

# Klasifikacija pokretnih korisničkih profila telekomunikacijske mreže metodologijom strojnog učenja

---

**Kušić, Kristina**

**Master's thesis / Diplomski rad**

**2020**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture / Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje**

*Permanent link / Trajna poveznica:* <https://urn.nsk.hr/urn:nbn:hr:235:381193>

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

*Download date / Datum preuzimanja:* **2024-12-26**

*Repository / Repozitorij:*

[Repository of Faculty of Mechanical Engineering and Naval Architecture University of Zagreb](#)



SVEUČILIŠTE U ZAGREBU  
FAKULTET STROJARSTVA I BRODOGRADNJE

# DIPLOMSKI RAD

**Kristina Kušić**

Zagreb, 2020.

SVEUČILIŠTE U ZAGREBU  
FAKULTET STROJARSTVA I BRODOGRADNJE

# DIPLOMSKI RAD

Mentor:

Doc. dr. sc. Tomislav Stipančić, dipl. ing.

Student:

Kristina Kušić

Zagreb, 2020.

*Izjavljujem da sam ovaj rad izradila samostalno koristeći znanja stečena tijekom studija i navedenu literaturu.*

*Zahvaljujem se svima koji su mi na bilo koji način pomogli prilikom izrade ovog rada.*

*Posebno se zahvaljujem svom mentoru doc. dr. sc. Tomislavu Stipančiću na ukazanom povjerenju, svoj pruženoj pomoći i savjetima tijekom izrade ovog rada te velikom razumijevanju. Također, zahvaljujem se kolegi dr. sc. Kristianu Skračiću na mnogim korisnim savjetima i pomoći pri izradi rada.*

*I na kraju, najveću zaslugu za ono što sam postigla pripisujem svojoj majci Jeli, koja je uvijek bila uz mene, bez obzira da li se radilo o teškim ili sretnim trenucima i bez čije podrške bi sve ovo bilo puno teže.*

*Veliko HVALA svima!*

Kristina Kušić



SVEUČILIŠTE U ZAGREBU  
FAKULTET STROJARSTVA I BRODOGRADNJE



Središnje povjerenstvo za završne i diplomske ispite  
Povjerenstvo za diplomske radove studija strojarstva za smjerove:  
proizvodno inženjerstvo, računalno inženjerstvo, industrijsko inženjerstvo i menadžment,  
inženjerstvo materijala te mehatronika i robotika

Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje	
Datum:	Prilog:
Klasa:	
Ur. broj:	

### DIPLOMSKI ZADATAK

Student: **KRISTINA KUŠIĆ** Mat. br.: 0035185396

Naslov rada na hrvatskom jeziku: **Klasifikacija pokretnih korisničkih profila telekomunikacijske mreže metodologijom strojnog učenja**

Naslov rada na engleskom jeziku: **Mobile user profile classification of a telecommunication network based on the machine learning methodology**

Opis zadatka:

Telekomunikacijski prijemnici i predajnici za mobilnu telefoniju sastoje se od većeg broja usmjerenih ćelija koje pokrivaju određena geoprostorna područja. Ovisno o fizičkom kretanju korisnika u mreži, svaka ćelija može imati drugačije vrijednosti parametara koji su za nju idealni, odnosno koji mogu rezultirati kvalitetnijim karakteristikama rada. Pretpostavka je da postoji konačan broj pokretnih (engl. mobility) korisničkih profila koji se koriste uslugama mreže te koji se ne mogu razlučiti jednostavnom geoprostornom analizom.

U radu je potrebno odrediti, prilagoditi i koristiti barem dva algoritama strojnog učenja te njima analizirati dostupnu bazu podataka s ciljem automatizirane identifikacije pokretnih profila pojedinih ćelija u mreži. Potrebno je uzeti u obzir učestalost pojavnosti profila s obzirom na doba dana i dostupnost usluga mreže. S obzirom na to da se profili mogu razlikovati ovisno o vremenskoj komponenti brzine, na taj je način potrebno temeljiti njihovu klasifikaciju (npr. spori, srednje brzi, brzi, jako brzi profili).


Dobivene rezultate je potrebno usporediti s rezultatima provođenja vanjske i neovisne evaluacije rada mreže.

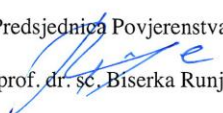
U radu je potrebno navesti korištenu literaturu te eventualno dobivenu pomoć.

Zadatak zadan:  
14. studenog 2019.

Rok predaje rada:  
16. siječnja 2020.

Predviđeni datum obrane:  
20. siječnja do 24. siječnja 2020.

Zadatak zadao:   
doc. dr. sc. Tomislav Stipančić

Predsjednica Povjerenstva:   
prof. dr. sc. Biserka Runje

## SADRŽAJ

<b>1. UVOD</b> .....	<b>1</b>
<b>2. MOBILNA MREŽA</b> .....	<b>3</b>
2.1. BAZNA STANICA .....	6
2.1.1. Domet ćelija .....	13
2.2. MOBILNE TEHNOLOGIJE .....	15
2.2.1. 1G - prva generacija .....	16
2.2.2. 2G - druga generacija .....	16
2.2.3. 3G - treća generacija .....	18
2.2.4. 4G – četvrta generacija .....	19
2.3. KEY PERFORMANCE INDICATORS - KPI .....	21
<b>3. STROJNO UČENJE I OBRADA PODATAKA</b> .....	<b>24</b>
3.1. PRIMJENA STROJNOG UČENJA .....	26
3.2. GRUPIRANJE (ENG. CLUSTERING) .....	28
3.2.1. K-means grupiranje .....	29
3.2.2. Odabir odgovarajućeg broja klastera .....	33
3.2.3. Gaussian Mixture metoda (GMM) .....	36
3.2.4. Usporedba K-means i GMM metode .....	40
<b>4. OPIS ZADATKA</b> .....	<b>41</b>
4.1. OBRADA PODATAKA .....	42
4.2. ANALIZA K-MEANS METODOM .....	47
4.2.1. Rezultati analize K-means metode .....	47
4.3. ANALIZA GAUSSIAN MIXTURE METODOM .....	51
4.3.1. Rezultati analize GMM metode .....	51
4.4. USPOREDBA REZULTATA K-MEANS I GMM METODE .....	54
4.5. USPOREDBA REZULTATA S NEOVISNIM MJERENJIMA .....	55
<b>5. ZAKLJUČAK</b> .....	<b>60</b>
<b>6. LITERATURA</b> .....	<b>62</b>
<b>7. PRILOZI</b> .....	<b>64</b>

**POPIS SLIKA**

Slika 1. Primjer bazne stanice.....	6
Slika 2. Prikaz antena za komunikaciju s mobilnim uređajima .....	7
Slika 3. Prikaz antena za komunikaciju.....	7
Slika 4. Ilustrativni prikaz pokrivenosti gradskog područja mrežom .....	8
Slika 5. Prikaz signalnih valova bazne stanice koje se dijelom reflektiraju od zgrada, a dijelom prolaze kroz nju .....	9
Slika 6. Primjer distrakcije signalnog vala .....	10
Slika 7. Prikaz moguće refleksije vala od više građevina.....	10
Slika 8. Skica tornjeva baznih stanica i relacija prikazani šesterokutom.....	11
Slika 9. Pokrivenost područja jedne bazne stanice sa tri sektorske antene .....	13
Slika 10. Prikaz „n“ broja podataka i željeni „k“ broj klastera.....	30
Slika 11. Prikaz nasumično dodijeljenih podatkovnih točaka klasteru .....	30
Slika 12. Prikaz izračunatih centroida klastera .....	30
Slika 15. Prikaz primjera grupiranja s većom količinom podataka.....	32
Slika 14. Primjer izračunavanja optimalnog broja klastera „metodom lakta“ .....	34
Slika 16. Krivulja normalne Gaussove razdiobe .....	36
Slika 17. 3D prikaz Gaussove krivulje .....	37
Slika 18. Primjer rezultata grupiranja K-means metodom.....	40
Slika 19. Primjer rezultata grupiranja GMM metodom .....	40
Slika 20. Određivanje raspona vremena .....	44
Slika 22. Provjera smislenosti zadanog raspona .....	45
Slika 23. Raspodjela korisnika unutar zadanih vremenskih raspona .....	45
Slika 24. Određivanje klastera K-means metodom.....	47
Slika 26. Klaster – 1 / K-means metoda.....	48
Slika 27. Klaster – 2 / K-means metoda.....	49
Slika 28. Klaster – 3 / K-means metoda.....	49
Slika 29. Grupacija klastera GMM metodom .....	51
Slika 30. Klaster – 0 / GMM metoda .....	52
Slika 31. Klaster – 1 / GMM metoda .....	52
Slika 32. Klaster – 2 / GMM metoda .....	53
Slika 33. Klaster – 3 / GMM metoda .....	53
Slika 34. Klaster – 0 / K-means, popodne.....	58
Slika 35. Klaster – 1 / K-means, popodne.....	58
Slika 36. Klaster – 2 / K-means, popodne.....	59
Slika 37. Klaster – 2 / K-means, popodne.....	59

## **POPIS TABLICA**

Tablica 1. Reprezentativni uzorak ulaznih podataka.....	43
Tablica 2. Usporedba rezultata različitih metoda.....	54
Tablica 3. Reprezentativni uzorci popodnevnih mjerenja.....	56
Tablica 4. Usporedba rezultata jutarnjih i popodnevnih mjerenja.....	57



**POPIS OZNAKA**

<b>Oznaka</b>	<b>Jedinica</b>	<b>Opis</b>
$\alpha$	[°]	Mjera kuta
d	[m]	Duljina
$v_1$	[bps]	Brzina prijenosa podataka u računalstvu
f	Hz	Frekvencija
$v_1$	[km/h]	Brzina
t	[s]	Vrijeme

## **SAŽETAK**

Tema ovog diplomskog rada je „Klasifikacija pokretnih korisničkih profila telekomunikacijske mreže metodologijom strojnog učenja“. U ovom radu opisani su koncepti rada mobilne telekomunikacijske mreže. Dvema metodama strojnog učenja analizirani su određeni parametri pokretnih korisničkih profila između dva europska grada. Korištena je dostupna baza podataka s ciljem identifikacije i grupacije pojedinih ćelija unutar zadanog područja. Cilj je pronaći ćelije visokog stupnja mobilnosti, odnosno ćelije čije je područje pokrivanja usmjereno na autoput i prugu. Korištene metode su K-means i Gaussian Mixture metoda. Provedena je usporedba rezultata dobivenih objema metodama. Također, dobiveni rezultati uspoređeni su s podacima o radu mreže iz drugog vremenskog razdoblja kako bi se dokazala ponovljivost i sezonalnost rezultata.

Ključne riječi: ćelija, bazna stanica, korisnički profil, grupiranje, klaster.

## SUMMARY

The topic of this master thesis is “Mobile user profiles classification of a telecommunication network based on the machine learning”. This project describes the concept of how mobile network works. This research uses two unsupervised machine learning methods in order to analyze cell *mobility* patterns. For this purpose, cell trace data from two European cities are used. The goal is to find “*high-mobility*” cells, or cells whose area of coverage is focused on the highway and track. The methods used were K-means and Gaussian Mixture method. A comparison of the results obtained with both methods is performed. Furthermore, the results were compared with the network performance data from other time interval to prove the repeatability and seasonality of the results.

Key words: cell, base station, user profile, clustering, cluster.

## 1. UVOD

Kako tehnologija napreduje, a svjetska populacija raste, tako raste i količina podataka koje svakodnevno stvaramo. Njihovo prikupljanje, sortiranje i analiziranje može biti težak zadatak. Istovremeno, razvoj mobilnih tehnologija i korištenje istih svakim danom je sve zastupljenije, kako u privatnim, tako i u poslovnim sferama.

Mobilne mreže danas su dio svakodnevnog života, vrlo su važne i značajne u svakodnevnom funkcioniranju stanovništva. U ovoj vrsti djelatnosti svoju potrebu i korist u različitom obujmu nalaze svi spektri društva, od pojedinaca do malih i velikih kompanija. Danas se gotovo svi služe nekim vidom mobilne komunikacije. Upravo radi potreba društva u čitavom svijetu (privatno, poslovno i globalno) mobilne telekomunikacije bilježe značajan razvoj unazad nekoliko desetljeća.

Mobilna mreža u svom značenju obuhvaća sučelje (eng. *hardver*) i programsku infrastrukturu (eng. *softver*). Do danas su razvijene i implementirane četiri ključne mobilne tehnologije – 1G, 2G, 3G i 4G. 5G mreža trenutno je u postupku razvoja i implementacije. Bazne stanice dio su hardverske infrastrukture mobilne mreže koje se sastoje se od većeg broja ćelija usmjerenih određenom geoprostornom području. Svaka se ćelija može podesiti na određene vrijednosti parametara u ovisnosti koji su za područje pokrivenosti idealni (npr. ruralno, planinsko, urbano područje, ili područje brzih cesta, vlakova, itd.). Time se postiže bolji i kvalitetniji rad pojedine ćelije što automatski rezultira boljim performansama cijele mreže. Pretpostavka je da postoji konačan broj pokretnih (eng. *mobility*) korisničkih profila ćelija te da se isti ne mogu razlučiti jednostavnom geoprostornom analizom.

Cilj ovog rada jest identificirati „*mobility*“ profile pojedinih ćelija u mreži između dva europska grada, kako bi se definirale „*high-mobility*“ ćelije. Ovaj cilj će se ostvariti analizom vremena zadržavanja korisnika u području pokrivanja pojedine ćelije. U radu su korištene dvije metode strojnog učenja pomoću kojih su se analizirala prikupljena mjerenja. S obzirom na to da se profili mogu razlikovati ovisno o vremenskom periodu provedenom u području neke ćelije, potrebno je same profile podijeliti u odgovarajuće vremenske periode i na temelju toga provesti analizu.

Dobiveni rezultati uspoređeni su s neovisnim mjerenjima iz drugog vremenskog perioda za iste ćelije i isto geoprostorno područje.

## 2. MOBILNA MREŽA

Mobilna mreža je komunikacijska mreža u kojoj je posljednja veza bežična. Mreža se rasprostire kroz površine koje pokrivaju bazne stanice, od kojih svaka koristi najmanje jedan (najčešće 3) primopredajnik signala s mobilnim uređajem i drugom baznom stanicom. Bazna stanica, cjelokupna infrastruktura unutar koje se nalaze i sve ćelije, na jednoj lokaciji pruža ćeliji mrežnu pokrivenost koja se može koristiti za prijenos glasa, podataka i drugih vrsta sadržaja. Stanica obično koristi drugačiji skup frekvencija od susjednih stanica, kako bi izbjegla smetnje i osigurala zajamčenu kvalitetu usluge unutar svake ćelije. Kada se spoje zajedno, te stanice pružaju radijsku pokrivenost na širokom zemljopisnom području što omogućuje velikom broju prijenosnih primopredajnika (npr. mobilni telefon, tablet i prijenosno računalo opremljeno mobilnim širokopojasnim modemima, *pager* itd.) međusobnu komunikaciju, s fiksnim primopredajnicima i telefonima bilo gdje u mreži, putem baznih stanica, čak i ako se neki od primopredajnika tijekom prijenosa kreću kroz više ćelija. [6]

Bazne stanice omogućuju niz poželjnih značajki unutar mobilne mreže. Jedna od njih je veći kapacitet od jednog velikog odašiljača, budući da se ista frekvencija može koristiti za višestruke veze sve dok su u različitim ćelijama. Mobilni uređaji troše manje energije u komunikaciji s baznom stanicom nego u komunikaciji s jednim odašiljačem ili satelitom, budući da su bazne stanice (konstrukcije) na maloj udaljenosti od korisnika, u bilo kojoj njegovoj poziciji tijekom kretanja. Veća površinska pokrivenost jednog zemaljskog odašiljača postiže se mogućnošću neograničenog dodavanja baznih stanica, prema potrebi, mogućnostima i odluci operatera – pružatelja telekomunikacijskih usluga.

Glavni pružatelji telekomunikacijskih usluga razmjestili su govorne i podatkovne mobilne mreže preko većine naseljenog kopnenog područja Zemlje. To omogućuje povezivanje mobilnih telefona i mobilnih računalnih uređaja s javnom komutiranom telefonskom mrežom i javnim internetom. Privatne mobilne mreže mogu se koristiti za istraživanje ili za velike organizacije, kao što je otprema za lokalne agencije za javnu sigurnost ili pojedine kompanije.

Povećani kapacitet u mobilnoj mreži, u usporedbi s mrežom s jednim odašiljačem, omogućio je višestrukim pozivateljima u istom području da koriste istu frekvenciju prebacivanjem poziva s iste frekvencije na najbližu dostupnu baznu stanicu koja ima tu istu

dostupnu frekvenciju. Ista se radiofrekvencija može ponovno upotrijebiti u drugom području za potpuno različit prijenos. Neizbježno, postoji određena razina smetnji od signala iz drugih stanica koje koriste istu frekvenciju, no ako postoji jedan običan odašiljač, samo se jedan prijenos može koristiti na bilo kojoj danoj frekvenciji. [4]

Ključna značajka mobilne mreže je mogućnost ponovnog korištenja frekvencija kako bi se povećala pokrivenost i kapacitet. Kao što je gore opisano, susjedne ćelije moraju koristiti različite frekvencije, međutim nema problema s dvije ćelije koje su dovoljno udaljene da djeluju na istoj frekvenciji, pod uvjetom da oprema korisnika (uređaj) mobilne mreže ne emitira previše energije.

Prva komercijalna mobilna mreža, generacija 1G, lansirana je u Japanu od strane Nippon Telegraph i Telephone (NTT) 1979. godine, najprije u gradskom području Tokija. U roku od pet godina, NTT mreža je proširena kako bi pokrila područje cijelog stanovništva Japana i postala prva 1G mreža u zemlji.

Mobilni telefoni rade slanjem i primanjem radio signala male snage. Signali se šalju i primaju od antena koje su priključene na radioodašiljače i prijemnike, tj. bazne stanice. Bazne stanice su povezane s ostalim mobilnim i fiksnim telefonskim mrežama i prenose signal, odnosno poziv na te mreže.

Najčešći primjer mobilne mreže je mreža mobilnog telefona (mobitela). Mobilni telefon je prijenosni telefon koji prima ili upućuje pozive putem stanice (bazne stanice) ili odašiljačkog tornja. Radio valovi koriste se za prijenos signala do i od mobitela. Moderne mreže mobilnih telefona koriste ćelije jer su radio frekvencije ograničene zajedničkim resursima. Svaka bazna stanica ima brojne radio stanice ili frekvencije za komunikaciju s mobilnim telefonima. Budući da je ovaj broj frekvencija ograničen, frekvencije se često ponovno koriste u susjednim stanicama. To se postiže smanjenjem razine snage bazne stanice kako bi se osiguralo minimalno ili nikakvo preklapanje pokrivenosti između ćelija. Stanice i uređaji mijenjaju frekvenciju pod kontrolom računala i koriste odašiljače male snage, tako da se obično ograničeni broj radio frekvencija može istovremeno koristiti od strane mnogih pozivatelja s manje smetnji. [2] [1]

Kada korisnik mobilnim uređajem pokušava uspostaviti poziv, prvi korak u postupku je da telefon provjeri postoji li pokrivenost u području u kojem je poziv upućen. Kada telefon potvrdi da postoji dovoljna jačina signala za upućivanje poziva, telefon uspostavlja vezu s

obližnjom baznom stanicom mobilnog telefona. Ova bazna stanica zatim uspostavlja poziv i zadržava poziv sve dok korisnik telefona ostane na pozivu i u dometu (području pokrivanja) te bazne stanice.

Praktično svaki mobilni sustav ima neku vrstu mehanizma emitiranja. To se može izravno koristiti za distribuciju informacija na više mobitela. Uobičajeno, na primjer u sustavima mobilne telefonije, najvažnija upotreba informacija emitiranja je postavljanje kanala za komunikaciju jedan-na-jedan između mobilnog primopredajnika i bazne stanice (eng. *paging*). Tri različite procedure pozivanja koje su općenito usvojene su sekvencijalni, paralelni i selektivni *paging*.

Pojedinosti o procesu *paginga* ponešto se razlikuju od tehnologije do tehnologije, ali obično znamo ograničen broj ćelija u kojima se telefon nalazi (GSM-2G, UTM-3G i LTE-4G). *Paging* se vrši slanjem emitirane poruke svim tim ćelijama. *Paging* poruke mogu se koristiti za prijenos informacija. U ovakvom sustavu, dok se distribuirani mobilni korisnici kreću od ćelije do ćelije tijekom kontinuirane komunikacije, prebacivanje s jedne stanične frekvencije na drugu staničnu frekvenciju vrši se elektronski bez prekida i bez operatora bazne stanice ili ručnog prebacivanja. To se naziva primopredaja. Obično se novi komunikacijski kanal (frekvencija) automatski bira za mobilnu jedinicu na novoj baznoj stanici koja će ga poslužiti. Mobilna jedinica se zatim automatski prebacuje s trenutnog kanala na novi kanal i komunikacija se nesmetano nastavlja.

Mobilni operater koristi mobilnu mrežu kako bi postigao pokrivenost i kapacitet svojih pretplatnika. Velika geografska područja podijeljena su u skupine manjih ćelija, tzv. klastere, kako bi se izbjegao gubitak signala u vidnom polju i podržao veliki broj aktivnih telefona u tom području. Sve su stanice spojene na telefonske centrale, koje se pak povezuju s javnom telefonskom mrežom. [2]

Budući da gotovo svi mobilni telefoni koriste mobilnu tehnologiju, pojam "mobitel" je u nekim regijama, osobito u SAD-u, zamjenjiv s "mobilnim telefonom". Međutim, satelitski telefoni su mobilni telefoni koji ne komuniciraju izravno sa zemaljskim baznim stanicama, ali to mogu učiniti posredno putem satelita.

