jednake duljine.



**Slika 4.22:** Rezultat testiranja TSM-a izrađenih metodom diskretizacije algoritmom  $k$  sredina za različit broj značajki

U Tablici 4.4 prikazani su rezultati testiranja tri vrste TSM-a: matrice izrađene metodom diskretizacije jednake duljine, jednake frekvencije te diskretizacija metodom  $k$  sredine. Za svaku vrstu TSM-a prikazana je matrica zabune za broj značajki koji je rezultirao najboljim omjerom točnosti i odziva što je redom 80, 30 i 40 značajki (Slike 4.20- 4.22). Podjednako dobar rezultat postignut je matricama izrađenim diskretizacijom jednake duljine i diskretizacijom jednake frekvencije. Razlika između mjera točnosti i mjera odziva je oko 3%, pri čemu se metodom diskretizacije jednake duljine postiže veća točnost, a metodom diskretizacije jednake frekvencije veći odziv. Ipak, metoda diskretizacije jednake duljine pokazuje bolji omjer između točnosti i odziva, pa je ta vrsta diskretizacije odabrana za višerazinsku metodu.

**Tablica 4.4:** Rezultati testiranja TSM na testnom skupu podataka

	Diskretizacija jednake dužine		Diskretizacija jednake frekvencije		Diskretizacija algoritmom $k$ sredina	
	Predviđena promjena	Predviđeno bez promjene	Predviđena promjena	Predviđeno bez promjene	Predviđena promjena	Predviđeno bez promjene
Stvarna promjena	128	11	133	6	133	6
Stvarno bez promjene	522	5.224	747	4.999	778	4.968
Točnost	90.94%		87.2%		86.67%	
Odziv	92.09%		95.68%		95.69%	

### 4.5.3 Rezultati metode za klasifikaciju trajektorije

Uspješnost klasifikacije trajektorije metodom RF ponajviše ovisi o parametrima stabala u šumi i značajkama kojima je opisana svaka klasa uključena u problem. Testiranje RF metode u sklopu ovog istraživanja temeljeno je na scenarijima testnog problema opisanog u potpoglavlju 3.3. Osim toga, koncept testiranja u testnom problemu uključuje ispitivanje značajki kroz uključivanje tri osjetila u tri zasebna koraka: značajke zasnovane na akceleraciji, značajke zasnovane na akceleraciji i žiroskopu te značajke zasnovane na akceleraciji, žiroskopu i magnetometru. Stoga su prikazani rezultati podijeljeni tako da se može raspoznati rezultat pojedinog scenarija i skupa značajki. Za potrebe testiranja prema strukturi predloženoj u testnom problemu razvijeno je 396 modela, od kojih je 144 za testiranje lokacijske ovisnosti, 144 za testiranje vremenske ovisnosti te 108 za testiranje korisničke ovisnosti. Kao primjer rezultirajućih hiperparametara za samo jedan od izrađenih modela prikazani su hiperparametri RF modela dobiveni opisanim optimizacijskim postupkom u potpoglavlju 4.3.4 u Tablici 4.5. Modeli koji su prikazani u Tablici 4.5 odnose se na testiranje lokacije nošenja uređaja u ruci te značajke zasnovane na akceleraciji za sve scenarije. Za izgradnju RF modela koristi se programska knjižnica *Scikit-Learn* te funkcija *RandomForestClassifier* kojoj su predani parametri: broj stabala, dubina stabla, mjera za podjelu čvora te najmanji potreban broj uzoraka za podjelu čvora i odlučivanje u listu [136]. Preostali parametri funkcije preuzimaju unaprijed definirane vrijednosti klase *RandomForestClassifier*. Broj stabala je u rasponu od 20 do 500, mjera za podjelu čvora je za sve scenarije entropija te je dubina stabla između 6 i 36. Najmanji broj uzoraka za podjelu i broj uzoraka

u listu ovisi o broju klasa po scenarijima tako da je njihov raspon optimalnih vrijednosti veći u odnosu na ostale parametre. U Tablici 4.5 je prikazana i veličina modela kao i trajanje učenja modela za svaki scenarij. Trajanje učenja i testiranja metode je znatno manje od modela u literaturi koji su temeljeni na metodama dubokog učenja što se i ističe kao osnovna prednost kada je klasifikator namijenjen za primjenu na mobilnim uređajima [137]. Ovo osobito dolazi do izražaja ako se model želi upotrebljavati za učenje tijekom prikupljanja podataka tzv. *online* učenje.

**Tablica 4.5:** Parametri RF modela za svaki scenarij

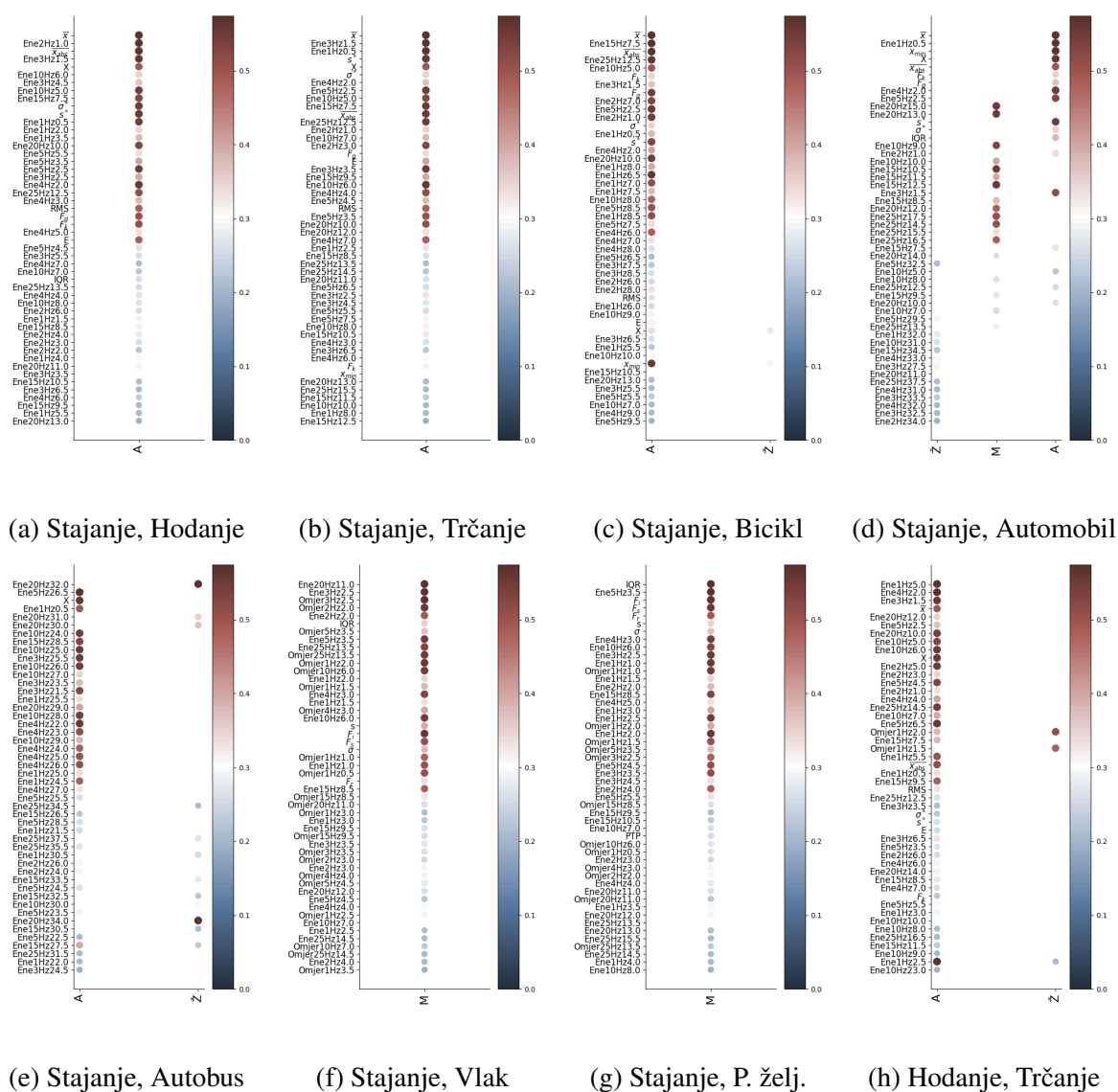
Scenarij	Parametri						Veličina modela
	Broj stabala	Dubina stabla	Mjera	Najmanji broj uzoraka za podjelu	Najmanji broj uzoraka u listu	Trajanje učenja	
1	20	25	Entropija	100	6	1 min	2 MB
2	200	21	Entropija	25	5	3 min	2 MB
3	50	30	Entropija	50	10	6 min	2 MB
4	100	35	Entropija	50	10	6 min	3 MB
5	500	36	Entropija	10	30	12 min	7 MB
6	200	6	Entropija	100	5	12 min	6 MB
7	400	28	Entropija	15	3	18 min	8 MB
8	100	18	Entropija	20	3	18 min	10 MB
9	300	6	Entropija	10	10	17 min	19 MB
10	500	25	Entropija	100	2	18 min	30 MB
11	200	31	Entropija	15	5	20 min	12 MB
12	400	24	Entropija	10	5	20 min	39 MB

Na Slici 4.23 prikazan je dio rezultata analize značajki opisanih u potpoglavlju 4.3.1, koji uključuje MI vrijednosti za svaki par modova prijevoza uključenih u problem. Razmatran je problem klasifikacije modova prijevoza u kojem je dopušteno korištenje značajki od sva tri osjetila: akcelerometra, magnetometra i žiroskopa. Za svaki par modova prijevoza je prikazano 50 značajki čija je MI vrijednost najveća. Važnost pojedine značajke prikazana je ispunjenim krugovima s dva atributa: (i) veličina kruga predstavlja važnost značajke, što je veća važnost, veći je krug, te (ii) boja kruga, od tamnoplave do tamnocrvene, predstavlja MI vrijednost. Na y-osi nalaze se značajke, a na x-osi osjetila. Osjetilima je pridružena oznaka A (akcelerometar), M (magnetometar) i Ž (žiroskop). Značajke navedene na y osi označene su jednakim oznakama koje su im pridružene u Tablici 4.1. Značajke koje se odnose na analizu različitih širina frekventijskog pojasa označene su trima informacijama: (i) mjera (Energija (Ene) ili Omjer); (ii) širina frekvencijskog pojasa i (iii) centar frekvencije.

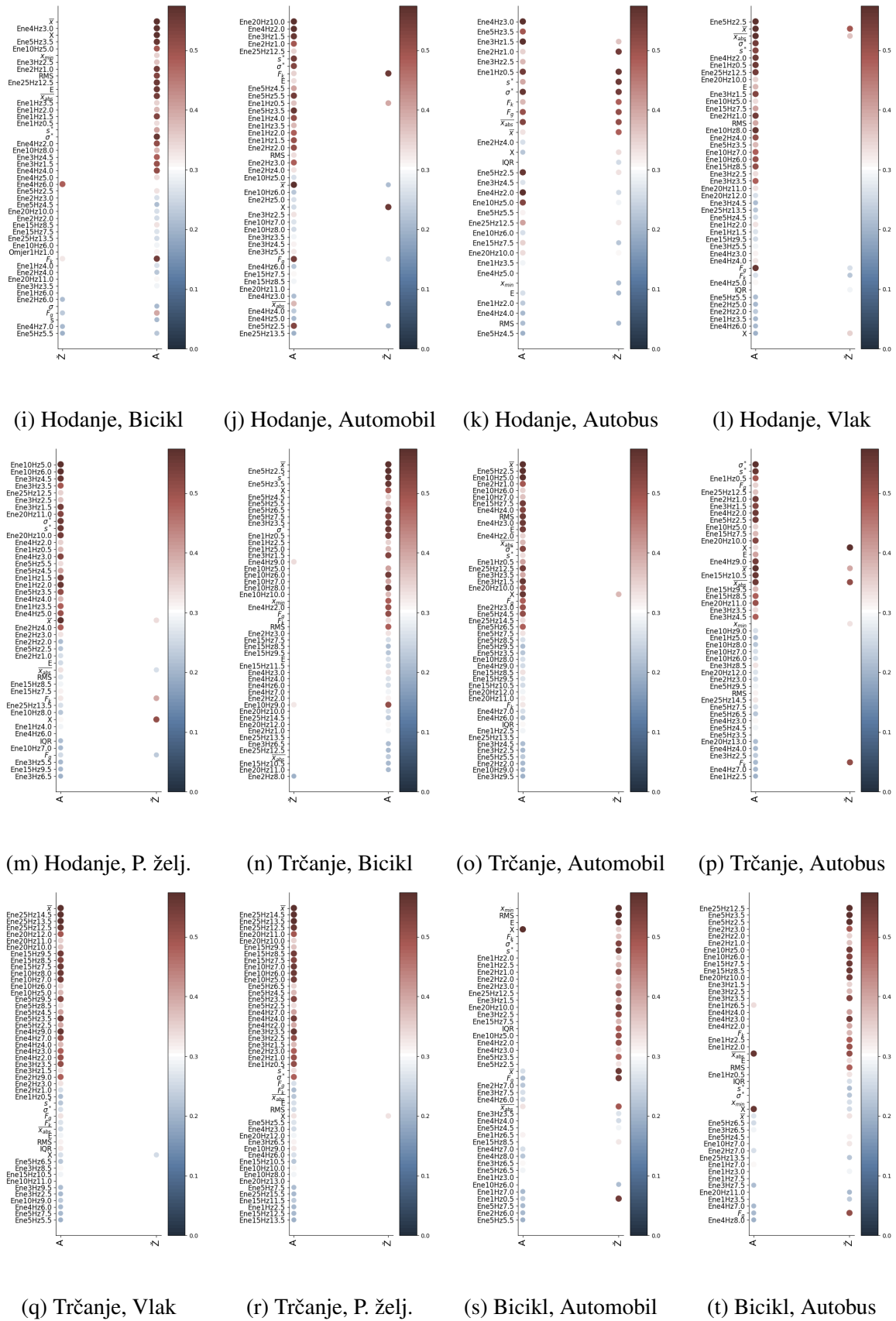
Različiti parovi pokazuju preferencu prema različitim osjetilima. Tako primjerice za razlikovanje modova prijevoza koji pripadaju skupini nemotoriziranih modova prijevoza uglavnom

#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

najviše doprinose značajke osjetila akcelerometra i žiroskopa, dok su kod značajki koje se odnose na razlikovanje između motoriziranih modova prijevoza može primijetiti veća zastupljenost značajki zasnovanih na vrijednostima magnetometra. Kod većine parova zastupljene su značajke koje se odnose na vremensku domenu podataka jednog vremenskog prozora poput srednje vrijednosti, medijana, standardne devijacije ili varijance. Značajke kojima se ispituje širina frekvencijskog pojasa pokazuju veću zastupljenost značajki koje se odnose na širinu frekvencijskog pojasa između 0-10 Hz za akcelerometar i žiroskop dok je za magnetometar nešto veća zastupljenost značajki koje se odnose na frekvencijski pojas od 20-35 Hz.

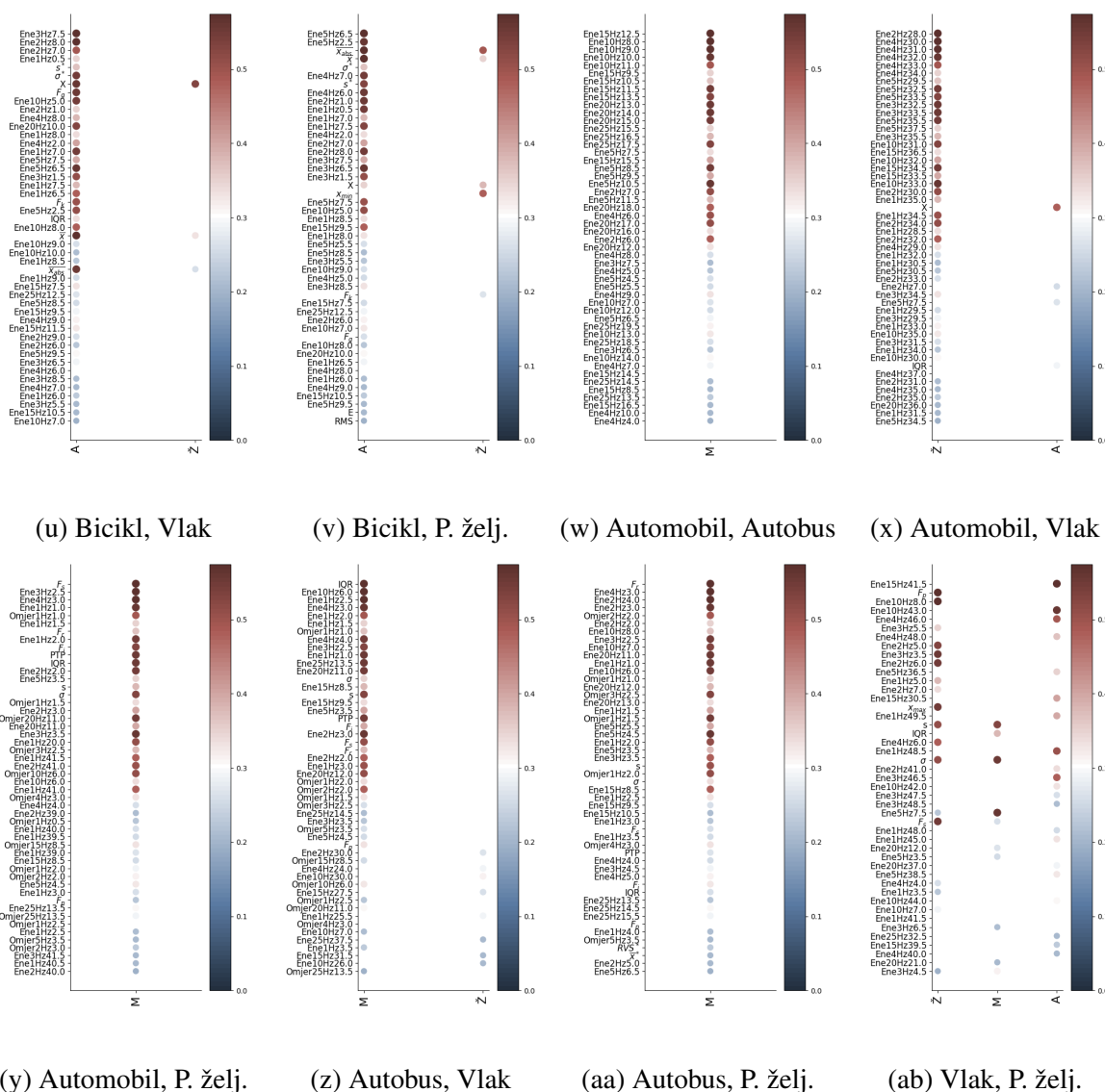


Slika 4.23: MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju



Slika 4.23: MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju

#### 4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu



**Slika 4.23:** MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju

Rezultati klasifikacije prikazani su u Tablicama 4.6, 4.7 i 4.8 kroz usporedbu rezultata ostvarenih na testnom problemu u literaturi [81] te rezultata ostvarenih u disertaciji. Rezultati koji su ostvareni u literaturi odnose se na model koji je naučen na cijelom skupu podataka, dok je u ovom istraživanju korišten samo objavljeni dio skupa podataka kojeg čini svega 8% cjelokupnog skupa podataka. Rezultati u tablicama su strukturirani tako da se za svaki podskup podataka na kojem je provedeno testiranje prikazuje rezultat postignut u literaturi, rezultat postignut korištenjem RF metode te poboljšanje RF rezultata primjenom višerazinske metode (oznaka *višerazinska* u tablicama).

Prikazano je 12 različitih scenarija od kojih svaki sadrži različit broj klasa uključenih u klasifikaciju. Svi scenariji testirani su na tri različita skupa značajki označenih u tablici: A, AŽ te AŽM. Scenariji 10, 11 i 12 predstavljaju problem klasifikacije šest ili više modova prijevoza na kojima je fokus ove doktorske disertacije.

U Tablici 4.6 je prikazano testiranje korisničke neovisnosti metode, tj. koliko uspješno metoda generalizira podatke s obzirom na pripadnost pojedinom korisniku. Nešto je veća točnost metoda koje koriste podatke s akcelerometra i žiroskopa u usporedbi s metodom koja koristi isključivo podatke s akcelerometra, ali se ipak najveće povećanje točnosti postiže uključivanjem podatka s magnetometra. Ako se promatra srednja vrijednost točnosti i F1 vrijednosti, bolja točnost postignuta je kod svih scenarija, dok je kod 30 od ukupno 36 scenarija postignuto poboljšanje F1 vrijednosti. Kada se uzmu u obzir scenariji u kojima je poboljšanje postignuto i oni u kojima poboljšanje nije postignuto onda je prosječno povećanje točnosti klasifikacije višerazinske metode 7,58%, a prosječno povećanje F1 vrijednosti 3,27%.

Isti podatci prikazani su u Tablici 4.7 za testiranje lokacijske neovisnosti metode, tj. koliko uspješno metoda generalizira podatke s obzirom na lokaciju na tijelu na kojoj je mobilni uređaj nošen. Sveukupno najveća točnost postiže se kombinacijom sva tri osjetila. Ako se promatra srednja vrijednost bolja točnost postignuta je u svim scenarijima dok je bolja F1 vrijednost postignuta u 22 od 36 scenarija. Uzimajući u obzir sve scenarije prosječno povećanje točnosti je 4,92%, a F1 vrijednosti 1,27%.

Posljednji rezultati testiranja prikazani su u Tablici 4.8 i odnose se na testiranje vremenske neovisnosti metode, tj. koliko uspješno metoda generalizira podatke s obzirom na period prikupljanja podataka. Jednako kao kod prethodna dva testiranja sveukupno najveća točnost postiže se kombinacijom sva tri osjetila. Višerazinskom metodom postignuta je bolja srednja vrijednost točnosti u svim scenarijima, a F1 vrijednost u 22 od 36 scenarija. Prosječno povećanje točnosti klasifikacije višerazinskom metodom je 7,5% dok je prosječno povećanje F1 vrijednosti približno 1%.

Sveukupno metoda pokazuje veću osjetljivost na lokaciju nošenja uređaja tijekom prikupljanja podataka u odnosu na period prikupljanja podataka i korisničko ponašanje. Dakle, izostavljanje pojedinih lokacija nošenja uređaja tijekom učenja modela pokazuje najveći utjecaj na ukupnu točnost modela. Konkretnije, najveći utjecaj na točnost metode unosi izostavljanje podataka prikupljenih s uređaja koji je nošen u ruci. Uzrok takvoj pogrešci su bitno različite vrijednosti podataka prikupljenih s akcelerometra kada je uređaj nošen u ruci. Primjerice tijekom putovanja korisnik može naglo podići ruku što će bitno promijeniti vrijednost akceleracije, a takav pokret nije povezan s akceleracijom uzrokovanom kretanjem nekim od modova prijevoza. Dok uređaji u torbi ili džepu imaju znatno manji broj takvih događaja.

Ovime se potvrđuje da veća raznolikost podataka, koja se može očitovati u broju korisnika te raznolikosti njihova ponašanja tijekom prikupljanja podataka, značajno pospješuje mogućnosti razlikovanja pojedinih modova prijevoza. Također, kada se usporede postignuti rezultat klasifikatora prema scenarijima, vidljivo je da se postižu veće točnosti kod scenarija koji razlikuju manji broj modova prijevoza. Dok se značajnije pogreške mogu primijetiti kod scenarija u kojima se uvodi razlikovanje između tračničkih sustava i cestovnih vozila.