Prijelaz s postojećeg analognog na digitalni standard slijedio je vrlo različit put u Europi i SAD-u. Kao posljedica toga, u SAD-u su se pojavili višestruki digitalni standardi, dok su Europa i mnoge druge zemlje izabrale GSM standard. [7]



## 2.1. Bazna stanica

Bazna stanica mobilne telekomunikacije (BSMT) je jedinstven naziv za lokaciju na kojoj se nalaze primopredajni uređaji i odgovarajuća telekomunikacijska oprema, koja služi za povezivanje bazne stanice s ostalim dijelovima javne mobilne mreže. Ona je pozicionirana u točki dodira svih ćelija koje sadrži, odnosno središtu kopnene površine unutar koje se distribuira mreža.



Slika 1. Primjer bazne stanice

Neki gradovi zahtijevaju da bazne stanice budu neprimjetne, moraju se „stopiti“ s okolinom, montirati na građevine ili reklamne konstrukcije. Ove se instalacije uglavnom nazivaju prikrivenim baznim stanicama ili nevidljivim stanicama.

Sastoji se od dva glavna dijela, radio uređaja i antene (*Slika 3.*). Radio uređaj služi za prijem i slanje signala mobilnim uređajima korisnika. Prijem i slanje signala vrše se na različitim frekvencijama. Stanica može istovremeno raditi na sve tri mobilne tehnologije jedne mreže, ali će se razlikovati opseg frekvencije koju koristi.

Na mjesto stanice postavljaju se antene i oprema za elektroničku komunikaciju - obično na radio jarbol, toranj ili drugu uzdignutu strukturu - radi stvaranja ćelije u mreži (*Slika 1.*). Bazna stanica konstruirana na tornju sadrži i jedan ili više skupa primopredajnika / prijemnika,

digitalne procesore signala, upravljačku elektroniku, GPS prijemnik za vremenski raspon, primarni i rezervni izvor električne energije i skloništa.



Slika 2. Prikaz antena za komunikaciju s mobilnim uređajima



Slika 3. Prikaz antena za komunikaciju

Svaka bazna stanica može nositi konačan broj poziva. U područjima visoke uporabe mobilnih telefona, kao što su središnji poslovni prostori i područja visoke gustoće, potrebno je više baznih stanica za upravljanje prometom poziva. U područjima visoke iskoristivosti često

postoji niz baznih stanica, od vrlo specifičnih rješenja u zgradi (dizajniranih da daju kvalitetnu pokrivenost unutar određene zgrade), do vrlo malih baznih stanica poznatih kao "mikročelije". Mikročelije pokrivaju malo geografsko područje i često se nalaze na raskrižjima i na područjima s višim protokom korisnika. U ruralnim područjima ili područjima gdje uporaba mobilnih telefona nije tako visoka, bazne stanice često će se nalaziti na brdima ili visokim strukturama kako bi se maksimiziralo područje pokrivenosti.

Položaj bazne stanice unutar ćelije određen je brojnim čimbenicima, uključujući topografiju i druga fizička ograničenja kao što su drveće i zgrade, kapacitet ćelije ili broj poziva za koje se očekuje da će biti napravljeni u ćeliji, te radio frekvencija na kojoj će raditi bazna stanica. Po obrascu koji ovisi o karakteristikama terena i karakteristikama prijema, ćelije mogu biti šesterokutnih, kvadratnih, kružnih ili nekih drugih pravilnih oblika, iako se šesterokutne najčešće primjenjuju. Mobilni telefon mora imati "pristup" do bazne stanice mobilnog telefona. Drugim riječima, radio signal s telefona na baznu stanicu mora biti neprekinut (Slika 4.). Brda, drveće i visoke građevine mogu zakloniti ovu liniju pristupa, tako da bazne stanice moraju biti vrlo pažljivo locirane kako bi se povećala dostupnost pokrivenosti. Svakoj ćeliji dodijeljene su višestruke frekvencije ( $f_1 - f_6$ ) koje imaju odgovarajuće radio baze. Skupina frekvencija može se ponovno upotrijebiti u drugim stanicama, pod uvjetom da iste frekvencije nisu korištene u susjednim stanicama, jer bi to uzrokovalo smetnje na zajedničkim kanalima.

[12]

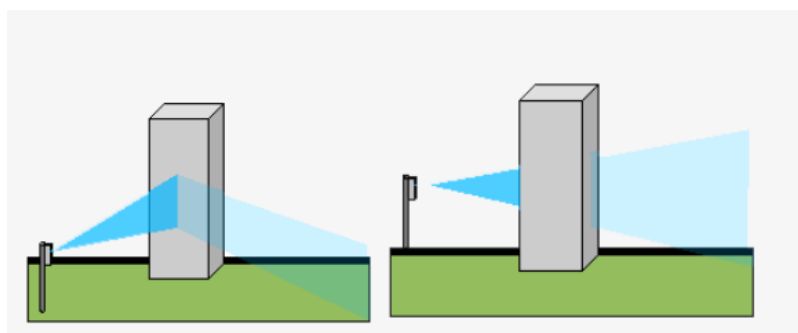


Slika 4. Ilustrativni prikaz pokrivenosti gradskog područja mrežom

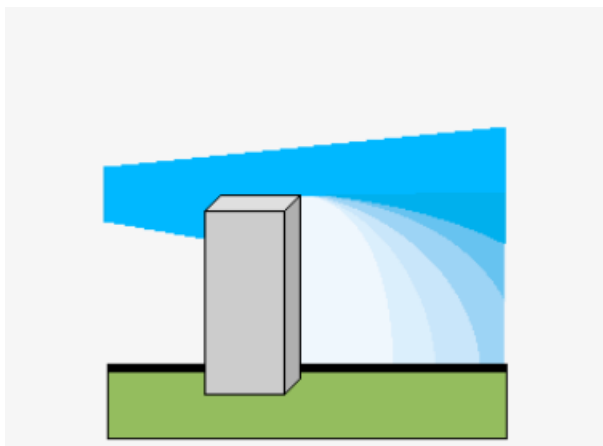
Stanična mreža je mreža ručnih mobilnih uređaja (mobitela) u kojoj svaki mobitel komunicira telefonskom mrežom radio valovima putem lokalne antene na staničnoj baznoj stanici.

Područje pokrivanja u kojem se pruža usluga podijeljeno je u mozaik malih geografskih područja koja se nazivaju "ćelije", a svako se od njih koristi zasebnim višekanalnim primopredajnikom i antenom na baznoj stanici. Svi mobiteli unutar stanice komuniciraju sa sustavom preko antene te stanice, na odvojenim frekvencijskim kanalima koje dodijeli bazna stanica iz zajedničkog baze frekvencija koje sustav koristi.

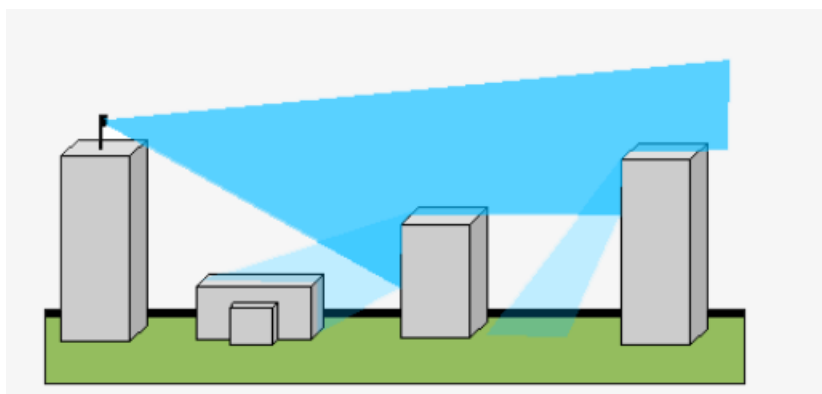
Mobitel se ponekad neće moći koristiti u svrhu komunikacije jer je predaleko od bazne stanice ili zato što je telefon na mjestu gdje su mobilni signali prigušeni i ograničeni zidovima zgrada, brežuljcima ili drugim građevinama. Signalima nije potrebna jasna linija vida, ali veća radio smetnja degradirat će ili eliminirati prijem. Kad mnogi ljudi pokušavaju istovremeno koristiti baznu stanicu, npr. tijekom zastoja u prometu ili sportskog događaja, na zaslonu telefona pojavit će se signal, ali onemogućeno je pokretanje nove veze. Drugi ograničavajući faktor za mobilne uređaje je mogućnost slanja signala iz baterije s malim naponom na mjesto ćelije. Neki mobiteli djeluju bolje od ostalih pod niskom potrošnjom energije ili slabom baterijom, obično zbog mogućnosti slanja dobrog signala s telefona na baznu stanicu. [13]



Slika 5. Prikaz signalnih valova bazne stanice koje se dijelom reflektiraju od zgradu, a dijelom prolaze kroz nju



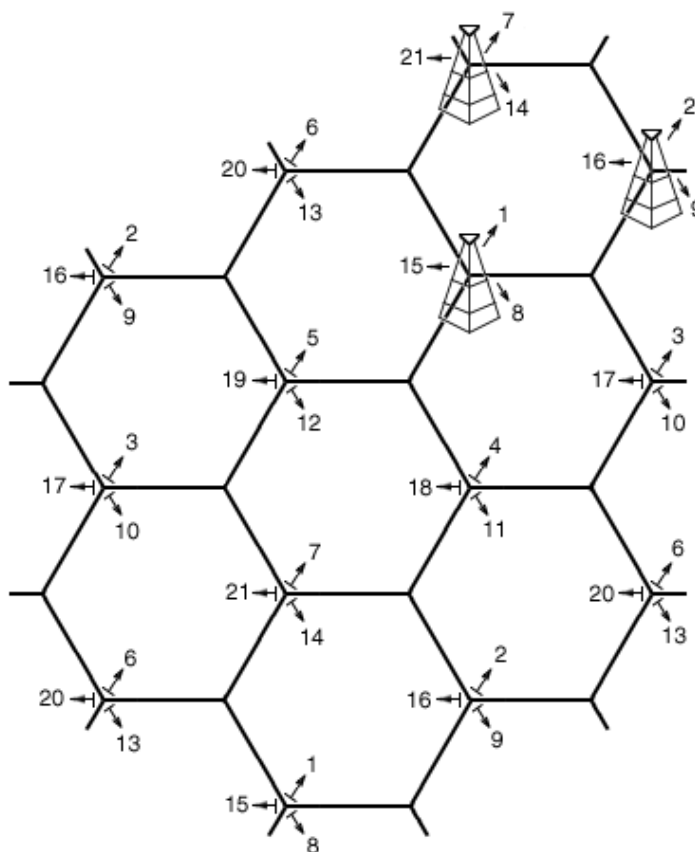
Slika 6. Primjer distrakcije signalnog vala



Slika 7. Prikaz moguće refleksije vala od više građevina

Svrha dobre organizacije baznih stanica je očuvanje radio opsega frekvencijskom upotrebom, radio signali male snage koji se koriste unutar svake ćelije ne putuju daleko izvan stanice, tako da se radio kanali mogu ponovo upotrijebiti u geografski razdvojenim ćelijama. Kada se mobilni korisnik prebaci iz jedne stanice u drugu, njegov se telefon automatski "predaje" anteni nove stanice i dodjeljuje mu novi skup frekvencija, a nakon toga komunicira s tom antenom. Ovaj postupak primopredaje pregledan je za korisnika i može se odvijati usred telefonskog poziva bez ikakvog prekida usluge. Svaki mobitel ima automatizirani dupleksni digitalni primopredajnik i komunicira sa staničnom antenom preko dva digitalna radio kanala.

Tipično se bazna stanica nalazi na rubu jedne ili više ćelija i prekriva više ćelija koristeći usmjerene antene. Uobičajena geometrija je locirati stanično mjesto na sjecištu triju susjednih ćelija, s tri antene pod kutom  $120^\circ$ , a svaka pokriva jednu ćeliju. Vrsta antene koja se koristi za bazne stanice (vertikalni bijeli pravokutnici na *Slici 3.*), naziva sektorska antena, obično se sastoji od vertikalnog kolinearnog niza dipola. Ima plosnati oblik zračenja u obliku ventilatora, koji je blago nagnut prema dolje kako bi prekrivao stanično područje bez zračenja pod većim kutovima u daljnje ćelije koje ponovno koriste iste frekvencije. Visinski kut antene mora biti pažljivo podešen, tako da snop pokriva cijelu ćeliju bez zračenja predaleko. U modernim sektorskim antenama nagib snopa obično se može podesiti elektroničkim putem, kako bi se izbjegla potreba da se tehničar penje na toranj i mehanički nagnje antenu kada je potrebno podešavanje.



Slika 8. Skica tornjeva baznih stanica i relacija prikazani šesterokutom

Iako su izvorne bazne stanice (tornjevi) stvorile ravnomjeren, višesmjerni signal, bazno stanična karta se može nacrtati pomoću baza koji se nalaze u kutovima šesterokuta gdje se konvergiraju tri stanice. Svaki bazna stanica ima tri seta antena usmjerenih u tri različita smjera sa  $120^\circ$  za svaku ćeliju (ukupno  $360^\circ$ ) i primanje / prijenos u tri različite stanice na različitim frekvencijama. To osigurava najmanje tri kanala i tri baze za svaku ćeliju i uvelike povećava šanse primanja upotrebljivog signala iz najmanje jednog smjera.

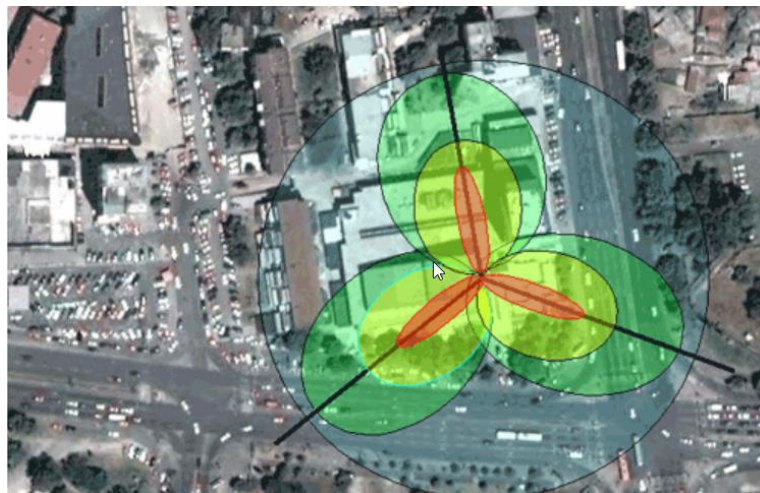
U mrežama globalnog sustava za mobilnu komunikaciju (GSM), točan je naziv za bazu primopredajnika (BTS), a kolokvijalni sinonimi su "toranj mobilnog telefona" ili "bazna stanica". Više mobilnih davatelja često štedi novac montirajući svoje antene na zajednički toranj. Budući da zasebni sustavi koriste različite frekvencije, antene se mogu nalaziti blizu jedna bez druge. Neke tvrtke pružatelji usluga (mobilni operateri) imaju više mobilnih mreža i na sličan način koriste obojene bazne stanice za dvije ili više mobilnih mreža (npr. 3G i 4G).  
[12]

### 2.1.1. Domet ćelija

Radni raspon ćelijskog mjesta (raspon u kojem se mobilni uređaji pouzdano povezuju s baznom stanicom) ovisi će o nizu faktora, a neki od njih su:

- visina antene nad okolnim terenom,
- učestalost signala koji se koristi,
- nazivna snaga odašiljača,
- potrebna brzina podataka veze pretplatničkog uređaja, odnosno korisnika,
- karakteristike usmjerenja nizova antene na mjestu,
- odbijanje i apsorpcija radio energije od zgrade ili vegetacije.

Također, domet može biti ograničen lokalnim geografskim ili regulatornim čimbenicima i vremenskim uvjetima.



Slika 9. Pokrivenost područja jedne bazne stanice sa tri sektorske antene

Uz to postoje ograničenja vremena u nekim tehnologijama (npr. u slobodnom prostoru GSM bi bio ograničen na 35 km iako je mogući domet čak 70 km uz posebnu opremu). Koncept "maksimalnog" dometa ćelije ipak je u praksi teško ostvariv. U gradovima, svaka stanica može imati domet do otprilike 0.8 km - 1,5 km, dok u ruralnim područjima raspon može biti i do 12



km. Moguće je da korisnik na jasnim otvorenim površinama može primati signale od mjesta na kojem je udaljen čak 40 km od bazne stanice.

Svakako je nužno osigurati dovoljno preklapanja ćelija za "primopredaju", tj. prebacivanje signala za mobilni uređaj s jedne bazne na drugu baznu stanicu, kao što je prijenos glasa putem GSM telefonskog poziva dok se korisnik kreće u automobilu ili vlaku. Potrebno područje preklapanja ne treba biti veliko kako bi se osigurala nesmetana „primopredaja“ uređaja.

U praksi se ćelije grupiraju u područjima velike gustoće stanovništva s najviše potencijalnih korisnika. Promet mobitela kroz jednu stranicu ograničen je kapacitetom bazne stanice, te postoji ograničen broj poziva ili podatkovnog prometa koji bazna stanica može istovremeno podnijeti. U prigradskim područjima bazne stanice su obično udaljene 2–3 km, a u gustim gradskim područjima bazne mogu biti udaljene čak 400–800 m.

Bazne stanice mogu podržati brojne razgovore s ograničenim brojem radio kanala (odsječci radio frekvencijskog spektra koji su potrebni za obavljanje jednog razgovora) licencirano operatera mobilnih usluga. Da bi se prevladalo ovo ograničenje, potrebno je ponoviti i ponovno koristiti iste kanale na različitim lokacijama. Kao što se autoradio mijenja s jedne lokalne postaje u potpuno drugu lokalnu postaju s istom frekvencijom kada putujete u drugi grad, isti se radio kanal ponovo koristi na baznoj stanici samo nekoliko kilometara dalje. Da bi to bilo moguće, signal ćelije namjerno se emitira malom snagom i u mnogim slučajevima je nagnut prema podu kako bi se ograničio njegov domet. To omogućava pokrivanje područja dovoljno malog da ne mora podržavati više razgovora nego što to mogu izdržati kanali. Zbog sektoriziranog rasporeda antena na baznoj stanici, moguće je mijenjati snagu i kut za svaki sektor, ovisno o pokrivenosti s drugim ćelijama na tom području. To su samo neki od parametara kojima operateri kreiraju kvalitetnu i visoko pristupačnu mobilnu mrežu. [13]

## 2.2. Mobilne tehnologije

"G" označi "generaciju" mobilne tehnologije. Svaka generacija je brža, sigurnija i pouzdanija. Faktor pouzdanosti je najteža prepreka koju je potrebno savladati. No važnije, svaka je generacija definirana kao skup standarda telefonske mreže, koji detaljno opisuju tehnološku implementaciju određenog sustava mobilne telekomunikacije. Brzina se povećava i tehnologija koja se koristi za postizanje te brzine također se mijenja. Na primjer, 1G nudi brzinu do 2,4 kbps, 2G do 64 kbps i temelji se na GSM-u, 3G do 144 kbps - 2 Mbps, dok 4G nudi brzinu prijenosa podataka do 100 Mbps - 1 Gbps i temelji se na LTE (eng. *long term evolution*) tehnologiji. Koju tehnologiju uređaj koristi u određenom trenutku moguće je vidjeti dok je uređaj spojen na internet, jer brzina interneta ovisi o jačini signala koja je prikazana kao 2G, 3G, 4G itd. tik do trake sa signalima na početnom zaslonu mobilnog uređaja, a to ujedno i označava korištenu tehnologiju. [3]

Cilj bežične komunikacije je pružiti visokokvalitetnu, pouzdanu komunikaciju poput žičane komunikacije, a svaka nova generacija usluga predstavlja veliki korak u tom smjeru.

Ovo evolucijsko putovanje započelo je 1979. godine iz 1G-a i još uvijek traje do 5G. Svaka od generacija ima standarde koji se moraju ispuniti kako bi službeno koristili G terminologiju. Postoje institucije zadužene za standardizaciju svake generacije mobilne tehnologije. Svaka generacija ima zahtjeve koji određuju parametre kao što su propusnost ili kašnjenje koje tehnologija treba ispuniti da bi se smatrala dijelom te generacije. Svaka je generacija izgrađena na istraživanju i razvoju koji se dogodio od posljednje generacije. 1G nije korišten za identifikaciju bežične tehnologije sve dok 2G, ili druga generacija, nije zaživjela među korisnicima. To je bio veliki skok u tehnologiji kada su bežične mreže prešle s analognog signala na digitalni. 3G je omogućio brži prijenos podataka za multimedijску upotrebu (najmanje 200 kbps) i dugo je držao standard za bežične prijenose podataka. [5]

### **2.2.1. 1G - prva generacija**

To je bila prva generacija tehnologije mobilne mreže. Jednostavni telefonski pozivi bili su sve što je tehnologija omogućavala. Prva generacija komercijalne mobilne mreže uvedena je krajem 1970-ih godina s potpuno implementiranim standardima koji su uspostavljeni tijekom 1980-ih. 1G je analogna tehnologija i telefoni su općenito imali slabi vijek trajanja baterije. Kvaliteta glasa bila je velika, no bez sigurnosti zadržavanja poziva, a nerijetko bi pozivi bili prekinuti. Radi se o telekomunikacijskim standardima koji su se koristili sve do zamjene digitalnim telekomunikacijama unutar 2G.

Prednosti ove tehnologije su komunikacijski sustavi s potpunim pristupom, no uz spomenute nedostatke, 1G tehnologija nije omogućavala korištenje „roaming“ sustava, postojao je ograničen broj korisnika i pokrivenost ćelija (znatno manji nego u tehnologijama viših generacija), mala je sigurnost korisnika u smislu da se pozivi lako detektiraju, mobilni uređaji su bili preveliki i s malim kapacitetom baterija. [5]

Svojstva 1G tehnologije:

- frekvencija 800 MHz i 900 MHz,
- širina pojasa: 10 MHz,
- tehnologija: analogno prebacivanje,
- usluga: samo glas,
- modulacija: frekvencijska modulacija (FM).

### **2.2.2. 2G - druga generacija**

GSM (Globalni sustav za mobilne komunikacije) je standard koji je razvijen od strane Europskog instituta za telekomunikacijske standarde (ETSI) kako bi opisao protokole za drugu generaciju (2G) mreža koje koriste mobilni uređaji poput mobilnih telefona i tableta. Druga generacija prvi put je razmještena u Finskoj u prosincu 1991. Od 2014. godine postala je globalni standard za mobilne komunikacije - s preko 90% tržišnog udjela, a djeluje u više od 193 zemlje. GSM tehnologija kasnije je postala osnovni standard za daljnji razvoj bežičnih

standarda. Ovaj je standard mogao podržavati do 14,4 do 64 kbps (maksimalnu) brzinu prijenosa podataka koja je dovoljna za SMS i e-mail usluge.

Prva velika promjena jest nadogradnja s 1G na 2G. Glavna razlika između dvaju mobilnih telefonskih sustava (1G i 2G) je u tome što su radio signali koje koristi 1G mreža analogni, dok su 2G mreže digitalne. Glavni motiv ove generacije bio je osigurati siguran i pouzdan komunikacijski kanal. Implementirao je koncept CDMA (eng. *code-division multiple access*) i GSM. Pružena je mala podatkovna usluga kao što su SMS i MMS. Napredak u tehnologiji od 1G do 2G uveo je mnoge temeljne usluge koje i danas koristimo, kao što su SMS, roaming, konferencijski pozivi, zadržavanje poziva i naplata na temelju usluga, npr. naknade temeljene na međugradskim pozivima i naplata u stvarnom vremenu. Maksimalna brzina 2G-a s GPRS-om je 50 kbps ili 1 Mbps s poboljšanom brzinom prijenosa podataka za GSM evoluciju EDGE (Enhanced Data Rate for GSM Evolution ili EGPRS). Prije nego što je napravio veliki skok od 2G do 3G bežičnih mreža, manje poznati 2.5G i 2.75G bili su privremeni standardi. Kako bi se podržala veća brzina prijenosa podataka, uvedena je i uspješno implementirana GPRS (General Packet Radio Service). GPRS je imao mogućnost prijenosa podataka do 171 kbps (maksimalno). EDGE (Enhanced Data GSM Evolution) je također razvijen za poboljšanje brzine prijenosa podataka za GSM mreže. EDGE je bio sposoban podržati do 473.6 kbps (maksimalno). [8]

Svojstva 2G tehnologije:

- digitalni sustav,
- omogućene SMS usluge,
- omogućen roaming,
- šifrirani prijenos glasa,
- prvi internet na nižim brzinama prijenosa podataka,
- ograničen broj korisnika i mogućnost infrastrukture,
- ograničena mobilnost korisnika.

### **2.2.3. 3G - treća generacija**

Univerzalni sustav mobilnih telekomunikacija (UMTS) je mobilni sustav treće generacije za mreže temeljene na GSM standardu, te predstavlja standarde mreže i usluga na koji su korisnici i danas naviknuti. Pretraživanje weba, e-pošta, preuzimanje videozapisa, dijeljenje slika i druge tehnologije pametnih telefona uvedeni su u treću generaciju. Komercijalno je uvedena 2001. godine. Ciljevi postavljeni pred mobilnu komunikaciju treće generacije bili su olakšavanje većeg kapaciteta glasa i podataka, podrška širokom rasponu aplikacija i povećanje prijenosa podataka uz niže troškove.

Ova mreža kombinira aspekte 2G mreže s nekim novim tehnologijama i protokolima kako bi pružila znatno višu brzinu prijenosa podataka.

Svojstva 3G tehnologije:

- veća brzina prijenosa podataka,
- omogućen video poziv,
- viša sigurnost korisnika,
- veći broj korisnika,
- veća pokrivenost područja,
- podrška za mobilne aplikacije,
- podrška za multimedijske poruke,
- lakše i brže pregledavanje internet stranica,
- TV streaming.

UMTS (Univerzalni sustav mobilnih telekomunikacija) koristi tehnologiju radijskog pristupa za širokopojasni pristup s višestrukim pristupom (WCDMA) radi pružanja veće učinkovitosti i propusnosti za mobilne mrežne operatore. UMTS specificira cjeloviti mrežni sustav koji uključuje radio pristupnu mrežu (UMTS - zemaljska radijska pristupna mreža ili UTRAN), jezgrenu mrežu (Mobile Application Part ili MAP) i provjeru autentičnosti korisnika putem SIM kartica (modul identiteta pretplatnika). Tehnologija opisana u UMTS-u ponekad se naziva i sloboda pristupa multimedijским sadržajima (FOMA) ili 3GSM.

Za razliku od EDGE (IMT Single-Carrier, na temelju GSM-a) i CDMA2000 (IMT Multi-Carrier), UMTS zahtijeva nove bazne stanice i nove raspodjele frekvencija, što je iziskivalo određene troškove. Samim time su i mobilni uređaji bili skuplji.

3G podržava konvencionalne mobilne glasovne, tekstualne i MMS usluge, ali također može prenositi podatke velikom brzinom, omogućujući mobilnim operaterima da isporuče aplikacije, uključujući streaming i širokopojasni pristup internetu. [10]

3G ili mreže treće generacije rade na drugačiji način od 2G mreža. Kada se poziv uspostavi na 2G, linija se drži otvorenom za razgovor korisnika tijekom cijelog trajanja poziva. S 3G mrežama, podaci poslani preko njih podijeljeni su u male „pakete podataka“ koji se ponovno sastavljaju u ispravnom redosljedju na kraju primitka. Ovo „pametno“ kodiranje znači da se može poslati više podataka i da se šalje učinkovitije i brže. Osim toga, 3G telefoni mogu biti u kontaktu s više od jedne bazne stanice u isto vrijeme i to osigurava bolje performanse u kvaliteti glasa i brzinama prijenosa podataka, odnosno više istovremenih poziva se može dogoditi u istom frekvencijskom području.

3.5G do 3.75G sustavi razvijeni su kako bi se povećala brzina prijenosa podataka u postojećim 3G mrežama, uvedena su još dva tehnološka poboljšanja u mreži. HSDPA (eng. *High Speed Downlink*) paketni pristup i HSUPA (eng. *High Speed Uplink Packet Access*), razvijen je i implementiran u 3G mreže. 3.5G mreža može podržavati brzinu prijenosa podataka do 2 Mbps. 3.75G sustav je poboljšana verzija 3G mreže s HSPA + High Speed paketnim pristupom. Kasnije će se ovaj sustav razviti u snažniji 3.9G sustav poznat kao LTE (*Long Term Evolution*), koji danas smatramo 4G tehnologijom. [9]

#### **2.2.4. 4G – četvrta generacija**

4G je vrlo različita tehnologija u usporedbi s 3G-om i omogućena je praktično samo zahvaljujući napretku tehnologije u posljednjih 10 godina. Ideja iza razvoja 4G bila je pružiti korisnicima veliku brzinu, visoku kvalitetu i visoki kapacitet, a istovremeno povećati sigurnost i smanjiti troškove glasovnih i podatkovnih usluga, multimedije i interneta putem IP-a. Potencijalne i trenutne aplikacije uključuju izmijenjeni pristup mobilnom webu, usluge igranja, mobilni TV visoke razlučivosti, video konferencije, 3D televiziju, itd.

Dva važna 4G standarda su WiMAX (do danas se prestao koristiti) i LTE (proširena s implementacijom). LTE je serija nadogradnji postojeće UMTS tehnologije uvedena na postojeći

Telstra frekvencijski pojas od 1800 MHz. Maksimalna brzina 4G mreže kada se uređaj kreće je 100 Mbps ili 1 Gbps za komunikaciju niske mobilnosti neovisno o kretanju korisnika, latencija smanjena s oko 300 ms na manje od 100 ms i postignuta znatno niža zagušenja. Također, potrebno je dijeliti mrežne resurse kako bi podržao više istodobnih veza na ćeliji. Kako se razvija, 4G bi mogao nadmašiti brzinu prosječne bežične širokopojasne kućne internetske veze. Nekoliko uređaja bilo je sposobno za punu brzinu kada je tehnologija prvi put izdana. Stvarna pokrivenost 4G u praksi bila je ograničena na velika gradska područja. Izvan pokrivenih područja, 4G telefoni su se povukli prema 3G standardima. Kada je 4G prvi put postao dostupan, jednostavno je bio malo brži od 3G. 4G nije isto što i 4G LTE koji je vrlo blizu ispunjavanju navedenih kriterija standarda za 4G tehnologiju.

Simultani prijenos glasa i podataka moguć je uz LTE sustav koji značajno poboljšava brzinu prijenosa podataka. Sve usluge, uključujući glasovne usluge, mogu se prenijeti putem IP paketa (paket podataka).

Poput 2G-a, 3G se razvio u 3.5G i 3.75G jer je uvedeno više značajki kako bi se donijela 4G tehnologija. 3G telefon ne može komunicirati putem 4G mreže, ali novije generacije telefona praktički su uvijek dizajnirane da budu kompatibilne unatrag, tako da 4G telefon može komunicirati putem 3G i 2G ako je potrebno. LTE napredna bežična tehnologija koja se koristi u sustavima 4G ima kompatibilnost s prethodnom verzijom, tako da je moguća lakša implementacija i nadogradnja LTE naprednih mreža u već postojeću infrastrukturu. Ipak široka primjena i nadogradnja oduzimaju mnogo vremena i novaca. Također, zakup frekvencijskog spektra u državama diljem svijeta je skup. [11][12]

#### Svojstva 4G tehnologije:

- veća brzina prijenosa podataka, do 1Gbps
- poboljšana sigurnost i mobilnost
- VoLTE putem LTE mreže (koristite IP paketa za glas)
- prijenos uživo (eng. *streaming*) videozapisa visoke kvalitete
- omogućen video poziv
- podrška za mobilne aplikacije
- podrška za multimedijske poruke

## 2.3. Key Performance Indicators - KPI

KPI-evi se smatraju ključnim pokazateljima mjerenja koja se koriste za identifikaciju i klasificiranje poslovnih rezultata, rada nekog sustava ili pogona. Iako se u ovom radu ne koriste direktno KPI podaci baznih stanica i cijele mreže, važno je objasniti iste radi lakšeg razumijevanja i kontrole rada bazne stanice, jer oni u konačnici pokazuju mogu li parametri kojima su definirane bazne stanice dati zadovoljavajuću kvalitetu mreže krajnjem korisniku.

Da bi se identificirali i uspostavili ključni KPI-i za određen sustav, proces koji se opisuje mora biti kreiran da zadovolji sljedeće zahtjeve:

- mora sadržavati jasne ciljeve
- proces mora biti mjerljiv, kvantitativno i kvalitativno
- mora biti moguće identificirati i riješiti organizacijska odstupanja

Ideja uvođenja KPI-eva u praćenje rada određenih procesa javlja se potrebnom da se isti lakše prate, kontroliraju i mijenjaju. Ono što se ne može izmjeriti i usporediti, ne može se niti poboljšati. Iako je ovaj pojam suštinski poslovno orijentiran, ima široku primjenu u opisu rada i kvalitete mobilnih mreža unutar svih korištenih tehnologija.

KPI se koriste u mnogim različitim djelatnostima kako bi se odredila učinkovitost poduzeća. Izbor i organizacija tih mjerenja su u djelokrugu uprave. Određeni KPI-evi mogu biti pregledani od strane operativnog osoblja kako bi se napravila poboljšanja. Druge se dostavljaju poslovnim jedinicama radi obavljanja izvršnih radnji. KPI-evi se također koriste za provjeru da su ugovori s kupcima pravilno ispunjeni. U bilo kojoj organizaciji te formule i vrijednosti mogu biti prilagodljive i imaju različite oblike i namjene.

KPI mjerenja nude kvantificirane podatke koji će pomoći da se donesu odluke o eventualnim promjenama.

Mrežni pružatelji usluga mjere učinkovitost kako bi osigurali najbolju moguću uslugu. Koriste ključne pokazatelje uspješnosti (KPI). To su vrijednosti utvrđene za kvantificiranje specifičnih aspekata funkcionalne mreže. KPI izvješća također su učinkoviti alati za inženjerske odjele u projektiranju, planiranju i upravljanju kapacitetima postojećih i budućih mreža.

KPI-evima se smatraju skup izračunatih pokazatelja rada mreže, prema unaprijed definiranim formulama. Svi parametri rada mreže se mjere i određenim formulama u



konačnici prikazuju kvantitetu i kvalitetu same mreže na određenom području. Svakodnevnim praćenjem KPI-eva i njihovih promjena omogućen je efektivan rad na optimizaciji mreže.

KPI koji se prate za rad mobilnih mreža u osnovi se dijele u 5 grupa:

- *Accessibility*
- *Retainability*
- *Mobility*
- *Integrity*
- *Availability*

„*Accessibility*“ predstavlja sposobnost mreže da se uspostavi poziv iz perspektive korisnika.

„*Retainability*“ jest sposobnost sustava da zadrži poziv i da ga održava uz što manje smetnji i komplikacija.

„*Mobility*“ predstavlja sposobnost mreže da prebacuje pozive između frekvencija i ćelija (prijenos glasa ).

„*Integrity*“ jest skupina KPI-eva koji daju podatke o brzini mreže te sposobnosti mreže da prenosi paketne podatke.

„*Availability*“ je zadnja, ali moglo bi se reći najvažnija skupina KPI-eva. Pokazuje dostupnost mreže krajnjem korisniku u svakom trenutku. Teži se vrijednosti od 100%, no uvijek se dopušta odstupanje od 5%. Dakle, 95% bi bila donja granica tolerancije zadovoljavajuće dostupnosti mreže, u ovisnosti o korištenju mreže, geografskom položaju, protoku korisnika kroz područje dometa bazne stanice i sl.

Upravljanje mrežom je složen pothvat i pojavljuju se problemi. Mrežni upravitelji - optimizatori često postavljaju osnovicu za operacije. Mrežni dizajneri određuju optimalne vrijednosti i pragove učinkovitosti, koji se zatim integriraju u alate za upravljanje mrežom.

Važno je zadržati mrežu u radu bez poteškoća i smetnji paralelno ju pokušavajući i unaprijediti. Tu dolazimo do optimizacije mreže. Dugoročna profitabilnost ovisi o pružanju najbolje moguće usluge kupcu na dogovorenoj razini. Iako timovi za kontrolu kvalitete ne

mogu biti pod stresom stalnog nadzora mreže, često su zaduženi za dubinsku analizu kako bi odredili najbolji put do poboljšanja mreže i visoku učinkovitost iste.

Bitno je napomenuti da vrijednost KPI poprilično varira ako se uzme u obzir u kojoj je fazi mreža koju pratimo, tako će npr. u fazi podizanja sustava i mreže KPI vrlo vjerojatno biti loši dok se mreža adekvatno ne optimizira i dok se ne ustabili njezin rad. Također, promjena KPI na nekom području mreže može se očitovati u specifičnim situacijama poput Nove godine, nogometnih utakmica, koncerata. Tamo gdje se očekuje veliki broj ljudi u isto vrijeme, a mreža je konfigurirana za prosječan (manji) broj ljudi, očekuju se privremeno lošije vrijednosti. Situacijama koje se mogu predvidjeti pristupa se poboljšanjem kapaciteta mreže unaprijed.

[4]

### 3. STROJNO UČENJE I OBRADA PODATAKA

Strojno učenje grana je umjetne inteligencije koja se bavi oblikovanjem algoritama koji svoju učinkovitost poboljšava na temelju empirijskih podataka, odnosno skupina algoritama za automatsku obradu podataka. Strojno učenje tvori temelj današnje računalne znanosti. Obrada podataka metodama strojnog učenja rezultira prediktivnim modelom, no primjene su puno šire od same predikcije, tako da se strojno učenje koristi za bilo koje preslikavanje ulaznih i izlaznih podataka koje je preteško ručno unijeti ili za koje ne postoje jasno precizirana pravila koja bi se unijela, ili se pak ta pravila prečesto mijenjaju.

Interdisciplinarnost strojnog učenja vidljiva je iz istovremene primjene računalne znanosti, statistike i vjerojatnosti, optimizacije i upravljanja, numeričke matematike i linearne algebre, teorije informacija, filozofije, psihologije, neurobiologije, kognitivne znanosti i dr.

Strojno učenje jedno je od danas najaktivnijih i najuzbudljivijih područja računske znanosti, ponajviše zbog brojnih mogućnosti primjene koje se protežu od raspoznavanja uzoraka i dubinske analize podataka do robotike, računalnog vida, bioinformatike, računalne lingvistike, itd.

Upoznatost s metodama nadziranog i nenadziranog strojnog učenja te zadacima klasifikacije i regresije, razumijevanje generativnih, diskriminativnih, parametarskih i neparametarskih modela, razumijevanje teorijskih osnova ovih modela, ugrađenih pretpostavki, njihovih prednosti i nedostataka omogućuje sposobnost oblikovanja i implementacije sustava za klasifikaciju, regresiju ili grupiranje te provođenja vrednovanja takvog sustava. [15]

Računala su u prošlosti mogla raditi samo ono za što su bila programirana. Strojno učenje omogućava im da uče na sličan način kako to rade ljudi, tako da stroj prikuplja znanje bazirano na prošlom iskustvu. Umjesto da mu se stalno mora ažurirati softverski kod, on je, kako vrijeme prolazi, samostalno sposoban poboljšavati svoj rad.

Za razliku od statistike koja se obično primjenjuje na „manji“ skup podataka i gdje je uloga samog statističara znatno veća od uloge računala i pomoćnih alata, strojno učenje se zasniva na automatiziranom korištenju i otkrivanju pravilnosti u podacima. Cijeli proces obrade podataka znatno ovisi o računalu i pomoćnim alatima, te o karakteriziranju podataka

u smislu što je „naučljivo“ i pod kojim uvjetima. Cilj je izgraditi model koji objašnjava podatke i omogućuje predviđanje i zaključivanje.

Učenjem se smatra memoriranje podataka s istovremenim zaključivanjem. Može se reći da je to učenje stroja (programiranje računala) da optimizira neki kriterij uspješnosti temeljem podatkovnih primjera ili iz prethodnih iskustva. Raspolaže se modelom koji je definiran parametrima (slobodnim) i optimizacijom tih parametara model mora moći predvidjeti svojstva novih, do sada neviđenih podataka.

Metodu, odnosno algoritam strojnog učenja, čine funkcija vrijednosti slobodnih parametara, algoritam pretraživanja i optimiziranja te metoda procjene greške. Funkcija može biti linearna, polinom, skup pravila, stabla odluka ili neuronska mreža. Evaluacija će se dobiti metodom procjene greške, odnosno skalarnom funkcijom koja će kvalificirati kvalitetu naučenog modela, tj. mjeru greške.

Opća podjela strojnog učenja jest na nadzirano i nenadzirano učenje. Pod nadzirano učenje (eng. *supervised learning*) smatra se obrada podataka kojoj je cilj napraviti model koji će raditi predviđanja za neke nove podatke (npr. klasifikacija objekata, regresija, predviđanje prodaje, cijena ili rizika, itd.). Ovim metodama će se eksplicitna informacija o danim primjerima i vrijednost njihove ciljne varijable koristiti za dobivanje cilja. [15]

Nenadzirano učenje (eng. *unsupervised learning*) bez anotacije ili povratne informacije o kategorizaciji ulaznih podataka, grupira prema nekoj strukturnoj pravilnosti unutar podataka koja će se odrediti ovisno o metodi. Nema učitelja niti kritičara, stoga je potrebno pronaći pravilnosti prema ulaznim podacima. Metoda neće otkriti što podaci znače i koja im je primjena, samo će ih organizirati prema izračunatoj pravilnosti.

Većina razvoja i probijanja u strojnom učenju jest u domeni nenadziranog učenja.

### 3.1. Primjena strojnog učenja

Primjena je vrlo široka. Rješavaju se složeni problemi koje nije moguće riješiti na klasičan algoritamski način, kao što su raspoznavanje uzorka, računalni vid, obrada prirodnog jezika, raspoznavanje emocija ili govora, robotika, itd. Koristi se kod sustava koji se moraju dinamički mijenjati i prilagođavati korisničkim sučeljima ili kada je potrebno obraditi velike količine podataka i provjeriti ima li znanja u njima.

Učenje se primjenjuje na različite baze podataka u raznim granama djelovanja kako bi se otkrile nepravilnosti implicitno sadržane u podacima.

Najčešće primjene algoritama strojnog učenja su u programskim implementacijama koje nije moguće napraviti na klasičan način, otkrivanje znanja unutar skupova te primjena u programima kojima je potrebno osigurati dinamičnu promjenu i prilagođavati ih određenim uvjetima.

Primjeri upotrebe strojnog učenja su klasa pacijenta s visokim rizikom za carski rez ili analizirati rizik banke kod dodjele kredita građanima. Neki od svakodnevnih primjera s kojima se većina današnjice susreće:

- *Google* koristi strojno učenje da poboljša preciznost rezultata traženja,
- *Facebook* prikazuje postove ovisno o vašim interesima i prošlom ponašanju na društvenoj mreži,
- *Netflix* vam predlaže popis sadržaja koje stvara zahvaljujući strojnom učenju,
- „samovozeći“ automobili prate objekte iz okruženja i koriste te podatke da bi poboljšali svoje vozačke sposobnosti,
- digitalni pomoćnici koriste strojno učenje da bi unaprijedili tehnologiju prepoznavanja govora.