4. Metoda za segmentaciju i klasifikaciju trajektorije u stvarnom vremenu

Tablica 4.6: Rezultati testiranja korisničke neovisnosti metode za klasifikaciju modova prijevoza

	Korisnik 1						Korisnik 2						Korisnik 3						Srednja vrijednost			
	[81]		RF		Višerazinska		[81]		RF		Višerazinska		[81]		RF		Višerazinska		[81]		Višerazinska	
O Sce.	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T
1	91,8	92,9	95,6	96,8	98,4	98,8	89,2	91,2	93,9	94,8	95,7	96,3	84,9	88,0	93,8	95,6	94,8	96,3	88,6	90,7	96,3	97,1
2	79,4	83,9	82,9	86,9	89,4	92,7	77,2	80,6	80,7	84,7	86,6	90,0	76,9	78,3	82,6	82,5	87,2	87,4	77,8	80,9	87,7	90,0
3	66,2	76,6	74,2	83,7	78,4	88,2	72,9	78,9	72,0	80,2	77,1	85,2	68,1	75,6	71,5	79,1	73,6	82,9	69,1	77,0	76,4	85,4
4	72,0	77,2	68,7	81,0	73,2	85,8	73,1	79,0	75,6	80,6	81,1	85,6	65,1	74,4	65,9	77,5	67,8	81,7	70,1	76,9	74,0	84,4
5	62,7	65,8	67,1	69,1	72,4	74,9	66,4	66,5	63,1	74,0	67,3	80,0	63,3	66,5	69,1	71,8	74,4	78,2	64,1	66,3	71,4	77,7
6	67,5	66,2	65,3	66,7	69,7	72,2	67,5	66,3	65,5	73,6	69,9	79,9	62,2	66,3	65,3	70,3	70,3	77,3	65,7	66,3	70,0	76,5
7	55,2	58,4	61,1	70,3	63,6	74,0	58,2	60,6	59,9	80,6	63,8	85,8	52,7	63,6	63,4	74,2	67,3	79,3	55,3	60,9	64,9	79,7
8	59,9	58,3	59,9	68,1	63,0	72,1	60,6	60,5	62,0	79,1	66,7	85,1	52,7	62,8	60,8	72,2	63,8	77,2	57,7	60,5	64,5	78,1
9	56,6	63,6	56,4	60,5	58,7	63,5	62,3	66,3	54,3	76,3	58,0	82,6	54,4	71,2	61,2	69,5	66,7	75,7	57,8	67,0	61,1	73,9
10	62,5	63,6	55,4	58,1	58,3	61,7	64,8	67,1	58,6	76,0	61,5	81,6	54,4	70,4	60,0	68,0	64,4	74,0	60,6	67,0	61,4	72,4
11	50,0	53,5	50,2	52,4	52,6	55,2	52,0	55,6	46,5	57,2	50,6	61,6	47,0	56,3	54,9	60,8	61,4	67,3	49,7	55,1	54,9	61,4
12	54,9	53,5	50,2	49,8	52,2	52,0	54,7	55,3	51,6	56,7	55,7	60,7	47,1	55,1	53,3	59,1	59,6	66,3	52,3	54,6	55,8	59,7
1	94,2	95,1	95,8	96,9	98,0	98,5	92,7	93,9	95,8	96,3	97,2	97,6	91,1	92,5	96,2	97,2	97,6	98,3	92,9	94,1	97,6	98,1
2	82,5	86,5	97,5	98,1	98,9	99,2	82,5	85,6	90,4	92,6	96,4	97,4	84,3	85,3	87,9	88,2	91,5	91,8	80,6	85,0	95,6	96,1
3	70,9	80,4	72,6	83,3	76,8	87,5	74,7	81,2	73,6	81,7	78,3	86,3	75,9	80,9	80,0	84,1	82,9	87,3	77,4	82,1	79,3	87,0
4	72,8	79,4	72,2	82,1	75,8	86,3	71,5	80,1	70,3	80,6	76,4	85,2	67,1	78,9	71,9	82,2	74,0	85,5	74,6	81,9	75,4	85,7
5	66,8	70,5	70,8	73,2	76,2	78,7	69,5	71,2	64,7	75,4	70,3	81,7	70,9	72,8	76,3	78,3	81,6	84,0	70,7	70,4	76,0	81,5
6	69,1	69,1	70,9	71,7	75,5	77,2	68,9	70,4	63,7	74,4	70,8	81,7	66,5	71,4	69,9	76,2	74,0	81,8	71,2	70,8	73,4	80,2
7	57,4	60,5	61,8	71,5	64,8	74,9	58,9	64,3	58,5	81,1	62,6	86,2	57,2	68,6	69,9	79,7	73,6	83,4	62,0	63,7	67,0	81,5
8	61,0	59,7	63,5	70,4	65,4	73,4	59,5	63,2	58,2	80,5	63,8	85,7	54,6	66,5	64,9	77,4	66,5	80,5	62,8	63,5	65,2	79,9
9	60,2	66,1	59,5	63,2	64,0	68,3	62,1	70,5	55,2	76,4	60,0	82,7	60,4	76,2	66,9	75,0	71,1	80,0	65,7	69,2	65,0	77,0
10	63,4	65,3	61,2	61,8	66,0	67,1	63,2	70,3	54,6	74,9	59,1	80,8	59,5	75,9	63,4	73,1	66,3	77,7	66,8	69,8	63,8	75,2
11	53,1	56,4	50,4	54,3	52,8	56,7	53,5	58,8	47,4	58,6	50,0	61,7	51,5	61,2	61,6	66,3	67,5	72,9	55,9	58,7	56,7	63,7
12	55,5	54,5	55,3	53,5	58,4	57,0	54,4	57,5	48,5	57,6	53,6	62,0	50,3	59,5	58,5	63,9	63,7	70,6	57,3	58,7	58,6	63,2
1	95,2	96,0	96,4	97,4	98,1	98,6	93,2	94,4	96,4	96,8	98,2	98,4	91,3	92,7	96,1	97,2	97,2	98,0	93,2	94,4	97,8	98,3
2	88,5	91,1	86,6	89,5	90,5	93,1	84,8	88,0	86,8	90,6	89,6	93,1	87,6	88,6	88,2	88,5	91,8	92,1	86,9	89,2	90,6	92,8
3	75,0	84,2	73,6	85,1	77,2	88,5	81,9	86,1	77,0	85,4	79,4	87,3	80,1	84,6	75,9	82,7	79,2	86,0	79,0	85,0	78,6	87,3
4	76,3	83,7	69,1	82,7	71,7	85,7	79,7	86,6	79,9	85,3	83,5	87,8	70,3	82,7	68,6	81,1	70,9	84,4	75,4	84,3	75,4	86,0
5	73,0	77,3	76,4	81,5	81,0	86,8	79,7	80,5	73,3	83,9	78,3	87,8	77,9	80,1	77,0	80,5	81,5	85,3	76,9	79,3	80,3	86,6
6	74,6	76,7	72,0	79,0	75,6	84,3	76,0	80,1	76,4	83,8	81,2	87,7	72,2	78,6	69,5	78,3	73,8	83,3	74,3	78,5	76,9	85,1
7	63,2	67,3	58,4	70,8	60,9	73,8	68,6	71,2	62,8	85,8	66,0	88,9	62,2	75,1	64,3	77,4	66,2	80,6	64,7	71,2	64,4	81,1
8	66,5	67,2	56,8	68,1	59,1	71,1	68,3	71,2	67,4	85,5	70,6	88,5	60,6	74,7	59,5	75,3	62,1	78,6	65,1	71,0	63,9	79,4
9	63,4	69,8	58,9	67,3	61,0	70,5	71,1	73,9	62,7	84,4	68,1	88,8	63,2	79,6	67,0	76,5	71,2	81,4	65,9	74,4	66,8	80,2
10	67,4	70,0	56,9	64,4	58,6	67,3	70,7	74,1	66,4	84,3	71,8	88,7	59,2	77,0	61,9	74,0	65,3	78,1	65,8	73,7	65,2	78,0
11	58,0	60,3	50,9	55,5	52,7	58,1	63,0	64,9	54,9	65,1	60,8	70,6	56,9	66,2	60,5	66,4	65,5	71,5	59,3	63,8	59,7	66,7
12	60,5	59,6	49,9	52,3	52,0	55,4	63,1	64,6	59,7	64,7	64,7	69,8	56,0	65,5	56,3	64,0	60,4	69,2	59,9	63,3	59,0	64,8

Tablica 4.7: Rezultati testiranja lokacijske neovisnosti metode za klasifikaciju modova prijevoza

O	Sec.	Lokacija 1			Lokacija 2			Lokacija 3			Lokacija 4			Srednja vrijednost															
		[81]		RF	Višerazinska		[81]		RF	Višerazinska		[81]		RF	Višerazinska		[81]		Višerazinska										
		F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T								
	1	86,3	88,5	90,4	92,3	94,5	95,7	93,6	94,7	94,9	96,1	95,6	96,6	91,7	93,1	96,1	97,0	97,3	97,9	92,9	94,1	94,2	95,7	94,9	96,2	91,1	92,6	95,6	96,6
	2	74,8	78,3	74,4	79,0	79,4	84,5	83,9	86,5	83,7	85,1	88,9	90,4	75,7	80,5	87,2	88,9	91,5	92,8	87,2	88,5	87,9	88,8	91,3	91,9	80,4	83,4	87,8	89,9
	3	66,1	73,0	60,4	73,9	63,4	78,6	70,4	77,7	78,9	82,2	83,4	87,0	75,6	79,7	69,3	82,7	70,6	85,7	84,4	87,4	74,6	85,2	76,8	88,1	74,1	79,4	73,6	84,9
	4	70,1	72,5	62,8	73,5	65,7	78,3	74,4	78,5	77,8	81,5	80,8	86,1	76,2	77,7	66,3	80,7	68,9	84,6	85,3	86,4	74,8	84,7	77,1	87,7	76,5	78,8	73,1	84,2
	5	60,4	61,4	58,7	59,3	65,9	65,7	67,7	68,4	71,1	70,9	76,2	76,3	69,5	68,9	64,6	70,0	68,9	75,4	79,3	78,9	72,0	76,8	74,6	80,8	69,2	69,4	71,4	74,6
A	6	64,8	60,7	55,1	56,5	59,1	61,2	73,6	71,0	72,3	70,3	76,6	76,0	72,4	68,2	65,9	69,2	68,7	74,4	81,4	78,5	71,5	76,2	74,5	81,1	73,1	69,6	69,7	73,2
	7	51,1	53,2	51,2	61,1	56,6	67,2	61,5	64,1	60,8	68,5	64,2	73,3	60,8	63,1	57,9	69,7	60,6	73,5	68,4	69,9	74,1	79,3	78,2	83,9	60,5	62,6	64,9	74,5
	8	57,2	54,6	53,8	60,6	57,4	66,4	65,1	63,9	63,4	68,3	66,3	73,3	64,5	62,5	60,5	68,3	63,1	72,1	72,2	70,2	73,2	78,4	76,2	82,5	64,8	62,8	65,8	73,6
	9	54,8	60,2	49,0	53,1	54,8	60,1	63,5	69,2	55,2	63,0	57,6	67,5	64,5	68,7	57,0	65,8	61,4	71,0	73,7	75,5	66,9	73,9	69,8	78,1	64,1	68,4	60,9	69,2
	10	62,0	61,4	43,2	53,3	51,0	62,6	65,3	66,9	58,3	62,4	59,3	65,9	68,2	68,4	59,3	64,5	62,4	68,9	77,1	75,5	66,9	72,8	69,4	77,0	68,1	68,1	60,5	68,6
	11	46,2	48,8	41,4	45,6	45,5	50,0	54,4	57,6	50,1	55,0	54,5	59,7	54,6	58,0	51,6	58,1	55,1	62,0	60,3	63,9	59,5	64,9	63,0	69,0	53,9	57,1	54,5	60,2
	12	52,5	50,3	45,4	45,2	49,2	49,8	59,2	58,2	53,5	54,2	57,0	58,6	58,1	57,0	53,0	56,2	56,9	60,2	65,5	64,4	59,8	64,1	62,9	67,7	58,8	57,5	56,5	59,1
	1	88,4	89,7	91,7	93,2	96,0	96,8	95,6	96,4	95,8	96,8	96,9	97,6	94,8	95,7	96,1	97,0	97,4	97,9	94,8	95,7	96,2	97,1	97,7	98,2	93,4	94,4	97,0	97,6
	2	75,5	79,6	78,4	82,3	83,8	87,9	86,8	89,4	87,7	89,2	92,8	93,9	81,9	85,7	88,0	89,7	92,4	93,7	89,4	90,9	88,2	89,6	92,6	93,5	83,4	86,4	90,4	92,3
	3	65,3	72,0	68,9	77,5	75,5	83,2	71,7	79,6	80,9	85,4	85,5	89,9	79,5	83,1	71,1	83,7	74,0	87,0	86,8	89,2	79,1	84,4	81,9	87,2	75,8	81,0	79,2	86,8
	4	58,0	70,6	72,8	77,2	80,0	83,5	72,0	78,5	81,7	85,4	84,7	89,4	81,0	82,5	73,9	83,4	76,6	86,6	87,9	89,2	82,8	87,3	84,7	90,4	74,7	80,2	81,5	87,5
	5	60,9	61,3	62,0	63,4	68,5	69,6	68,7	70,4	80,1	81,3	84,9	86,8	74,8	74,1	67,6	73,1	71,7	78,0	81,9	81,5	82,4	83,0	87,8	89,3	71,6	71,8	78,2	80,9
A	6	54,2	59,1	65,9	63,2	71,6	69,0	69,1	68,5	80,0	80,8	84,0	86,4	76,5	73,1	71,7	72,8	75,7	78,1	83,5	81,1	80,1	81,3	83,9	86,6	70,8	70,4	78,8	80,0
Ž	7	53,5	55,1	56,0	65,3	59,7	69,8	59,8	63,3	66,3	73,9	69,0	77,7	65,2	67,4	59,6	71,1	62,6	75,0	71,5	72,6	76,7	79,5	80,6	83,9	62,5	64,6	68,0	76,6
	8	49,3	52,6	61,9	65,4	66,2	70,5	61,2	60,9	69,4	73,5	71,5	77,5	68,4	66,6	64,7	70,8	67,3	74,3	74,4	72,6	76,5	77,6	79,4	81,8	63,4	63,2	71,1	76,0
	9	56,3	60,8	52,0	57,5	56,4	62,8	60,8	68,0	63,4	70,8	66,6	74,9	69,6	73,2	58,2	67,2	62,3	72,3	76,0	77,6	74,2	77,1	80,0	83,4	65,7	69,9	66,3	73,4
	10	51,6	59,5	56,5	57,3	60,3	62,6	62,7	66,0	65,9	70,0	69,0	74,3	72,8	72,8	62,2	66,7	66,7	72,0	78,9	78,0	75,0	77,2	79,5	83,2	66,5	69,1	68,9	73,0
	11	45,6	48,5	46,7	50,8	50,8	55,2	54,6	57,5	57,5	61,6	61,8	65,9	58,9	61,5	50,3	57,0	54,0	61,1	63,6	66,9	64,8	68,4	68,6	72,5	55,7	58,6	58,8	63,7
	12	44,7	48,2	51,4	50,7	54,7	54,8	55,4	54,2	60,3	60,9	64,0	65,0	62,1	60,6	54,9	56,1	58,3	60,1	67,6	67,4	64,9	65,7	68,6	70,2	57,5	57,6	61,4	62,5
	1	90,5	91,6	93,1	94,4	96,9	97,5	95,9	96,5	95,9	96,8	96,8	97,5	95,5	96,3	96,5	97,3	97,7	98,2	95,2	96,0	96,3	97,1	98,0	98,4	94,3	95,1	97,4	97,9
	2	81,5	84,9	80,1	84,1	83,7	88,2	89,5	91,7	87,5	89,4	92,0	93,4	88,5	90,5	90,2	91,8	93,8	95,0	90,8	92,3	92,3	93,2	96,0	96,6	87,6	89,8	91,4	93,3
	3	71,1	77,5	63,8	77,1	66,8	81,0	77,0	84,4	78,9	85,2	82,2	88,4	85,4	88,7	73,0	85,6	75,8	88,7	87,4	90,3	82,7	88,9	85,9	91,7	80,2	85,2	77,7	87,5
	4	62,0	75,8	67,1	76,1	69,7	79,5	74,4	81,7	77,4	84,3	80,2	87,4	84,9	87,4	65,6	81,7	67,7	84,6	88,2	90,0	79,1	87,5	82,2	90,4	77,4	83,7	75,0	85,5
	5	70,2	71,3	63,9	69,2	69,4	75,8	75,8	78,6	82,4	85,9	87,0	90,4	83,8	84,5	74,8	82,3	78,5	86,5	86,1	86,5	84,4	88,3	88,3	92,2	79,0	80,2	80,8	86,2
A	6	61,6	68,8	63,5	68,0	68,3	75,2	74,0	75,8	80,5	84,8	84,6	89,4	84,1	83,4	69,2	78,7	72,6	83,3	86,7	86,0	82,0	87,4	84,2	91,3	76,6	78,5	77,4	84,8
Ž	7	60,5	63,1	53,0	64,6	55,6	68,3	66,4	70,3	62,4	72,7	65,7	76,2	74,9	77,4	58,8	72,4	60,2	75,0	75,1	77,1	74,1	78,3	77,5	81,6	69,2	72,0	64,8	75,3
M	8	55,4	60,9	54,2	63,6	55,1	66,5	66,4	67,5	64,0	71,7	67,0	75,2	76,7	76,9	55,8	69,1	57,5	71,8	77,3	76,9	73,2	77,6	76,5	81,1	69,0	70,5	64,0	73,7
	9	61,5	66,1	56,0	63,5	61,8	69,4	66,8	72,9	62,7	72,7	66,3	76,6	75,8	79,1	60,3	71,2	63,5	74,8	78,0	79,8	78,1	82,1	82,8	86,4	70,5	74,4	68,6	76,8
	10	55,6	64,7	60,1	62,7	64,9	68,1	65,7	69,4	64,5	72,0	66,9	75,3	78,1	78,6	58,6	67,7	62,7	72,8	80,4	79,8	76,4	81,2	79,8	85,1	70,0	73,1	68,6	75,3
	11	55,0	55,9	51,2	54,9	58,0	61,9	61,3	63,3	55,6	61,4	58,7	64,7	67,9	69,7	53,3	59,9	56,0	63,2	68,8	70,6	65,9	69,4	68,0	71,9	63,2	64,9	60,2	65,4
	12	51,8	54,5	53,3	54,3	58,7	60,7	61,5	59,9	57,6	60,2	61,3	64,5	69,5	68,2	51,9	56,0	54,9	59,4	70,7	69,3	66,8	69,4	69,0	72,1	63,4	63,0	61,0	64,2

**Tablica 4.8:** Rezultati testiranja vremenske neovisnosti metode za klasifikaciju modova prijevoza

		Period 1			Period 2			Period 3			Period 4			Srednja vrijednost															
		[81]	RF	Višerazinska	[81]	RF	Višerazinska	[81]	RF	Višerazinska	[81]	RF	Višerazinska	[81]	Višerazinska														
O	Sec.	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T	F1	T														
A	1	92,3	93,7	98,0	98,1	98,8	98,9	93,2	94,3	99,7	99,8	100,0	100,0	93,3	94,1	98,9	99,3	99,8	99,9	92,4	94,1	98,7	98,8	99,9	99,9	92,8	94,1	99,6	99,7
A	2	83,8	85,8	92,1	93,1	95,8	96,4	83,9	86,9	74,9	96,1	75,1	96,3	82,6	85,6	90,6	94,9	94,6	97,4	82,4	84,6	91,6	95,2	94,3	97,3	83,2	85,7	90,0	96,9
A	3	81,5	84,4	89,2	90,0	93,2	93,7	81,8	85,8	61,5	85,2	61,8	85,6	81,3	83,9	89,0	93,8	94,0	96,6	79,6	83,2	76,7	85,4	78,1	86,6	81,0	84,2	81,8	90,6
A	4	82,4	83,7	89,7	89,7	94,6	94,4	84,3	85,1	61,5	85,2	61,8	85,6	83,8	83,5	91,8	94,4	96,8	97,8	82,0	82,9	79,4	86,1	81,4	87,9	83,1	83,8	83,7	91,4
A	5	75,8	75,2	88,5	88,5	93,7	93,8	76,6	76,4	61,4	85,0	61,4	85,9	75,4	74,7	87,6	88,5	93,3	93,5	73,0	72,5	61,1	78,7	63,5	84,2	75,2	74,7	78,0	89,4
A	6	76,6	73,6	89,9	89,1	94,9	94,4	79,5	76,1	61,4	85,0	61,4	85,9	78,5	74,5	89,9	88,7	95,6	94,3	75,8	71,7	65,5	78,5	68,2	84,1	77,6	74,0	80,0	89,7
A	7	67,5	67,9	77,5	90,3	75,1	93,8	68,8	69,5	71,3	84,9	71,4	85,9	66,8	68,8	78,2	74,8	80,7	76,9	64,6	69,5	62,3	62,8	63,5	63,5	66,9	68,9	72,7	80,0
A	8	70,3	67,7	75,8	90,2	78,7	93,8	72,4	69,2	68,4	85,0	68,6	85,9	70,7	68,7	82,0	75,1	85,8	77,9	69,0	69,6	65,9	61,6	65,0	59,9	70,6	68,8	74,5	79,4
A	9	72,8	73,5	80,9	87,4	88,3	92,2	72,4	73,7	76,5	90,0	78,4	92,9	71,4	74,0	74,9	72,7	80,5	78,6	69,1	76,4	55,8	64,1	59,0	67,5	71,4	74,4	76,6	82,8
A	10	74,9	72,7	82,3	87,5	86,1	91,9	76,6	74,0	77,2	85,0	78,6	87,8	75,1	73,8	79,3	75,1	86,5	82,4	73,7	76,7	60,7	65,0	65,3	70,6	75,1	74,3	79,1	83,2
A	11	59,9	64,2	73,3	79,6	79,6	86,3	61,7	64,8	48,9	62,0	52,0	69,1	60,0	64,0	63,9	64,9	70,5	70,7	56,9	60,5	48,8	68,6	52,1	72,9	59,6	63,4	63,6	74,8
A	12	63,0	63,7	76,2	79,6	83,3	84,7	65,6	64,5	47,7	60,6	50,3	68,6	63,9	63,3	72,2	70,3	81,7	81,6	61,5	60,4	54,3	68,9	58,5	75,7	63,5	63,0	68,5	77,7
A	1	95,3	96,1	98,3	98,4	99,0	99,1	95,3	96,1	99,7	99,8	100,0	100,0	95,2	95,7	99,1	99,5	99,8	99,9	95,1	96,1	98,8	98,9	99,9	99,9	95,2	96,0	99,7	99,7
A	2	87,0	89,1	92,0	93,2	95,1	95,9	85,7	89,1	75,0	96,3	65,2	96,4	85,1	88,3	90,3	94,6	96,1	98,0	85,4	87,6	91,1	95,1	94,3	97,3	85,8	88,5	87,7	96,9
A	3	86,1	88,0	89,6	90,1	92,2	92,8	84,4	87,6	61,0	83,4	67,4	88,4	83,9	86,4	86,9	92,6	92,9	96,6	83,7	86,1	76,5	85,4	79,5	87,3	84,5	87,0	83,0	91,3
A	4	85,5	87,0	89,8	89,7	93,0	92,9	86,7	87,6	61,6	85,3	61,9	85,8	85,9	86,2	89,2	92,8	94,0	96,6	85,1	85,8	73,6	83,1	75,2	84,4	85,8	86,6	81,0	89,9
A	5	80,1	79,6	89,8	90,0	93,0	93,1	79,7	79,6	61,6	85,3	62,0	85,9	78,5	77,9	84,2	84,8	89,7	88,5	77,4	76,2	60,5	77,4	63,9	84,2	78,9	78,3	77,2	87,9
A	6	80,4	78,2	90,1	89,3	94,3	93,9	82,4	79,4	61,5	85,2	62,5	86,0	81,1	77,6	87,3	85,4	92,7	91,1	79,5	75,6	61,1	75,6	64,2	81,0	80,8	77,7	78,4	88,0
A	7	71,8	71,9	73,1	91,0	75,2	93,6	72,0	72,4	71,6	85,3	71,9	86,8	70,4	72,0	77,8	77,2	84,9	83,8	68,9	73,2	59,2	59,9	60,5	60,4	70,8	72,4	73,1	81,2
A	8	74,1	71,6	76,2	90,5	78,9	93,8	75,2	72,2	61,5	85,2	61,8	85,6	73,5	71,7	82,8	77,6	89,6	84,8	72,6	73,2	59,7	58,5	60,3	58,8	73,9	72,2	72,7	80,8
A	9	77,3	77,1	78,8	89,0	83,4	92,7	75,8	76,5	71,5	85,2	71,5	85,4	74,9	76,9	68,9	73,6	74,8	80,8	73,7	79,8	50,2	58,2	54,2	62,9	75,4	77,6	71,0	80,5
A	10	78,5	76,2	79,4	88,5	80,3	93,1	79,3	76,5	71,7	85,4	72,0	85,9	77,9	76,5	75,2	73,7	83,9	84,3	77,0	79,6	52,5	56,4	55,0	59,9	78,2	77,2	72,8	80,8
A	11	64,3	68,3	68,7	80,5	73,2	86,6	65,3	68,1	42,1	53,2	46,6	58,9	63,5	67,1	60,8	65,1	66,8	72,1	61,6	64,5	45,7	62,6	48,3	65,8	63,7	67,0	58,7	70,9
A	12	67,0	67,7	71,0	79,2	75,9	86,0	68,7	67,6	43,6	53,4	44,0	61,2	67,3	66,9	67,3	66,2	73,1	72,8	64,9	63,8	47,7	62,9	51,8	69,3	67,0	66,5	61,2	72,3
A	1	95,5	96,2	98,0	98,1	99,3	99,3	95,7	96,4	99,7	99,8	100,0	100,0	95,4	95,9	99,4	99,6	99,8	99,9	95,5	96,4	98,8	98,9	99,9	99,9	95,5	96,2	99,8	99,8
A	2	90,1	91,6	95,1	95,5	96,6	97,0	89,1	91,6	84,9	96,7	93,1	99,0	88,4	90,9	94,7	97,0	98,0	99,0	89,7	91,5	91,1	95,2	94,5	97,3	89,3	91,4	95,6	98,1
A	3	87,8	90,0	94,3	94,4	96,3	96,3	86,7	89,9	67,3	88,4	72,6	90,5	86,5	89,0	91,8	96,0	95,6	98,2	87,2	90,1	76,4	85,3	78,9	86,9	87,1	89,7	85,9	93,0
A	4	87,3	89,3	94,2	94,0	96,0	95,8	88,2	89,8	70,2	89,3	75,2	91,1	87,7	88,7	92,5	95,8	96,6	98,3	87,8	89,7	73,5	82,6	76,6	85,0	87,7	89,4	86,1	92,6
A	5	85,9	86,1	94,6	94,7	97,1	97,2	85,2	85,6	67,6	88,3	75,8	90,7	83,7	83,8	92,4	94,5	96,1	97,5	84,8	85,3	62,5	84,3	65,2	88,0	84,9	85,2	83,6	93,4
A	6	85,4	85,1	94,8	94,4	97,1	96,9	86,8	85,4	62,6	91,1	86,0	94,5	85,3	83,5	93,2	94,4	96,8	97,5	85,1	84,4	63,3	82,2	65,5	85,8	85,7	84,6	86,4	93,7
A	7	76,6	77,1	75,5	93,9	77,2	95,8	77,0	77,8	72,8	90,1	79,7	92,5	75,2	77,2	85,0	85,0	90,5	90,1	74,5	80,7	60,5	61,9	62,0	62,7	75,8	78,2	77,4	85,3
A	8	78,4	77,3	78,9	93,8	80,3	95,4	79,4	77,5	63,4	92,5	85,8	94,6	77,7	76,9	89,0	86,9	93,8	91,9	77,6	80,9	61,7	59,6	63,1	60,0	78,3	78,1	80,8	85,5
A	9	79,3	79,5	85,7	93,9	91,1	96,4	78,6	79,3	76,9	88,3	81,0	90,8	77,8	80,0	84,9	85,5	91,8	92,4	77,2	83,9	54,3	64,6	58,4	69,4	78,2	80,7	80,6	87,3
A	10	80,6	79,1	85,5	93,7	89,6	96,1	81,2	79,0	74,5	92,8	93,6	97,5	80,5	79,7	85,4	84,5	92,7	92,3	80,0	83,9	53,1	58,3	56,9	62,8	80,6	80,4	83,2	87,2
A	11	70,0	72,2	76,6	84,2	76,8	89,9	70,9	72,3	48,0	59,1	50,4	60,8	68,7	71,0	77,5	78,1	84,5	85,3	67,8	70,3	46,1	65,2	50,1	70,1	69,4	71,4	65,5	76,5
A	12	72,1	72,1	77,2	83,6	78,4	88,6	73,6	71,9	50,3	60,1	55,9	63,6	71,7	70,6	78,8	74,5	84,5	80,6	70,6	69,6	48,4	63,1	51,2	66,6	72,0	71,1	67,5	74,9

---

# Validacija metode

---

Validacija metode je proces utvrđivanja koliko dosljedno model opisuje ponašanje sustava. Valjanost modela procjenjuje se operativno i konceptualno. Operativna procjena utvrđuje slaže li se izlaz modela s promatranim podacima dok konceptualna procjena uključuje jesu li pretpostavke na kojima se zasniva model opravdane. Kod procesa validacije važno je uzeti u obzir sličnost sustava i podataka koji se uspoređuju [138]. Validacija je u ovom istraživanju provedena na potpuno neovisnom vlastitom skupu podataka. Skup podataka prikupljen je mobilnom aplikacijom te sadrži podatke s osjetila mobilnog uređaja. Svaki zapis u skupu podataka označen je modom prijevoza koji korisnik koristi. Za proces validacije korišteni su samo modovi prijevoza koji su zastupljeni u SHL skupu podataka. Također, skup podataka prikupljen je na prometnoj mreži Republike Hrvatske, koja ima bitno različito ponašanje od područja na kojem je prikupljen SHL skup podataka, pa se neizravno validacijom ispituje utjecaj ponašanja prometne mreže na rezultate klasifikacije.