Zanimljivo je napomenuti da i zdravstvena industrija eksperimentira s velikim brojem primjena, uključujući predviđanje životnog vijeka, organiziranje pacijentovih podataka, čak i dijagnosticiranje određenih bolesti. Prikupljanjem podataka poput rendgenskih slika, genetskih profila i krvnih testova, pokušava se postići da računala daju bržu i sigurniju dijagnozu.

Poslovni svijet također koristi benefite strojnog učenja. Razna poduzeća se odlučuju za prikupljanje i obradu podataka u svrhu boljeg poslovanja i boljeg razumijevanja tržišta.

Strojno učenje posjeduje vrlo široku primjenu metoda (umjetne inteligencije) za rješavanje složenih problema na gotovo svim područjima ljudske djelatnosti. Toliko je postalo zanimljivo da neki jednostavno individualno eksperimentiraju s tom vrstom znanja samo kako bi preispitali vlastite dosege.

Osim očitih prednosti razvoja strojnog učenja, paralelno s vrlo brzim razvojem istog, postavljaju se razna pitanja etičke prirode. Ljudi su zabrinuti za svoja radna mjesta i brinu da će zbog strojnog učenja gubiti poslove. Takve promjene značajno ovise o industriji, ali potencijal za dramatičnu promjenu mnogih poslova, poput vozača, bankara, čak i određenih liječnika, zasigurno postoji.

Također, često se postavljaju pitanja moralnog karaktera u vezi strojnog učenja i naglašava se njegov potencijal da naruši ljudsku privatnost. Primjerice, ne bi bilo teško stvoriti program koji bi prikupljao bilo čije poruke, podatke i informacije čime se otvara mogućnost zlouporabe istih. Strojno učenje može biti vrlo moćan alat, no važno je imati na umu i njegove negativne posljedice i efekte. [15][16]

### 3.2. Grupiranje (eng. *clustering*)

Grupiranje je proces podjele populacije ili podatkovnih točaka u više skupina tako da su podatkovne točke u istim skupinama sličnije drugim točkama podataka u istoj skupini od onih u drugim skupinama. Jednostavnim riječima, cilj je razdvojiti skupine sa sličnim osobinama i dodijeliti ih klasterima. Ideja grupiranja jest stvaranja klastera podataka, tako da svaki pojedinačni klaster ima najbližnje točke. To je u osnovi zbirka predmeta na temelju sličnosti i različitosti među njima. Ovo je najpopularnija tehnika nenadziranog učenja. Može koristiti dovoljno jednostavne algoritme za lako shvaćanje i razumijevanje, ali samo grupiranje ima mnogo težinskih slojeva. Grupiranje nije ograničeno na jednostavne algoritme.

Grupiranje se može podijeliti u „tvrdo“ i „meko“. Razlika je u tome što se u tzv. „tvrdom“ grupiranju svaka podatkovna točka (podatak) nastoji dodijeliti nekoj grupi tako da joj potpuno pripada. Dok u „mekom“ grupiranju umjesto svrstavanja svake podatkovne točke u zaseban klaster, dodjeljuje se vjerojatnost da će određena točka pripasti određenom klasteru.

Ne postoje kriteriji za dobro grupiranje. Ovisno o korisniku, samoj prirodi podataka i očekivanim rezultatima, mogu se koristiti sve metode koji zadovoljavaju njihove potrebe. [18]

Metode grupiranja:

1. Metode temeljene na gustoći

Ove metode povezuju guste regije područja koja imaju neku sličnost i razlikuje se od druge guste regije prostora. Ova metoda ima dobru točnost i sposobnost spajanja dva klastera.

Primjeri: Prostorno grupiranje aplikacija zasnovanih na gustoći, Točke za određivanje radi identificiranja strukture klastera, itd.

2. Hijerarhijske metode

Klasteri formirani ovom metodom formiraju strukturu tipa stabla koja se temelji na hijerarhiji. Novi klasteri se formiraju koristeći prethodno formirani klaster.

Primjeri: grupiranje pomoću zastupnika, uravnoteženo iterativno smanjenje grupiranje i korištenje hijerarhije) itd.

### 3. Metode particioniranja

Ove metode dijele objekte na „k“ klastera i svaki dio čini jedan klaster. Spada pod „tvrde“ metode.

### 4. Metode temeljene na mreži

U ovoj se metodi prostor podataka formulira u konačni broj ćelija koje tvore mrežnu strukturu. Sve operacije klastera izvedene na ovim mrežama su brze i neovisne o broju objekata podataka.

Primjeri: Statistička informacijska mreža, valni klaster, CLIQUE (*eng. clustering in quest*), itd.

#### **3.2.1. K-means grupiranje**

Budući da je zadatak grupiranja subjektivan, sredstava koja se mogu koristiti za postizanje tog cilja je mnogo. Svaka metodologija slijedi drugačiji skup pravila za definiranje "sličnosti" među točkama podataka. Poznato je više od 100 algoritama grupiranja, no pojedini se algoritmi upotrebljavaju vrlo često u praksi. „K-means“ grupiranje je vrlo često korišten i relativno jednostavan algoritam strojnog učenja bez nadzora temeljen na daljini. Za razliku od tradicionalnih nadziranih algoritama strojnog učenja, K-means pokušava klasificirati podatke bez da se prethodno obučio s označenim podacima. Nakon što je algoritam pokrenut i grupe definirane, svi novi podaci mogu se dodijeliti najrelevantnijoj grupi. Podaci se mogu grupirati na temelju centroida, distribucije, gustoće, itd. K-means to čini pomoću centroida.

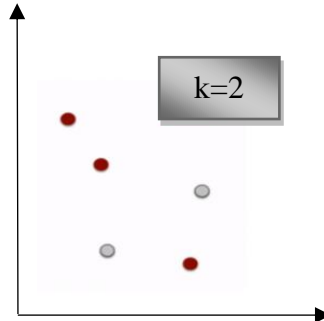
Centroidni modeli su iterativni algoritmi grupiranja u kojima je pojam sličnosti izveden iz blizine podatkovne točke u središtu klastera. K-means algoritam grupiranja popularan je algoritam koji spada u ovu kategoriju. U tim modelima, broj klastera potrebnih na kraju moraju biti prethodno spomenuti, zbog čega je važno imati prethodno znanje o skupu podataka. Ovi se modeli iterativno prikazuju kako bi pronašli lokalni optimumi.

Minimiziranjem sume kvadrata udaljenosti unutar klastera, K-means pokušava rasporediti „n“ broj mjerenja u „k“ broj klastera tako da se svako mjerenje usmjeri u klaster s najbližom srednjom vrijednošću. K-means je iterativni algoritam grupiranja kojem je cilj pronaći lokalne maksimume u svakoj iteraciji. [17]



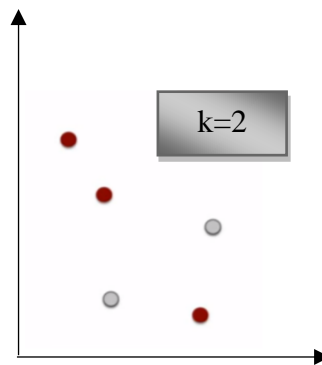
Provodi se kroz sljedećih 5 koraka:

1. Za određen broj podataka, potrebno je odabrati željeni broj klastera, odnosno grupa kao što je prikazano na *Slici 10*.



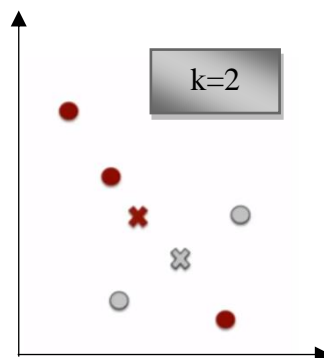
Slika 10. Prikaz „n“ broja podataka i željeni „k“ broj klastera

2. Nasumično svaku podatkovnu točku dodjeljuje jednom od klastera kao što je prikazano sa *Slici 11*.



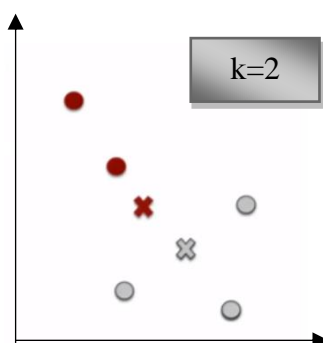
Slika 11. Prikaz nasumično dodijeljenih podatkovnih točaka klasteru

3. Izračunavanje centroida klastera. Oni su označeni „X“ na *Slici 12*.



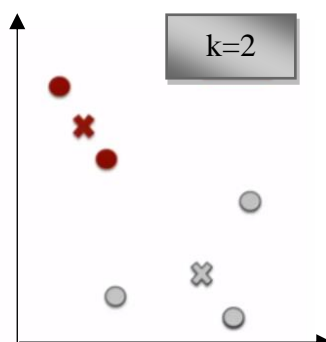
Slika 12. Prikaz izračunatih centroida klastera

4. Ponovo se vrši dodjela svake podatkovne točke, ali ovaj put isključivo najbližem, prethodno određenom središtu klastera.



Slika 13. Prikaz ponovljene podjele podatkovne točke najbližem od svih određenih centara klaster

5. Ponovo izračunavanje centroida za svaki od „k“ klastera

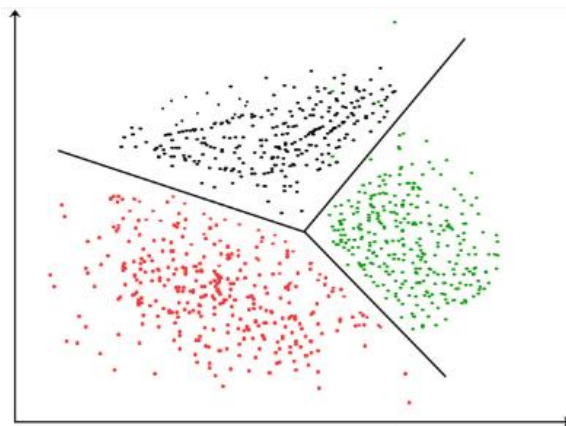


Slika 14. Prikaz ponovo izračunatih centroida klastera

Podatkovne točke dodjeljuju se najbližem centroidu i formira se klaster. Centroidi se zatim ažuriraju, a podatkovne točke preraspodjeljuju.

Zadnja dva koraka se ponavljaju dok se ne dođe do minimalnih promjena unutar kojih više poboljšanja nisu moguća, odnosno dok se postigne globalna optimala. Taj se proces nastavlja iterativno sve dok se lokacija centroida više ne promijeni. To je trenutak kada se više neće mijenjati položaj podatkovne točke između dva klastera za dva uzastopna ponavljanja.

Kreirani klasteri imaju tendenciju prema kružnom obliku radi korištenja srednje vrijednosti tijekom iteracije, što je potrebno uzimati u obzir pri odabiru ove metode grupiranja.[21]



Slika 15. Prikaz primjera grupiranja s većom količinom podataka

Neke praktične primjene K-means metode :

- profiliranje kupca - koristiti za karakterizaciju i otkrivanje korisničkih segmenata u marketinške svrhe,
- segmentacija tržišta,
- računalni vid,
- osiguranje: koristi se za priznavanje kupaca, njihovih namjera, identifikiranje prijevara i izračun rizika ulaganja,
- internet tražilice,
- astronomija,
- biologija: koristiti se za razvrstavanje među različite vrste biljaka i životinja,
- knjižnice: koristi se u grupiranju različitih knjiga na temelju tema i informacija,
- geografija: učenjem područja pogođenih potresom možemo utvrditi opasne zone,
- itd.

K-means konvergira u ograničenom broju iteracija. Budući da algoritam ponavlja funkciju čija je domena konačni skup, iteracijom se postiže konvergencija. U usporedbi s drugim metodama grupiranja, tehnika grupiranja K-means brza je i učinkovita s obzirom na njezine formule. Teško je predvidjeti optimalan broj klastera, tj. vrijednost  $k$ . Za određivanje broja klastera mora pokrenuti algoritam grupiranja za raspon  $k$  vrijednosti, osim ako se unaprijed ne zna željeni broj klastera.

Metoda je široko korištena za analizu klastera, lako ju je razumijeti i vrlo se brzo nauči koristiti. Međutim, izračunata udaljenost nije dovoljno precizna u mnogim praktičnim primjenama, metoda ne može konkurirati najboljim metodama grupiranja. [18]

### 3.2.2. Odabir odgovarajućeg broja klastera

Vrlo često podaci s kojima se radi imaju višestruke dimenzije koje znatno otežavaju preglednosti, a time i određivanje optimalnog broja klastera. Postoji više matematičkih metoda za određivanje broja klastera. Jedna od njih je kreiranje „WCSS“ grafa preko formule:

$$WSS = \sum_{n=1}^{\infty} (x_i + x_i)^2 \quad (F1)$$

WCSS se definira kao zbroj kvadratne udaljenosti između svakog člana klastera i njegovog centroida, a metoda je dobila ime po samom konačnom izgledu grafa.

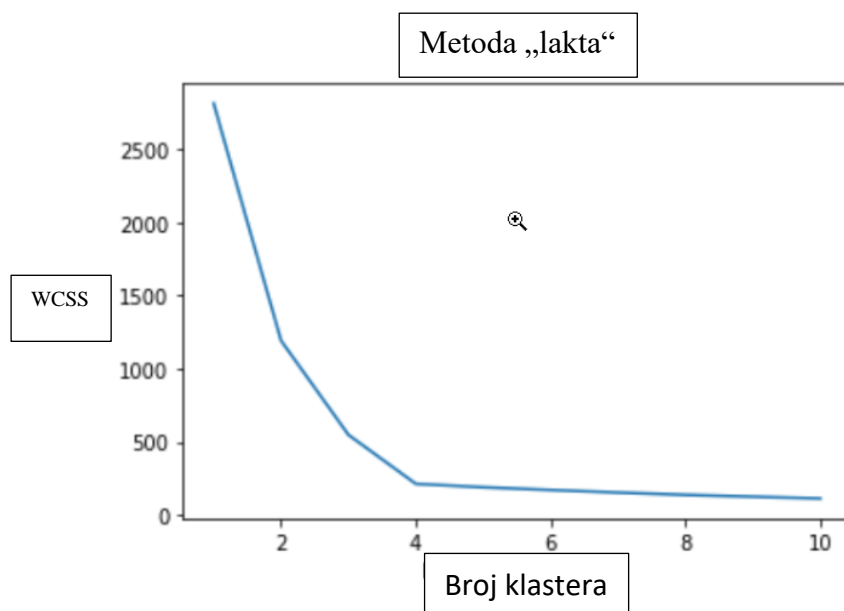
Za implementaciju WCSS grafa „metodom lakta“, potrebno je koristiti niže navedeni kôd u nekim od programskih jezika (Python):

```
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300,
n_init=10, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
```

```
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

WCSS graf prikazuje odnos između podatkovnih točaka i zbroja klastera unutar klastera (WCSS), zatim odabiremo broj klastera kod kojih promjena WCSS-a počinje izravnavati („metoda lakta“ – radi izgleda samog grafa). Nagla promjena na grafu koja ukazuje na optimalan broj klastera, formirat će se upravo na rednom broju točke zaključno s predloženim brojem klastera. U primjeru na *Slici 14.* izračunat je optimalan broj klastera  $k=3$ , sukladno ranije navedenom kôdu.

„Elbow metoda“, tj „metoda lakta“ promatra ukupni WSS kao funkciju broja klastera: potrebno je odabrati niz klastera tako da dodavanje drugog klastera ne poboljšava znatno ukupni WSS. Iz primjera na *Slici 14.* odgovarajući broj klastera bi bio 4. [18]



Slika 14. Primjer izračunavanja optimalnog broja klastera „metodom lakta“

Programski kôd K-means metode koja generira podatke funkcije “make\_blobs” iz modula „sklearn.datasets“, a parametar centra određuje broj klastera:

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.datasets.samples_generator import make_blobs
from sklearn.cluster import Kmeans
X, y = make_blobs(n_samples=300, centers=4, cluster_std=0.60,
random_state=0)plt.scatter(X[:,0], X[:,1])

wcss = []for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300,
n_init=10, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()

kmeans = KMeans(n_clusters=4, init='k-means++', max_iter=300,
n_init=10, random_state=0)
pred_y = kmeans.fit_predict(X)plt.scatter(X[:,0], X[:,1])
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:,
1], s=300, c='red')
plt.show()
```

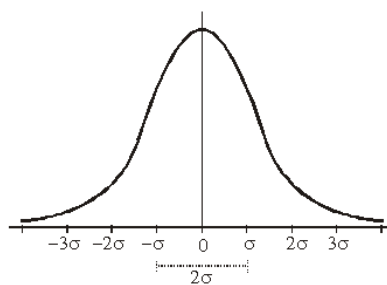
### 3.2.3. Gaussian Mixture metoda (GMM)

Gaussian Mixture model smatra se vrlo moćnim algoritmom grupiranja. GMM pretpostavlja da su sve podatkovne točke nastale iz mješavine konačnog broja Gaussovih distribucija s nepoznatim parametrima. Svaka od tih distribucija predstavlja tzv. „grozd“, stoga GMM nastoji zajedno grupirati podatkovne točke koje pripadaju jednoj distribuciji. GMM može se smatrati generaliziranjem K-means metode.

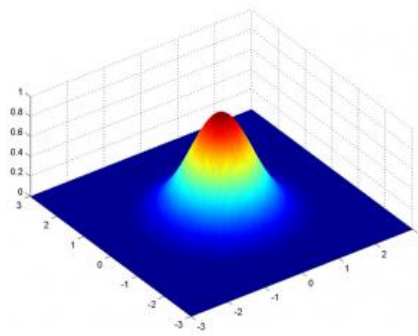
Ukoliko se pretpostavi da postoje tri Gaussove distribucije (GD1, GD2 i GD3) i da one imaju određenu srednju vrijednost ( $\mu_1$ ,  $\mu_2$ ,  $\mu_3$ ) i varijancu ( $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$ ), za dani skup podatkovnih točaka, GMM identificira vjerojatnost da svaka podatkovna točka pripada svakoj od tih distribucija. GMM su modeli temeljeni na vjerojatnosti i koriste postupke „mekog“ grupiranja raspodjelu točaka po klasterima. [24]

Na primjer, pretpostavka je da postoje tri skupine podataka označene trima bojama – plavom, zelenom i crvenom. Podatkovna točka označena crvenom bojom imat će vjerojatnost 1 da pripadne „crvenom“ klasteru, a za ostala dva klastera vjerojatnost će biti 0. No, ako podatkovna točka sadrži podatke koji ne pripadaju 100% niti jednom klasteru, kao npr. ljubičasta boja. Ta točka će imati vjerojatnost 0,5 za „plavi“ i 0,5 za „crveni“ klaster, a za „zeleni“ 0.

Gaussova distribucija može se shvatiti kao 3D krivulja u obliku zvona stvora prema uzoru na „normalnu Gaussovu razdiobu“, 2D krivulja u obliku zvona.



Slika 16. Krivulja normalne Gaussove razdiobe



Slika 17. 3D prikaz Gaussove krivulje

Funkcija gustoće vjerojatnosti jest:

$$f(x | \mu, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp \left[ -\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu) \right] \quad (F2)$$

gdje je  $x$  ulazni vektor,  $\mu$  je 2D srednji vektor, a  $\Sigma$  2 koverzijska matrica  $2 \times 2$ . Kovarijancija definira oblik ove krivulje. Na taj način moguće je generirati vektor za  $d$ -dimenzije. Dakle, GMM sadrži  $x$  i  $\mu$  kao vektore duljine  $d$ , a  $\Sigma$  bi bio kovarijantna matrica  $d \times d$ . Za skup podataka  $s$   $d$  značajkama GMM s „ $k$ “ brojem raspodjela, gdje je  $k$  ekvivalent broju klastera, a svaka ima određenu matricu srednjeg vektora i varijance. Srednji vektor i varijacija određuju se *Expectation-Maximization* (EM) metodom.

EM je statistički algoritam koji pronalazi ispravne parametre modela, a osnova je mnogih algoritama. Ova metoda obično se koristi kada nedostaju podaci ili su nepotpuni. Takve varijable koje nedostaju nazivaju se latentne (skrivenne) varijable. Teško je odrediti ispravne parametre modela kada postoje varijable koje nedostaju. Problematika leži u tome što latentnim varijablama nije poznata matrica srednjih vrijednosti i varijacije.

Algoritam EM sastoji se od 2 koraka, E-korak i M-korak. [24]

E-korak pomoću postojećih podataka određuje optimalne vrijednosti varijabli koje nedostaju, a M-korak na temelju procijenjenih vrijednosti generiranih u E-koraku kreira potpune odgovarajuće parametre modela.

Za svaku točki  $x_i$  potrebno je izračunati vrijednost pripadanja klaster/distribuciji  $c_1, c_2, c_3, \dots, c_k$  preko omjera vjerojatnosti da  $x_i$  pripada klasteru  $c_i$  i sume vjerojatnosti da  $x_i$  pripada svim klasterima.



Nakon toga, ažuriraju se vrijednosti  $\Pi$ ,  $\mu$  i  $\Sigma$ . Nova gustoća definira se omjerom broja dodijeljenih klastera podatkovnoj točki i ukupnog broja klastera, preko sljedeće formule:

$$\Pi = \frac{\text{Broj točaka dodijeljenih klasteru}}{\text{Ukupan broj klastera}} \quad (\text{F3})$$

Srednja vrijednost i matrica kovarijacije ažuriraju se na temelju vrijednosti dodijeljenih distribucijom, razmjerno s vrijednostima vjerojatnosti za tu podatkovnu točku. Dakle, podatkovna točka koja ima veću vjerojatnost da će biti dio te distribucije, bit će joj i dodijeljena.