Ovo poglavlje je strukturirano u tri osnovne cjeline. U potpoglavlju 5.1 opisana je mobilna aplikacija *Collecty* koja se koristi za prikupljanje podataka kretanja korisnika prometnom mrežom. Prikazane su osnovne komponente podataka, koji se prikupljaju, mogućnosti mobilne aplikacije te dva moguća načina prikupljanja i pohranjivanja podataka. Potpoglavlje 5.2 daje osnovne informacije o prikupljenom skupu podataka. Rezultati validacije prikazani su u potpoglavlju 5.3. Zasebno su prikazani validacijski rezultati metode za segmentaciju trajektorije te rezultati višerazinske metode za klasifikaciju trajektorije. U ovom poglavlju prikazan je dodatno rezultat metode za klasifikaciju modova prijevoza kada je jedan od uključenih modova prijevoza električni romobil. U dostupnoj literaturi nema modela niti javno objavljenih podataka koji uključuju električni romobil u modove prijevoza.

## 5.1 Mobilna aplikacija Collecty

Mobilni uređaji zajedno s aplikacijama zasnovanim na lokaciji značajno su promijenili smjer istraživanja, koja sadrže prostornu komponentu, kao i način korištenja mobilnih uređaja. Mobilna aplikacija razvijena u ovom istraživanju može se svrstati u aplikacije zasnovane na lokaciji jer je jedan od podataka koji se prikuplja GPS lokacija korisnika. Aplikacija je razvijena za mobilne uređaje s ugrađenim operacijskim sustavom *Android* te je za razvoj korišteno integrirano razvojno okruženje *Android Studio* i programski jezik *Java*. Razvijena aplikacija organizirana je u 6 aktivnosti za čiji je razvoj korišten programski okvir *Sensors* koji se koristi za prikupljanje podataka s osjetila mobilnog uređaja, dok je za prikupljanje GPS lokacije uređaja korištena klasa *LocationManager*. Za uspješan rad aplikacije potrebna je komunikacija aplikacije s poslužiteljem, za čiju uspostavu je korišten ASP.NET okvir.

### 5.1.1 Programski okvir *Sensors*

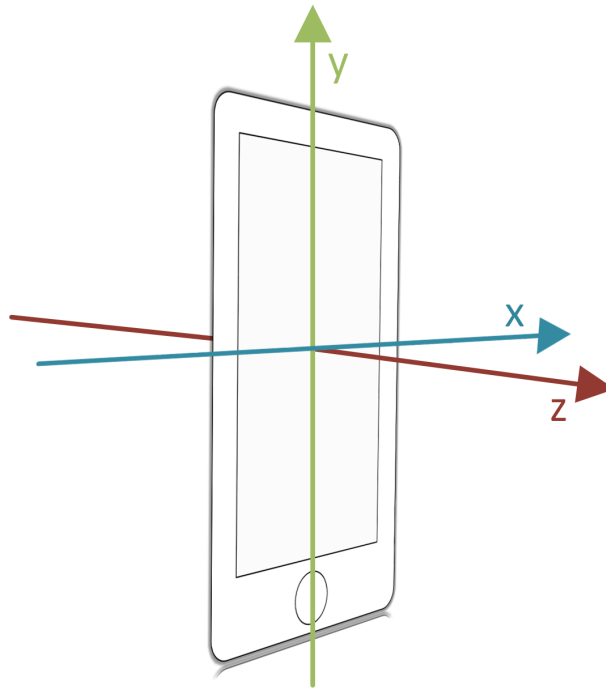
Većina uređaja s ugrađenim operacijskim sustavom *Android* ima ugrađena osjetila koja mjere kretanje i orijentaciju uređaja. Neobrađeni podatci s tih osjetila su visoko precizni i mogu se koristiti za praćenje kretanja ili pozicioniranje uređaja u tri dimenzije: po  $x$ ,  $y$  i  $z$  osi osjetila. U različitim područjima mogu se primijeniti različita osjetila i njihova mjerenja. Tako, primjerice za razvoj igara koriste se mjerenja gravitacije kako bi se identificirali složeni pokreti mobilnog uređaja poput naginjanja ili zamaha dok su za aplikacije u području mobilnosti korisnika značajnija osjetila poput akcelerometra, žiroskopa, magnetometra, podatci s osjetila za mjerenje temperature uređaja ili podatci s osjetila vlažnosti ako se proučava utjecaj vremenskih uvjeta na mobilnost korisnika.

Programski okvir sadrži više skupina osjetila, od kojih su u ovom istraživanju korištene dvije skupine:

1. Osjetila pokreta - osjetila koja mjere sile ubrzanja i rotacije. Za razvoj aplikacije korišteno je osjetilo akcelerometra, gravitacije, žiroskopa i rotacije.
2. Osjetila položaja - mjere fizički položaj mobilnog uređaja. Za razvoj aplikacije je korišteno osjetilo magnetometar.

Programski okvir *Sensors* koristi standardni 3-osni koordinatni sustav za prikaz mjerenja s osjetila. Koordinatni sustav je definiran u odnosu na zaslon uređaja kada se uređaj nalazi u predefiniranoj orijentaciji, Slika 5.1. Dakle, os  $x$  je vodoravna i usmjerena je u desno, os  $y$  je okomita i usmjerena je prema gore, dok je os  $z$  usmjerena prema vanjskoj strani mobilnog uređaja. Predefinirana orijentacija uređaja nije uvijek jednaka, primjerice za većinu tablet uređaja predefinirana je orijentacija vodoravna dok je za mobilne uređaje okomita. Za obradu mjerenja s osjetila bitna je činjenica da se položaj koordinatnog sustava ne mijenja bez obzira na promjenu položaja mobilnog uređaja. Za sva osjetila čija su mjerenja prikupljana aplikacijom koristi se

opisani koordinatni sustav.



**Slika 5.1:** Koordinatni sustav programskog okvira *Sensors* u odnosu na mobilni uređaj

Za identifikaciju osjetila, koja se nalaze na mobilnom uređaju, korištena je klasa *SensorManager* koja unutar sebe ima više metoda za prikupljanje podataka. Dvije metode koje su korištene u tu svrhu su *onAccuracyChanged()* i *onSensorChanged()*. Metoda *onAccuracyChanged()* se poziva u slučajevima u kojima se promjeni točnost osjetila dok se metoda *onSensorChanged()* poziva kada osjetilo promjeni vrijednost. U slučaju promjene vrijednosti osjetila informacije o novim podacima osjetila uključuju: točnost podataka, osjetilo koje je generiralo podatke, vremensku oznaku zapisa i nova mjerenja koja je osjetilo zabilježilo. U aplikaciji nisu zapisivani podatci s osjetila u trenutku kada se promijenila vrijednost svakog osjetila već je implementiran brojač vremena koji je svakih 10 ms aktivirao prikupljanje zadnjih vraćenih vrijednosti metode *onSensorChanged()* za svako osjetilo.

### Akcelerometar

Osjetilo akcelerometra mjeri ubrzanje primijenjeno na uređaj, uključujući silu gravitacije. Iz tog razloga, ako uređaj miruje na stolu, akcelerometar očitava vrijednosti  $9,81 \text{ m/s}^2$ . Za mjerenje stvarnog ubrzanja uređaja, doprinos sile gravitacije mora se ukloniti iz podataka akcelerometra. U Collecty aplikaciji korištena je vrsta mjerenja s akcelerometra s već uklonjenom komponentom gravitacije. Gotovo svaki Android uređaj ima ugrađen akcelerometar te je njegova osnovna prednost da uzrokuje oko 10 puta manju potrošnju energije od ostalih osjetila pokreta.

### Žiroskop

Žiroskop mjeri brzinu rotacije u  $rad/s$  oko  $x$ ,  $y$  i  $z$  osi uređaja. Koordinatni sustav osjetila jednak je standardnom koordinatnom sustavu, a rotacija je pozitivna u smjeru suprotnom od kazaljke na satu. Drugim riječima, ako se uređaj postavljen na ishodištu koordinatnog sustava i promatra se s nekog pozitivnog mjesta na osi  $x$ ,  $y$  ili  $z$ , rotacija ima pozitivnu vrijednost ako se uređaj rotira suprotno od kazaljke na satu.

### Gravitacija

Osjetilo gravitacije daje trodimenzionalni vektor koji pokazuje smjer i veličinu gravitacije. Obično se ovo osjetilo koristi za određivanje relativne orijentacije uređaja u prostoru.

### Rotacija

Vektor rotacije predstavlja orijentaciju uređaja kao kombinaciju kuta i osi, u kojoj se uređaj zarotirao za kut  $\theta$  oko  $x$ ,  $y$  ili  $z$  osi. U mobilnim uređajima ne postoji osjetilo rotacije već rotacija predstavlja virtualno osjetilo koje nastaje kombinacijom podataka s osjetila akcelerometra, žiroskopa i magnetometra. Matrica rotacije je koncept linearne algebre koji prevodi podatke osjetila iz jednog koordinatnog sustava u drugi. U ovom slučaju, iz koordinatnog sustava uređaja u Zemljin koordinatni sustav.

Iz podatka o rotaciji može se izračunati orijentacija koja predstavlja kutove orijentacije uređaja. Svi kutovi su u radijanima te su prikazani s tri karakteristike: azimut (kut između trenutnog smjera kompasa uređaja i magnetskog sjevera), nagib (kut između ravnine paralelne sa zaslonom uređaja i ravnine paralelne s tlom) i zakret (kut između ravnine okomite na zaslon uređaja i ravnine okomite na tlo).

### Magnetometar

Osjetilo magnetometra omogućuje praćenje promjena u magnetskom polju zemlje. Rezultat su podatci o jakosti polja u  $\mu T$  za svaku od tri koordinatne osi.

#### 5.1.2 Klasa *LocationManager*

Klasa *LocationManager* omogućuje pristup uslugama lokacije mobilnog uređaja. Te usluge omogućuju aplikacijama ažuriranja geografskih koordinata položaja mobilnog uređaja sa zadanim periodom uzorkovanja ili generiranje obavijesti kada se mobilni uređaj nalazi u blizini unaprijed zadane geografske lokacije.

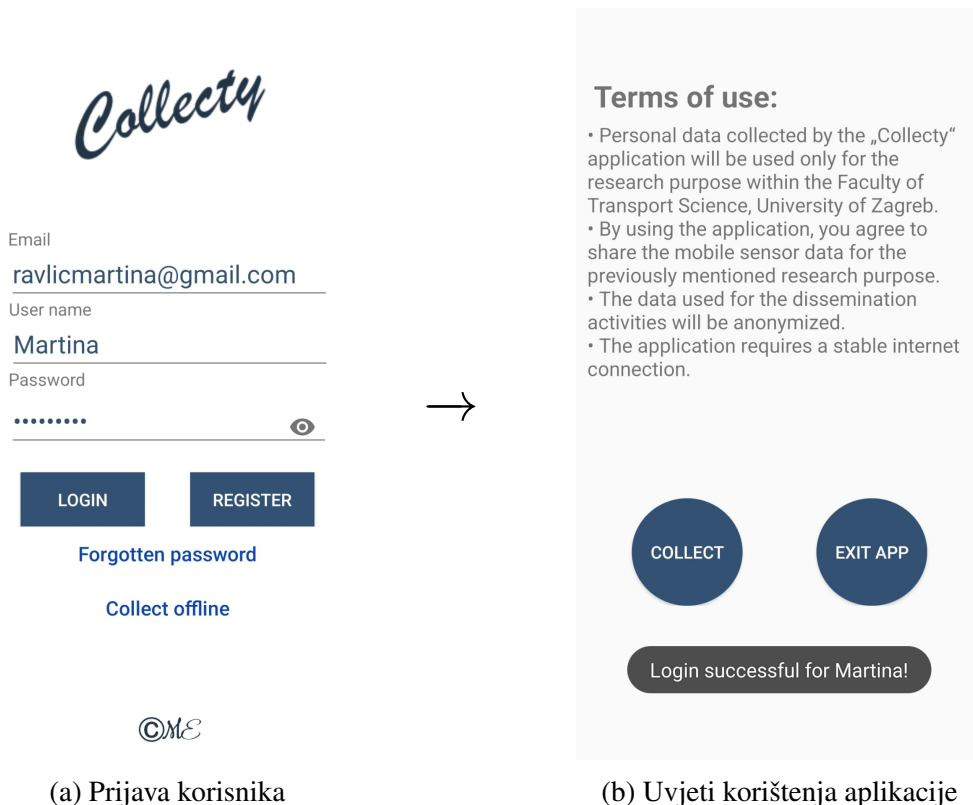
Prije nego aplikacija započne primati ažuriranja lokacije mobilnog uređaja, potrebno je deklarirati ispravna dopuštenja u aplikaciji. Dio dopuštenja o prikupljanju lokacije mobilnog ure-

đaja uključuje korisničko odobrenje pri prvom pokretanju aplikacije. Također, potrebno je definirati davatelja lokacija. Većina suvremenih Android uređaja može primiti ažuriranja lokacije putem više tehnologija, tj. davatelja usluga. Davatelji usluge lokacije mogu imati različite karakteristike izvedbe u smislu točnosti, srednjeg vremena oporavka sustava, potrošnje energije i slično. U mobilnoj aplikaciji *Collecty* davatelj usluge lokacije je GPS.

Lokacija mobilnog uređaja prikuplja se u obliku geografske dužine i širine čije veličine se izračunavaju na temelju primarne GNSS antene unutar uređaja i GNSS satelita. Korištena je metoda *onLocationChanged* za prikupljanje geografskih lokacija. Metoda *onLocationChanged* prima parametre minimalnog vremena i interval udaljenosti u kojima će se primiti ažuriranja lokacije. U aplikaciji *Collecty* interval je definiran vremenski i iznosi 10 ms kao kod preostalih osjetila.

### 5.1.3 Struktura mobilne aplikacije

Mobilna aplikacija *Collecty* podijeljena je u 6 slijednih aktivnosti. Prva aktivnost prikazana je na Slici 5.2 i odnosi se na prijavu ili registraciju korisnika u aplikaciju te odabir načina prikupljanja podataka, Slika 5.2a.



**Slika 5.2:** Početna aktivnost za prijavu korisnika u mobilnu aplikaciju Collecty

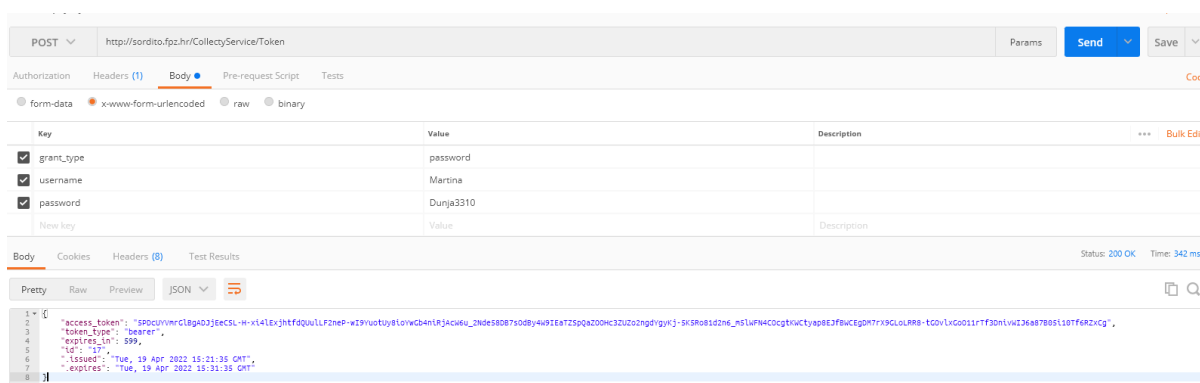
Ako korisnik prvi put pokreće aplikaciju potrebno se registrirati u sustav. Za registraciju, korisnik upisuje svoju adresu elektroničke pošte, odabrano korisničko ime i zaporku. Zaporka



## 5. Validacija metode

mora sadržavati 4 znaka, jedno veliko slovo, jedno malo slovo i jedan broj. Odabrani korisnički podatci se zapisuju u bazu podataka te se korisničkim podacima dodjeljuje token koji će pri svakom novom pokretanju aplikacije biti pridružen istim korisničkim podacima. Pri ponovnom pokretanju aplikacije korisnik upisuje samo korisničko ime i zaporku koje se provjeravaju u bazi podataka i ako su uneseni podatci točni započinje sljedeća aktivnost.

Korisnički podatci pohranjuju se u bazi podataka koja je smještena na lokalnom poslužitelju. Za komunikaciju mobilne aplikacije s poslužiteljem korišten je ASP.NET okvir u formi aplikacijskog programskog sučelja (engl. *Application Program Interface*, API). Cijela aplikacija može se podijeliti u skup APIa za rad s korisničkim podacima i skup APIa za rad s podacima s mobitela. Poslužitelj korišten u ovoj disertaciji je [sordito.fpz.hr](http://sordito.fpz.hr) nabavljen u sklopu projekta **SORDITO**. Za provjeru korisnika koristi se token autentifikacija čiji sadržaj čine tri parametra u formatu Ključ:Vrijednost: *grant\_type: password, username: korisničko ime, password: zaporka*, kao u primjeru na Slici 5.3. Za autentifikaciju i autorizaciju korišten je *OAuth* i *OWIN*. *OAuth* je otvoreni protokol koji omogućuje sigurnu autorizaciju jednostavnom i standardnom metodom iz web, mobilnih i desktop aplikacija. *OWIN* definira standardno sučelje između .NET web poslužitelja i web aplikacija. Unutar ASP.NET programa implementirane su metode unutar klase *ApplicationOAuthProvider* koje omogućuju autentifikaciju i autorizaciju nad vlastitom bazom podataka u kojoj se između ostalog nalazi i tablica s korisničkim podacima *tblUserMaster*. Početni ASP.NET projekt sadrži implementiranu autentifikaciju i autorizaciju, no koristi vlastite tablice korisnika i uloga, što nije primjenljivo u ovom slučaju. Ako korisnik postoji u bazi podataka, rezultat POST akcije je pristupni token tipa *bearer* izdan od strane poslužitelja. *Bearer* token je jednostavni pristupni token čiji je sadržaj zapisan u obliku znakovnog niza. U ovom slučaju token je validan 599 dana, Slika 5.3.



Slika 5.3: POST akcija za dohvat tokena unutar aplikacije Postman

Najbitnije klase za rad poslužiteljskog programa prikazane su na Slici 5.4. Za rad s korisničkim podacima implementirano je ukupno 5 APIa, koje većinom reprezentiraju operacije za kreiranje, čitanje, ažuriranje i brisanje (engl. Create, Read, Update and Delete, CRUD) korisničkih podataka:

- *Token* - predefinicirana POST metoda koja dohvaća token na temelju unesenih korisničkih podataka;
- *RegisterNewUser* (POST - CREATE) - POST metoda koja kao parametre prima korisničko ime, adresu elektroničke pošte i zaporku za registraciju novog korisnika u bazi podataka. Ako u bazi podataka ne postoji korisnik sa zadanim korisničkim imenom i elektroničkom poštom, kreira se novi korisnik u tablici s korisničkim podacima. U protivnom metoda vraća grešku putem poruke i zastavice;
- *SendPassword* - GET metoda koja kao parametre prima korisničko ime i adresu elektroničke pošte. Ako u bazi podataka postoji korisnik s istim korisničkim imenom i adresom elektroničke pošte, metoda na adresu elektroničke pošte korisnika šalje poruku s naslovom *[Collecty] - password reset!* koji sadrži podatke o zaporci s porukom *"Your password for Collecty app is: zaporka! Best regards, ME"*. Za slanje poruke koristi se gmail SMTP klijent i adresa elektroničke pošte *drscnotification@gmail.com*;
- *GetUserData* (GET - READ) - GET metoda koja kao parametar prima *ID* korisnika u bazi podataka i ako korisnik postoji vraća korisničke podatke traženog korisnika;
- *EditUserData* (PUT - UPDATE) - POST metoda koja kao parametre prima *ID* korisnika, te nove korisničke podatke (korisničko ime, adresu elektroničke pošte i zaporku) te ih postavlja ako korisnik postoji u bazi podataka.

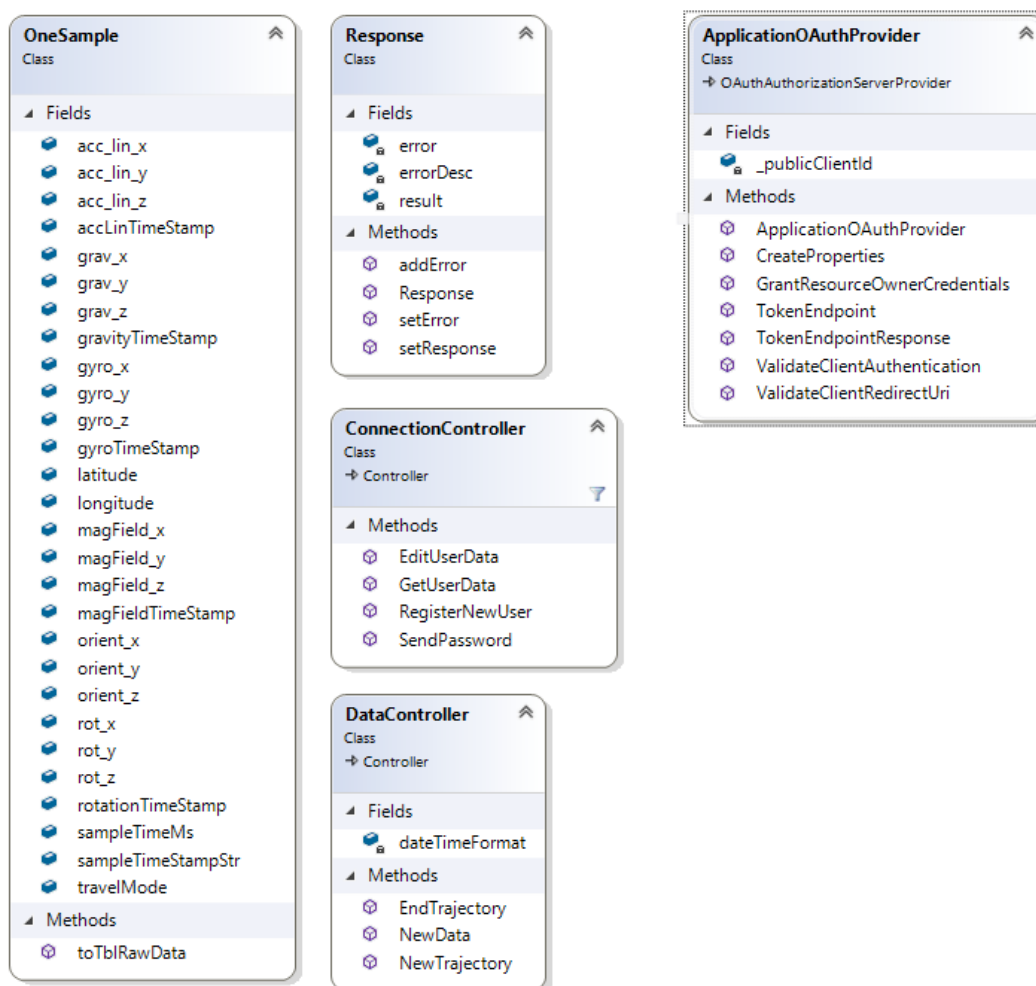
Osim korisničkih podataka u bazu podataka se spremaju podatci s osjetila mobilnog uređaja. Podatci se spremaju u dvije tablice *tblTrajectory* i *tblRawData*. Za rad s podacima s osjetila mobilnog uređaja implementirano je ukupno tri APIa:

- *NewTrajectory* - POST metoda koja kao parametar prima *ID* korisnika i vrijeme početka trajektorije, te ako korisnik postoji u bazi podataka u tablici *tblTrajectory* dodaje se nova trajektorija;
- *NewData* - POST akcija koja preko parametra prima *ID* trajektorije za koju se podatci šalju i same podatke u JSON (engl. Java Script Object Notation) formatu kao listu instanci klase *OneSample*. Primjer *OneSample* klase koja odgovara klasi *oneSample* unutar Android Studia (Slika 5.8) prikazan je na Slici 5.4. Klasa sadrži sve podatke koji se mogu poslati iz mobilne aplikacije. Ako nema grešaka u podacima, u tablicu *tblRawData* će se spremati sirovi mobilni podatci;
- *EndTrajectory* - POST metoda koja prima četiri parametra: *ID* trajektorije koja je završila, vrijeme završetka trajektorije, oznaku kraja trajektorije i podatke u JSON formatu. Ako postoji trajektorija s traženim *ID*em u bazi podataka, metoda ažurira kraj trajektorije i oznaku kraja u tablici *tblTrajectory*. Oznaka kraja može biti: 1 - trajektorija je validna, 2 - trajektorija nije validna ili 3 - korisnik je zaboravio ugasiti aplikaciju. Automatsko isključivanje aplikacije pokreće se 2 h nakon početka korištenja.

Svim metodama ne može se pristupiti ako korisnik nije autoriziran, odnosno ako u zaglavlju

## 5. Validacija metode

zahtjeva prema poslužitelju ne postoji *Ključ:Vrijednost* par *Autorizacija:bearer token*. Također, sve CRUD operacije nad tablicama baze podataka stavljene su *try-catch* blok i popraćene odgovarajućim transakcijama i izolacijskim nivoom *READ COMMITED*. Za rad s tablicama baze podataka korišten je [Entity Framework](#) koji omogućuje lakši rad s tablicama bez potrebe za izravnim pisanjem koda strukturiranim upitnim jezikom (engl. Structured Query Language, SQL).

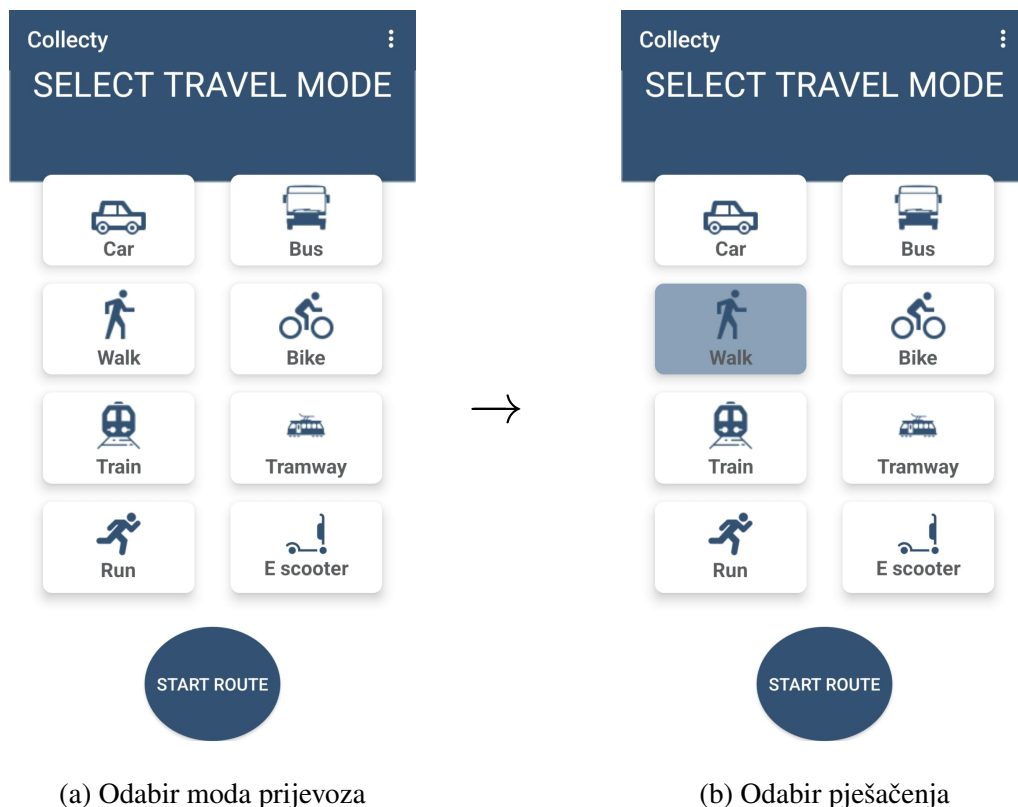


Slika 5.4: Klase u poslužiteljskom programu

U prvoj aktivnosti se osim opisanog odabire način prikupljanja podataka. Dostupna su dva načina prikupljanja podataka: mrežno i izvan mrežno prikupljanje podataka. Osnovna razlika je da se kod mrežnog prikupljanja podataka, svi podatci s osjetila mobilnog uređaja odmah šalju na poslužitelj, dok se kod izvan mrežnog prikupljanja podataka, podatci s osjetila spremaju u unutarnju memoriju mobilnog uređaja. Podatci se spremaju prema putanji "Moji dokumenti/Unutarnja memorija/Preuzimanja" u datoteku naziva Collecty\_TID\_redni broj trajektorije korisnika. Ako dostupnost i količina mobilnih podataka nije neograničena tijekom prikupljanja podataka, pouzdanije je prikupljati podatke u izvan mrežnom načinu kako bi se osigurali kontinuirani zapisi u trajektoriji.

Nakon prijave korisnik može prihvatiti pravila korištenja aplikacije ili zatvoriti aplikaciju, ako se s pravilima ne slaže, Slika 5.2b. Ističu se dvije stavke u pravilima korištenja: (i) stavka 1 u kojoj korisnik dopušta korištenje podataka za istraživanja provedena u ovoj doktorskoj disertaciji te (ii) stavka 2 i 3 kojima korisnik dopušta objavu vlastitih podataka u istraživačke svrhe u anonimiziranom formatu. Prihvatanje ovih pravila je važno kako bi podatci mogli biti korišteni u istraživačke svrhe, a obvezno je trenutnom Općom uredbom o zaštiti podataka [139].

Druga aktivnost u aplikaciji koristi se za odabir trenutnog moda prijevoza, prikaz na Slici 5.5. Korisnik može odabrati jedan od osam modova prijevoza: automobili, autobus, hodanje, bicikl, vlak, tramvaj, trčanje ili električni romobil. Odabran mod prijevoza je istaknut plavom bojom, Slika 5.5b. Nakon odabira, pritiskom na tipku *Start Route* započinje prikupljanje podataka s osjetila mobilnog uređaja pri čemu se svakom zapisu pridružuje odabrani mod prijevoza.



**Slika 5.5:** Aktivnost za odabir trenutnog moda prijevoza

Treća aktivnost odnosi se na praćenje parametara s osjetila mobilnog uređaja korisnika i služi isključivo za vizualizaciju podataka korisniku. U pozadini se podatci šalju na poslužitelj ili spremaju u unutarnju memoriju mobilnog uređaja. Na Slici 5.6 prikazan je zaslon uređaja s mrežom tipki od kojih svaka tipka predstavlja jedno osjetilo. Odabrano osjetilo je istaknuto plavom bojom, Slika 5.6b. U središnjem dijelu se na grafu iscrtavaju vrijednosti mjerenja s odabranog osjetila. Nova vrijednost na grafu prikazuju se svake sekunde. Za akcelometar, gravitaciju, žiroskop i magnetometar prikazane su vrijednosti magnitude istih. Za rotaciju su prikazane vrijednosti po svim osima te su za lokaciju prikazana geografska dužina i širina.

Osim vizualnog prikaza, u ovoj aktivnosti korisnik može promijeniti mod prijevoza (pritiskom na tipku *Change mode*) ili označiti kraj rute (pritiskom na tipku *End route*). Ako korisnik pritisne na tipku za promjenu moda prijevoza, vraća se na prethodnu aktivnost u kojoj može ponovno odabrati mod prijevoza, kao što je to prikazano na Slici 5.5. Nakon promjene moda prijevoza, korisniku se ponovno prikazuje treća aktivnost s vizualizacijom podataka koji se prikupljaju. Druga mogućnost je završetak trajektorije čijim se odabirom prelazi na sljedeću aktivnost.



Slika 5.6: Aktivnost za prikaz vrijednosti s osjetila tijekom korištenja aplikacije

U posljednjoj aktivnosti korisniku se prikazuje trajektorija na digitalnoj karti. Za prikaz trajektorije korištena je digitalna karta OSM (engl. *Open Street Map*). OSM je projekt za stvaranje besplatne geografske baze podataka svijeta koja se može uređivati i nudi dva gotova rendera, koji sadrže karte. U aplikaciji Collecty korišten je render Mapnik koji daje karte u piksel formatu. Na Slici 5.7 su prikazana dva zadnja prikaza u aplikaciji. Na prvom prikazu na Slici 5.7a podloga je OSM na kojoj je iscrtana korisnička trajektorija pri čemu se dio trajektorije iscrtava ikonoma koja odgovara modu prijevoza koji je na tom dijelu trajektorije korišten. Prikazana je trajektorija u kojoj su korištena dva moda prijevoza. Crvenim ikonama pješaka je označen dio trajektorije u kojoj je korisnik hodao, a narančasta oznaka električnog romobila označava preostali dio trajektorije. Korisnik ima dvije mogućnosti: potvrditi trajektoriju (pritiskom na tipku *Confirm route*) ili odbaciti trajektoriju (pritiskom na tipku *Reject route*). Ovisno o odabiru korisnika, svi zapisi prikazane trajektorije dobivaju oznaku 1 (potvrđena ruta) ili 2 (odbijena ruta).

Na digitalnoj karti ne iscrtavaju se sve GPS točke već se nova GPS točka iscrtava na karti ako je od zadnje GPS točke isteklo 5 s ili prošlo više od 100 m. Nakon korisničke potvrde prikazuje se poruka korisniku i tipka za gašenje aplikacije (*Exit app*), kao što je prikazano na Slici 5.7b.



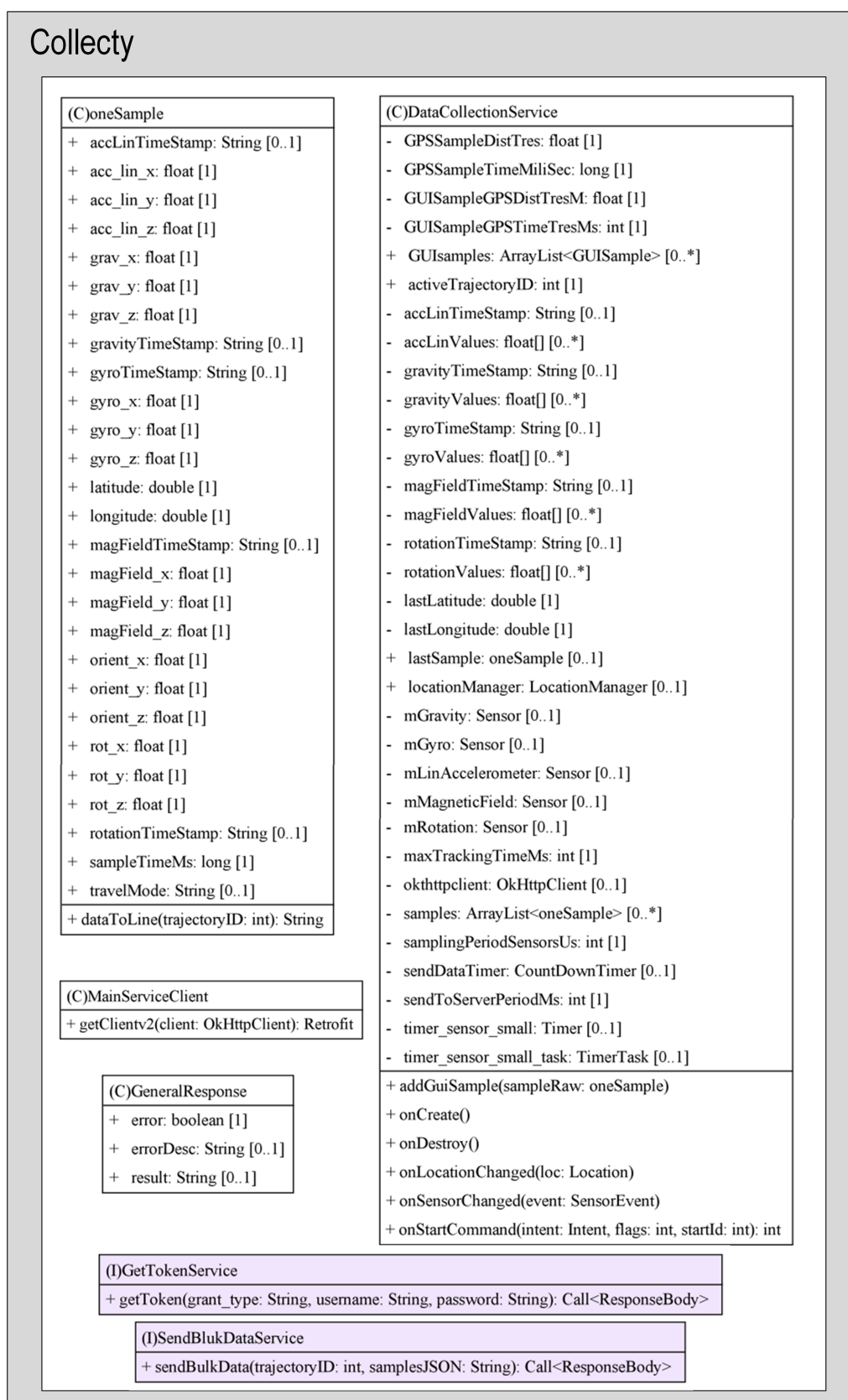
(a) Prikaz trajektorije

(b) Izlaz iz aplikacije

**Slika 5.7:** Aktivnost za potvrdu i provjeru ispravnosti prikupljenih podataka

Mobilna aplikacija može raditi i u pozadini što znači da korisnik može paralelno koristiti druge aplikacije ili imati zaključan zaslon na mobilnom uređaju. U tu svrhu je implementirana notifikacija koja upozorava korisnika da postoji aplikacija u pozadini koja je aktivna. Implementacija notifikacije je neophodna za prikupljanje GPS podataka čije prikupljanje u suprotnom prestaje kada se zaključa zaslon mobilnog uređaja.

Najvažnije klase u Android studiu za rad s aplikacijom prikazane su na Slici 5.8. Klasa *OneSample* koristi se za pohranu jednog uzorka podataka. Vidljivo je da su atributi klase oni podaci koji su navedeni da se prikupljanju s osjetila mobilnog uređaja (po jedan atribut za svaku os), mod prijevoza te geografska dužina i širina. *DataCollectionService* je glavna klasa u aplikaciji te su unutar nje pohranjene sve metode koje se koriste za uzorkovanje podataka za slanje (*onSensorChanged()*, *onLocationChanged()*) te prikaz na grafičkom sučelju aplikacije (*addGUISample()*). Za slanje podataka nakon isteka brojača vremena *sendDataTimer* koristi se metoda *SendBulkDataService()*. Klasa *MainServiceClient* koristi se formatiranje zaglavlja za komunikaciju poslužitelja i mobilne aplikacije pri čemu je korištena biblioteka *Retrofit* za povezivanje aplikacije s lokalnim poslužiteljem. *GeneralResponse* je klasa za dohvaćanje pogrešaka ako se dogode na poslužitelju tijekom spremanja podataka. Za dohvaćanje tokena s poslužitelja tijekom prijave korisnika u aplikaciju koristi se metoda *GetTokenService()*.

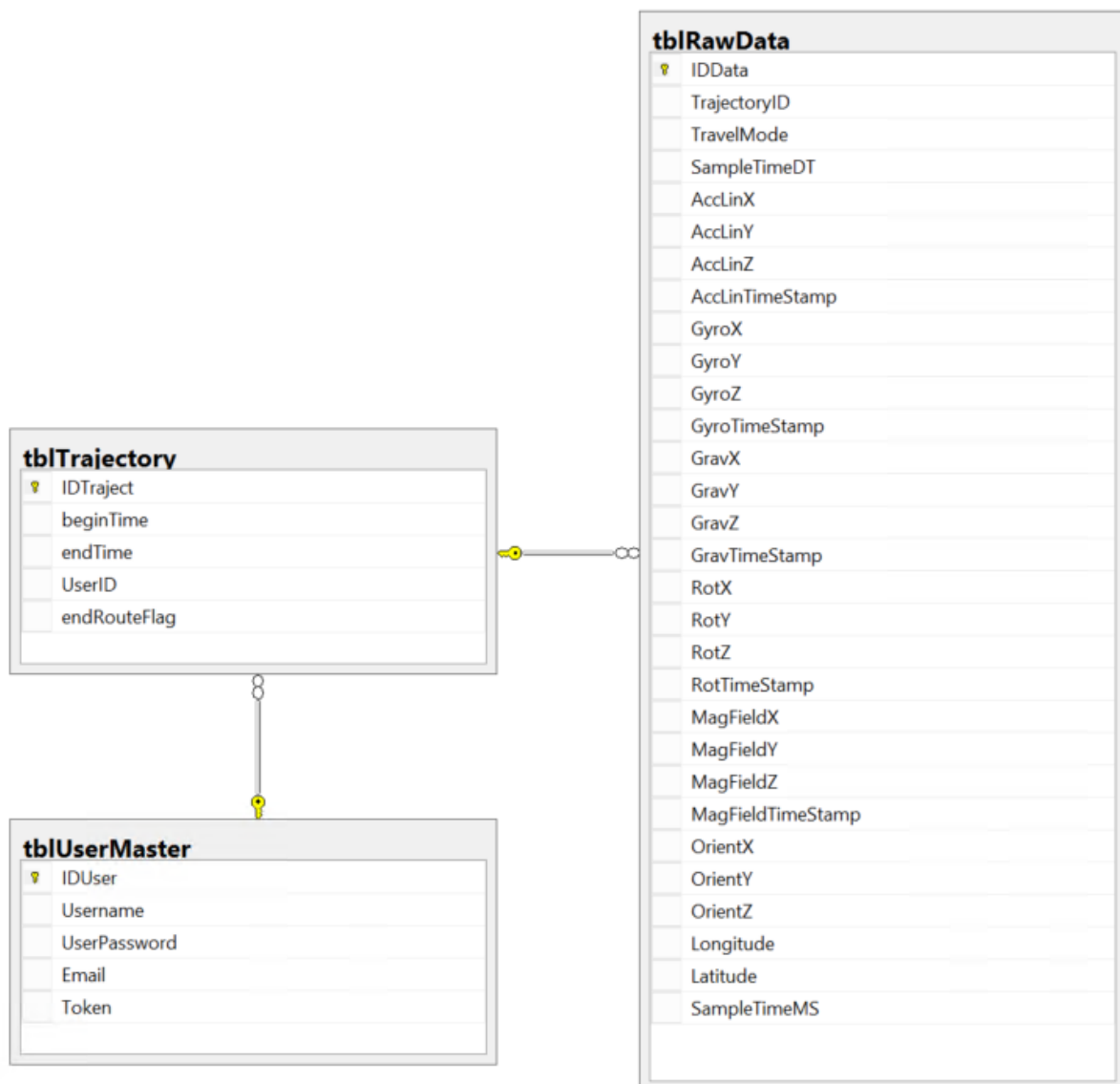
Slika 5.8: Dijagram izdvojenih klasa u mobilnoj aplikaciji *Collecty*

### 5.1.4 Pohrana podataka

Podatci koji se prikupljaju mobilnom aplikacijom pohranjuju se na dva načina ovisno o načinu rada aplikacije, pa su prisutne i dvije strukture za pohranu podataka. Za izvanmrežni način rada

podatci su strukturirani u tekstualne datoteke, a za prikupljanje podataka u mrežnom načinu rada kreirana je relacijska baza podataka na poslužitelju. Zapisivanje podataka u bazu podataka ili u tekstualnu datoteku nije jednako vremenu uzorkovanja podataka već se podatci prikupljaju tijekom perioda od 5 s te se onda svi prikupljeni podatci zapisuju u memoriju.

Baza podataka kreirana je pomoću SQL jezika u SQL Server 2014 Management Studio-u. Dijagram baze podataka u koji su podatci pohranjeni prikazan je na Slici 5.9.



Slika 5.9: Dijagram baze podataka

Baza podataka sadrži tri tablice: *tblUserMaster*, *tblTrajectory* i *tblRawData* međusobno povezane stranim i primarnim ključevima. Tablica *tblUserMaster* sadrži 5 atributa koji se koriste za pohranu korisničkih podataka o korisničkom imenu, zaporci, elektroničkoj pošti te dodijeljenom tokenu za prijavu u aplikaciju. U tablicu *tblTrajectory* pohranjuju se podatci o trajektoriji kroz vrijeme početaka i kraja rute te oznaku rute (prihvaćena ili odbijena). Također u ovoj tablici se nalazi strani ključ korisničkog IDa kojim je svaka trajektorija povezana s korisnikom



koji je trajektoriju prikupio. Podatci s osjetila mobilnog uređaja za svaku trajektoriju pohranjeni su u tablici *tblRawData* koja sadrži 30 atributa, a jedan od njih je i strani ključ ID trajektorije kojim se svaki sirovi podatak pridružuje jednoj trajektoriji. Prikupljeni podatci tijekom perioda od 5 s koji se nalaze u klasi *OneSample* formatiraju se u mobilnoj aplikaciji u JSON format te se tako strukturirani podatci šalju na poslužitelj gdje se ponovno ekstrahiraju u bazu podataka.

Tekstualna datoteka sadrži sve podatke koji su sadržani u bazi podataka samo je format zapisa podataka drugačiji. Na Slici 5.10 prikazan je isječak jedne tekstualne datoteke. Iz datoteke su izbrisani središnji podatci te su na slici prikazani s tri točke. Dakle, vidljivo je da prva linija datoteke počinje ključnom riječi *Begin* koju slijede ID trajektorije te vrijeme početka rute. Nakon toga svaka linija reprezentira jedan uzorak podataka u kojoj su sadržani svi podatci sadržani u tablici *tblRawData* te su međusobno odijeljeni znakom točka-zarez (;). Posljednja linija tekstualne datoteke započinje ključnom riječi *End* koju slijedi oznaka trajektorije te vrijeme završetka rute. U trajektoriji na Slici 5.10 oznaka trajektorije je 1 što ukazuje da je trajektorija prihvaćena.

Ako se podatci prikupljaju u obliku tekstualne datoteke, onda korisnik treba datoteku učitati na za to predviđeno mjesto na servis za pohranu podataka u oblaku.

```
Begin;45;16/02/2022 01:30:10.833
45;16/02/2022 01:30:10.982;Tramway;0,02936512;0,23280001;-0,33311462;16/02/2022 01:30:10.975;0,32184279;-
0,01291632;0,14727269;16/02/2022 01:30:10.975;-0,71402347;2,22093773;9,5251236;16/02/2022 01:30:10.975;0,
0,04740912;-0,66298282;16/02/2022 01:30:10.975;-8,90625;-26,28750038;-19,35000038;16/02/2022 01:30:10.975
0,25992194;0,07437665;15.97918434;45.80524266
45;16/02/2022 01:30:10.992;Tramway;0,01114291;0,28244781;-0,44327068;16/02/2022 01:30:10.985;0,32290804;-
0,06191845;0,112119;16/02/2022 01:30:10.985;-0,70589113;2,25860262;9,51687145;16/02/2022 01:30:10.985;0,1
0,04878975;-0,66242427;16/02/2022 01:30:10.985;-8,88750076;-25,68750191;-19,18125153;16/02/2022 01:30:10.
0,27200115;0,07642654;15.97918434;45.80524266
45;16/02/2022 01:30:11.002;Tramway;-0,05144572;0,34141374;-0,41559124;16/02/2022 01:30:10.995;0,31758171;
0,14181322;0,08761793;16/02/2022 01:30:10.995;-0,6916756;2,29638839;9,50886822;16/02/2022 01:30:10.995;0,
0,05036496;-0,6620338;16/02/2022 01:30:10.995;-9,07500076;-25,44375038;-19,08750153;16/02/2022 01:30:10.9
0,28113657;0,08150239;15.97918434;45.80524266
.
.
.
45;16/02/2022 01:40:24.909;Walk;-0,36462665;0,04125476;-3,69645548;16/02/2022 01:40:24.909;-0,1181111;0,0
0,25859275;16/02/2022 01:40:24.909;-0,3071402;2,11756349;9,57036972;16/02/2022 01:40:24.909;0,09383851;-0
0,63356668;16/02/2022 01:40:24.906;-22,10625076;-5,0625;-40,21875;16/02/2022 01:40:24.906;1,27466249;-
0,35041323;0,11409733;15.99615525;45.80748511
45;16/02/2022 01:40:24.921;Walk;-0,15430978;0,23873305;-3,84793425;16/02/2022 01:40:24.919;-0,16072164;0,
0,24580958;16/02/2022 01:40:24.919;-0,32145664;2,10401511;9,57288742;16/02/2022 01:40:24.919;0,09368285;-
0,63430178;16/02/2022 01:40:24.916;-22,06875038;-5,19375038;-40,27500153;16/02/2022 01:40:24.916;1,205893
0,38787767;0,08314992;15.99615525;45.80748511
End;1;16/02/2022 01:40:28.196
```

Slika 5.10: Tekstualna datoteka za pohranu podataka

## **5.2 Analiza istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama**

Istraživački skup podataka prikupljen u ovom istraživanju sadrži 243 sata prikupljenih podataka za 8 različitih modova prijevoza. Skup podataka prikupljalo je 15 korisnika. Veličina skupa podataka je 37 GB, od čega je 26 GB sadržano u tekstualnim datotekama (prikupljano izvan mrežnim načinom), a 11 GB je pohranjeno u relacijskoj bazi podataka (prikupljano mrežnim načinom). Jedan zapis zauzima 156 B memorijskog prostora.

Korisnici koji su sudjelovali u prikupljanju podataka dobili su uputu o načinu prikupljanja podataka u trenutku instalacije aplikacije na vlastite mobilne uređaje. Proces prikupljanja podataka generira nekoliko zahtjeva za korisnike:

1. Uključivanje mobilne aplikacije tijekom kretanja prometnom mrežom.
2. Održavanje mobilne aplikacije uključenom tijekom prikupljanja podataka te postupanje tijekom označavanja podataka multimodalnim karakteristikama prema dobivenim uputama.
3. Provjera ispravnosti prikupljenih geografskih podataka na digitalnoj karti prema dobivenim uputama.
4. Isključivanje mobilne aplikacije.
5. Postavljanje tekstualne datoteke na servis za pohranu podataka u oblaku te uklanjanje tekstualne datoteke s mobilnog uređaja.

Upute korisnicima za korištenje mobilne aplikacije zasnivaju se na tri osnovne informacije: trenutak promjene moda prijevoza, kategorizacija stajanja te pregled rute prije potvrde. Naime, tijekom promjene moda prijevoza, korisnici bi mogli različito odabrati vremenski trenutak u kojem će promijeniti mod prijevoza. Tako bi dio korisnika promijenio mod prijevoza kod, primjerice, dolaska vlaka na stajalište, dok bi dio korisnika promjenu učinilo tek kada bi započeli vožnju u vlaku. Kako bi se osiguralo da korisnici jednako označavaju promjenu moda prijevoza, svima je dana uputa da promjene mod prijevoza u trenutku početka korištenja novog moda prijevoza. Drugi dio upute odnosi se na stajanje, za koje su korisnici upućeni da stajanje pripada hodajućem segmentu. Za provjeru ispravnosti trajektorije korisnici trebaju obratiti pozornost na više segmenata u trajektoriji. Prva provjera odnosi se na točnost geografskih lokacija prikazane trajektorije. Druga provjera se odnosi na provjeru korištenih modova prijevoza, tj. odgovaraju li ikone moda prijevoza na digitalnoj karti korištenom modu prijevoza. Zadnja provjera odnosi se na geografske lokacije promjene moda prijevoza gdje korisnici trebaju utvrditi jesu li na tim lokacijama zbilja promijenili mod prijevoza.

Posljednji korak u procesu prikupljanja podataka (točka 5) odnosi se isključivo na izvan

mrežno prikupljanje podataka. Svim korisnicima je ustupljena poveznica na koju mogu postaviti tekstualnu datoteku te putanja na mobilnom uređaju na koju aplikacija zapisuje podatke.