Srednja vrijednost i matrica kovarijacije računaju se preko sljedećih formula:

$$\mu = \frac{1}{\text{Broj točaka dodijeljenih klasteru}} \sum_i r_{ic} x_i \quad (\text{F4})$$

$$\Sigma_{ic} = \frac{1}{\text{Broj točaka dodijeljenih klasteru}} \sum_i r_{ic} (x_i - \mu_c)^T (x_i - \mu_c) \quad (\text{F5})$$

Na temelju vrijednosti dobivenih iz opisanih koraka, izračunavaju se nove vjerojatnosti za svaku točku. Postupak se iterativno ponavlja i sukladno tome ažuriraju podatke. [25] [26]

*Sklearn.mixture* je programski paket koji omogućuje učenje i istovremeno korištenje kombinacija Gaussovih modela (podržane dijagonalne, sferne, vezane i pune kovarijantne matrice). Uzorkovanje i procjena dostupni su iz ulaznih podataka (podatkovnih točaka). Omogućeni su i objekti za određivanje odgovarajućeg broja klastera.

Primjer kôda:

```
import pandas as pd
data = pd.read_csv('Clustering_gmm.csv')

plt.figure(figsize=(7,7))
plt.scatter(data["Weight"],data["Height"])
plt.xlabel('Weight')
plt.ylabel('Height')
plt.title('Data Distribution')
plt.show()

# training gaussian mixture model
from sklearn.mixture import GaussianMixture
gmm = GaussianMixture(n_components=4)
gmm.fit(data)

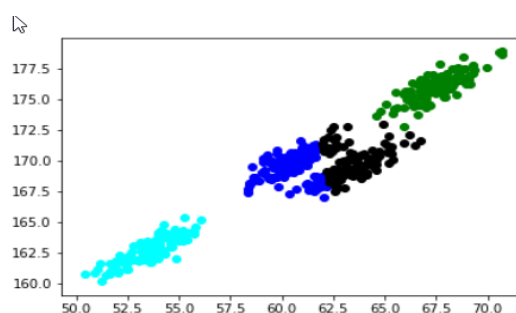
#predictions from gmm
labels = gmm.predict(data)
frame = pd.DataFrame(data)
frame['cluster'] = labels
frame.columns = ['Weight', 'Height', 'cluster']

color=['blue','green','cyan', 'black']
for k in range(0,4):
    data = frame[frame["cluster"]==k]
    plt.scatter(data["Weight"],data["Height"],c=color[k])
plt.show()
```

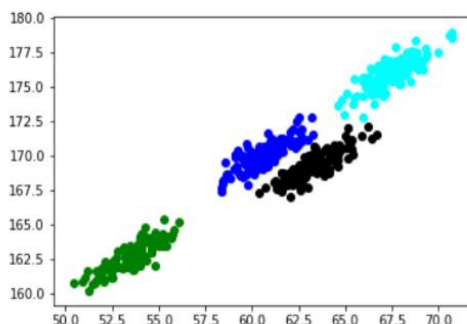
### 3.2.4. Usporedba K-means i GMM metode

Razlika između K-means i GMM metode je u tome što K-means uzima u obzir srednju vrijednost kako bi se izračunao centroid, a GMM uz srednju vrijednost uzima u obzir i varijacije podataka. Stoga je potreban drugačiji način dodjeljivanja klastera podatkovnim točkama. Umjesto korištenja modela temeljenog na daljini, GMM koristi model temeljen na distribuciji. [22]

Iz *Slike 18.* i *Slike 19.* vidljivo je kako se u praksi očituje razlika između ove dvije metode. K-means metodom dobiveni su klaster oblika elipse, iako za prikazani slučaj to nije praktično. GMM metodom došlo se do „točnije“ i prihvatljivije podjele u klastere oblika elipse, kako je vizualno i predvidljivo. Različite boje predstavljaju različitu raspodjelu točaka u rezultatne klastere, iako je ilustrativno vidljivo koja je grupacija „točnija“ u odnosu na sam položaj točaka. Za iste ulazne podatke, metode su dale različitu podjelu unutar klastera.



Slika 18. Primjer rezultata grupiranja K-means metodom



Slika 19. Primjer rezultata grupiranja GMM metodom

## 4. OPIS ZADATKA

Zadatak je identificirati „*mobility*“ profile ćelija u mreži. Ovisno o prirodi i potrebama kretanja korisnika unutar mreže, svaka ćelija može imati drugačije vrijednosti parametara koji su za nju najpovoljniji. Težnja tome da svaka ćelija bude efektivno optimizirana rezultira boljim performansama mreže, što svaki operater nastoji pružiti svojim korisnicima. Uz pretpostavku da postoji konačan broj profila, te da se isti ne mogu grupirati jednostavnim geoprostornom analizom, cilj projekta je određivanje „*mobility*“ profila pojedinih ćelija u mreži.

Potrebno je pronaći „*high-mobility*“ ćelije, odnosno ćelije koje pokrivaju područja brzog prolaska korisnika kroz nju na području između dva grada. Analizirat će se područje rada mreže između dva europska grada udaljena oko 160 km, povezana autoputom, brzom cestom, lokalnim cestama i prugom. S obzirom na to da je to područje raznolikih geografskih i demografskih svojstava, ne može se jednostavnim metodama odrediti koje ćelije su usmjerene k brzim cestama i pruzi. Gradovi su povezani brzom cestom i prugom, ideja je izdvojiti „*high-mobility*“ ćelije koje pokrivaju to područje kako bi se lakše optimizirale (promjena parametara u svrhu boljeg rada). Analiza uključuje 2099 ćelija.

Potreba za ovim pojavila se kada se ustanovilo da korisnicima na tom području kretanja vrlo često „puca“ veza, iz razloga što se oni jednostavno prebrzo kreću kroz mrežu i ćelije ne stignu nesmetano obaviti primopredaju korisnika. Korisnici se kreću brzinama između 100 km/h i 200 km/h na autoputu, a u vlakovima između 200 km/h i 300 km/h. Zbog ćelijskih parametara se kretanje korisnika može drugačije odraziti na primopredaju korisnika između ćelija (eng. *handover*). Ono što je bitno jest posljedica kretanja korisnika na količinu primopredaje, a ne samo kretanje istog.

Na temelju ulaznih podataka i provedbom dvije metode strojnog učenja pokušat će se dobiti grupacija svih ćelija u 4 klastera. Nema potrebe da se traži grupacija u više klastera, jer je poznata priroda podataka koji se analiziraju. Iz znanja o mobilnim mrežama, baznim stanicama i ćelijama te optimizaciji istih, 4 klastera će biti dovoljna za olakšavanje promjene parametara mreže u svrhu njezinog boljeg rada.

Skup podataka koji se analizira jest vrijeme (s) korisnika provedeno u nekoj ćeliji.

## 4.1. Obrada podataka

U Tablici 1. prikazan je dio ulazni podaci. U lijevom stupcu nalaze se imena ćelija u anonimiziranom prikazu, a u desnom stupcu vrijeme koje je korisnik proveo u ćeliji. S obzirom na to da se radi o 1 048 576 uzoraka, u tablici su reprezentativno prikazani samo neki od njih. Podaci su prikupljeni na radni dan u siječnju 2019. godine, između 6:00 h i 8:00 h.

Ćelija	Vrijeme (s)	Ćelija	Vrijeme (s)	Ćelija	Vrijeme (s)
cell-1	7.863	cell-1	2.249	cell-4	1.34
cell-2	11.923	cell-5	3.276	cell-1	11.884
cell-1	13.214	cell-1	2.089	cell-5	5.323
cell-2	7.133	cell-5	0.321	cell-4	2.117
cell-3	3.906	cell-0	0.679	cell-1	5.643
cell-1	2.35	cell-4	0.542	cell-2	2.869
cell-4	1.581	cell-1	2.026	cell-3	29.887
cell-1	5.379	cell-4	3.968	cell-1	13.617
cell-5	27.947	cell-1	15.579	cell-5	13.057
cell-3	5.932	cell-1	8.8	cell-0	9.476
cell-2	20.18	cell-1	1.462	cell-3	3.485
cell-4	2.264	cell-4	26.947	cell-5	3.423
cell-1	11.268	cell-5	0.697	cell-2	3.294
cell-2	1.968	cell-5	10.668	cell-4	14.122
cell-4	2.302	cell-5	4.243	cell-2	6.517
cell-5	3.004	cell-1	5.409	cell-0	1.116
cell-1	6.43	cell-5	4.492	cell-5	2.877
cell-5	12.383	cell-1	9.823	cell-2	8.194
cell-1	7.104	cell-0	3.763	cell-1	3.237
cell-4	7.488	cell-1	2.224	cell-2	1.535
cell-5	0.384	cell-3	1.787	cell-5	1.687
cell-2	11.837	cell-2	12.302	cell-1	3.214
cell-1	1.378	cell-1	24.242	cell-1	3.568
cell-2	2.123	cell-1	15.175	cell-0	4.448
cell-5	3.855	cell-5	2.847	cell-1	5.778
cell-1	15.558	cell-1	2.899	cell-0	1.933
cell-2	0.403	cell-3	3.841	cell-1	18.31
cell-2	9.659	cell-5	7.73	cell-0	1.355
cell-1	5.924	cell-1	3.121	cell-3	3.367
cell-1	1.056	cell-0	8.574	cell-1	1.321
cell-2	1.954	cell-1	1.499	cell-3	1.764
cell-2	2.553	cell-4	6.847	cell-0	4.219
cell-0	49.384	cell-1	1.544	cell-1	20.748
cell-1	2.399	cell-4	2.116	cell-1	5.473
cell-1	9.261	cell-5	27.029	cell-3	5.751
cell-1	10.718	cell-4	2.677	cell-1	4.35
cell-5	0.624	cell-5	1.932	cell-1	4.215
cell-1	1.447	cell-1	6.599	cell-4	10.758
cell-0	7.357	cell-5	4.81	cell-5	3.782
cell-0	2.5	cell-3	4.263	cell-3	3.596
cell-1	2.01	cell-5	11.781	cell-1	2.02
cell-0	1.205	cell-0	9.735	cell-4	1.599
cell-5	2.653	cell-5	1.604	cell-1	1.203
cell-5	20.536	cell-2	1.998	cell-3	2.57
cell-1	5.839	cell-1	1.049	cell-1	6.107

cell-5	5.094
cell-0	1.113
cell-1	9.424
cell-4	1.34
cell-1	11.884
cell-5	5.323
cell-4	2.117
cell-1	5.643
cell-2	2.869
cell-3	29.887
cell-1	13.617
cell-5	13.057
cell-0	9.476
cell-3	3.485
cell-5	3.423
cell-2	3.294
cell-4	14.122
cell-2	6.517
cell-0	1.116
cell-5	2.697
cell-4	7.221
cell-4	0.705
cell-1	11.114
cell-1	1.645
cell-0	2.229
cell-1	9.25
cell-3	5.223
cell-4	0.774
cell-5	0.257
cell-1	0.816
cell-3	6.723
cell-4	2.937
cell-5	7.159
cell-1	2.577
cell-4	0.673
cell-1	1.932
cell-1	7.473
cell-4	3.276
cell-2	1.631
cell-2	4.329
cell-5	0.97
cell-5	4.598
cell-1	2.941
cell-0	5.426
cell-0	6.284
cell-1	3.96
cell-4	12.775
cell-4	11.402

cell-2	8.364
cell-3	3.338
cell-5	2.877
cell-2	8.194
cell-1	3.237
cell-2	1.535
cell-5	1.687
cell-1	3.214
cell-1	3.568
cell-0	4.448
cell-1	5.778
cell-0	1.933
cell-1	18.31
cell-0	1.355
cell-3	3.367
cell-1	1.321
cell-3	1.764
cell-0	4.219
cell-1	20.748
cell-1	5.473
cell-3	5.751
cell-1	4.35
cell-1	4.215
cell-4	10.758
cell-5	3.782
cell-3	3.596
cell-1	2.02
cell-2	0.257
cell-4	1.599
cell-1	1.203
cell-3	4.238
cell-3	2.57
cell-1	6.107
cell-1	0.537
cell-5	0.322
cell-2	2.192
cell-1	4.786
cell-2	2.98
cell-1	6.832
cell-1	6.2
cell-4	0.901
cell-4	1.037
cell-5	9.966
cell-1	16.221
cell-5	3.765
cell-4	0.658
cell-1	3.134
cell-2	0.285

cell-1	0.537
cell-5	0.322
cell-1	0.342
cell-1	21.953
cell-1	4.474
cell-1	2.486
cell-4	7.824
cell-5	3.161
cell-1	12.284
cell-1	0.622
cell-1	1.685
cell-0	3.419
cell-5	2.063
cell-2	0.6
cell-1	3.526
cell-2	5.703
cell-0	7.993
cell-1	1.692
cell-4	3.333
cell-4	0.879
cell-1	5.745
cell-4	3.07
cell-1	6.666
cell-4	1.225
cell-2	4.057
cell-3	6.027
cell-0	5.996
cell-0	5.506
cell-5	2.056
cell-5	6.302
cell-0	30.246
cell-5	6.259
cell-0	2.331
cell-2	0.298
cell-5	7.04
cell-2	16.461
cell-4	0.718
cell-3	4.424
cell-0	1.744
cell-3	7.74
cell-1	4.569
cell-3	4.213
cell-4	12.597
cell-1	5.102
cell-1	11.335
cell-1	5.134
cell-5	19.59
cell-5	25.038

Tablica 1. Reprezentativni uzorak ulaznih podataka

Ulazni podaci prikazani su u Tablici 1.. Posljedica kratkog zadržavanja unutar područja pokrivenosti ćelije (iz nekog razloga), jest povećanje količine signalizacije u mreži. Stoga je prvi korak upravo određivanje razumnog raspona vremena. [Prilog I]

```

28 # compute geometric mean
29 # through formula antilog
30 # (((log(x)) + log(2)) + ...
31 # ... + log(n))/log(2)
32 # and return the value to
33 # main function:
34 sum = sum / n;
35
36 return math.exp(sum);
37
38 con = sqlite3.connect('C:/Users/ekikrsu/Desktop/Diplomski rad/ctr_mobility_v3_agg_WORKING_COPY.db')
39 cur = con.cursor()
40
41 cqi_group_query = '''
42     SELECT TIME_IN_CELL_s
43     FROM ctr_mobility
44     ...
45     ...
46
47 cur.execute(cqi_group_query)
48
49 data = []
50 for values in cur.fetchall():
51     if values[0] != '' and float(values[0]) < 26 :
52         data.append(float(values[0]))
53
54
55 for q in [50, 90, 95, 100]:
56     print("{}% percentile: {}".format(q, np.percentile(data, q)))
57
58 num_bins = 10
59 counts, bin_edges = np.histogram(data, bins=num_bins, normed=True)
60 cdf = np.cumsum(counts)
61 plt.plot(bin_edges[1:], cdf/cdf[-1])
62
63 plt.hist(data, bins=15)
64
65 ecdf = ECDF(data)
66 # get cumulative probability for values
67 print('P(x<2): %.3f' % ecdf(2))
68 print('P(x<3): %.3f' % ecdf(3))
69 print('P(x<5): %.3f' % ecdf(5))
70 print('P(x<10): %.3f' % ecdf(10))
71 print('P(x<15): %.3f' % ecdf(15))
72 print('P(x<20): %.3f' % ecdf(20))
73 print('P(x<40): %.3f' % ecdf(40))
74 print('P(x<60): %.3f' % ecdf(60))
75 print('P(x<80): %.3f' % ecdf(80))
76 print('P(x<100): %.3f' % ecdf(100))
77 # plot the cdf
78 plt.plot(ecdf.x, ecdf.y)
79
80 gmean = geometric_mean(data, len(data))
81 print("Geometric mean: ", gmean)
82
83
84
85 plt.show()
86

```

Python console output:

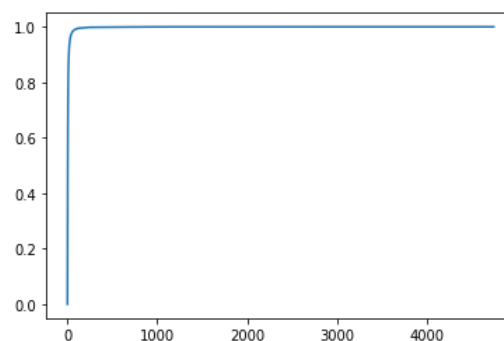
```

In [1]: runfile('C:/Users/ekikrsu/Desktop/Diplomski rad/Zadatak/metoda 1 - k-means/l.py', wdir='C:/Users/ekikrsu/Desktop/Diplomski rad/Zadatak/metoda 1 - k-means')
50% percentile: 3.823
90% percentile: 15.729
95% percentile: 26.993
100% percentile: 4740.855
P(x<2): 0.269; C:/Users/ekikrsu/Desktop/Diplomski rad/Zadatak/metoda 1 - k-means/l.py:68: VisibleDeprecationWarning: Passing normed=True on non-uniform bins has always been broken, and computes neither the probability density function nor the probability mass function. The result is only correct if the bins are uniform, when density=True will produce the same result anyway. The argument will be removed in a future version of numpy.
counts, bin_edges = np.histogram(data, bins=num_bins, normed=True)
P(x<3): 0.428
P(x<5): 0.595
P(x<10): 0.823
P(x<15): 0.894
P(x<20): 0.926
P(x<40): 0.971
P(x<60): 0.984
P(x<80): 0.990
P(x<100): 0.994
Geometric mean: 3.8560979751583453

```

Slika 20. Određivanje raspona vremena

Statistika ulaznih podataka pokazala je da se više od 92% korisnika kretalo u ćeliji manje od 20 sekundi, što je vidljivo u Slici 21. [Prilog I]

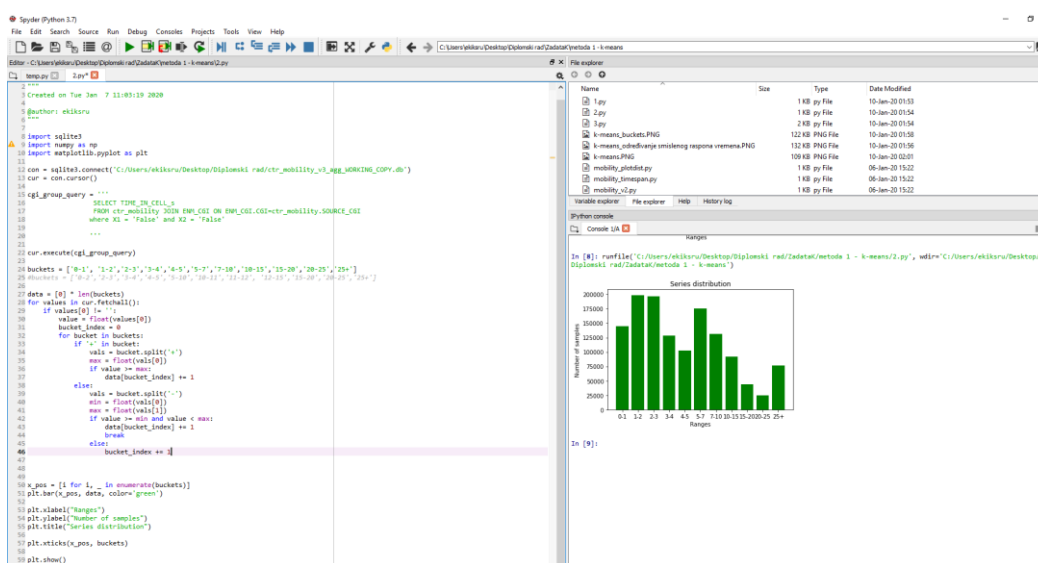


Slika 21. CDF graf

Iz *Slike 21.* vidi se da u području pokrivanja ćelije postoje korisnici koji se u njoj nalaze po 3000, 4000 sekundi. Zaključak je da se radi o korisnicima koji se ne kreću unutar istraživanog vremena.

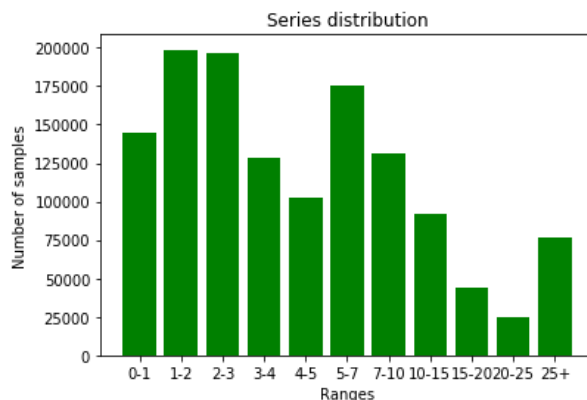
S obzirom na to da je 20 sekundi vrijeme koje većina korisnika provedu na području određene ćelije, raspon podataka odabran je logički, s obzirom na poznavanje prirode istih, a on je:

- ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-25', '25+']



Slika 22. Provjera smislenosti zadanog raspona

Na *Slici 23.* vidljivo je na koji način uzorci padaju u rasponima [Prilog II].



Slika 23. Raspodjela korisnika unutar zadanih vremenskih raspona



Iz histograma je vidljivo kako se najveći broj uzoraka ( u konkretnom zadatku – korisnika) kreće u rasponu od 1-2, 2-3 i 5-7 sekundi. Određivanjem logičnog raspona i potvrdom da je većina korisnika unutar njega, pristupa se trećem koraku analize.