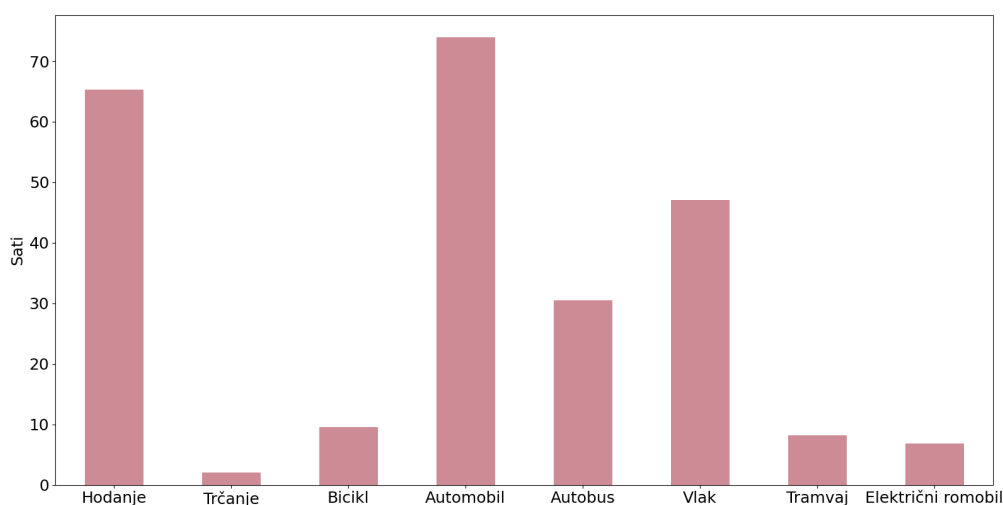
U Tablici 5.1 prikazani su svi atributi koji čine zaglavlje tekstualne datoteke s pripadajućim mjernim jedinicama. Sve oznake vremena zapisane su u 12 satnom sustavu u formatu dan/mjesec/godina sat:minute:sekunde.milisekunde. Jedan zapis sadrži više podataka o vremenu, pri čemu se prvi vremenski zapis (redak 2 Tablica 5.1) odnosi na vremenski trenutak uzorkovanja dok se ostale vremenske oznake odnose na vremenski trenutak kada se dogodila zadnja promjena vrijednosti osjetila prije uzorkovanja. Ta dva podataka ne moraju uvijek biti jednaka jer je okidajući događaj za prikupljanje podataka istek vremena mjerača, a ne promjena osjetila kako je to opisano u potpotpoglavlju 5.1.1. Navedene vrijednosti odgovaraju opisima osjetila iz potpotpoglavlja 5.1.1.

**Tablica 5.1:** Sadržaj zaglavlja tekstualne datoteke u kojoj su zapisani podatci

Redni broj	Naziv atributa	Mjerna jedinica
1	IDTrajektorije	cijeli broj
2	Vrijeme	Datum i vrijeme
3	Mod prijevoza	Walk=Hodanje, Run=Trčanje, Bike=Bicikl, Car=Automobil, Bus=Autobus, Tramway=Tramvaj, Train=Vlak, E scooter= Električni romobil
4	Akceleracija x os	$m/s^2$
5	Akceleracija y os	$m/s^2$
6	Akceleracija z os	$m/s^2$
7	Vrijeme zadnje promjene akceleracije u uzorku	Datum i vrijeme
8	Stopa rotacije po x osi	$rad/s$
9	Stopa rotacije po y osi	$rad/s$
10	Stopa rotacije po z osi	$rad/s$
11	Vrijeme zadnje promjene žiroskopa u uzorku	Datum i vrijeme
12	Jakost magnetskog polja duž x osi	$\mu T$
13	Jakost magnetskog polja duž y osi	$\mu T$
14	Jakost magnetskog polja duž z osi	$\mu T$
15	Vrijeme zadnje promjene žiroskopa u uzorku	Datum i vrijeme
16	Gravitacijsko ubrzanje po x osi	$m/s^2$
17	Gravitacijsko ubrzanje po y osi	$m/s^2$
18	Gravitacijsko ubrzanje po z osi	$m/s^2$
19	Vrijeme zadnje promjene gravitacijskog ubrzanja u uzorku	Datum i vrijeme
20	Rotacija oko x osi	Stupnjevi
21	Rotacija oko y osi	Stupnjevi
22	Rotacija oko z osi	Stupnjevi
23	Vrijeme zadnje promjene rotacije u uzorku	Datum i vrijeme
24	Geografska širina	Decimalni stupnjevi
25	Geografska dužina	Decimalni stupnjevi

Podatci zapisani u relacijskoj bazi podataka imaju strukturu kao što je prikazano na Slici 5.9 dok su podatci pohranjeni u tekstualne datoteke strukturirani u mape prema sljedećem pravilu: *Korijenska mapa/ID korisnika/Collecty\_TID\_redni broj.txt*. Na ovaj način označena je pripadnost svake datoteke određenom korisniku te je trajektoriji dodijeljen redni broj. Neke od tekstualnih datoteka označene su oznakom 2 koja signalizira da je trajektorija tijekom provjere ispravnosti rute odbijena.

Istraživački skup podataka prikupljan je tijekom razdoblja od 5 mjeseci, od studenog 2021. do ožujka 2022. godine. Ukupna količina podataka prikupljena za različite modove prijevoza nije jednaka te je raspodjela količine podataka prikazana na Slici 5.11 pri čemu su na  $x$  osi prikazani razmatrani modovi prijevoza, a na  $y$  osi sati prikupljenih podataka. Količina podataka za modove prijevoza automobil, hodanje, vlak i autobus je veća u odnosu na količinu podataka prikupljenu za trčanje, električni romobil, tramvaj i bicikl. Ukupno je prikupljeno oko 243 sata podataka za sve modove prijevoza što je usporedivo s referentnim skupovima podataka: SHL skup podataka za koji je objavljeno 236 sati korisničkih trajektorija te TMD skup podataka koji sadrži 31 sat korisničkih trajektorija.



**Slika 5.11:** Raspodjela količine podataka prema modovima prijevoza

Preciznija raspodjela podataka prikazana je u Tablici 5.2 gdje je naznačena količina podataka po korisniku za svaki mod prijevoza u satima. Vidljivo je da je količina podataka različita za svakog korisnika. Tako se u skupu podataka nalazi veća količina podataka koje su prikupili korisnici s ID-evima 29, 23 i 18 dok su znatno manje količine podataka korisnika s ID-evima 28, 26 i 20. Kod većine modova prijevoza zastupljeni su podatci raznolikih korisnika dok su u podacima za modove prijevoza trčanje, bicikl i električni romobil prisutni podatci dva korisnika. Manja zastupljenost navedenih modova prijevoza u korisničkim trajektorijama može biti uzrokovana zimskim mjesecima tijekom kojih je prikupljan skup podataka.

**Tablica 5.2:** Raspodjela podataka prema modovima prijevoza po korisniku izražena u minutama

ID korisnika	Hodanje	Trčanje	Bicikl	Automobil	Autobus	Vlak	Tramvaj	Električni romobil	Ukupno
17	8,49	0,0	0,0	6,77	0,0	2,84	0,0	0,2	18,3
23	10,89	0,06	0,18	0,55	22,73	0,35	5,06	0,0	39,82
24	0,25	0,0	0,0	1,96	0,0	0,0	0,0	0,0	2,21
25	0,71	0,0	0,0	9,11	0,0	0,0	0,75	0,0	10,57
29	29,94	0,0	0,0	4,52	0,92	39,33	0,0	0,0	74,71
37	0,0	0,0	0,0	30,31	0,0	0,0	0,0	0,0	30,31
39	1,28	0,0	0,0	1,48	0,82	3,78	0,0	0,0	7,36
40	6,26	0,02	0,0	0,27	5,66	0,0	2,27	0,0	14,48
26	0,03	0,0	0,0	0,12	0,0	0,0	0,0	0,0	0,15
18	5,32	0,0	9,39	15,64	0,0	0,0	0,0	6,61	36,96
28	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,4
27	1,87	0,0	0,0	0,83	0,0	0,78	0,0	0,0	3,48
31	0,16	0,0	0,0	0,28	0,39	0,0	0,0	0,0	0,83
35	0,09	0,0	0,0	2,07	0,0	0,0	0,0	0,0	2,16
20	0,02	0,0	0,0	0,0	0,0	0,0	0,14	0,0	0,16
<b>Ukupno</b>	65,71	2,08	9,57	73,91	30,52	47,08	8,22	6,81	243,9

U Tablici 5.3 prikazana je raspodjela količine podataka po korisnicima za različite modove prijevoza izražena u prijeđenoj udaljenosti, mjerna jedinica su kilometri. Za izračun udaljenosti između dvije točke u trajektoriji korištena je *Haversine* formula. *Haversine* formula često se koristi za izračun geografske udaljenosti na Zemlji pri čemu se računa najkraća udaljenost između dvije točke na površini sfere. U Izrazu 5.1 prikazan je izraz za izračun udaljenosti pri čemu je  $r$  polumjer Zemlje,  $d$  udaljenost između dvije točke,  $x_1$  i  $x_2$  geografska širina promatranih točaka te  $y_1$  i  $y_2$  geografska dužina promatranih točaka.

$$\text{haversin}\left(\frac{d}{r}\right) = \text{haversin}(x_2 - x_1) + \cos(x_1)\cos(x_2)\text{haversin}(y_2 - y_1) \quad (5.1)$$

Na raspodjelu količine podataka iskazanu u kilometrima značajan je utjecaj pripadanja moda prijevoza motoriziranim ili ne motoriziranim modovima prijevoza. U pravilu, motoriziranim modovima prijevoza se korisnici kreću većim brzinama, pa prijeđu veću udaljenost u manjoj jedinici vremena.

Jednaka razlika primjećuje se kod prijeđenih udaljenosti po korisnicima. Tako korisnici kod kojih su zastupljeniji ne motorizirani modovi prijevoza imaju manju ukupnu prijeđenu udaljenost u usporedbi s korisnicima motoriziranih modova prijevoza. Primjerice ako se uspoređuju korisnici s IDem 18 i 37, vremenski oba korisnika imaju približno jednaku količinu sati podataka, štoviše korisnik 18 ima veći broj sati podataka. Međutim, količina podataka korisnika 37 izražena u prijeđenoj udaljenosti je za otprilike 60% veća u odnosu na prijeđenu udaljenost

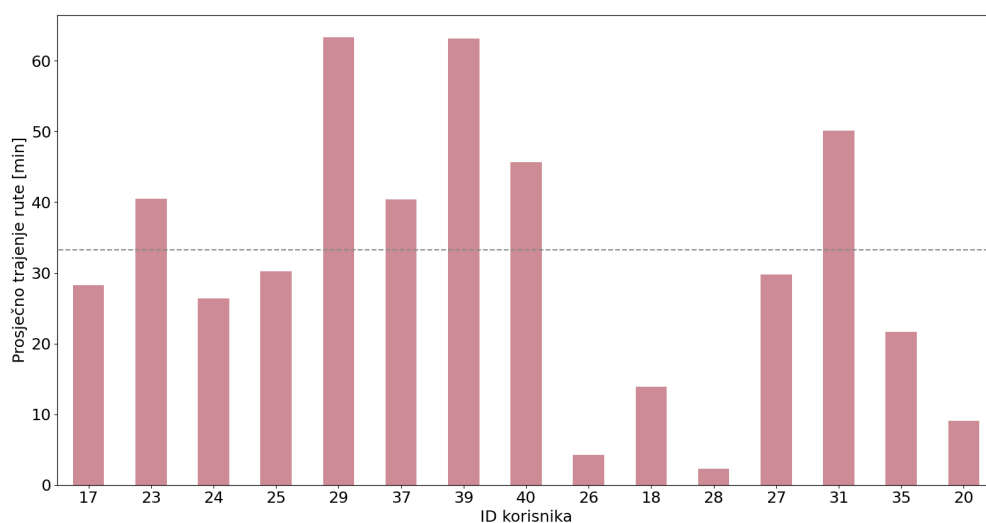
5.2. Analiza istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama

korisnika 18.

**Tablica 5.3:** Raspodjela podataka prema modovima prijevoza po korisniku izražena u kilometrima

ID korisnika	Hodanje	Trčanje	Bicikl	Automobil	Autobus	Vlak	Tramvaj	Električni romobil	Ukupno
17	0,56	0,0	0,0	3,66	0,0	2,4	0,01	0,03	6,66
23	0,78	0,0	0,03	0,36	9,93	0,26	1,3	0,0	12,66
24	0,08	0,0	0,0	0,64	0,0	0,0	0,0	0,0	0,72
25	0,04	0,0	0,0	3,64	0,0	0,0	0,15	0,0	3,83
29	2,48	0,0	0,0	3,63	0,61	34,63	0,0	0,02	41,37
37	0,0	0,0	0,0	29,95	0,0	0,0	0,0	0,0	29,95
39	0,07	0,0	0,0	0,83	0,43	2,77	0,0	0,0	4,1
40	0,27	0,0	0,0	0,13	2,08	0,0	0,64	0,0	3,12
26	0,0	0,0	0,0	0,04	0,0	0,0	0,0	0,0	0,04
18	0,26	0,0	2,14	14,97	0,0	0,0	0,0	1,32	18,69
28	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2
27	0,06	0,0	0,0	0,43	0,0	0,5	0,0	0,0	0,99
31	0,0	0,0	0,0	0,17	0,17	0,0	0,0	0,0	0,34
35	0,12	0,0	0,0	0,5	0,0	0,0	0,0	0,0	0,62
20	0,1	0,0	0,0	0,0	0,0	0,0	0,03	0,0	0,13
<b>Ukupno</b>	<b>283,86</b>	<b>30,4</b>	<b>130,59</b>	<b>3537,37</b>	<b>792,96</b>	<b>2433,61</b>	<b>127,16</b>	<b>82,09</b>	<b>7418,04</b>

Trajanje pojedine trajektorije nije jednako te postoje odstupanja trajanja trajektorije između korisnika. Na Slici 5.12 prikazano je prosječno trajanje rute za svakog korisnika. Na x osi prikazani su ID-ovi korisnika, a na y osi prikazano je prosječno trajanje trajektorija u minutama. Siva isprekidana linija označava prosječno trajanje trajektorija svih korisnika.

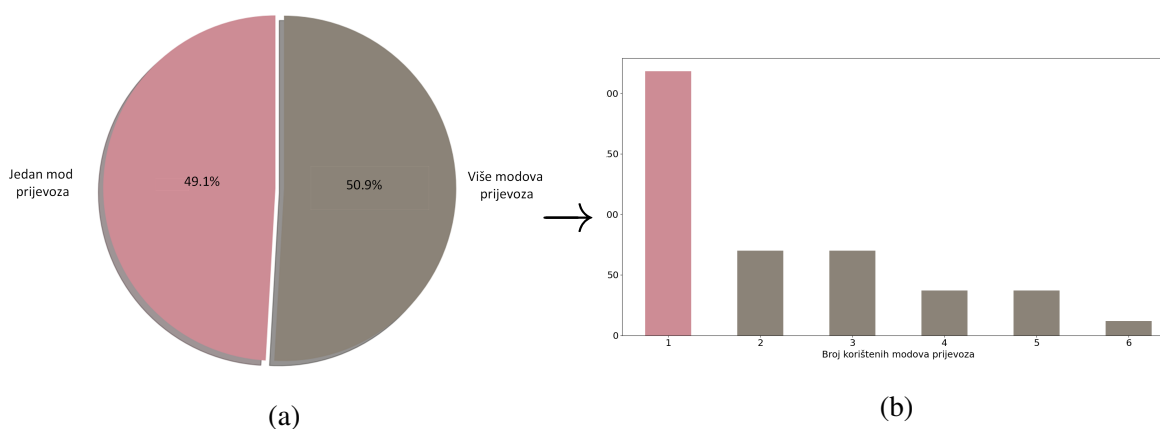


**Slika 5.12:** Prosječno trajanje trajektorija svakog korisnika

Prosječno trajanje trajektorija za korisnike 23, 29, 37, 39, 40 i 31 je veće od prosječnog

trajanja svih trajektorija preostalih korisnika. Vremenski kratke rute zastupljene su u podacima korisnika 26 i 28.

U ovom istraživanju razvijena je uostalom metoda za segmentaciju trajektorija s obzirom na promjenu moda prijevoza, pa je važno da skup podataka sadrži dio trajektorija u kojima je korišteno više uzastopnih modova prijevoza. Na Slici 5.13 je prikazana raspodjela trajektorija prema broju korištenih modova prijevoza. Omjer trajektorija s jednim ili više modova prijevoza približno je jednak, Slika 5.13a. Raspodjela broja trajektorija u kojima je korišteno od 1-6 modova prijevoza je prikazana na Slici 5.13b. Kod višemodalnih trajektorija najveći je broj trajektorija koje sadrže 2 i 3 moda prijevoza, dok je broj trajektorija s većim brojem modova prijevoza manji.

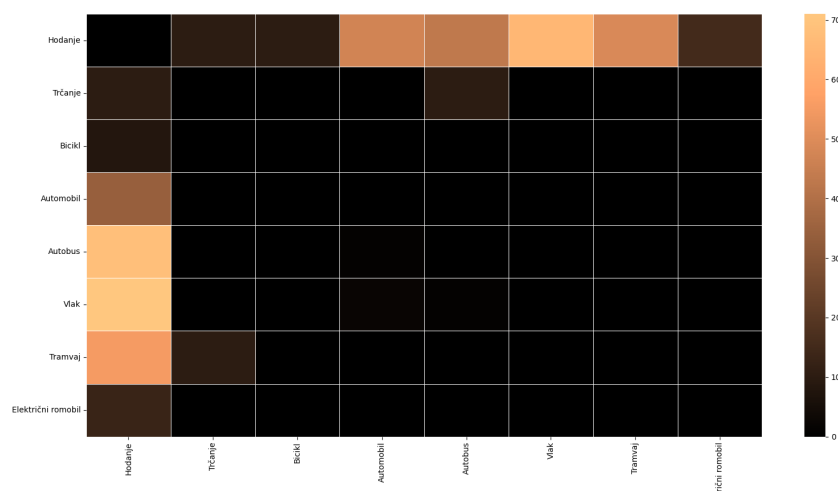


**Slika 5.13:** Raspodjela trajektorije prema broju korištenih modova prijevoza u trajektoriji

U višemodalnim trajektorijama mogu se promatrati prijelazi između uzastopnih modova prijevoza. Na Slici 5.14 prikazana je upravo matrica prijelaza između pojedinih modova prijevoza. U redcima i stupcima matrice navedeni su modovi prijevoza, a jedan element matrice prikazuje broj prijelaza između određenog para modova prijevoza. Boja elementa matrice od crne do žute određuje frekvenciju prijelaza između modova prijevoza. Vidljivo je da su najzastupljeniji prijelazi između hodanja i svih ostalih modova prijevoza. Osim hodanja, prisutan je manji broj prijelaza između trčanja i ostalih modova prijevoza. Prijelazi između preostalih modova prijevoza nisu zastupljeni u skupu podataka.

Za segmentaciju trajektorije u nekim istraživanjima se postavlja pravilo da za promjenu moda prijevoza mora biti detektirano pješaćenje, inače promjena moda prijevoza nije izvediva [37, 85]. Ovaj prikaz potvrđuje pretpostavke korištene u tim istraživanjima jer je zapravo kod svih prijelaza između modova prijelaza postojao segment u kojem je korisnik hodao ili trčao.

Skup podataka prikupljen je na prometnoj mreži u Republici Hrvatskoj pri čemu je najviše podataka prikupljeno u Gradu Zagrebu i Zagrebačkoj županiji. Manja količina podataka prikupljena je u Sisačko-moslavačkoj, Krapinsko-zagorskoj, Varaždinskoj i Primorsko-goranskoj



Slika 5.14: Matrica zastupljenosti prijelaza između modova prijevoza

županiji. Na Slikama 5.15- 5.17 je na digitalnoj karti prikazana gustoća prikupljenih podataka za odabrane modove prijevoza. Kao podloga je korištena digitalna karta OSM. Intenzitet boje od plave prema crvenoj pokazuje gustoću podataka na odabranom području. Slika 5.15 prikazuje prostornu gustoću svih prikupljenih podataka te podataka za mod prijevoza automobil. Ako se promatraju svi podatci, Slika 5.15a, vidljivo je da se najviše podataka odnosi na područje grada Zagreba i državne ceste D30 i D36 prema Sisku. Gustoća podataka za mod prijevoza automobil prikazana na Slici 5.15b ima vrlo sličnu strukturu prostornoj gustoći svih podataka. Na nekim cestovnim segmentima očituju se točkasta crvena područja koja se mogu interpretirati kao zagušenja, radovi na cestovnoj infrastrukturi ili raskrižja. Dakle, već i na skupu podataka ove veličine mogu se detektirati žarišne točke na prometnicama koje mogu ukazivati na zahtjeve za intervencijom na prometnoj infrastrukturi.

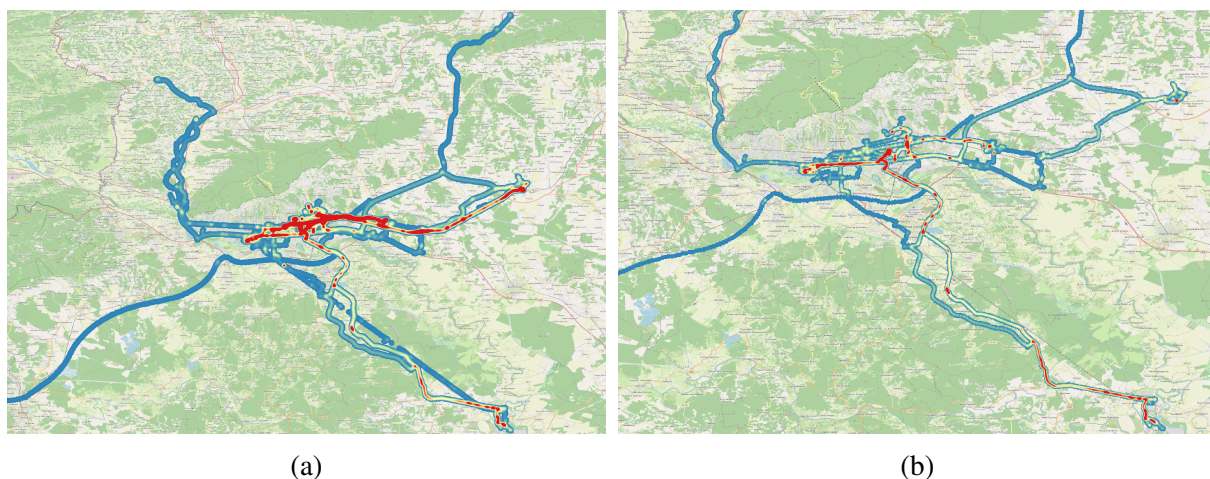
Na Slici 5.16a prikazana je gustoća podataka koji se odnose na vlak. Iz ovakvog prikaza mogu se prostorno odrediti javne željezničke linije. Najveća koncentracija podataka odnosi se na Zagrebačku prigradsku željeznicu. Gustoća podataka za električni romobil prikazana je na Slici 5.16b. Infrastruktura koja je korištena za ovaj mod prijevoza uglavnom se odnosi na kvartovske prometnice i pokazuje sličnost s infrastrukturom koja se koristi za hodanje i bicikliranje. Također, unutar trajektorija se ne pokazuju područja koja bi mogla ukazivati na žarišta, pa se može zaključiti da električni romobil pripada modovima prijevoza koji su manje podložni zagušenjima.

Gustoća podataka za mod prijevoza autobus prikazana je na Slici 5.17 pri čemu su na Slici 5.17a prikazani svi podatci koji se odnose na mod prijevoza autobus, a na Slici 5.17b samo podatci koji se odnose na prikazanu rutu. Jednako kao kod prikaza za vlak, gustoća podataka prikazana prostorno jasno ocrta javne autobusne linije. Najveća količina podataka prikupljena je za autobusne linije u užem centru grada Zagreba. Ako se prikaže gustoća po-

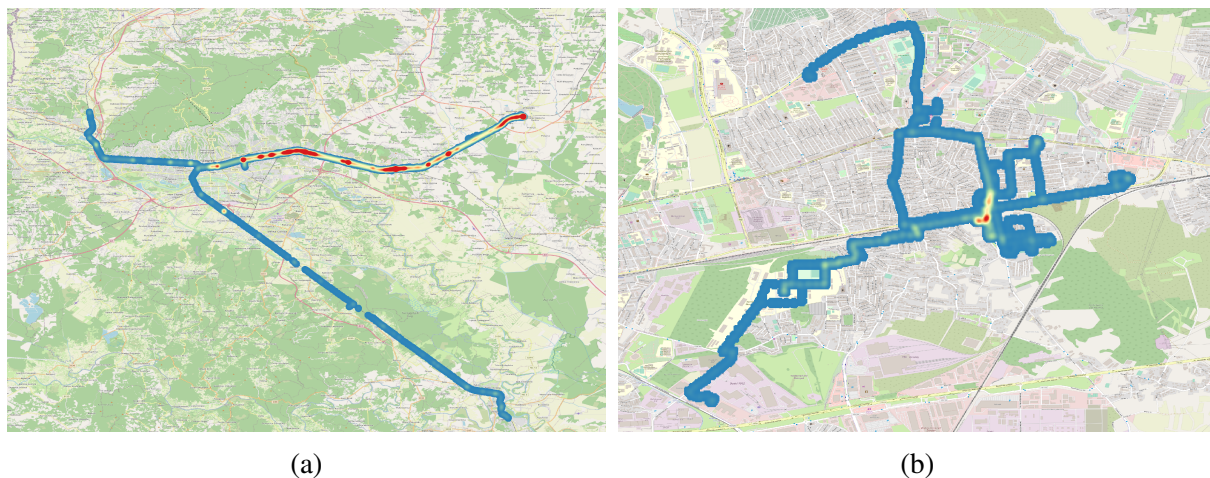


## 5. Validacija metode

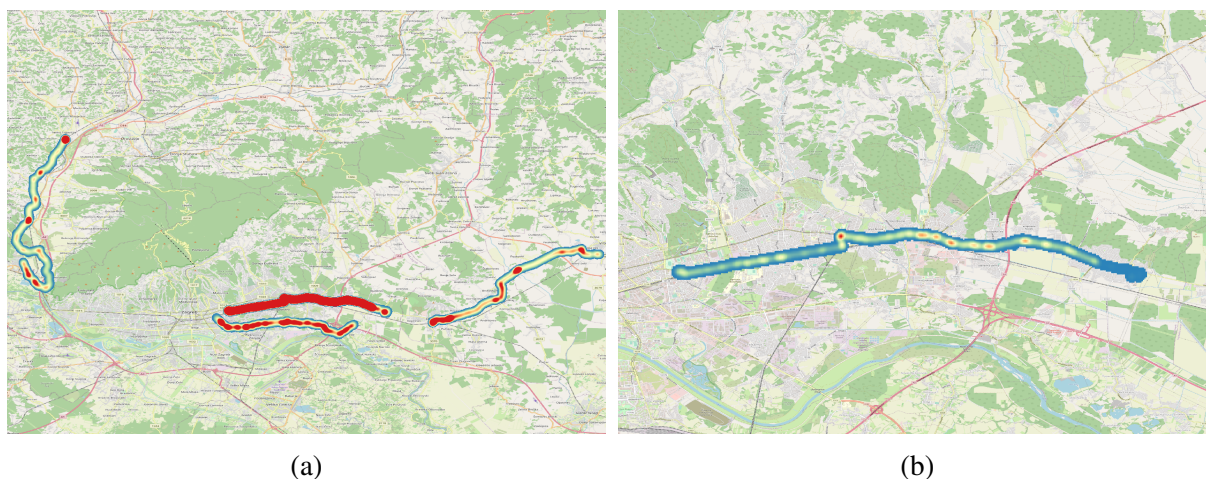
dataka samo za jednu autobusnu liniju, kao na Slici 5.17b , ističu se žarišta (crvena boja) koja mogu signalizirati autobusna stajališta u nekoj autobusnoj liniji. Analizom gustoće podataka na nekom dijelu prometne mreže može se detektirati korisnička preferencija prema nekom modu prijevoza na nekom geografskom području. U tom slučaju podatci se trebaju ciljano prikupljati na ograničenom geografskom području.