## 4.2. Analiza K-means metodom

S obzirom na poznavanje opisa i značenja ulaznih podataka, unaprijed će se odabrati broj klastera. Četiri klastera su dovoljna grupacija prema zahtjevima optimizacijskog tima. [Prilog III]

```

1  """
2  Created on Tue Jan 7 11:23:27 2020
3
4  @author: ekikrsu
5  """
6
7  import sqlite3
8  import numpy as np
9  from sklearn.cluster import KMeans
10
11 con = sqlite3.connect('C:/Users/ekikrsu/Desktop/Diplomski rad/ctr_mobility_v3_egg_MORKING_COPY.db')
12 cur = con.cursor()
13
14 cgi_group_query = """
15     SELECT SOURCE CGI, group_concat(TIME_IN_CELL_s)
16     FROM ctr_mobility
17     GROUP BY SOURCE CGI
18 """
19
20 cur.execute(cgi_group_query)
21
22 buckets = ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-25', '25+']
23
24 data = []
25 cells = []
26
27 for cgi, values in cur.fetchall():
28     time_in_cell_for_cgi = values.split(',')
29     feature_vector = [] * len(buckets)
30     for current_time in time_in_cell_for_cgi:
31         if current_time == "":
32             continue
33
34         working_time = float(current_time)
35
36         bucket_index = 0
37         for bucket in buckets:
38             if working_time in bucket:
39                 vals = bucket.split(',')
40                 max = float(vals[0])
41                 feature_vector[bucket_index] += 1
42                 if working_time >= max:
43                     feature_vector[bucket_index] += 1
44             else:
45                 vals = bucket.split(',')
46                 min = float(vals[0])
47                 max = float(vals[1])
48                 if working_time >= min and working_time < max:
49                     feature_vector[bucket_index] += 1
50                     break
51             else:
52                 bucket_index += 1
53
54 data.append(feature_vector)
55 cells.append(cgi)
56
57 X = np.array(data)
58
59 kmeans = KMeans(n_clusters=4, random_state=23).fit(X)
60

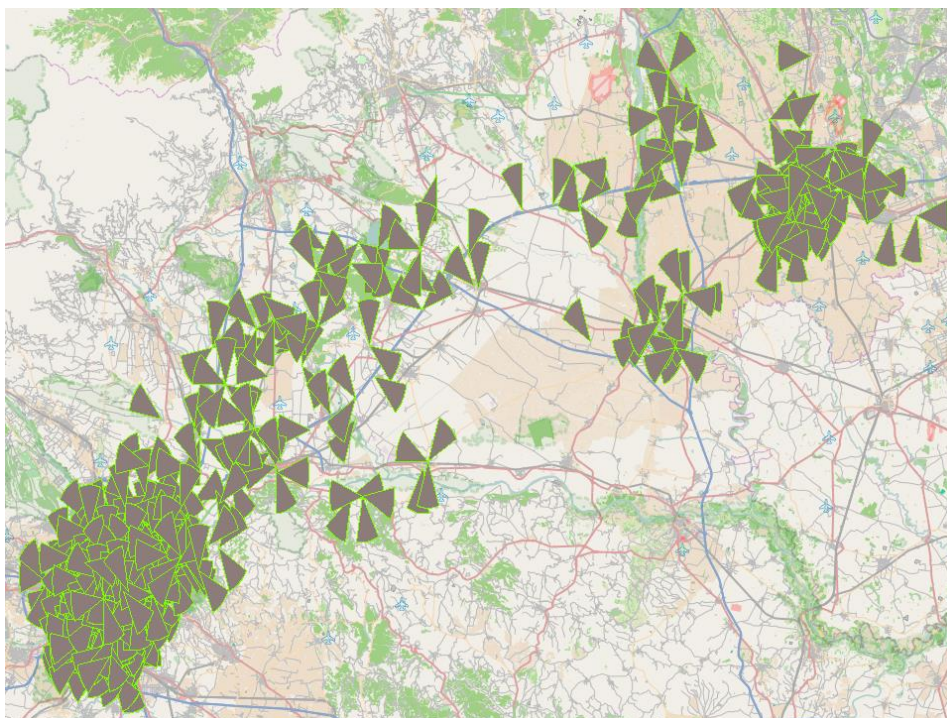
```

Slika 24. Određivanje klastera K-means metodom

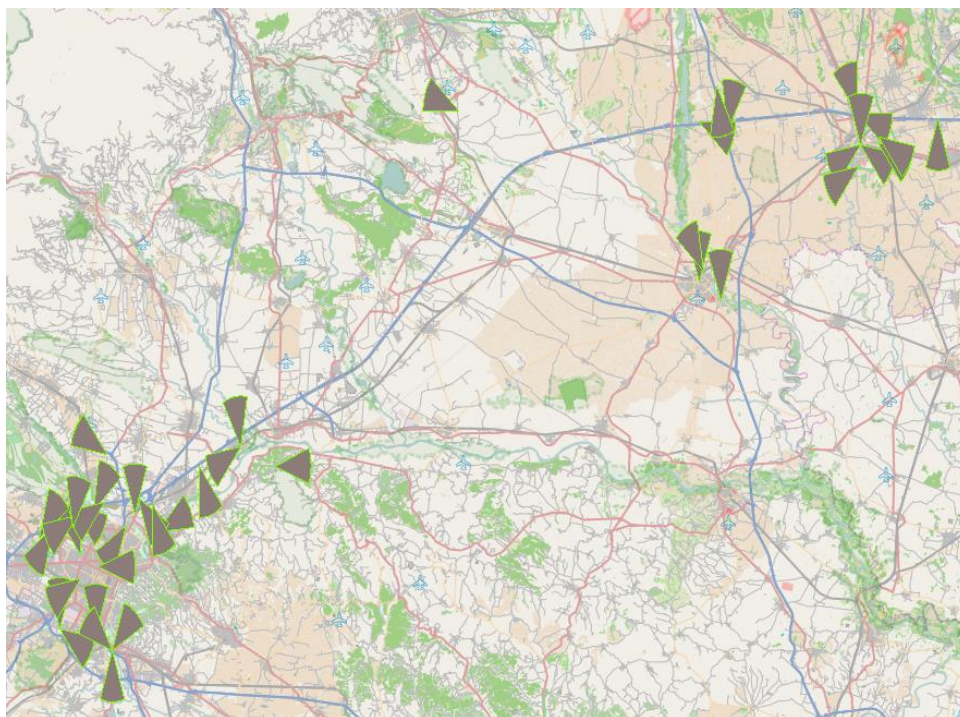
### 4.2.1. Rezultati analize K-means metode

Rezultati K-means metode i podjela ćelija u 4 tražena klastera prikazani su na Slici 25., Slici 26., Slici 27. i Slici 28.

Dobivene ćelije unutar klastera prikazane na karti:

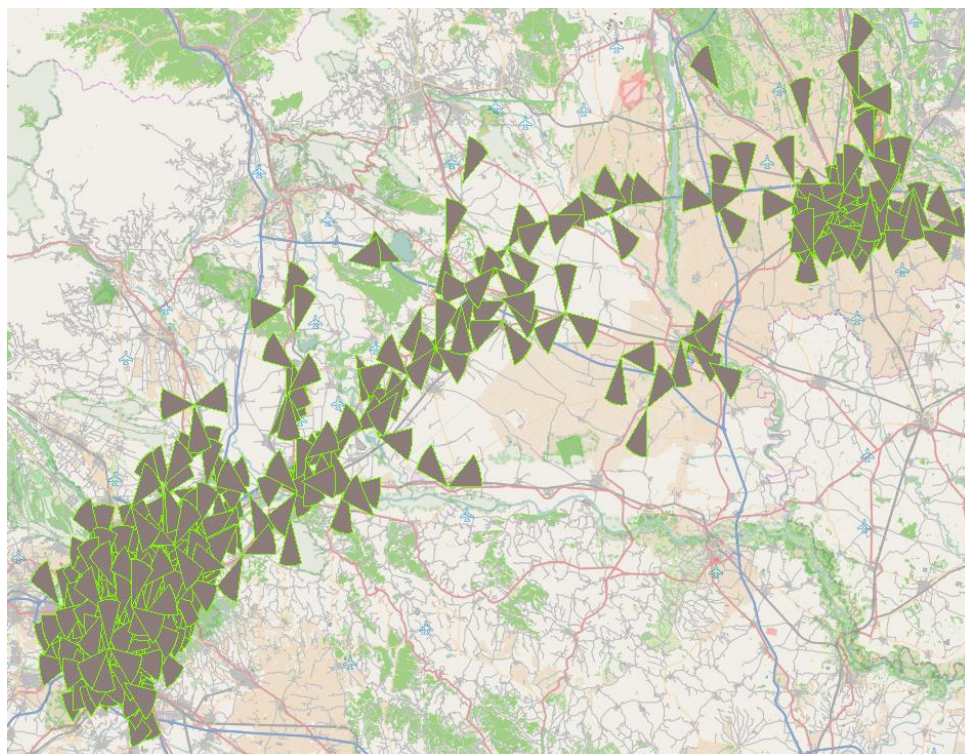


Slika 25 . Klaster – 0 / K-means metoda

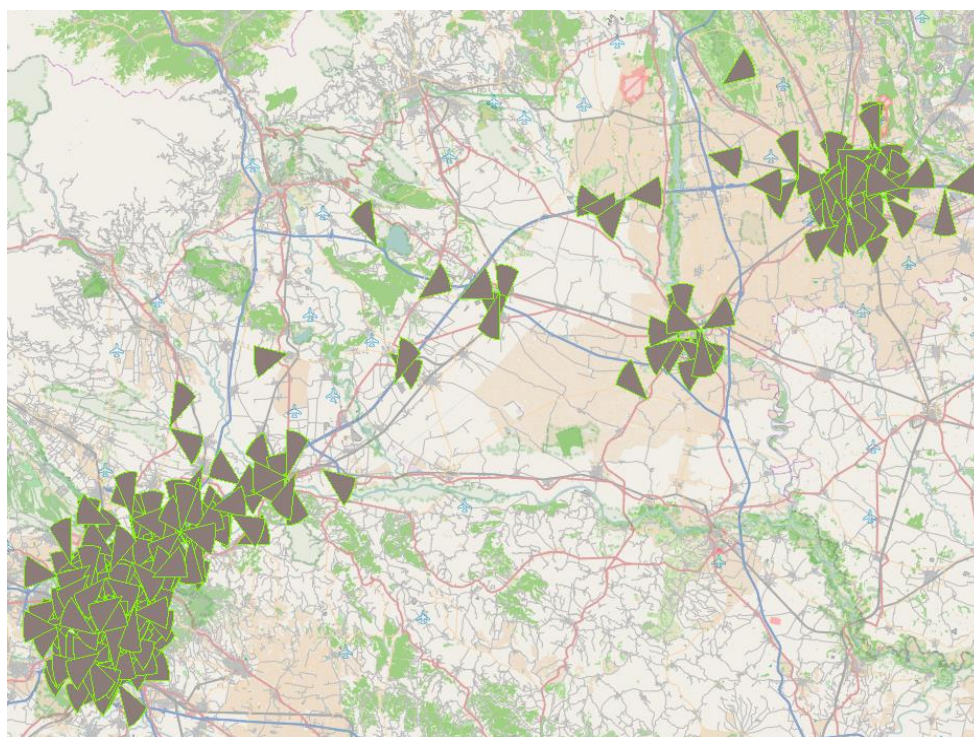


Slika 26. Klaster – 1 / K-means metoda





Slika 27. Klaster – 2 / K-means metoda



Slika 28. Klaster – 3 / K-means metoda

Na *Slici 25.* i *Slici 27.* vidljivo je kako se klasificiranje „*high-mobility*“ ćelije nalaze duž autoputa i pruge. Vizualno se prepoznaje upravo ta putanja.

Ostale grupacije mogu se objasniti na sljedeći način. Vidljivo je koliko je veliki broj „*high-mobility*“ ćelija u svakom klasteru pozicionirano u samim gradovima. Obzirom na ideju ovog zadatka, taj podatak se ignorira i objašnjava na sljedeći način. Bazne stanice u gradovima su znatno bliže pozicionirane jedna drugoj nego što je to duž autoputa, pruga i manje naseljenih područja. Radi veće koncentracije korisnika u gradovima i veće opterećenosti mreže, one mogu biti postavljene na udaljenost od 200 m do 400 m, a više im se preklapaju područja pokrivenosti. Korisnici kroz grad ne moraju se nužno brzo kretati da bi brzo prelazili iz jedne ćelije u drugu. Dovoljno je kretati se dozvoljenom brzinom u automobilu i kroz par kilometara vrlo brzo mijenjati područja pokrivenosti ćelija. Također, istu situaciju je moguće postići bržom šetnjom kroz centar grada gdje su bazne stanice gusto raspoređene, ili vožnjom biciklom ili nekim drugim prijevoznim sredstvom.

### 4.3. Analiza Gaussian Mixture metodom

Isti podaci odrađeni su i GMM algoritmom. Prva dva koraka se ne razlikuju u odnosu na K-means metodu, s obzirom na to da oni nisu dio metode grupiranja, nego priprema podataka za provedbu samih metoda.

Zadnji korak analize, korištenje GMM metode dao je podjelu ćelija po klasterima prikazanu na *Slici 29*. [Prilog IV]

```

21 buckets = ['1-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-25', '25+']
22
23 data = []
24 for rgl in range(1, len(buckets)):
25     time_in_cell_for_rgl = values.split(',')
26     feature_vector = [0] * len(buckets)
27     for current_time in time_in_cell_for_rgl:
28         if current_time == '':
29             continue
30         working_time = float(current_time)
31
32         bucket_index = 0
33         for bucket in buckets:
34             if '1' in bucket:
35                 value = bucket.split('-')[1]
36                 if working_time >= float(value):
37                     feature_vector[bucket_index] += 1
38             else:
39                 value = bucket.split('-')[1]
40                 min = float(value)
41                 max = float(value)
42                 if working_time >= min and working_time < max:
43                     feature_vector[bucket_index] += 1
44             else:
45                 bucket_index += 1
46
47         data.append(feature_vector)
48         cell_index.append(rgl)
49
50 X = np.array(data)
51
52 # metode
53 # metoda 1
54 # metoda 2
55 # metoda 3
56 # metoda 4
57 # metoda 5
58 # metoda 6
59 # metoda 7
60 # metoda 8
61 # metoda 9
62 # metoda 10
63 # metoda 11
64 # metoda 12
65 # metoda 13
66 # metoda 14
67 # metoda 15
68 # metoda 16
69 # metoda 17
70 # metoda 18
71 # metoda 19
72 # metoda 20
73 # metoda 21
74 # metoda 22
75 # metoda 23
76 # metoda 24
77 # metoda 25
78 # metoda 26
79 # metoda 27
80 # metoda 28
81 # metoda 29
82 # metoda 30
83 # metoda 31
84 # metoda 32
85 # metoda 33
86 # metoda 34
87 # metoda 35
88 # metoda 36
89 # metoda 37
90 # metoda 38
91 # metoda 39
92 # metoda 40
93 # metoda 41
94 # metoda 42
95 # metoda 43
96 # metoda 44
97 # metoda 45
98 # metoda 46
99 # metoda 47
100 # metoda 48
101 # metoda 49
102 # metoda 50
103 # metoda 51
104 # metoda 52
105 # metoda 53
106 # metoda 54
107 # metoda 55
108 # metoda 56
109 # metoda 57
110 # metoda 58
111 # metoda 59
112 # metoda 60
113 # metoda 61
114 # metoda 62
115 # metoda 63
116 # metoda 64
117 # metoda 65
118 # metoda 66
119 # metoda 67
120 # metoda 68
121 # metoda 69
122 # metoda 70
123 # metoda 71
124 # metoda 72
125 # metoda 73
126 # metoda 74
127 # metoda 75
128 # metoda 76
129 # metoda 77
130 # metoda 78
131 # metoda 79
132 # metoda 80
133 # metoda 81
134 # metoda 82
135 # metoda 83
136 # metoda 84
137 # metoda 85
138 # metoda 86
139 # metoda 87
140 # metoda 88
141 # metoda 89
142 # metoda 90
143 # metoda 91
144 # metoda 92
145 # metoda 93
146 # metoda 94
147 # metoda 95
148 # metoda 96
149 # metoda 97
150 # metoda 98
151 # metoda 99
152 # metoda 100
153 # metoda 101
154 # metoda 102
155 # metoda 103
156 # metoda 104
157 # metoda 105
158 # metoda 106
159 # metoda 107
160 # metoda 108
161 # metoda 109
162 # metoda 110
163 # metoda 111
164 # metoda 112
165 # metoda 113
166 # metoda 114
167 # metoda 115
168 # metoda 116
169 # metoda 117
170 # metoda 118
171 # metoda 119
172 # metoda 120
173 # metoda 121
174 # metoda 122
175 # metoda 123
176 # metoda 124
177 # metoda 125
178 # metoda 126
179 # metoda 127
180 # metoda 128
181 # metoda 129
182 # metoda 130
183 # metoda 131
184 # metoda 132
185 # metoda 133
186 # metoda 134
187 # metoda 135
188 # metoda 136
189 # metoda 137
190 # metoda 138
191 # metoda 139
192 # metoda 140
193 # metoda 141
194 # metoda 142
195 # metoda 143
196 # metoda 144
197 # metoda 145
198 # metoda 146
199 # metoda 147
200 # metoda 148
201 # metoda 149
202 # metoda 150
203 # metoda 151
204 # metoda 152
205 # metoda 153
206 # metoda 154
207 # metoda 155
208 # metoda 156
209 # metoda 157
210 # metoda 158
211 # metoda 159
212 # metoda 160
213 # metoda 161
214 # metoda 162
215 # metoda 163
216 # metoda 164
217 # metoda 165
218 # metoda 166
219 # metoda 167
220 # metoda 168
221 # metoda 169
222 # metoda 170
223 # metoda 171
224 # metoda 172
225 # metoda 173
226 # metoda 174
227 # metoda 175
228 # metoda 176
229 # metoda 177
230 # metoda 178
231 # metoda 179
232 # metoda 180
233 # metoda 181
234 # metoda 182
235 # metoda 183
236 # metoda 184
237 # metoda 185
238 # metoda 186
239 # metoda 187
240 # metoda 188
241 # metoda 189
242 # metoda 190
243 # metoda 191
244 # metoda 192
245 # metoda 193
246 # metoda 194
247 # metoda 195
248 # metoda 196
249 # metoda 197
250 # metoda 198
251 # metoda 199
252 # metoda 200
253 # metoda 201
254 # metoda 202
255 # metoda 203
256 # metoda 204
257 # metoda 205
258 # metoda 206
259 # metoda 207
260 # metoda 208
261 # metoda 209
262 # metoda 210
263 # metoda 211
264 # metoda 212
265 # metoda 213
266 # metoda 214
267 # metoda 215
268 # metoda 216
269 # metoda 217
270 # metoda 218
271 # metoda 219
272 # metoda 220
273 # metoda 221
274 # metoda 222
275 # metoda 223
276 # metoda 224
277 # metoda 225
278 # metoda 226
279 # metoda 227
280 # metoda 228
281 # metoda 229
282 # metoda 230
283 # metoda 231
284 # metoda 232
285 # metoda 233
286 # metoda 234
287 # metoda 235
288 # metoda 236
289 # metoda 237
290 # metoda 238
291 # metoda 239
292 # metoda 240
293 # metoda 241
294 # metoda 242
295 # metoda 243
296 # metoda 244
297 # metoda 245
298 # metoda 246
299 # metoda 247
300 # metoda 248
301 # metoda 249
302 # metoda 250
303 # metoda 251
304 # metoda 252
305 # metoda 253
306 # metoda 254
307 # metoda 255
308 # metoda 256
309 # metoda 257
310 # metoda 258
311 # metoda 259
312 # metoda 260
313 # metoda 261
314 # metoda 262
315 # metoda 263
316 # metoda 264
317 # metoda 265
318 # metoda 266
319 # metoda 267
320 # metoda 268
321 # metoda 269
322 # metoda 270
323 # metoda 271
324 # metoda 272
325 # metoda 273
326 # metoda 274
327 # metoda 275
328 # metoda 276
329 # metoda 277
330 # metoda 278
331 # metoda 279
332 # metoda 280
333 # metoda 281
334 # metoda 282
335 # metoda 283
336 # metoda 284
337 # metoda 285
338 # metoda 286
339 # metoda 287
340 # metoda 288
341 # metoda 289
342 # metoda 290
343 # metoda 291
344 # metoda 292
345 # metoda 293
346 # metoda 294
347 # metoda 295
348 # metoda 296
349 # metoda 297
350 # metoda 298
351 # metoda 299
352 # metoda 300
353 # metoda 301
354 # metoda 302
355 # metoda 303
356 # metoda 304
357 # metoda 305
358 # metoda 306
359 # metoda 307
360 # metoda 308
361 # metoda 309
362 # metoda 310
363 # metoda 311
364 # metoda 312
365 # metoda 313
366 # metoda 314
367 # metoda 315
368 # metoda 316
369 # metoda 317
370 # metoda 318
371 # metoda 319
372 # metoda 320
373 # metoda 321
374 # metoda 322
375 # metoda 323
376 # metoda 324
377 # metoda 325
378 # metoda 326
379 # metoda 327
380 # metoda 328
381 # metoda 329
382 # metoda 330
383 # metoda 331
384 # metoda 332
385 # metoda 333
386 # metoda 334
387 # metoda 335
388 # metoda 336
389 # metoda 337
390 # metoda 338
391 # metoda 339
392 # metoda 340
393 # metoda 341
394 # metoda 342
395 # metoda 343
396 # metoda 344
397 # metoda 345
398 # metoda 346
399 # metoda 347
400 # metoda 348
401 # metoda 349
402 # metoda 350
403 # metoda 351
404 # metoda 352
405 # metoda 353
406 # metoda 354
407 # metoda 355
408 # metoda 356
409 # metoda 357
410 # metoda 358
411 # metoda 359
412 # metoda 360
413 # metoda 361
414 # metoda 362
415 # metoda 363
416 # metoda 364
417 # metoda 365
418 # metoda 366
419 # metoda 367
420 # metoda 368
421 # metoda 369
422 # metoda 370
423 # metoda 371
424 # metoda 372
425 # metoda 373
426 # metoda 374
427 # metoda 375
428 # metoda 376
429 # metoda 377
430 # metoda 378
431 # metoda 379
432 # metoda 380
433 # metoda 381
434 # metoda 382
435 # metoda 383
436 # metoda 384
437 # metoda 385
438 # metoda 386
439 # metoda 387
440 # metoda 388
441 # metoda 389
442 # metoda 390
443 # metoda 391
444 # metoda 392
445 # metoda 393
446 # metoda 394
447 # metoda 395
448 # metoda 396
449 # metoda 397
450 # metoda 398
451 # metoda 399
452 # metoda 400
453 # metoda 401
454 # metoda 402
455 # metoda 403
456 # metoda 404
457 # metoda 405
458 # metoda 406
459 # metoda 407
460 # metoda 408
461 # metoda 409
462 # metoda 410
463 # metoda 411
464 # metoda 412
465 # metoda 413
466 # metoda 414
467 # metoda 415
468 # metoda 416
469 # metoda 417
470 # metoda 418
471 # metoda 419
472 # metoda 420
473 # metoda 421
474 # metoda 422
475 # metoda 423
476 # metoda 424
477 # metoda 425
478 # metoda 426
479 # metoda 427
480 # metoda 428
481 # metoda 429
482 # metoda 430
483 # metoda 431
484 # metoda 432
485 # metoda 433
486 # metoda 434
487 # metoda 435
488 # metoda 436
489 # metoda 437
490 # metoda 438
491 # metoda 439
492 # metoda 440
493 # metoda 441
494 # metoda 442
495 # metoda 443
496 # metoda 444
497 # metoda 445
498 # metoda 446
499 # metoda 447
500 # metoda 448
501 # metoda 449
502 # metoda 450
503 # metoda 451
504 # metoda 452
505 # metoda 453
506 # metoda 454
507 # metoda 455
508 # metoda 456
509 # metoda 457
510 # metoda 458
511 # metoda 459
512 # metoda 460
513 # metoda 461
514 # metoda 462
515 # metoda 463
516 # metoda 464
517 # metoda 465
518 # metoda 466
519 # metoda 467
520 # metoda 468
521 # metoda 469
522 # metoda 470
523 # metoda 471
524 # metoda 472
525 # metoda 473
526 # metoda 474
527 # metoda 475
528 # metoda 476
529 # metoda 477
530 # metoda 478
531 # metoda 479
532 # metoda 480
533 # metoda 481
534 # metoda 482
535 # metoda 483
536 # metoda 484
537 # metoda 485
538 # metoda 486
539 # metoda 487
540 # metoda 488
541 # metoda 489
542 # metoda 490
543 # metoda 491
544 # metoda 492
545 # metoda 493
546 # metoda 494
547 # metoda 495
548 # metoda 496
549 # metoda 497
550 # metoda 498
551 # metoda 499
552 # metoda 500
553 # metoda 501
554 # metoda 502
555 # metoda 503
556 # metoda 504
557 # metoda 505
558 # metoda 506
559 # metoda 507
560 # metoda 508
561 # metoda 509
562 # metoda 510
563 # metoda 511
564 # metoda 512
565 # metoda 513
566 # metoda 514
567 # metoda 515
568 # metoda 516
569 # metoda 517
570 # metoda 518
571 # metoda 519
572 # metoda 520
573 # metoda 521
574 # metoda 522
575 # metoda 523
576 # metoda 524
577 # metoda 525
578 # metoda 526
579 # metoda 527
580 # metoda 528
581 # metoda 529
582 # metoda 530
583 # metoda 531
584 # metoda 532
585 # metoda 533
586 # metoda 534
587 # metoda 535
588 # metoda 536
589 # metoda 537
590 # metoda 538
591 # metoda 539
592 # metoda 540
593 # metoda 541
594 # metoda 542
595 # metoda 543
596 # metoda 544
597 # metoda 545
598 # metoda 546
599 # metoda 547
600 # metoda 548
601 # metoda 549
602 # metoda 550
603 # metoda 551
604 # metoda 552
605 # metoda 553
606 # metoda 554
607 # metoda 555
608 # metoda 556
609 # metoda 557
610 # metoda 558
611 # metoda 559
612 # metoda 560
613 # metoda 561
614 # metoda 562
615 # metoda 563
616 # metoda 564
617 # metoda 565
618 # metoda 566
619 # metoda 567
620 # metoda 568
621 # metoda 569
622 # metoda 570
623 # metoda 571
624 # metoda 572
625 # metoda 573
626 # metoda 574
627 # metoda 575
628 # metoda 576
629 # metoda 577
630 # metoda 578
631 # metoda 579
632 # metoda 580
633 # metoda 581
634 # metoda 582
635 # metoda 583
636 # metoda 584
637 # metoda 585
638 # metoda 586
639 # metoda 587
640 # metoda 588
641 # metoda 589
642 # metoda 590
643 # metoda 591
644 # metoda 592
645 # metoda 593
646 # metoda 594
647 # metoda 595
648 # metoda 596
649 # metoda 597
650 # metoda 598
651 # metoda 599
652 # metoda 600
653 # metoda 601
654 # metoda 602
655 # metoda 603
656 # metoda 604
657 # metoda 605
658 # metoda 606
659 # metoda 607
660 # metoda 608
661 # metoda 609
662 # metoda 610
663 # metoda 611
664 # metoda 612
665 # metoda 613
666 # metoda 614
667 # metoda 615
668 # metoda 616
669 # metoda 617
670 # metoda 618
671 # metoda 619
672 # metoda 620
673 # metoda 621
674 # metoda 622
675 # metoda 623
676 # metoda 624
677 # metoda 625
678 # metoda 626
679 # metoda 627
680 # metoda 628
681 # metoda 629
682 # metoda 630
683 # metoda 631
684 # metoda 632
685 # metoda 633
686 # metoda 634
687 # metoda 635
688 # metoda 636
689 # metoda 637
690 # metoda 638
691 # metoda 639
692 # metoda 640
693 # metoda 641
694 # metoda 642
695 # metoda 643
696 # metoda 644
697 # metoda 645
698 # metoda 646
699 # metoda 647
700 # metoda 648
701 # metoda 649
702 # metoda 650
703 # metoda 651
704 # metoda 652
705 # metoda 653
706 # metoda 654
707 # metoda 655
708 # metoda 656
709 # metoda 657
710 # metoda 658
711 # metoda 659
712 # metoda 660
713 # metoda 661
714 # metoda 662
715 # metoda 663
716 # metoda 664
717 # metoda 665
718 # metoda 666
719 # metoda 667
720 # metoda 668
721 # metoda 669
722 # metoda 670
723 # metoda 671
724 # metoda 672
725 # metoda 673
726 # metoda 674
727 # metoda 675
728 # metoda 676
729 # metoda 677
730 # metoda 678
731 # metoda 679
732 # metoda 680
733 # metoda 681
734 # metoda 682
735 # metoda 683
736 # metoda 684
737 # metoda 685
738 # metoda 686
739 # metoda 687
740 # metoda 688
741 # metoda 689
742 # metoda 690
743 # metoda 691
744 # metoda 692
745 # metoda 693
746 # metoda 694
747 # metoda 695
748 # metoda 696
749 # metoda 697
750 # metoda 698
751 # metoda 699
752 # metoda 700
753 # metoda 701
754 # metoda 702
755 # metoda 703
756 # metoda 704
757 # metoda 705
758 # metoda 706
759 # metoda 707
760 # metoda 708
761 # metoda 709
762 # metoda 710
763 # metoda 711
764 # metoda 712
765 # metoda 713
766 # metoda 714
767 # metoda 715
768 # metoda 716
769 # metoda 717
770 # metoda 718
771 # metoda 719
772 # metoda 720
773 # metoda 721
774 # metoda 722
775 # metoda 723
776 # metoda 724
777 # metoda 725
778 # metoda 726
779 # metoda 727
780 # metoda 728
781 # metoda 729
782 # metoda 730
783 # metoda 731
784 # metoda 732
785 # metoda 733
786 # metoda 734
787 # metoda 735
788 # metoda 736
789 # metoda 737
790 # metoda 738
791 # metoda 739
792 # metoda 740
793 # metoda 741
794 # metoda 742
795 # metoda 743
796 # metoda 744
797 # metoda 745
798 # metoda 746
799 # metoda 747
800 # metoda 748
801 # metoda 749
802 # metoda 750
803 # metoda 751
804 # metoda 752
805 # metoda 753
806 # metoda 754
807 # metoda 755
808 # metoda 756
809 # metoda 757
810 # metoda 758
811 # metoda 759
812 # metoda 760
813 # metoda 761
814 # metoda 762
815 # metoda 763
816 # metoda 764
817 # metoda 765
818 # metoda 766
819 # metoda 767
820 # metoda 768
821 # metoda 769
822 # metoda 770
823 # metoda 771
824 # metoda 772
825 # metoda 773
826 # metoda 774
827 # metoda 775
828 # metoda 776
829 # metoda 777
830 # metoda 778
831 # metoda 779
832 # metoda 780
833 # metoda 781
834 # metoda 782
835 # metoda 783
836 # metoda 784
837 # metoda 785
838 # metoda 786
839 # metoda 787
840 # metoda 788
841 # metoda 789
842 # metoda 790
843 # metoda 791
844 # metoda 792
845 # metoda 793
846 # metoda 794
847 # metoda 795
848 # metoda 796
849 # metoda 797
850 # metoda 798
851 # metoda 799
852 # metoda 800
853 # metoda 801
854 # metoda 802
855 # metoda 803
856 # metoda 804
857 # metoda 805
858 # metoda 806
859 # metoda 807
860 # metoda 808
861 # metoda 809
862 # metoda 810
863 # metoda 811
864 # metoda 812
865 # metoda 813
866 # metoda 814
867 # metoda 815
868 # metoda 816
869 # metoda 817
870 # metoda 818
871 # metoda 819
872 # metoda 820
873 # metoda 821
874 # metoda 822
875 # metoda 823
876 # metoda 824
877 # metoda 825
878 # metoda 826
879 # metoda 827
880 # metoda 828
881 # metoda 829
882 # metoda 830
883 # metoda 831
884 # metoda 832
885 # metoda 833
886 # metoda 834
887 # metoda 835
888 # metoda 836
889 # metoda 837
890 # metoda 838
891 # metoda 839
892 # metoda 840
893 # metoda 841
894 # metoda 842
895 # metoda 843
896 # metoda 844
897 # metoda 845
898 # metoda 846
899 # metoda 847
900 # metoda 848
901 # metoda 849
902 # metoda 850
903 # metoda 851
904 # metoda 852
905 # metoda 853
906 # metoda 854
907 # metoda 855
908 # metoda 856
909 # metoda 857
910 # metoda 858
911 # metoda 859
912 # metoda 860
913 # metoda 861
914 # metoda 862
915 # metoda 863
916 # metoda 864
917 # metoda 865
918 # metoda 866
919 # metoda 867
920 # metoda 868
921 # metoda 869
922 # metoda 870
923 # metoda 871
924 # metoda 872
925 # metoda 873
926 # metoda 874
927 # metoda 875
928 # metoda 876
929 # metoda 877
930 # metoda 878
931 # metoda 879
932 # metoda 880
933 # metoda 881
934 # metoda 882
935 # metoda 883
936 # metoda 884
937 # metoda 885
938 # metoda 886
939 # metoda 887
940 # metoda 888
941 # metoda 889
942 # metoda 890
943 # metoda 891
944 # metoda 892
945 # metoda 893
946 # metoda 894
947 # metoda 895
948 # metoda 896
949 # metoda 897
950 # metoda 898
951 # metoda 899
952 # metoda 900
953 # metoda 901
954 # metoda 902
955 # metoda 903
956 # metoda 904
957 # metoda 905
958 # metoda 906
959 # metoda 907
960 # metoda 908
961 # metoda 909
962 # metoda 910
963 # metoda 911
964 # metoda 912
965 # metoda 913
966 # metoda 914
967 # metoda 915
968 # metoda 916
969 # metoda 917
970 # metoda 918
971 # metoda 919
972 # metoda 920
973 # metoda 921
974 # metoda 922
975 # metoda 923
976 # metoda 924
977 # metoda 925
978 # metoda 926
979 # metoda 927
980 # metoda 928
981 # metoda 929
982 # metoda 930
983 # metoda 931
984 # metoda 932
985 # metoda 933
986 # metoda 934
987 # metoda 935
988 # metoda 936
989 # metoda 937
990 # metoda 938
991 # metoda 939
992 # metoda 940
993 # metoda 941
994 # metoda 942
995 # metoda 943
996 # metoda 944
997 # metoda 945
998 # metoda 946
999 # metoda 947
1000 # metoda 948
1001 # metoda 949
1002 # metoda 950
1003 # metoda 951
1004 # metoda 952
1005 # metoda 953
1006 # metoda 954
1007 # metoda 955
1008 # metoda 956
1009 # metoda 957
1010 # metoda 958
1011 # metoda 959
1012 # metoda 960
1013 # metoda 961
1014 # metoda 962
1015 # metoda 963
1016 # metoda 964
1017 # metoda 965
1018 # metoda 966
1019 # metoda 967
1020 # metoda 968
1021 # metoda 969
1022 # metoda 970
1023 # metoda 971
1024 # metoda 972
1025 # metoda 973
1026 # metoda 974
1027 # metoda 975
1028 # metoda 976
1029 # metoda 977
1030 # metoda 978
1031 # metoda 979
1032 # metoda 980
1033 # metoda 981
1034 # metoda 982
1035 # metoda 983
1036 # metoda 984
1037 # metoda 985
1038 # metoda 986
1039 # metoda 987
1040 # metoda 988
1041 # metoda 989
1042 # metoda 990
1043 # metoda 991
1044 # metoda 992
1045 # metoda 993
1046 # metoda 994
1047 # metoda 995
1048 # metoda 996
1049 # metoda 997
1050 # metoda 998
1051 # metoda 999
1052 # metoda 1000

```

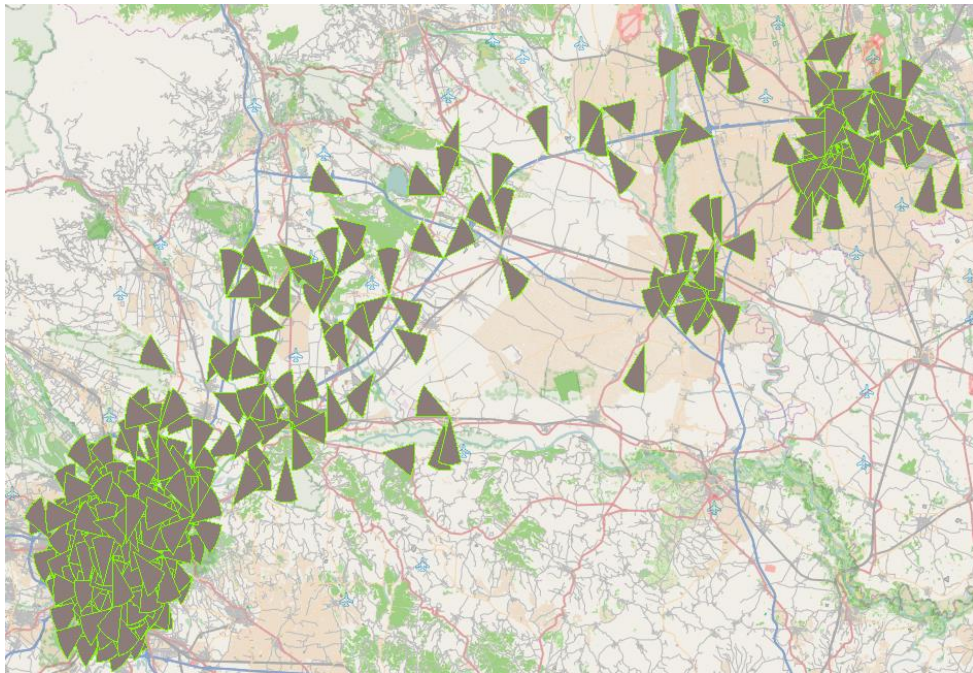
Slika 29. Grupacija klastera GMM metodom

#### 4.3.1. Rezultati analize GMM metode

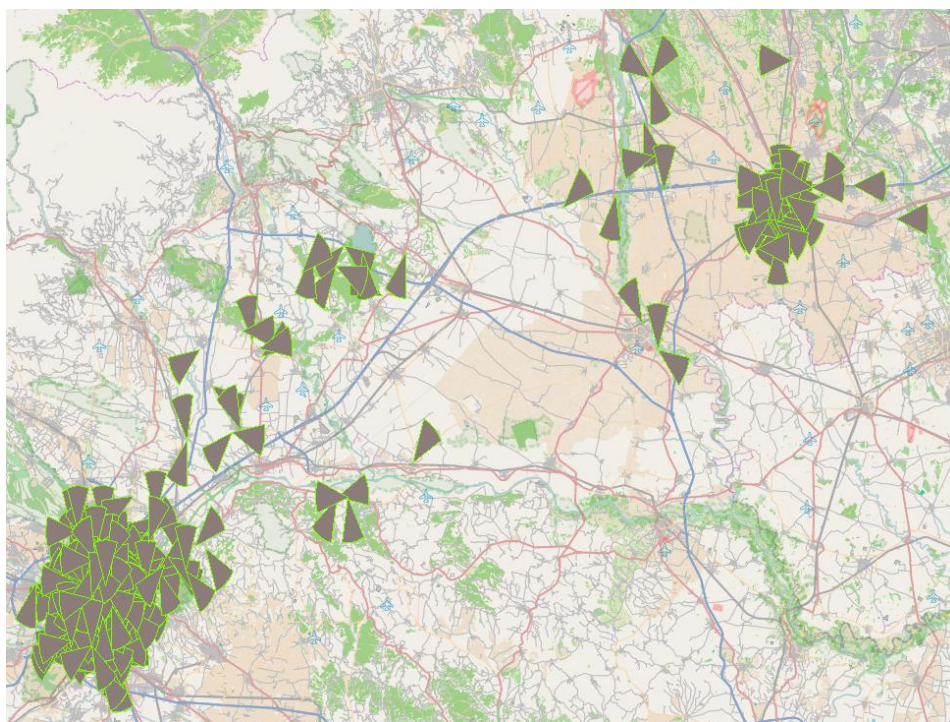
Geografski prikaz dobivenih klastera slijedi na *Slici 30*., *Slici 31*., *Slici 32*. i *Slici 33*.

Gusta raspodjela klasificiranih „high-mobility“ ćelija unutar samih gradova dogodila se iz istog razloga kao i u K-means metodi. Priroda ulaznih podataka objašnjava ovu pojavu. Prema *Slici 32*. može se primijetiti klasificiranje ćelija duž prometnica, što je i bio cilj analize.



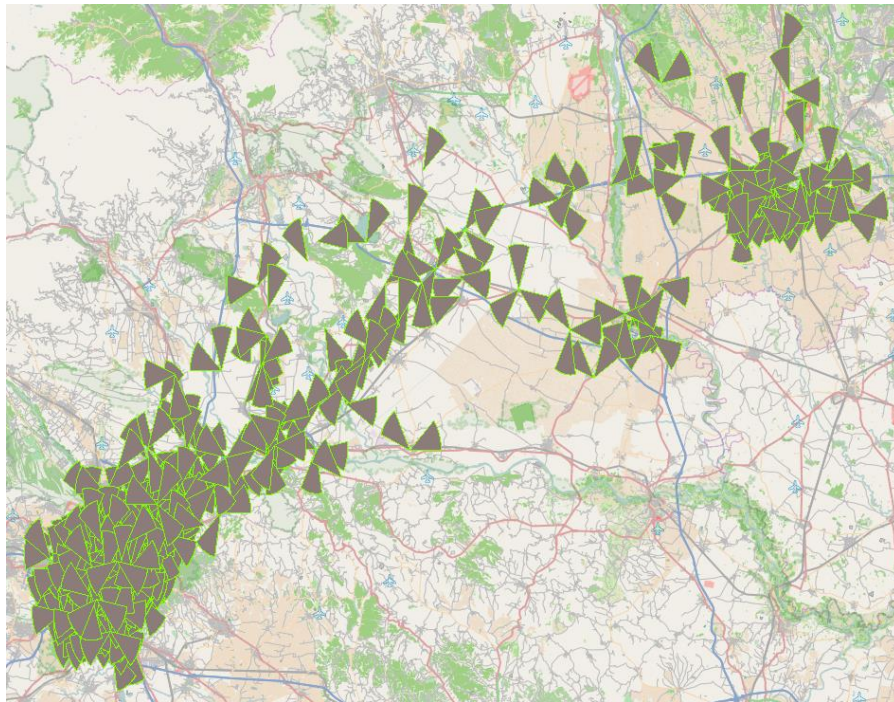


Slika 30. Klaster – 0 / GMM metoda

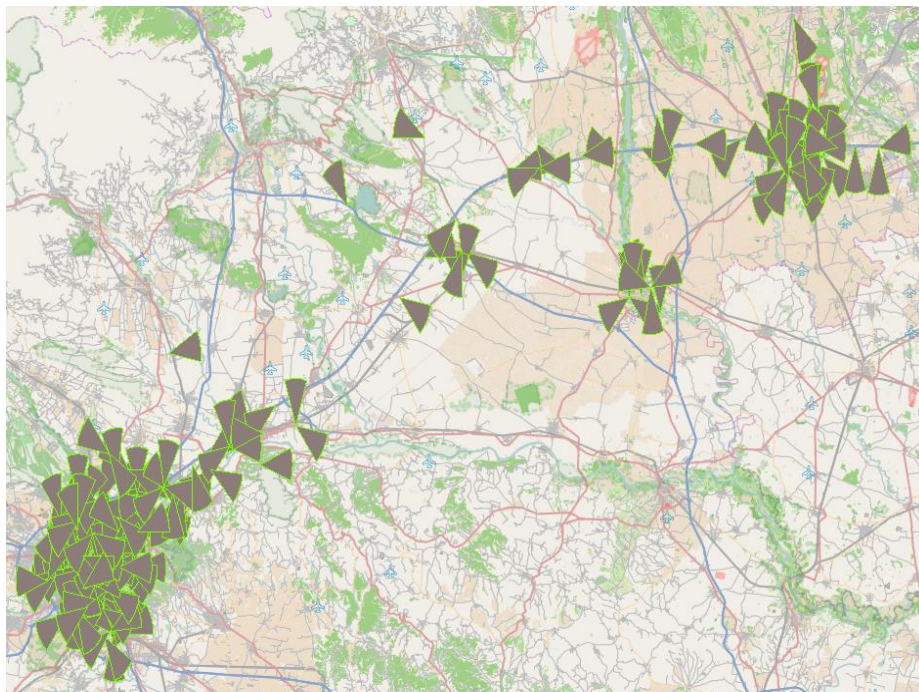


Slika 31. Klaster – 1 / GMM metoda





Slika 32. Klaster – 2 / GMM metoda



Slika 33. Klaster – 3 / GMM metoda



#### 4.4. Usporedba rezultata K-means i GMM metode

Obje metode su dale isti rezultat, ali ne i istu geografsku raspodjelu, što se može objasniti činjenicom da su broj klastera i maksimalan broj iteracija postavljeni isto. Usporedbom rezultata prikazanom u Tablici 2., vidljivo je da su obje jednak broj ćelija raspodijelile u zadane klastere. Tablica 2. prikazuje broj ćelija koje su svrstane u istu grupaciju klastera različitim metodama, što znači da su obje metode jednak broj ćelija svrstale u zadana 4 klastera. [Prilog IV]

K-means klaster / GMM klaster	GMM-3	GMM-1	GMM-0	GMM-2
K-means-3	624	0	0	0
K-means-1	0	1152	0	0
K-means-0	0		281	0
K-means-2	0	0	0	42

Tablica 2. Usporedba rezultata različitih metoda

## 4.5. Usporedba rezultata s neovisnim mjerenjima

Za provjeru sezonalnosti i ponovljivost rezultata, provedena je analiza istih ćelija na istom području, ali u drugo vrijeme. Ulazni podaci prikazani u Tablici 3. sadrže periode korisnika unutar ćelije izražena u sekundama, za isti dan, u razdoblju od 16:00 h do 18:00 h.