**Slika 5.15:** Karta gustoće prikupljenih podataka za sve podatke i za mod prijevoza automobil



**Slika 5.16:** Karta gustoće prikupljenih podataka za mod prijevoza vlak i električni romobil



**Slika 5.17:** Karta gustoće prikupljenih podataka za sve podatke moda prijevoza autobus i za jednu autobusnu liniju

### 5.3 Rezultati validacije

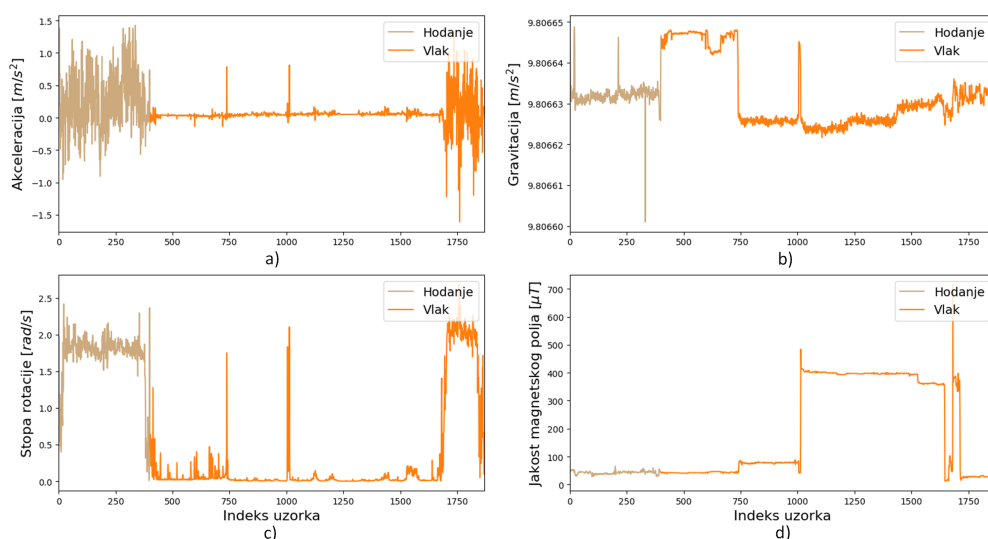
Validacija metode opisane u poglavlju 4 provedena je koristeći prikupljeni istraživački skup podataka. Podatci su obrađeni prema opisanoj proceduri u prethodnim poglavljima te su korišteni samo oni podatci s osjetila i značajke koje su korišteni u SHL skupu podataka. U nastavku su prikazani rezultati metode za segmentaciju trajektorije te rezultati višerazinske metode za segmentaciju i klasifikaciju trajektorije. Također, tijekom validacije višerazinske metode iz prikupljenog istraživačkog skupa podataka su isključeni svi zapisi koji se odnose na mod prijevoza tramvaj i električni romobil jer podatci tih modova prijevoza nisu bili dostupni tijekom procesa učenja metode.

Testiranje je provedeno za tri komponente koje definiraju skup podataka: vrijeme kada je skup podataka prikupljen, lokaciju nošenja uređaja te korisnika koji je podatke prikupljao. Validacijom metode na skupu podataka koji je prikupljen na različitom geografskom području od testnog skupa podataka može se kazati da se uz tri navedene komponente ispituje prostorna ovisnost metode, tj. koliko geografsko područje na kojem su podatci prikupljeni utječe na točnost metode. Stoga, obzirom da je istraživački skup podataka prikupljen na različitom geografskom području u odnosu na SHL skup podataka može se zaključiti da se validacijom ostvaruje proširenje testnog problema na ispitivanje prostorne neovisnosti metode.

Na pouzdanost validacije može utjecati zastupljenost pojedinog moda prijevoza. Naime, ukoliko je mala zastupljenost moda prijevoza koji uzrokuje najveću pogrešku klasifikatora onda točnost klasifikatora u procesu validacije može biti prividno usporediva s točnošću u procesu testiranja. Međutim, u ovom primjeru je najmanja zastupljenost modova prijevoza trčanje i bicikl, čija je točnost predviđanja među najvećima, pa na pouzdanost procesa validacije u ovom primjeru nema značajnog utjecaja veličina uzorka za svaki mod prijevoza.

### 5.3.1 Rezultati validacije metode za segmentaciju i klasifikaciju trajektorije

Neobrađeni podatci u prikupljenom skupu podataka pokazuju sličnosti s podacima u SHL skupu podataka. Slika 5.18 prikazuje neobrađene podatke s osjetila na y-osi u vremenskim uzorcima čiji indeks je prikazan na x-osi. Prikazani su podatci jedne trajektorije iz skupa podataka za validaciju u kojem su korištena dva povezana moda prijevoza: hodaње i vlak s jednom MTP (promjena boje krivulje na slici). Na slici su prikazane vrijednosti za sva četiri osjetila mobilnog uređaja. Tijekom promjene moda prijevoza vidljiva je značajna promjena vrijednosti akceleracije, gravitacije i stope rotacije (Slike 5.18a,b,c), dok jakost magnetskog polja (Slika 5.18d) ne pokazuje značajne promjene vrijednosti tijekom promjene moda prijevoza. Za vrijeme trajanja trajektorije vidljiv je različit raspon izmjerenih vrijednosti sa svih osjetila tijekom vožnje u vlaku i hodaња.

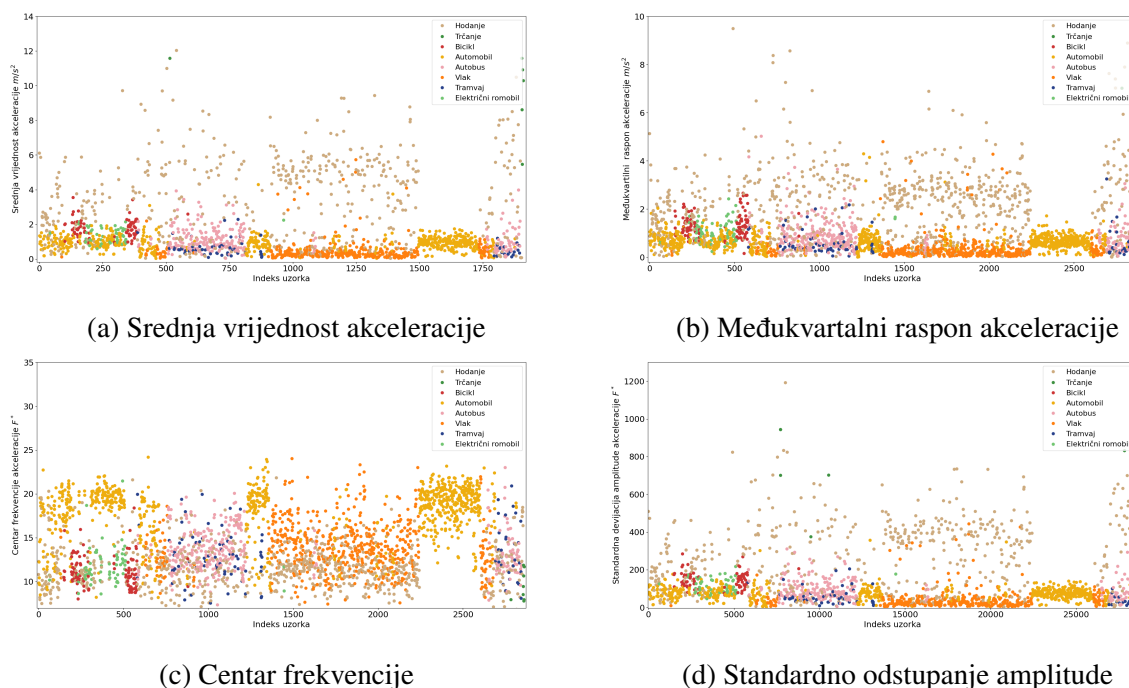


**Slika 5.18:** Podatci jedne trajektorije s osjetila: a) akcelerometar b) gravitacija c) stopa rotacije d) jakost magnetskog polja

Rezultat obrade podataka nad prikupljenim skupom podataka su značajke za učenje metode. Odabrane značajke za magnitudu akceleracije: srednja vrijednost i međukvartilni raspon u vremenskoj domeni te centar frekvencije i standardno odstupanje u frekvencijskoj domeni prikazane su na Slici 5.19. Iste značajke prikazane su na Slici 4.14 za SHL skup podataka. Različiti modovi prijevoza označeni su različitim bojama te je prikazan samo dio podataka radi jasnije vizualizacije podataka.

Na Slikama 5.19a, 5.19b, 5.19c i 5.19d između ostalog nalaze se podatci za modove prijevoza tramvaj i električni romobil. Mod prijevoza tramvaj ima slične vrijednosti u prikazanim značajkama modu prijevoza vlak. Ista sličnost je prisutna u SHL skupu podataka modu prije-

voza podzemna željeznica, pa se može zaključiti da modovi prijevoza koji pripadaju skupini tračničkih vozila pokazuju slično ponašanje tijekom kretanja prometnom mrežom. Električni romobil je mod prijevoza koji nije analiziran u literaturi u kontekstu razlikovanja modova prijevoza. Vidljivo je na prikazanim značajkama da električni romobil pokazuje najveću sličnost s modom prijevoza bicikl, a prema tome najviše se razlikuje od ostalih motoriziranih modova prijevoza.



**Slika 5.19:** Vrijednosti odabranih značajki akceleracije u vremenskim prozorima za prikupljeni skup podataka

U Tablici 5.4 prikazani su rezultati validacije TSM-a na prikupljenom skupu podataka u obliku matrice zabune te ukupne točnosti i odziva. Od sveukupno 472 prijelaza između modova prijevoza detektirano je 416 prijelaza što čini 88,51%. Ukupna postignuta točnost metode tijekom testiranja je 90,94%, a tijekom validacije 94,36%, što je razlika od 3,4%. Razlika u odzivu tijekom testiranja i validacije manja je od 4%. Ovime se potvrđuje prostorna neovisnost metode za segmentaciju trajektorije.

**Tablica 5.4:** Rezultati validacije TSM-a na prikupljenom skupu podataka

	Predviđena promjena	Predviđeno bez promjene
Stvarna promjena	416	56
Stvarno bez promjene	1.526	26.059
Točnost		94,36%
Odziv		88,51%

U Tablici 5.5 su prikazani rezultati validacije višerazinske metode za segmentaciju i klasifikaciju trajektorije na prikupljenom skupu podataka. Za validaciju je odabrano 11 scenarija prisutnih u testnom problemu. Zadnji scenarij se ne može testirati jer u prikupljenom skupu podataka ne postoji mod prijevoza podzemna željeznica koji nije u upotrebi na području Republike Hrvatske. Rezultat validacije nalazi se u posljednja dva stupca tablice, a u prethodnim stupcima su navedene prosječne vrijednosti točnosti i odziva modela izrađenih za testiranje korisničke, vremenske i lokacijske neovisnosti.

Kako validacijski skup podataka nisu prikupljali isti korisnici niti je označavana pozicija nošenja uređaja, za validaciju metode korišteni su modeli izrađeni za testiranje vremenske neovisnosti.

Rezultati validacije pokazuju usporedive vrijednosti točnosti metode onima postignutima u fazi testiranja metode. Štoviše u nekim scenarijima se pokazuje da je metoda manje prostorno ovisna nego korisnički ili lokacijski. Primjerice za scenarij 4 rezultat validacije metode je za 1% veći od korisničke neovisnosti, a oko 2% veći u odnosu na testiranje lokacijske neovisnost. Stoga se može zaključiti da je metoda prostorno neovisna te da promjena prometne mreže na kojoj se podatci prikupljaju utječe približno jednako kao i promjena korisnika koji podatke prikuplja ili lokacije na kojoj se uređaj nosi.

**Tablica 5.5:** Rezultati višerazinske metode za segmentaciju i klasifikaciju trajektorije na prikupljenom skupu podataka

	Korisnička neovisnost		Vremenska neovisnost		Lokacijska neovisnost		Rezultat validacije	
	F1	T	F1	T	F1	T	F1	T
<b>Scenarij 1</b>	93,2	94,4	99,8	99,8	94,3	95,1	81,4	89,0
<b>Scenarij 2</b>	86,9	89,2	95,6	98,0	87,6	89,8	83,2	89,1
<b>Scenarij 3</b>	79,0	85,0	85,6	93,0	80,2	85,2	79,3	85,2
<b>Scenarij 4</b>	75,4	84,3	86,1	92,6	77,4	83,7	82,2	85,3
<b>Scenarij 5</b>	76,9	79,3	83,6	93,4	79,0	80,2	70,5	76,9
<b>Scenarij 6</b>	74,3	78,5	86,4	93,7	76,6	78,5	76,6	77,8
<b>Scenarij 7</b>	64,7	71,2	77,4	85,3	69,2	72,0	72,3	72,2
<b>Scenarij 8</b>	65,1	71,0	80,8	85,5	69,0	70,5	71,2	71,7
<b>Scenarij 9</b>	65,9	74,4	80,6	87,3	70,5	74,4	60,5	68,9
<b>Scenarij 10</b>	65,8	73,7	83,2	87,1	70,0	73,1	66,4	69,2
<b>Scenarij 11</b>	59,3	63,8	65,5	76,5	63,2	64,9	64,2	66,3

### 5.3.2 Klasifikacija modova prijevoza u prikupljenom skupu podatka

Kako električni romobil nije zastupljen u do sada razvijenim metodama za klasifikaciju modova prijevoza, izrađen je model za klasifikaciju modova prijevoza na temelju podatka prikupljenih u

sklopu ovog istraživanja s ciljem identifikacije modova prijevoza s kojima se električni romobil zamjenjuje. Dakle, cjelokupni skup podataka podijeljen je na skup podataka za učenje i skup podataka za testiranje tako da podatci jedne trajektorije nisu razdvojnivi. SHL skup podataka nije korišten kod ovog modela jer električni romobil nije zastupljen u tom skupu podataka. Za predviđanje je korištena višerazinska metoda za segmentaciju i klasifikaciju trajektorije opisana u poglavlju 4. Također primijenjena je jednaka metodologija za obradu podataka i optimizaciju parametara RF modela. Za izradu metode korištene su vrijednosti za osjetila akcelerometra, žiroskopa i magnetometra.

U Tablici 5.6 je prikazana matrica zabune te ukupna točnost metode. Vidljivo je da je najbolji rezultat postignut za električni romobil i automobil, dok je najmanja točnost klasifikacije za mod prijevoza tramvaj. Jednako kao i kod testiranja na SHL skupu podataka, najčešće zamjene događaju se kod motoriziranih modova prijevoza, automobil i autobus te između tračničkih vozila. Za električni romobil (kratica E. romobil u Tablici 5.6) postignuta je visoka točnost klasifikacije, a najčešće je zamijenjen s hodanjem, biciklom i automobilom što potvrđuje zaključke o sličnosti primijećene tijekom analize podataka.

**Tablica 5.6:** Rezultati modela naučenog na *Collecty* podacima za klasifikaciju modova prijevoza

	Hodanje	Trčanje	Bicikl	E. romobil	Automobil	Autobus	Tramvaj	Vlak
Hodanje	3.734	0	3	2	199	8	7	108
Trčanje	0	6	0	0	0	0	0	0
Bicikl	59	0	466	1	8	0	0	0
E. romobil	22	0	0	324	2	0	0	0
Automobil	55	0	0	0	4.056	83	3	55
Autobus	80	0	0	0	59	1.488	1	51
Tramvaj	87	0	0	0	36	46	271	42
Vlak	173	0	0	0	186	5	6	2.478
F1 vrijednost				77,69%				
Točnost				90,29%				



---

# Zaključak

---

Problem klasifikacije modova prijevoza pripada skupini problema HAR područja. Uključivanjem mobilnih uređaja i podataka s njihovih osjetila u praćenje kretanja korisnika ovaj problem dobiva nova obilježja, veće mogućnosti, ali i veće zahtjeve nad rješenjima problema. Korisničke trajektorije mogu sadržavati jedan ili više uzastopnih modova prijevoza. Ako trajektorija sadrži više uzastopnih modova prijevoza razmatrani problem može se podijeliti na dva osnovna zadatka: segmentacija trajektorije i klasifikacija segmenata trajektorije s obzirom na mod prijevoza. Zadatak segmentacije je podijeliti trajektoriju na segmente u kojima je prisutan jedan mod prijevoza, a zatim se metodama klasifikacije detektira korišteni mod prijevoza na segmentu. U procesu klasifikacije razmatraju se uglavnom odabrani motorizirani te odabrani nemotorizirani modovi prijevoza.

Tijekom klasifikacije modova prijevoza vrlo često se ne provodi odvojen proces za segmentaciju trajektorije. Stoga ne postoji veći broj istraživanja koji razvija metode zasnovane na podatkovnom učenju kao rješenju problema segmentacije trajektorije. Razvijene metode uglavnom su zasnovane na nizu iskustvenih pravila za čiji je razvoj potrebno poznavanje specifičnosti sustava. U ovom istraživanju razvijena je metoda za segmentaciju trajektorije koja se zasniva na TSM matricama kako bi se prepoznao obrazac ponašanja korisnika tijekom prelaska iz jednog u drugi mod prijevoza.

Razvijene metode za klasifikaciju trajektorije zasnivaju se na metodama strojnog učenja pri čemu su najbolji rezultati postignuti metodama koje sadrže stabla odluke. Višerazinske metode uglavnom se odnose na odvojeno odlučivanje o skupinama modova prijevoza, a zatim o modovima prijevoza unutar skupine. Primjerice prva odluka se odnosi na pripadanje moda prijevoza skupini motoriziranih ili nemotoriziranih modova prijevoza, a sljedeća odluka definira mod prijevoza unutar skupine. Klasifikacija se provodi na povijesnim podacima ili u stvarnom



vremenu. Ako se provodi u stvarnom vremenu, onda je najčešća pretpostavka u istraživanjima da se mod prijevoza ne mijenja unutar jednog vremenskog prozora. U ovom istraživanju razvijena je višerazinska metoda koja istovremeno segmentira i klasificira trajektoriju s obzirom na mod prijevoza. Metoda je prilagođena za rad u stvarnom vremenu te je testirana na testnom skupu podataka.

U HAR području za klasifikaciju modova prijevoza dostupna su tek tri skupa podataka, od čega se ne odnose svi na podatke s mobilnih uređaja. Stoga je za prikaz prostorne neovisnosti metode prikupljen vlastiti istraživački skup podataka na području Republike Hrvatske. Skup podataka je prikupljan mobilnom aplikacijom Collecty razvijenom za potrebe ovog istraživanja. U skupu podataka su zastupljeni podatci o osam modova prijevoza, uostalom i za mod prijevoza električni romobil za koji do sada ne postoje dostupni skupovi podataka niti rezultati klasifikacije. Višerazinska metoda za segmentaciju i klasifikaciju trajektorije testirana je na prikupljenom skupu podataka.

U ovoj doktorskoj disertaciji cilj istraživanja je razviti metodu za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu te prikupiti skup podataka s osjetila mobilnih uređaja o kretanju korisnika prometnom mrežom i na tim podatcima validirati metodu. Za ostvarenje prve komponente razvijena je višerazinska metoda za segmentaciju i klasifikaciju trajektorije, a za validaciju metode prikupljen je istraživački skup podataka mobilnom aplikacijom Collecty.

### 6.1 Hipoteze i ostvareni znanstveni doprinosi istraživanja

Prva hipoteza istraživanja je da se višerazinska metoda za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza može primijeniti za rješavanje problema klasifikacije šest modova prijevoza u stvarnom vremenu. Ova hipoteza istraživanja potvrđena je razvojem višerazinske metode za segmentaciju i klasifikaciju trajektorije koja je prilagođena za klasifikaciju u stvarnom vremenu. Metoda je testirana na testnom problemu koji se sastoji od 12 scenarija, pri čemu svaki scenarij sadrži različit broj modova prijevoza, od najmanje dva do najviše sedam modova prijevoza. Ostvareni rezultati točnosti i odziva razvijene metode pokazuju usporedive rezultate s rezultatima u literaturi.

Druga hipoteza istraživanja je da rezultati testiranja i validacije predložene metode na javno objavljenom i prikupljenom skupu podataka pokazuju usporedivu točnost s rezultatima dostupnim u literaturi. Ova hipoteza je potvrđena testiranjem i validacijom razvijene metode na testnom SHL skupu podataka te vlastitom istraživačkom skupu podataka označenim multimodalnim karakteristikama. Ostatak poglavlja daje prikaz osnovnih znanstvenih doprinosa kao i zaključaka o svakom doprinosu pojedinačno.

Tijekom razvoja metode postavljene su pretpostavke o ponašanju korisnika tijekom kretanja

prometnom mrežom te pretpostavke o načinu pristizanja podataka. Ove pretpostavke ujedno predstavljaju i ograničenja za primjenu metode na neki drugi skup podataka. Osnovna pretpostavka opisane metodologije je da podatci stižu na poslužitelj, gdje se obrađuju, u vremenskom slijedu i odnose se na jednog korisnika ili uređaj. Dakle, podatci jednog korisnika šalju se na poslužitelj te se obrađuju svakih 30 sek. Druga pretpostavka odnosi se na vremensko trajanje promjene moda prijevoza. Naime, u ovom istraživanju je postavljena pretpostavka da korisnik ne može učiniti dvije promjene moda prijevoza za 30 sek. Dakle, završiti korištenje prvog moda prijevoza, koristiti drugi mod prijevoza te započeti korištenje trećeg moda prijevoza. Takav tip promjene moda prijevoza smatra se u ovom istraživanju neizvedivim.

### **6.1.1 Identifikacija značajki iz skupa podataka o kretanju korisnika prometnom mrežom u vremenskoj i frekvencijskoj domeni u svrhu segmentacije trajektorije s obzirom na mod prijevoza**

U svrhu segmentacije trajektorije korišteni su podatci s četiri osjetila mobilnog uređaja: akcelerometar, magnetometar, gravitacija i žiroskop. Svako osjetilo sadrži podatke mjerenja za  $x$ ,  $y$  i  $z$  os. Kako bi se izračunale orijentacijski neovisne značajke, skupu podataka pridružena je magnituda mjerenja za svako osjetilo. Za provedbu segmentacije u stvarnom vremenu podatci su podijeljeni u vremenske prozore veličine 30 s. Nad svakim vremenskim prozorom i svakim osjetilom izračunato je 27 različitih statističkih mjera od čega se njih 20 odnosi na vremensku domenu, a 7 na frekvencijsku domenu. Za pretvorbu vremenske u frekvencijsku domenu korišten je FFT algoritam za DFT. Za odabir relevantnih značajki provedeno je sortiranje značajki prema ANOVA F vrijednosti te su iz skupa podataka uklonjene značajke koje uzrokuju lažno pozitivna predviđanja.

Na temelju rezultata koji su postignuti metodom za segmentaciju trajektorije koja koristi identificiran skup značajki može se zaključiti da izračunate i odabrane značajke mogu opisati uzorak ponašanja korisnika tijekom promjene moda prijevoza.

### **6.1.2 Razvoj višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza u stvarnom vremenu**

Razvijena višerazinska metoda za segmentaciju i klasifikaciju trajektorije koristi TSM za segmentaciju trajektorije te RF za klasifikaciju trajektorije. TSM su izrađene korištenjem tri različite metode diskretizacije podataka: metoda jednake duljine, metoda jednake frekvencije te metoda  $k$  sredine. Nakon testiranja TSM-a, najbolji rezultat je postignut matricama izrađenim metodom jednake duljine. Za klasifikaciju modova prijevoza korištena je metoda RF s hiperparametrima optimiziranim algoritmom TPE. Značajke koje se koriste za klasifikaciju mo-

dova prijevoza dodatno sadrže značajke temeljene na različitim širinama frekvencijskog pojasa i centru frekvencije. U ovom slučaju, značajke su odabrane algoritmom RFE u kombinaciji s uzajamnom informacijom za svaku značajku i svaki par modova prijevoza. Obje metode prilagođene su za obradu podataka u stvarnom vremenu, pri čemu je korištena veličina vremenskog prozora od 30 s. Višerazinska metoda predstavlja način komunikacije između ove dvije metode u procesu donošenja odluke. Dodatno, u proces odlučivanja su uključena dva pravila koja se zasnivaju na pravilima slijednih zapisa.

Za testiranje metode korišten je SHL skup podataka za koji je dostupna struktura problema. Naime, problem je strukturiran za ispitivanje triju komponenti metode: korisnička neovisnost, neovisnost lokacije nošenja uređaja te neovisnost vremena prikupljanja podataka. Svaka komponenta ispituje se kroz 12 različitih scenarija od kojih svaki sadrži različite modove prijevoza. Tako je izrađeno 396 različitih modela višerazinske metode te su rezultati postignute točnosti i odziva uspoređeni s rezultatima u literaturi.

### **6.1.3 Izrada istraživačkog skupa podataka mjerenja s osjetila mobilnog uređaja označenog multimodalnim karakteristikama prikladnim za prepoznavanje aktivnosti i moda prijevoza korisnika**

Istraživački skup podataka prikupljen je mobilnim uređajima na koje je instalirana aplikacija *Collecty*. Mobilna aplikacija *Collecty* razvijena je u sklopu ovog istraživanja, a služi za prikupljanje mjerenja s osjetila mobilnog uređaja te zapisivanje tih podataka u bazu podataka ili tekstualnu datoteku. Svaki zapis označen je modom prijevoza koji je korisnik odabrao kao trenutni mod prijevoza. U prikupljanju podataka je sudjelovalo 15 korisnika, a prikupljeno je oko 243 sata podataka za 8 različitih modova prijevoza. Izrađena je struktura istraživačkog skupa podataka pri čemu su podatci pohranjeni na dva načina: u tekstualne datoteke s pripadajućim atributima za svaku trajektoriju zasebno ili u relacijskoj bazi podataka.

Dodatno, u istraživačkom skupu podataka nalazi se dio podataka koji se odnosi na mod prijevoza električni romobil, čije ponašanje u odnosu na ostale modove prijevoza nije do sada interpretirano u literaturi. Za analizu navedenog moda prijevoza prikazano je nekoliko značajki akceleracije te je identificirana sličnost s modom prijevoza bicikl. Također, na prikazanoj prostornoj analizi vidljivo je da se za električni romobil koristi uglavnom ista prometna infrastruktura kao za nemotorizirane modove prijevoza.

## **6.2 Buduća istraživanja**

Tijekom izrade doktorske disertacije zapaženo je nekoliko otvorenih pitanja u području klasifikacije modova prijevoza. U ovom poglavlju istaknute su neke smjernice za buduća istraživanja.

Prvi mogući smjer budućeg istraživanja odnosi se na implementaciju metode za segmentaciju i klasifikaciju trajektorije u mobilni uređaj. Na ovaj način omogućuje se stvarno vremenska validacija metode u kojoj može sudjelovati i korisnik, primjerice potvrdom predviđanja. Osim toga, mali je broj istraživanja koji proučava utjecaj aplikacija ovog tipa na mobilni uređaj u vidu potrošnje baterije, pa bi se implementacijom metode predviđanja te istovremenim prikupljanjem označenih podataka dobio potpuni uvid u utjecaj mobilne aplikacije na mobilni uređaj i ponašanje korisnika.

Drugi mogući smjer budućeg istraživanja je implementacija metode za klasifikaciju modova prijevoza na mobilnim uređajima koja ima mogućnost učenja tijekom prikupljanja podataka. Često se u literaturi postižu značajna poboljšanja predviđanja kada se omogući ovaj tip učenja. Na primjeru modova prijevoza može se omogućiti korisniku da u aplikaciji odabere mod prijevoza, ako je korisnik izvršio odabir, zapis se koristi za učenje, a u suprotnom metoda vrši predviđanje. Na ovaj način omogućuje se metodi da dopuni svoje znanje, tj. pravila, novim podacima. Osim toga, ovime se može neutralizirati ovisnost metode o korisničkom ponašanju tijekom prikupljanja podataka.

Treći mogući smjer istraživanja odnosi se na primjenu većeg skupa podataka za učenje. Većim skupovima podataka tijekom učenja modela postiže se bolja generalizacija metode zbog poznavanja većeg broja različitih korisnika, lokacija, perioda, prostora i slično. Također, izradom novih istraživačkih skupova podataka omogućuje se proširenje do sada dostupnih skupova podataka novim skupovima podataka, koji potencijalno mogu rezultirati boljim rezultatima predviđanja.



# Bibliografija

- [1] Butler, L., Yigitcanlar, T., Paz, A., “Smart urban mobility innovations: A comprehensive review and evaluation”, *IEEE Access*, Vol. 8, 2020, pp. 196 034-196 049.
- [2] Treiber, M., Kesting, A., “Trajectory and floating-car data”, in *Traffic flow dynamics*. Springer, 2013, pp. 7–12.
- [3] Thomas, T., Puello, L. L. P., Geurs, K., “Intrapersonal mode choice variation: Evidence from a four-week smartphone-based travel survey in the Netherlands”, *Journal of Transport Geography*, 2018.
- [4] Incel, O. D., Kose, M., Ersoy, C., “A review and taxonomy of activity recognition on mobile phones”, *BioNanoScience*, Vol. 3, No. 2, 2013, pp. 145–171.
- [5] Xiao, Z., Wang, Y., Fu, K., Wu, F., “Identifying different transportation modes from trajectory data using tree-based ensemble classifiers”, *ISPRS International Journal of Geo-Information*, Vol. 6, 02 2017, pp. 57.
- [6] Prelipcean, A. C., Gidófalvi, G., Susilo, Y. O., “Meili: A travel diary collection, annotation and automation system”, *Computers, Environment and Urban Systems*, Vol. 70, 2018, pp. 24 - 34, available at: <http://www.sciencedirect.com/science/article/pii/S0198971517305240>
- [7] van Dijk, J., “Identifying activity-travel points from GPS-data with multiple moving windows”, *Computers Environment and Urban Systems*, 02 2018.
- [8] Wang, B., Gao, L., Juan, Z., “Travel mode detection using GPS data and socioeconomic attributes based on a random forest classifier”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 5, May 2018, pp. 1547-1558.
- [9] Biljecki, F., Ledoux, H., van Oosterom, P., “Transportation mode-based segmentation and classification of movement trajectories”, *International Journal of Geographical Information Science*, Vol. 27, No. 2, 2013, pp. 385-407, available at: <https://doi.org/10.1080/13658816.2012.692791>

- [10] Pereira, F., Carrion, C., Zhao, F., Cottrill, C. D., Zegras, C., Ben-Akiva, M., “The future mobility survey: Overview and preliminary evaluation”, in Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 9, 2013.
- [11] Gonzalez, P. A., Weinstein, J. S., Barbeau, S. J., Labrador, M. A., Winters, P. L., Georggi, N. L., Perez, R., “Automating mode detection for travel behaviour analysis by using global positioning systems enabled mobile phones and neural networks”, IET Intelligent Transport Systems, Vol. 4, No. 1, March 2010, pp. 37-49.
- [12] Wang, H., Liu, G., Duan, J., Zhang, L., “Detecting transportation modes using deep neural network”, IEICE Transactions on Information and Systems, Vol. E100.D, 05 2017, pp. 1132-1135.
- [13] Dabiri, S., Heaslip, K., “Inferring transportation modes from GPS trajectories using a convolutional neural network”, Transportation Research Part C: Emerging Technologies, Vol. 86, 2018, pp. 360-371, available at: <http://www.sciencedirect.com/science/article/pii/S0968090X17303509>
- [14] Li, L., Zhu, J., Zhang, H., Tan, H., Du, B., Ran, B., “Coupled application of generative adversarial networks and conventional neural networks for travel mode detection using GPS data”, Transportation Research Part A: Policy and Practice, Vol. 136, 2020, pp. 282-292, available at: <https://www.sciencedirect.com/science/article/pii/S0965856420305607>
- [15] Shafique, M. A., Hato, E., “Use of acceleration data for transportation mode prediction”, Transportation, Vol. 42, No. 1, 2015, pp. 163–188.
- [16] Yu, M.-C., Yu, T., Wang, S.-C., Lin, C.-J., Chang, E. Y., “Big data small footprint: The design of a low-power classifier for detecting transportation modes”, Proc. VLDB Endow., Vol. 7, No. 13, Aug. 2014, pp. 1429–1440, available at: <http://dx.doi.org/10.14778/2733004.2733015>
- [17] Feng, T., Timmermans, H. J., “Transportation mode recognition using GPS and accelerometer data”, Transportation Research Part C: Emerging Technologies, Vol. 37, 2013, pp. 118 - 130, available at: <http://www.sciencedirect.com/science/article/pii/S0968090X13002039>
- [18] Dabiri, S., Lu, C., Heaslip, K., Reddy, C. K., “Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data”, IEEE Transactions on Knowledge and Data Engineering, 2019, pp. 1-1.

- [19] James, J., “Semi-supervised deep ensemble learning for travel mode identification”, *Transportation Research Part C: Emerging Technologies*, Vol. 112, 2020, pp. 120–135.
- [20] Markos, C., James, J., “Unsupervised deep learning for GPS-based transportation mode identification”, in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.
- [21] Zheng, Y., Xie, X., Ma, W. Y., “Geolife: A collaborative social networking service among user, location and trajectory”, *IEEE Data(base) Engineering Bulletin*, June 2010, available at: <https://www.microsoft.com/en-us/research/publication/geolife-a-collaborative-social-networking-service-among-user-location-and-trajectory/>
- [22] Gjoreski, H., Ciliberto, M., Wang, L., Ordonez Morales, F. J., Mekki, S., Valentin, S., Roggen, D., “The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics With Mobile Devices”, *IEEE Access*, Vol. 6, 2018, pp. 42 592-42 604.
- [23] Carpineti, C., Lomonaco, V., Bedogni, L., Felice, M., Bononi, L., “Custom dual transportation mode detection by smartphone devices exploiting sensor diversity”, in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. Los Alamitos, CA, USA: IEEE Computer Society, mar 2018, pp. 367-372, available at: <https://doi.ieeecomputersociety.org/10.1109/PERCOMW.2018.8480119>
- [24] Wolf, J., “Using GPS data loggers to replace travel diaries in the collection of travel data”, Phd thesis, 05 2000.
- [25] Schöenfelder, S., Axhausen, K., Antille, N., Bierlaire, M., “Exploring the potentials of automatically collected GPS data for travel behaviour analysis - a Swedish data source”, *GI-Technologien für Verkehr und Logistik*, Vol. 13, 01 2002, pp. 155–179, available at: <http://infoscience.epfl.ch/record/117085>
- [26] Axhausen, K. W., Schöenfelder, S., Wolf, J., Oliveira, M., Samaga, U., “80 weeks of GPS-traces: approaches to enriching the trip information. submitted to the 83rd transportation research board meeting”, 2003.
- [27] Draijer, G., Kalfs, N., Perdok, J., “Global positioning system as data collection method for travel research”, *Transportation Research Record*, Vol. 1719, 01 2000, pp. 147-153.
- [28] Bao, L., Intille, S. S., “Activity recognition from user-annotated acceleration data”, in *Pervasive Computing*, Ferscha, A., Mattern, F., (ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–17.



- [29] Chung, E.-H., Shalaby, A., “A trip reconstruction tool for GPS-based personal travel surveys”, *Transportation Planning and Technology*, Vol. 28, No. 5, 2005, pp. 381-401.
- [30] Consolvo, S., W. McDonald, D., Toscos, T., Y. Chen, M., Froehlich, J., L. Harrison, B., V. Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., Landay, J., “Activity sensing in the wild: A field trial of ubifit garden”, in *Conference on Human Factors in Computing Systems - Proceedings*, 01 2008, pp. 1797-1806.
- [31] Brezmes, T., Gorricho, J.-L., Cotrina, J., “Activity recognition from accelerometer data on a mobile phone”, in *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, Omatu, S., Rocha, M. P., Bravo, J., Fernández, F., Corchado, E., Bustillo, A., Corchado, J. M., (ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 796–799.
- [32] Bieber, G., Voskamp, J., Urban, B., “Activity recognition for everyday life on mobile phones”, in *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, Stephanidis, C., (ed.). Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 289–296.
- [33] Bolbol, A., Cheng, T., Tsapakis, I., Haworth, J., “Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification”, *Computers, Environment and Urban Systems*, Vol. 36, No. 6, 2012, pp. 526 - 537, special Issue: *Advances in Geocomputation*, available at: <http://www.sciencedirect.com/science/article/pii/S0198971512000543>
- [34] Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M., “Using mobile phones to determine transportation modes”, *ACM Trans. Sen. Netw.*, Vol. 6, No. 2, Mar. 2010, pp. 13:1–13:27, available at: <http://doi.acm.org/10.1145/1689239.1689243>
- [35] Shen, L., Stopher, P. R., “Review of GPS travel survey and GPS data-processing methods”, *Transport Reviews*, Vol. 34, No. 3, 2014, pp. 316-334, available at: <https://doi.org/10.1080/01441647.2014.903530>
- [36] Nick, T., Coersmeier, E., Geldmacher, J., Goetze, J., “Classifying means of transportation using mobile sensor data”, in *The 2010 International Joint Conference on Neural Networks (IJCNN)*, July 2010, pp. 1-6.
- [37] Zheng, Y., Chen, Y., Li, Q., Xie, X., Ma, W.-Y., “Understanding transportation modes based on GPS data for web applications”, *ACM Trans. Web*, Vol. 4, No. 1, 2010, pp. 1:1–1:36, available at: <http://doi.acm.org/10.1145/1658373.1658374>

- [38] Endo, Y., Toda, H., Nishida, K., Kawanobe, A., “Deep feature extraction from trajectories for transportation mode estimation”, in *Advances in Knowledge Discovery and Data Mining*, Bailey, J., Khan, L., Washio, T., Dobbie, G., Huang, J. Z., Wang, R., (ed.). Cham: Springer International Publishing, 2016, pp. 54–66.
- [39] Xiao, G., Juan, Z., Zhang, C., “Travel mode detection based on gps track data and bayesian networks”, *Computers, Environment and Urban Systems*, Vol. 54, 2015, pp. 14-22, available at: <https://www.sciencedirect.com/science/article/pii/S0198971515000587>
- [40] Stenneth, L., Wolfson, O., Yu, P. S., Xu, B., “Transportation mode detection using mobile phones and GIS information”, in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '11. New York, NY, USA: ACM, 2011, pp. 54–63, available at: <http://doi.acm.org/10.1145/2093973.2093982>
- [41] Semanjski, I., Gautama, S., Ahas, R., Witlox, F., “Spatial context mining approach for transport mode recognition from mobile sensed big data”, *Computers, Environment and Urban Systems*, Vol. 66, 2017, pp. 38-52, available at: <https://www.sciencedirect.com/science/article/pii/S0198971516304367>
- [42] Gong, H., Chen, C., Bialostozky, E., Lawson, C. T., “A GPS/GIS method for travel mode detection in New York City”, *Computers, Environment and Urban Systems*, Vol. 36, No. 2, 2012, pp. 131 - 139, special Issue: Geoinformatics 2010, available at: <http://www.sciencedirect.com/science/article/pii/S0198971511000536>
- [43] Wang, L., Ciliberto, M., Gjoreski, H., Lago, P., Murao, K., Okita, T., Roggen, D., *Locomotion and Transportation Mode Recognition from GPS and Radio Signals: Summary of SHL Challenge 2021*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 412–422, available at: <https://doi.org/10.1145/3460418.3479373>
- [44] Wang, L., Gjoreski, H., Ciliberto, M., Lago, P., Murao, K., Okita, T., Roggen, D., “Summary of the sussex-huawei locomotion-transportation recognition challenge 2020”, in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, ser. UbiComp-ISWC '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 351–358, available at: <https://doi.org/10.1145/3410530.3414341>
- [45] Fang, S.-H., Liao, H.-H., Fei, Y.-X., Chen, K.-H., Huang, J.-W., Lu, Y.-D., Tsao, Y., “Transportation modes classification using sensors on smartphones”, *Sensors*, Vol. 16, No. 8, 2016, available at: <https://www.mdpi.com/1424-8220/16/8/1324>

- [46] Fang, S.-H., Fei, Y.-X., Xu, Z., Tsao, Y., “Learning transportation modes from smartphone sensors based on deep neural network”, *IEEE Sensors Journal*, Vol. 17, No. 18, 2017, pp. 6111-6118.
- [47] Wang, S., Chen, C., Ma, J., “Accelerometer based transportation mode recognition on mobile phones”, in *2010 Asia-Pacific Conference on Wearable Computing Systems*, April 2010, pp. 44-46.
- [48] Jahangiri, A., Rakha, H. A., “Applying machine learning techniques to transportation mode recognition using mobile phone sensor data”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 5, Oct 2015, pp. 2406-2417.
- [49] Hemminki, S., Nurmi, P., Tarkoma, S., “Accelerometer-based transportation mode detection on smartphones”, in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, ser. *SenSys '13*. New York, NY, USA: ACM, 2013, pp. 13:1–13:14, available at: <http://doi.acm.org/10.1145/2517351.2517367>
- [50] Su, X., Caceres, H., Tong, H., He, Q., “Online travel mode identification using smartphones with battery saving considerations”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 10, Oct 2016, pp. 2921-2934.
- [51] Siirtola, P., Rönning, J., “Recognizing human activities user-independently on smartphones based on accelerometer data”, *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 1, No. 5, 06/2012 2012, pp. 38-45, available at: [http://www.ijimai.org/journal/sites/default/files/IJIMAI20121\\_5\\_5.pdf](http://www.ijimai.org/journal/sites/default/files/IJIMAI20121_5_5.pdf)
- [52] Zhang, Z., Poslad, S., “A new post correction algorithm (pocoa) for improved transportation mode recognition”, in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 1512-1518.
- [53] Yang, J., “Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones”, *ACM International Workshop on Interactive Multimedia for Consumer Electronics*, Vol. 39, 01 2009, pp. 1-10.
- [54] Sonderen, T., “Detection of transportation mode solely using smartphones”. *Computer Science*, 2016.
- [55] Widhalm, P., Nitsche, P., Brändle, N., “Transport mode detection with realistic smartphone sensor data”, *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 573-576.

- [56] Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., Maurer, P., “Supporting large-scale travel surveys with smartphones – a practical approach”, *Transportation Research Part C: Emerging Technologies*, Vol. 43, 2014, pp. 212 - 221, special Issue with Selected Papers from Transport Research Arena.
- [57] Nour, A., Hellinga, B., Casello, J., “Classification of automobile and transit trips from smartphone data: Enhancing accuracy using spatial statistics and GIS”, *Journal of Transport Geography*, Vol. 51, 2016, pp. 36 - 44, available at: <http://www.sciencedirect.com/science/article/pii/S0966692315002070>
- [58] Ashqar, H. I., Almannaa, M. H., Elhenawy, M., Rakha, H. A., House, L., “Smartphone transportation mode recognition using a hierarchical machine learning classifier and pooled features from time and frequency domains”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 1, 2018, pp. 244–252.
- [59] Zhu, Y., Luo, H., Guo, S., Zhao, F., “Data mining for transportation mode recognition from radio-data”, in *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 423–427.
- [60] Wu, J., Jafari, R., “Orientation independent activity/gesture recognition using wearable motion sensors”, *IEEE Internet of Things Journal*, Vol. 6, No. 2, 2019, pp. 1427-1437.
- [61] Ahmed, M., Antar, A. D., Hossain, T., Inoue, S., Ahad, M. A. R., “Poiden: Position and orientation independent deep ensemble network for the classification of locomotion and transportation modes”, in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, ser. UbiComp/ISWC '19 Adjunct. New York, NY, USA: Association for Computing Machinery, 2019, pp. 674–679, available at: <https://doi.org/10.1145/3341162.3345570>
- [62] Ld, N., Nguyen, N., Nguyen, T.-H., Nguyen, H.-N., “Vehicle mode and driving activity detection based on analyzing sensor data of smartphones”, *Sensors*, Vol. 18, 03 2018, pp. 1036.
- [63] Gong, L., Kanamori, R., Yamamoto, T., “Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons”, *Travel Behaviour and Society*, Vol. 11, 2018, pp. 131 - 140, available at: <http://www.sciencedirect.com/science/article/pii/S2214367X1730039X>
- [64] Erdelić, T., Vrbančić, S., Rožić, L., “A model of speed profiles for urban road networks using g-means clustering”, in *2015 38th International Convention on Information and*

- Communication Technology, Electronics and Microelectronics (MIPRO), May 2015, pp. 1081-1086.
- [65] Zheng, Y., Liu, L., Wang, L., Xie, X., “Learning transportation mode from raw GPS data for geographic applications on the web”, in Proceeding of the 17th International Conference on World Wide Web 2008, WWW’08, 01 2008, pp. 247-256.
- [66] Zheng, Y., Xie, X., Ma, W.-Y., “Understanding mobility based on GPS data”, in Proceedings of the 10th ACM conference on Ubiquitous Computing (UbiComp 2008), September 2008, available at: <https://www.microsoft.com/en-us/research/publication/understanding-mobility-based-on-gps-data/>
- [67] Etemad, M., Soares Júnior, A., Matwin, S., “Predicting transportation modes of GPS trajectories using feature engineering and noise removal”, in Advances in Artificial Intelligence, Bagheri, E., Cheung, J. C., (ed.). Cham: Springer International Publishing, 2018, pp. 259–264.
- [68] Jahangiri, A., Rakha, H., “Applying machine learning techniques to transportation mode recognition using mobile phone sensor data”, IEEE Transactions on Intelligent Transportation Systems, 10 2015, pp. 1-12.
- [69] Jahangiri, A., Rakha, H., “Developing a support vector machine (SVM) classifier for transportation mode identification by using mobile phone sensor data”. Transportation Research Board 93rd Annual Meeting, 01 2014.
- [70] Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., Korhonen, I., “Activity classification using realistic data from wearable sensors”, IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 1, Jan 2006, pp. 119-128.
- [71] Shafique, M., Hato, E., “Formation of training and testing datasets, for transportation mode identification”, Journal of Traffic and Logistics Engineering, Vol. 3, 06 2015, pp. 77-80.
- [72] Júnior, A. S., Moreno, B. N., Times, V. C., Matwin, S., dos Anjos Formiga Cabral, L., “Grasp-uts: an algorithm for unsupervised trajectory segmentation”, International Journal of Geographical Information Science, Vol. 29, No. 1, 2015, pp. 46-68.
- [73] Nikitopoulos, P., Paraskevopoulos, A., Doulkeridis, C., Pelekis, N., Theodoridis, Y., “Hot spot analysis over big trajectory data”, in 2018 IEEE International Conference on Big Data (Big Data). Los Alamitos, CA, USA: IEEE Computer Society, dec 2018, pp. 761-770, available at: <https://doi.ieeecomputersociety.org/10.1109/BigData.2018.8622376>

- [74] Cai, G., Lee, K., Lee, I., “Mining mobility patterns from geotagged photos through semantic trajectory clustering”, *Cybernetics and Systems*, Vol. 49, No. 4, 2018, pp. 234-256.
- [75] Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E., Zegras, P. C., “Future mobility survey: Experience in developing a smartphone-based travel survey in Singapore”, *Transportation Research Record*, Vol. 2354, No. 1, 2013, pp. 59-67, available at: <https://doi.org/10.3141/2354-07>
- [76] Prelipcean, A. C., Gidofalvi, G., Susilo, Y. O., “Mobility collector”, *Journal of Location Based Services*, Vol. 8, No. 4, 2014, pp. 229-255, available at: <https://doi.org/10.1080/17489725.2014.973917>
- [77] Geurs, K. T., Thomas, T., Bijlsma, M., Douhou, S., “Automatic Trip and Mode Detection with Move Smarter: First Results from the Dutch Mobile Mobility Panel”, *Transportation Research Procedia*, Vol. 11, 2015, pp. 247 - 262, *transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods 16-21 November 2014, Leura, Australia*, available at: <http://www.sciencedirect.com/science/article/pii/S2352146515003130>
- [78] Montini, L., Prost, S., Schrammel, J., Rieser-Schussler, N., Axhausen, K. W., “Comparison of travel diaries generated from smartphone data and dedicated GPS devices”, *Transportation Research Procedia*, Vol. 11, 2015, pp. 227 - 241, *transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods 16-21 November 2014, Leura, Australia*, available at: <http://www.sciencedirect.com/science/article/pii/S2352146515003117>
- [79] Reddy, S., Burke, J., Estrin, D., Hansen, M., Srivastava, M., “Determining transportation mode on mobile phones”, in *2008 12th IEEE International Symposium on Wearable Computers*, Sep. 2008, pp. 25-28.
- [80] Bloch, A., Erdin, R., Meyer, S., Keller, T., Spindler, A., “Battery-efficient transportation mode detection on mobile devices”, in *2015 16th IEEE International Conference on Mobile Data Management*, Vol. 1, June 2015, pp. 185-190.
- [81] Wang, L., Gjoreski, H., Ciliberto, M., Mekki, S., Valentin, S., Roggen, D., “Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition With the Sussex-Huawei Dataset”, *IEEE Access*, Vol. 7, 2019, pp. 10 870-10 891.

- [82] Rasmussen, T. K., Ingvardson, J. B., Halldorsdottir, K., Nielsen, O. A., “Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area”, *Computers, Environment and Urban Systems*, Vol. 54, No. Complete, 2015, pp. 301-313.
- [83] Schuessler, N., Axhausen, K., “Processing raw data from global positioning systems without additional information”, *Transportation Research Record*, Vol. 2105, 12 2009, pp. 28-36.
- [84] Schüssler, N., Montini, L., Dobler, C., “Improving post-processing routines for GPS observations using prompted-recall data”, in *Arbeitsberichte Verkehrs- und Raumplanung*, Vol. 724. Zürich: International Steering Committee for Travel Survey Conferences (IS-CTSC), 2011, 9th International Conference on Transport Survey Methods; Conference Location: Termas de Puyehue, Chile; Conference Date: November 14-18, 2011.
- [85] Tsui, S., Shalaby, A., “Enhanced system for link and mode identification for personal travel surveys based on global positioning systems”, *Transportation Research Record*, Vol. 1972, 01 2006, pp. 38-45.
- [86] Bohte, W., Maat, K., “Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands”, *Transportation Research Part C-Emerging Technologies*, Vol. 17, 06 2009, pp. 285-297.
- [87] Xiao, G., Juan, Z., Gao, J., “Inferring Trip Ends from GPS Data Based on Smartphones in Shanghai”, in *TRB 94th Annual Meeting Compendium of Papers*. Washington DC, United States: Transportation Research Board 94th Annual Meeting, 01 2015.
- [88] Zheng, Y., Zhang, L., Ma, Z., Xie, X., Ma, W.-Y., “Recommending friends and locations based on individual location history”, *ACM Transactions on the Web*, Vol. 5, No. 1, feb 2011, available at: <https://doi.org/10.1145/1921591.1921596>
- [89] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., “A density-based algorithm for discovering clusters in large spatial databases with noise”, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, pp. 226–231.
- [90] Palma, A. T., Bogorny, V., Kuijpers, B., Alvares, L. O., “A clustering-based approach for discovering interesting places in trajectories”, in *Proceedings of the 2008 ACM Symposium on Applied Computing*, ser. SAC '08. New York, NY, USA: Association for Computing Machinery, 2008, pp. 863–868, available at: <https://doi.org/10.1145/1363686.1363886>

- [91] Chen, W., Ji, M., Wang, J., “T-dbscan: A spatiotemporal density clustering for GPS trajectory segmentation”, *International Journal of Online and Biomedical Engineering*, Vol. 10, 2014, pp. 19-24.
- [92] Leiva, L. A., Vidal, E., “Warped k-means: An algorithm to cluster sequentially-distributed data”, *Information Sciences*, Vol. 237, 2013, pp. 196-210.
- [93] Jain, A. K., Dubes, R. C., *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.
- [94] Etemad, M., Júnior, A. S., Hoseyni, A., Rose, J., Matwin, S., “A trajectory segmentation algorithm based on interpolation-based change detection strategies”, in *Proceedings of the Workshops of the EDBT/ICDT 2019 Joint Conference, EDBT/ICDT 2019, Lisbon, Portugal, March 26, 2019*, ser. CEUR Workshop Proceedings, Vol. 2322. CEUR-WS.org, 2019, available at: [http://ceur-ws.org/Vol-2322/BMDA\\_4.pdf](http://ceur-ws.org/Vol-2322/BMDA_4.pdf)
- [95] Etemad, M., Soares, A., Etemad, E., Rose, J., Torgo, L., Matwin, S., “SWS: an unsupervised trajectory segmentation algorithm based on change detection with interpolation kernels”, *GeoInformatica*, Vol. 25, 2021, pp. 269–289.
- [96] Soares Junior, A., Cesario Times, V., Renso, C., Matwin, S., Cabral, L. A., “A semi-supervised approach for the semantic segmentation of trajectories”, in *2018 19th IEEE International Conference on Mobile Data Management (MDM)*, 2018, pp. 145-154.
- [97] Akbari, A., Wu, J., Grimsley, R., Jafari, R., “Hierarchical signal segmentation and classification for accurate activity recognition”, in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, ser. UbiComp '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1596–1605, available at: <https://doi.org/10.1145/3267305.3267528>
- [98] Aha, D., Kibler, D., “Instance-based learning algorithms”, *Machine Learning*, Vol. 6, 1991, pp. 37-66.
- [99] Peterson, L. E., “K-nearest neighbor”, *Scholarpedia*, Vol. 4, No. 2, 2009, pp. 1883, revision #137311.
- [100] Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., *Classification and regression trees*. Routledge, 2017.
- [101] Fayyad, U. M., Irani, K. B., “The attribute selection problem in decision tree generation”, in *Proceedings of the Tenth National Conference on Artificial Intelligence*, ser. AAAI'92. AAAI Press, 1992, pp. 104–110.



- [102] Cheushev, V., Simovici, D. A., Shmerko, V., Yanushkevich, S., “Functional entropy and decision trees”, in Proceedings. 1998 28th IEEE International Symposium on Multiple-Valued Logic (Cat. No. 98CB36138). IEEE, 1998, pp. 257–262.
- [103] Dayhoff, J. E., Neural network architectures: An introduction. Van Nostrand Reinhold Co., 1990.
- [104] Müller, B., Reinhardt, J., Strickland, M. T., Neural networks: An introduction. Springer Science & Business Media, 1995.
- [105] Fine, S., Singer, Y., Tishby, N., “The hierarchical hidden markov model: Analysis and applications”, Machine learning, Vol. 32, No. 1, 1998, pp. 41–62.
- [106] Lappalainen, H., Miskin, J. W., “Ensemble learning”, in Advances in Independent Component Analysis. Springer, 2000, pp. 75–92.
- [107] Breiman, L., “Random forests”, Machine Learning, Vol. 45, No. 1, Oct 2001, pp. pp. 5–32, available at: <https://doi.org/10.1023/A:1010933404324>
- [108] Margineantu, D. D., Dietterich, T. G., “Pruning adaptive boosting”, in Proceedings of the Fourteenth International Conference on Machine Learning, ser. ICML '97. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997, pp. 211–218.
- [109] Alasadi, S. A., Bhaya, W. S., “Review of data preprocessing techniques in data mining”, Journal of Engineering and Applied Sciences, Vol. 12, No. 16, 2017, pp. 4102–4107.
- [110] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A. T., “A survey of mobile phone sensing”, IEEE Communications Magazine, Vol. 48, No. 9, 2010, pp. 140-150.
- [111] Smith, S. W., “Chapter 8 - the discrete fourier transform”, in Digital Signal Processing, Smith, S. W., (ed.). Boston: Newnes, 2003, pp. 141-168, available at: <https://www.sciencedirect.com/science/article/pii/B9780750674447500455>
- [112] Ravi, D., Wong, C., Lo, B., Yang, G.-Z., “Deep learning for human activity recognition: A resource efficient implementation on low-power devices”, in 2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN), 2016, pp. 71-76.
- [113] Ha, S., Yun, J.-M., Choi, S., “Multi-modal convolutional neural networks for activity recognition”, in 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015, pp. 3017-3022.

- [114] Nussbaumer, H. J., “The fast fourier transform”, in Fast Fourier Transform and Convolution Algorithms. Springer, 1981, pp. 80–111.
- [115] Thibault, K., Euler’s Formula: A Complete Guide. Math Value - e-book.
- [116] Cooley, J. W., Tukey, J. W., “An algorithm for the machine calculation of complex fourier series”, Mathematics of Computation, Vol. 19, No. 90, 1965, pp. 297–301, available at: <http://www.jstor.org/stable/2003354>
- [117] Huynh, T., Schiele, B., “Analyzing features for activity recognition”, in Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies. New York, NY, USA: Association for Computing Machinery, 2005, pp. 159–163, available at: <https://doi.org/10.1145/1107548.1107591>
- [118] Ertuğrul, Ö. F., Dal, S., Hazar, Y., Aldemir, E., “Determining relevant features in activity recognition via wearable sensors on the myo armband”, Arabian Journal for Science and Engineering, Vol. 45, 2020, pp. 10 097–10 113.
- [119] Gjoreski, H., Gams, M., “Accelerometer data preparation for activity recognition”, in Proceedings of the International Multiconference Information Society, Ljubljana, Slovenia, Vol. 1014, 2011, pp. 1014.
- [120] Bayat, A., Pomplun, M., Tran, D. A., “A study on human activity recognition using accelerometer data from smartphones”, Procedia Computer Science, Vol. 34, 2014, pp. 450-457, the 9th International Conference on Future Networks and Communications (FNC’14)/The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC’14)/Affiliated Workshops, available at: <https://www.sciencedirect.com/science/article/pii/S1877050914008643>
- [121] Friedrich, B., Lübbe, C., Hein, A., “Combining LSTM and CNN for mode of transportation classification from smartphone sensors”, in Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, ser. UbiComp-ISWC ’20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 305–310, available at: <https://doi.org/10.1145/3410530.3414350>
- [122] Yurtman, A., Barshan, B., “Activity recognition invariant to sensor orientation with wearable motion sensors”, Sensors, Vol. 17, 08 2017.
- [123] Fu, R., Wang, H., Zhao, W., “Dynamic driver fatigue detection using hidden markov model in real driving condition”, Expert Systems with Applications, Vol. 63,

- 2016, pp. 397-411, available at: <https://www.sciencedirect.com/science/article/pii/S0957417416303293>
- [124] Li, Z., Fang, H., Huang, M., “Diversified learning for continuous hidden markov models with application to fault diagnosis”, *Expert Systems with Applications*, Vol. 42, 12 2015, pp. 9165-9173.
- [125] Asghari, P., Soelimani, E., Nazerfard, E., “Online human activity recognition employing hierarchical hidden markov models”, *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, 2020, pp. 1141-1152.
- [126] Freedman, D., Diaconis, P., “On the histogram as a density estimator: L<sub>2</sub> theory”, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. 57, No. 4, 1981, pp. 453–476.
- [127] Sturges, H. A., “The choice of a class interval”, *Journal of the american statistical association*, Vol. 21, No. 153, 1926, pp. 65–66.
- [128] Clarke, E. J., Barton, B. A., “Entropy and mdl discretization of continuous variables for bayesian belief networks”, *International Journal of Intelligent Systems*, Vol. 15, No. 1, 2000, pp. 61–92.
- [129] Joița, D., “Unsupervised static discretization methods in data mining”, Titu Maiorescu University, Bucharest, Romania, 2010.
- [130] Sawilowsky, S. S., “Fermat, schubert, einstein, and behrens-fisher: the probable difference between two means when  $s_1^2 = s_2^2$ ”, *Journal of Modern Applied Statistical Methods*, Vol. 1, 2002, pp. 461–472.
- [131] Vinh, N. X., Epps, J., Bailey, J., “Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance”, *The Journal of Machine Learning Research*, Vol. 11, 2010, pp. 2837–2854.
- [132] Guyon, I., Weston, J., Barnhill, S., Vapnik, V., “Gene selection for cancer classification using support vector machines”, *Machine Learning*, Vol. 46, 01 2002, pp. 389-422.
- [133] Mani, I., Zhang, I., “knn approach to unbalanced data distributions: a case study involving information extraction”, in *Proceedings of workshop on learning from imbalanced datasets*, Vol. 126. ICML, 2003, pp. 1–7.
- [134] Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., “Algorithms for hyper-parameter optimization”, in *Advances in Neural Information Processing Systems*, Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., (ed.), Vol. 24. Curran

- Associates, Inc., 2011, available at: <https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf>
- [135] Schakenbos, R., Paix, L. L., Nijenstein, S., Geurs, K. T., “Valuation of a transfer in a multimodal public transport trip”, *Transport Policy*, Vol. 46, 2016, pp. 72-81, available at: <https://www.sciencedirect.com/science/article/pii/S0967070X15300743>
- [136] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., “Scikit-learn: Machine learning in Python”, *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.
- [137] Wang, L., Gjoreski, H., Ciliberto, M., Lago, P., Murao, K., Okita, T., Roggen, D., “Summary of the sussex-huawei locomotion-transportation recognition challenge 2020”, in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 2020, pp. 351–358.
- [138] Kerr, L. A., Goethel, D. R., “Chapter twenty one - simulation modeling as a tool for synthesis of stock identification information”, in *Stock Identification Methods (Second Edition)*, second edition ed., Cadrin, S. X., Kerr, L. A., Mariani, S., (ed.). San Diego: Academic Press, 2014, pp. 501-533, available at: <https://www.sciencedirect.com/science/article/pii/B9780123970039000217>
- [139] “Zakon o provedbi opće uredbe o zaštiti podataka”, [https://narodne-novine.nn.hr/clanci/sluzbeni/2018\\_05\\_42\\_805.html](https://narodne-novine.nn.hr/clanci/sluzbeni/2018_05_42_805.html), pristupljeno: 2022-04-20.



# Popis slika

2.1. Vrijednosti mjerenja akceleracije u smjeru x osi kretanja korisnika a) jednim ili b) više modova prijevoza . . . . .	10
2.2. Mjerenja s osjetila za jednu trajektoriju korisnika s više uzastopnih modova prijevoza . . . . .	17
2.3. Proces klasifikacije modova prijevoza u stvarnom vremenu . . . . .	18
3.1. Proces donošenja odluke tijekom klasifikacije kNN algoritmom . . . . .	26
3.2. Osnovni dijelovi DT . . . . .	27
3.3. Primjer donošenja odluke na osnovi jedne značajke skupa podataka . . . . .	28
3.4. Osnovna struktura neuronske mreže . . . . .	29
3.5. Struktura neurona . . . . .	30
3.6. Struktura HMM . . . . .	31
3.7. Osnovna struktura metode RF . . . . .	32
3.8. Osnovna struktura višeslojnih metoda . . . . .	33
3.9. Zastupljenost modova prijevoza u SHL skupu podataka . . . . .	34
4.1. Dijagram procesa izrade višerazinske metode za segmentaciju i klasifikaciju trajektorije s obzirom na mod prijevoza . . . . .	38
4.2. Proces podijele sirovih podataka u vremenske prozore . . . . .	40
4.3. Preslikavanje podataka iz vremenske u frekvencijsku domenu . . . . .	42
4.4. Proces pretvorbe sirovih podataka u vektor značajki . . . . .	42
4.5. Formiranje TSM za jednu značajku u vektoru značajki . . . . .	47
4.6. Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda jednakog razmaka . . . . .	49
4.7. Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda jednake frekvencije . . . . .	49
4.8. Diskretizacija srednje vrijednosti magnitude akceleracije u 15 razreda grupiranih algoritmom $k$ sredine . . . . .	50
4.9. Testiranje TSM-a . . . . .	52
4.10. Proces smanjenja broja značajki . . . . .	53

4.11. Postupak ujednačavanja broja uzoraka . . . . .	57
4.12. Rezultat predviđanja RF metode prikazan u prvom retku i rezultat nakon primjene prvog pravila u drugom retku . . . . .	60
4.13. Rezultat predviđanja RF metode prikazan u prvom retku i rezultat nakon primjene drugog pravila u drugom retku . . . . .	61
4.14. Vrijednosti mjerenja akceleracije za vremenske prozore prikazano kroz više značajki . . . . .	64
4.15. Vrijednosti mjerenja žiroskopa za vremenske prozore prikazano kroz više značajki . . . . .	65
4.16. Prikaz ovisnosti značajki s najvećom MI vrijednosti . . . . .	66
4.17. TSM za (a) značajku asimetrije podataka i akceleraciju po z-osi za vremenske prozore s promjenom moda prijevoza; (b) značajku asimetrije podataka i akceleraciju po z-osi za vremenske prozore bez promjene moda prijevoza, (c) značajku centar frekvencije i magnitudu akceleracije za vremenske prozore s promjenom moda prijevoza; i (d) značajku centar frekvencije i magnitudu akceleracije za vremenske prozore bez promjene moda prijevoza . . . . .	68
4.18. Vrijednost sirovih podataka u trajektoriji po x osi akcelerometra i srednja vrijednost akceleracije . . . . .	69
4.19. Važnost značajki prema F vrijednosti . . . . .	70
4.20. Rezultat testiranja TSM-a izrađenih metodom diskretizacije jednake duljine za različit broj značajki . . . . .	70
4.21. Rezultat testiranja TSM-a izrađenih metodom diskretizacije jednake frekvencije za različit broj značajki . . . . .	71
4.22. Rezultat testiranja TSM-a izrađenih metodom diskretizacije algoritmom k sredina za različit broj značajki . . . . .	71
4.23. MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju . . . . .	74
4.23. MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju . . . . .	75
4.23. MI vrijednost za svaki par modova prijevoza uključenih u klasifikaciju . . . . .	76
5.1. Koordinatni sustav programskog okvira <i>Sensors</i> u odnosu na mobilni uređaj . . . . .	83
5.2. Početna aktivnost za prijavu korisnika u mobilnu aplikaciju <i>Collecty</i> . . . . .	85
5.3. POST akcija za dohvata tokena unutar aplikacije <i>Postman</i> . . . . .	86
5.4. Klase u poslužiteljskom programu . . . . .	88
5.5. Aktivnost za odabir trenutnog moda prijevoza . . . . .	89
5.6. Aktivnost za prikaz vrijednosti s osjetila tijekom korištenja aplikacije . . . . .	90
5.7. Aktivnost za potvrdu i provjeru ispravnosti prikupljenih podataka . . . . .	91
5.8. Dijagram izdvojenih klasa u mobilnoj aplikaciji <i>Collecty</i> . . . . .	92
5.9. Dijagram baze podataka . . . . .	93
5.10. Tekstualna datoteka za pohranu podataka . . . . .	94

---

5.11. Raspodjela količine podataka prema modovima prijevoza . . . . .	97
5.12. Prosječno trajanje trajektorija svakog korisnika . . . . .	99
5.13. Raspodjela trajektorije prema broju korištenih modova prijevoza u trajektoriji .	100
5.14. Matrica zastupljenosti prijelaza između modova prijevoza . . . . .	101
5.15. Karta gustoće prikupljenih podataka za sve podatke i za mod prijevoza automobil	102
5.16. Karta gustoće prikupljenih podataka za mod prijevoza vlak i električni romobil	102
5.17. Karta gustoće prikupljenih podataka za sve podatke moda prijevoza autobus i za jednu autobusnu liniju . . . . .	103
5.18. Podatci jedne trajektorije s osjetila: a) akcelerometar b) gravitacija c) stopa rotacije d) jakost magnetskog polja . . . . .	104
5.19. Vrijednosti odabranih značajki akceleracije u vremenskim prozorima za prikup- ljeni skup podataka . . . . .	105





# Popis tablica

2.1. Istraživanja u području klasifikacije modova prijevoza na povijesnim podacima	12
2.2. Značajke u području klasifikacije modova prijevoza	14
3.1. Primijenjene metode na problemu klasifikacije modova prijevoza	25
3.2. Osnovne karakteristike objavljenog skupa podataka SHL	33
3.3. Svojstva testnog skupa podataka	35
3.4. Vremenska podjela SHL skupa podataka	36
4.1. Značajke u vremenskoj i frekvencijskoj domeni	44
4.2. Značajke frekvencijskih potpojasa	45
4.3. Broj zapisa nakon brisanja podataka	63
4.4. Rezultati testiranja TSM na testnom skupu podataka	72
4.5. Parametri RF modela za svaki scenarij	73
4.6. Rezultati testiranja korisničke neovisnosti metode za klasifikaciju modova prijevoza	78
4.7. Rezultati testiranja lokacijske neovisnosti metode za klasifikaciju modova prijevoza	79
4.8. Rezultati testiranja vremenske neovisnosti metode za klasifikaciju modova prijevoza	80
5.1. Sadržaj zaglavlja tekstualne datoteke u kojoj su zapisani podaci	96
5.2. Raspodjela podataka prema modovima prijevoza po korisniku izražena u minutama	98
5.3. Raspodjela podataka prema modovima prijevoza po korisniku izražena u kilometrima	99
5.4. Rezultati validacije TSM-a na prikupljenom skupu podataka	105
5.5. Rezultati višerazinske metode za segmentaciju i klasifikaciju trajektorije na prikupljenom skupu podataka	106
5.6. Rezultati modela naučenog na <i>Collecty</i> podacima za klasifikaciju modova prijevoza	107



# Popis algoritama

3.1. Metoda za detekciju stajanja, [88]	21
3.2. Algoritam DBSCAN, [89]	22
3.3. Algoritam $k$ -sredina, [93]	23
3.4. GRASP-UTS, [72]	24
4.1. Metoda rekurzivnog uklanjanja značajki, [132]	54
4.2. Algoritam slučajne šume, [107]	55
4.3. CART, [100]	56
4.4. TPE, [134]	58
4.5. Višerazinska metoda za segmentaciju i klasifikaciju trajektorije	61



# Popis kratica

AdaBoost - adaptivno pojačanje - Adaptive Boosting

ANOVA - analiza varijance - ANalysis Of VAriance

API - aplikacijsko programsko sučelje - Application Program Interface

CART - klasifikacijska i regresijska stabla - Classification And Regression Trees

CB-SMoT - prepoznavanje stajanja i kretanja u trajektoriji pomoću grupiranja - Clustering-Based Stops and Moves of Trajectories

CNN - konvolucijska neuronska mreža - Convolutional Neural Network

CRUD - kreiraj, pročitaj, promjeni i izbriši - Create, Read, Update and Delete

DBSCAN - prostorno grupiranje zasnovano na gustoći - Density-Based Spatial Clustering of Applications with Noise

DFT - diskretna Fourierova transformacija - Discrete Fourier Transform

DNN - duboka neuronska mreža - Deep Neural Network

DT - stablo odluke - Decision Tree

EoG - kraj stajanja - End of Gap

EoW - kraj hodanja - End of Walk

FFT - brza Fourierova transformacija - Fast Fourier Transform

FL - neizrazita logika - Fuzzy Logic

FT - Fourierova transformacija - Fourier Transform

GIS - geografski informacijski sustav - Geographic Information System

GNSS - globalni sustav za satelitsku navigaciju - Global Navigation Satellite System

GPS - globalni položajni sustav - Global Positioning System

GRASP-UTS - nenadziran pohlepan algoritam za segmentaciju trajektorije - Greedy Randomized Adaptive Search Procedure for Unsupervised Trajectory Segmentation

HAR - prepoznavanje ljudskih aktivnosti - Human Activity Recognition

HMM - skriveni Markovljevi modeli - Hidden Markov Model

ID3 - iterativno dijeljenje na dva dijela - Iterative Dichotomiser 3

JSON - format za prijenos strukture podataka - Java Script Object Notation

kNN -  $k$  najbližih susjeda -  $k$  Nearest Neighbor

LTE - dugoročna evolucija mobilnih sustava - Long Term Evolution

MDL - najmanja dužina opisa - Minimum Description Length

MI - uzajamna informacija - Mutual Information

MTP - točka promjene moda prijevoza - Mode Transfer Point

NB - Bayesov klasifikator - Naive Bayes

NN - neuronska mreža - Neural Network

OWS - oktalna segmentacija vremenskih prozora - Octal Window Segmentation

PBC - klasifikacija zasnovana na podacima prikupljenim na jednoj geografskoj lokaciji - Point-Based Classification

RF - slučajna šuma - Random Forest

RFE - rekurzivno smanjenje broja značajki - Recursive Feature Elimination

RGRASP-SemTS - polunadziran pohlepan algoritam za segmentaciju trajektorije - Reactive GRASP for semantic Semi-supervised Trajectory Segmentation

SBC - klasifikacija zasnovana na podacima prikupljenim na jednom segmentu trajektorije - Segment-Based Classification

SHL - Sussex-Huawei skup podataka o kretanju i prijevozu korisnika - The Sussex-Huawei Locomotion and Transportation Dataset

SoW - početak hodanja - Start of Walk

SQL - upitni jezik SQL - Structured Query Language

STP - detekcija točke stajanja - Staying Point Detection

SVM - metoda potpornih vektora - Support Vector Machine

SWS - segmentacija vremenskih prozora s pomičnim vremenom početka vremenskog prozora  
- Sliding Window Segmentation

T-DBSCAN - prostorno grupiranje u trajektoriji zasnovano na gustoći - Trajectory DBSCAN

TMD - detekcija moda prijevoza - Transport Mode Detection

TPE - predviđanje zasnovano na stablastoj strukturi modela - Tree based Parzen Estimator

TSM - matrice prijelaznih stanja - Transition State Matrices

Wk-Means - izmijenjen algoritam k sredina - Warped-kMeans





# Životopis

**Martina Erdelić** rođena je 16.8.1990. u Zagrebu gdje je završila osnovnoškolsko obrazovanje te Opću gimnaziju u Sesvetama. Nakon srednjoškolskog obrazovanja upisuje Fakultet prometnih znanosti gdje završava preddiplomski studij 2012. godine na smjeru informacijsko komunikacijski promet. Titulu magistre inženjerke prometa stječe 2014. godine obranom diplomskog rada na temu „Metode procjene kvalitete usluge prijenosa zvuka strujanjem“ (mentorica: prof. dr. sc. Štefica Mrvelj). U studenom 2014. godine zapošljava se na Fakultetu prometnih znanosti u sklopu projekta „SORDITO-Sustav za Optimizaciju Ruta u DINamičkom Transportnom Okruženju“ na radnom mjestu asistenta na projektu. Tijekom 2015. godine upisuje poslijediplomski doktorski studij „Tehnološki sustavi u prometu i transportu“ na Fakultetu prometnih znanosti te je izabrana u suradničko zvanje asistenta iz područja tehničkih znanosti, polje tehnologija prometa i transporta na Fakultetu prometnih znanosti, Sveučilište u Zagrebu.

Trenutno je zaposlena na Fakultetu prometnih znanosti na Zavodu za inteligentne transportne sustave, te je nastavno angažirana na kolegijima Algoritmi i programiranje, Baze podataka, Napredne baze podataka, Rudarenje podataka i Optimizacija prometnih procesa. Sudjelovala je na znanstvenim projektima te je bila neposredni voditelj na više završnih studentskih radova. U njezine vještine ubrajaju se programiranje u nekoliko programskih jezika C#, R, Python i Java, te znanje rada u sljedećim programskim alatima: MATLAB, Android Studio, Microsoft Office, Microsoft SQL, Apache Flink i dr.

Njezino područje interesa uključuje obradu i analizu velikih skupova podataka te primjenu metoda strojnog učenja na velikim skupovima podataka iz područja prometa i transporta. U dosadašnjem istraživanju objavila je dva rada u časopisima indeksiranim u Web of Science Core Collection i šest radova na međunarodnim znanstvenim konferencijama.

## Popis objavljenih djela

### Radovi u časopisima

1. Erdelić, M., Carić, T., Erdelić, T., Tišljarić, L., "Transition State Matrices Approach for Trajectory Segmentation Based on Transport Mode Change Criteria", Sustainability, Vol.

14, 2022., str.2756

2. Erdelić, T., Carić, T., Erdelić, M., Tišljarić, L., Turković, A., Jelušić, N., "Estimating congestion zones and travel time indexes based on the floating car data", *Computers, Environment and Urban Systems*, Vol. 87, 2021, str. 101604

### **Radovi na međunarodnim znanstvenim skupovima**

1. Tišljarić, L., Carić, T., Erdelić, T., Erdelić, M., "Traffic state estimation using speed profiles and convolutional neural networks", in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2020, str. 2147-2152, doi:10.23919/MIPRO48935.2020.9245177
2. Erdelić, M., Carić, T.; Ivanjko, E., Jelušić, N., "Classification of Travel Modes Using Streaming GNSS Data", *Transportation Research Procedia*, Vol. 40, 2019, str. 209-216, TRANSCOM 2019 13th International Scientific Conference on Sustainable, Modern and Safe Transport, doi:10.1016/j.trpro.2019.07.032
3. Erdelić, T., Carić, T., Erdelić, M., Tišljarić, L., "Electric vehicle routing problem with single or multiple recharges", *Transportation Research Procedia*, Vol. 40, 2019, str. 217 - 224, TRANSCOM 2019 13th International Scientific Conference on Sustainable, Modern and Safe Transport, doi: 10.1016/j.trpro.2019.07.033
4. Erdelić, T., Carić, T., Ravlić, M., "Predicting waiting times at intersections", in *2017 International Symposium ELMAR*, 2017, str. 31-34., doi: 10.23919/ELMAR.2017.8124428
5. Ravlić, M., Erdelić, T., Carić, T., "Optimizing charging station locations for fleet of electric vehicles using multi-source weber problem", in *2016 International Symposium ELMAR*, 2016, str. 111-114., doi: 10.1109/ELMAR.2016.7731766
6. Erdelić, T., Ravlić, M., Carić, T., "Travel time prediction using speed profiles for road network of Croatia", in *2016 International Symposium ELMAR*, Sept 2016, str. 97-100., doi: 10.1109/ELMAR.2016.7731763