Ćelija	Vrijeme (s)	Ćelija	Vrijeme (s)	Ćelija	Vrijeme (s)
cell-1	1.66	cell-347	6.284	cell-1148	0.546
cell-2	8.712	cell-351	6.55	cell-1142	20.083
cell-2	27	cell-349	1.13	cell-1140	1.626
cell-3	8.815	cell-349	0.684	cell-1145	5.467
cell-2	2.476	cell-349	0.77	cell-1142	6.687
cell-1	1.921	cell-352	9.389	cell-1144	16.243
cell-0	2.906	cell-350	21.779	cell-1146	2.408
cell-1	1.877	cell-352	0.76	cell-1142	22.879
cell-1	1.91	cell-349	1.323	cell-1144	8.976
cell-3	12.497	cell-349	14.109	cell-1143	5.903
cell-1	5.016	cell-349	3.112	cell-1148	4.942
cell-1	3.626	cell-347	1.735	cell-1145	1.477
cell-1	5.913	cell-348	0.347	cell-1147	19.454
cell-3	5.608	cell-352	3.925	cell-1142	57.013
cell-3	10.773	cell-352	0.767	cell-1145	4.674
cell-2	10.872	cell-352	22.889	cell-1145	18.883
cell-1	10.624	cell-352	42.8	cell-1145	14.257
cell-1	9.36	cell-351	0.869	cell-1145	1.79
cell-4	5.002	cell-349	0.949	cell-1148	3.783
cell-0	3.477	cell-347	1.191	cell-1146	8.879
cell-1	1.482	cell-350	0.432	cell-1148	0.902
cell-0	1.198	cell-352	1.854	cell-1144	6.95
cell-1	9.915	cell-352	6.402	cell-1144	2.235
cell-4	6.847	cell-349	5.981	cell-1141	3.867
cell-0	2.496	cell-352	8.204	cell-1146	7.421
cell-5	1.578	cell-352	5.43	cell-1145	9.182
cell-4	34.693	cell-352	2.074	cell-1144	2.712
cell-1	6.924	cell-351	5.029	cell-1146	7.679
cell-1	5.447	cell-351	1.628	cell-1147	7.015
cell-1	1.205	cell-349	5.519	cell-1142	2.905
cell-0	0.92	cell-348	1.171	cell-1147	15.727
cell-2	2.516	cell-352	1.615	cell-1146	8.022
cell-3	4.434	cell-350	1.205	cell-1146	1.268
cell-5	2.821	cell-352	12.111	cell-1145	3.327

cell-0	21.729	cell-349	0.977	cell-1144	1.813
cell-4	0.459	cell-352	4.332	cell-1142	1.577
cell-3	12.362	cell-351	3.886	cell-1145	10.278
cell-5	2.689	cell-349	1.271	cell-1147	23.541
cell-5	0.753	cell-350	2.086	cell-1144	1.083
cell-3	4.693	cell-348	16.419	cell-1147	4.959
cell-4	1.17	cell-351	11.654	cell-1142	2.575
cell-1	0.715	cell-351	1.815	cell-1142	2.317
cell-3	2.032	cell-349	7.827	cell-1140	59.875
cell-3	4.702	cell-347	13.926	cell-1146	1.937
cell-3	2.492	cell-347	2.119	cell-1147	1.827
cell-1	5.068	cell-348	0.708	cell-1144	3.62
cell-0	15.303	cell-347	4.499	cell-1146	2.113
cell-3	5.89	cell-352	2.789	cell-1145	8.757
cell-1	2.728	cell-352	16.571	cell-1146	2.328
cell-3	2.227	cell-352	14.795	cell-1144	2.001
cell-3	17.087	cell-347	6.739	cell-1148	0.916
cell-3	1.964	cell-347	7.802	cell-1145	6.363
cell-5	13.996	cell-352	4.72	cell-1145	30.47
cell-3	9.68	cell-352	2.251	cell-1146	5.67
cell-0	3.71	cell-350	7.72	cell-1144	10.981
cell-0	1.593	cell-350	3.994	cell-1148	4.262
cell-3	3.99	cell-351	57.864	cell-1144	3.077
cell-3	11.199	cell-348	17.917	cell-1146	4.047
cell-3	1.068	cell-351	63.91	cell-1144	1.133
cell-1	3.537	cell-351	9.803	cell-1145	3.852
cell-0	4.263	cell-349	1.888	cell-1146	5.992
cell-3	0.845	cell-351	3.233	cell-1145	2.252
cell-3	11.392	cell-349	0.585	cell-1140	3.831
cell-1	132.395	cell-352	5.562	cell-1147	10.695
cell-5	1.048	cell-348	5.547	cell-1144	4.451
cell-4	8.938	cell-350	7.916	cell-1146	10.896
cell-0	9.071	cell-349	2.276	cell-1141	1.98
cell-4	17.52	cell-351	2.743	cell-1151	6.967
cell-1	9.198	cell-348	1.265	cell-1140	0.633
cell-1	2.353	cell-348	6.908	cell-1146	1.573
cell-1	1.39	cell-349	2.961	cell-1143	17.149
cell-1	72.305	cell-349	0.753	cell-1142	2.332
cell-0	92.48	cell-349	1.779	cell-1148	1.386
cell-1	1.793	cell-347	2.717	cell-1147	1.777

Tablica 3. Reprezentativni uzorci popodnevnih mjerenja

Usporedba jutarnjih i popodnevnih analiza prikazana je u Tablici 4. [Prilog IV] :

	UJUTRO-0	UJUTRO-3	UJUTRO-2	UJUTRO-1
POPODNE-1	497	52	98	0
POPODNE-2	74	255	3	9
POPODNE-3	57	2	972	0
POPODNE-0	1	7	0	46

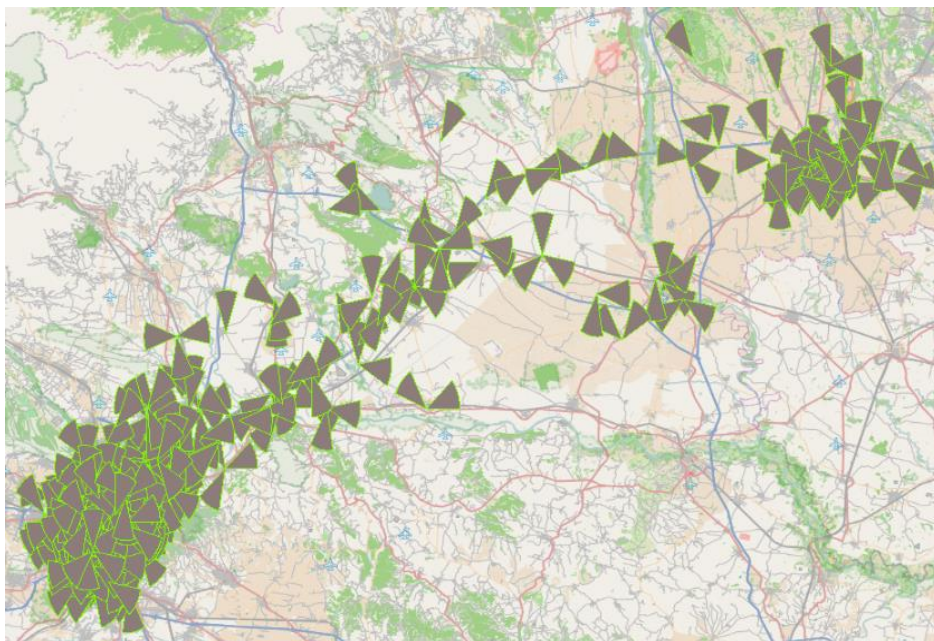
Tablica 4. Usporedba rezultata jutarnjih i popodnevnih mjerenja

U tablici su prikazane količine jednako klasificiranih ćelija u različitom periodu. Neke ćelije su različito klasificirane (označeno narančasto).

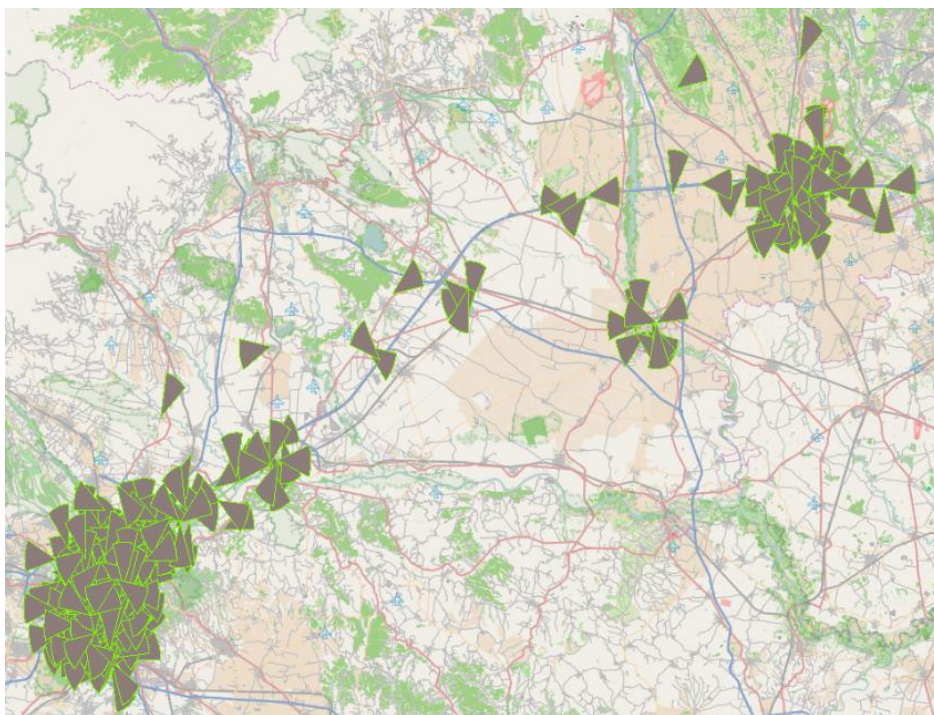
Usporedbom različitih perioda rada ćelija potvrđuje se pretpostavka da ćelija ne mora biti tijekom cijelog dana klasificirana kao „*high-mobility*“ ćelija. Njezino stanje ovisi o količini i načinu kretanja korisnika, što se mijenja kroz period sata, dana, tjedna, itd.

Ipak, istaknuta dijagonala (označena zeleno) unutar Tablice 4. pokazuje kako je veći dio ćelija ispravno klasificiran kao ćelija visokog stupnja mobilnosti, čime se dokazala sezonalnost i ponovljivost rezultata.

Rezultati popodnevnog mjerenja prikazani na karti:

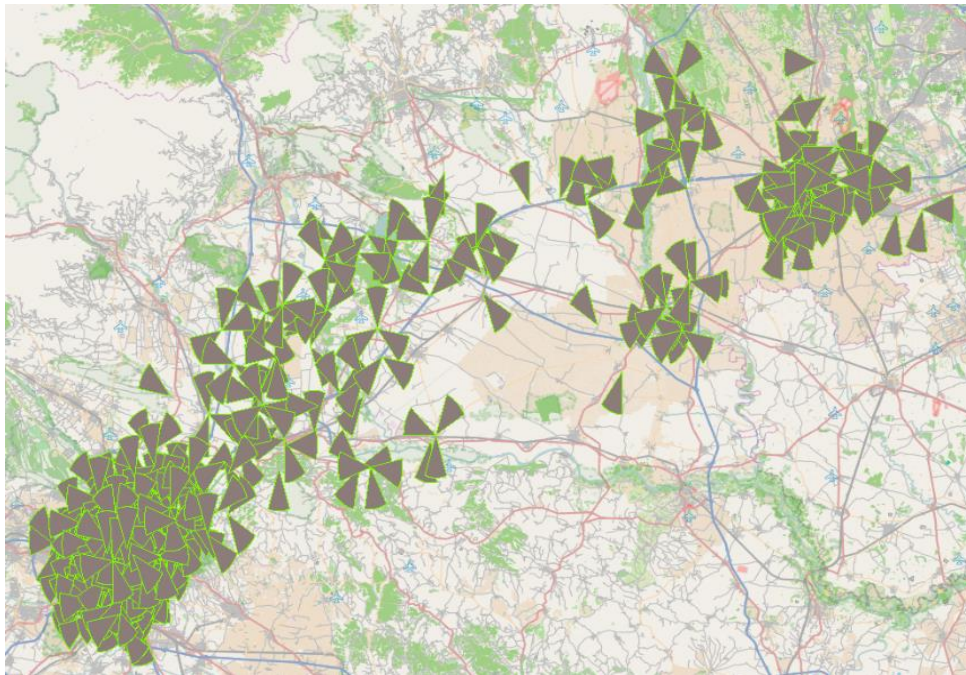


Slika 34. Klaster – 0 / K-means, popodne

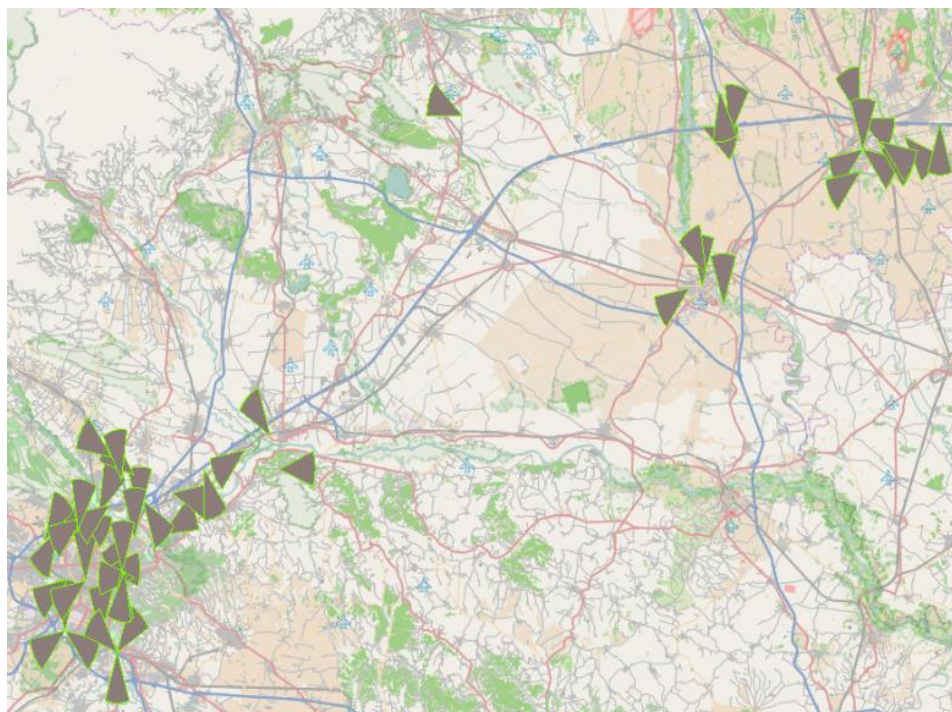


Slika 35. Klaster – 1 / K-means, popodne





Slika 36. Klaster – 2 / K-means, popodne



Slika 37. Klaster – 2 / K-means, popodne

## 5. ZAKLJUČAK

Zadatak ovog diplomskog rada je klasifikacija profila ćelija iz podataka korisničkih profila telekomunikacijske mreže, što je i realizirano. Implementirane su i uspoređene dvije metode strojnog učenja (grupiranja): K-means i GMM metoda. Usporedbom rezultata dobivenih putem obje metode može se zaključiti da između njih gotovo i nema razlike. Obje metode rezultirale su podjelom jednakog broja ćelije u klasterima, iako se vidi neznatna razlika u geografskom prikazu. Bez obzira na različit način obrade podataka, kojih i nije bilo „puno“ u usporedbi s podacima na projektima za koje se ove metode koriste u praksi, obje metode su trajale oko 2 – 3 min, pri čemu je programski kôd K-means metode neznatno jednostavniji. Prema tome, ne može se ocijeniti koja je od tih dviju metoda bolja. Kao što je i navedeno u radu, ne postoji nužno bolja i lošija metoda, sve ovisi o potrebama korisnika i prirodi podataka. Iz vizualizacije rezultata na karti, intuitivno se može zaključiti da je K-means metoda dala „točnije rezultate“, iako se u svim rezultatima vizualno istaknuo barem jedan klaster ćelija koji pokriva autoput i prugu, kako je i bilo očekivano. Pod „točnije rezultate“ smatraju se rezultati bliski očekivanima. U ovom slučaju, bilo bi zanimljivo pokušati s hijerarhijskim grupiranjem ili MeanShift metodom. Može se reći da je u ovom zadatku klasificiranja bilo vrlo jednostavno izabrati metodu, jer je bio poznat broj željenih klastera.

Ćelije se ne mogu striktno definirati kao „*high-mobility*“. „*Mobility*“ svojstvo ćelije je vremenski zavisno. Ćelije koje su svrstane u „*high-mobility*“ profil, ne moraju nužno imati „*high-mobility*“ ponašanje 24/7 jer se primjerice vikendima manje putuje, postoje i tzv. „špice“ kada je povećanje svojstva mobilnosti uzrokovano korisnicima koji odlaze i dolaze s posla. „*Mobility*“ svojstvo ovisi o tome koje je doba dana, kreće li se kroz ćeliju veći ili manji broj korisnika. Dakle, postoji sezonalnost u ponašanju. Jedini način da se potpuno točno provjeri metoda jest ručna analiza svake ćelije zasebno, direktnim pogledom u mrežu, u njezine performanse i statistiku rada, što je praktički nemoguće i nitko to ne bi radio na razini mreže.

Kako bi se sa sigurnošću moglo reći da je ćelija „*high-mobility*“, podaci bi se trebali periodički prikupljati, analizirati i uspoređivati.

Ovaj rad je rezultirao uspješnim klasificiranjem profila, što može poslužiti kao osnova za daljnju analizu čime bi se identificirao postotak stanja visoke mobilnosti pojedine ćelije. Daljnjom analizom moglo bi se definirati postotak vremena ćelije (npr. 30-50%, 50-70%, 70-

90%, 90%+) kada ona ima svojstvo „*high-mobility*“ i koji je to interval dana ili tjedna. Na taj način bi se došlo do podataka o tome kakav je obrazac ponašanja pojedinih ćelija.



## 6. LITERATURA

- [1] R. C. Raciti: *Cellular technology*, New Southeastern University, 2012.
- [2] D. Yu. Ignatov, A. N. Filippov, A. D. Ignatov, X. Zhang: *Homogenous Network Optimization*, 2016.
- [3] Golio, Mike: *RF and Microwave Passive and Active Technologies*, 2018.
- [4] Guowang Miao, Jens Zander, Ki Won Sung, Ben Slimane: *Fundamentals of Mobile Data Networks*, Cambridge University, 2016.
- [5] N. Nasser, A. Hasswa, H. Hassanein: *Handovers in fourth generation heterogeneous networks*, IEEE Commun. Mag., vol. 44, Oct. 2006.
- [6] G. L. Stuber: *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic Publishers, 2001.
- [7] Leslie D. Fife, Le Gruenwald: *Research Issues for Data Communication in Mobile Ad-Hoc Network Database Systems*, Brigham Young University – Hawaii, The University of Oklahoma, School of Computer Science, Norman, 2013.
- [8] K. Tutschku: *Demand-based Radio Network Planning of Cellular Mobile Communication Systems*, University of Wurzburg Institute of Computer Science Research Report Series, 1997.
- [9] Yufei Wu and Samuel Pierre: *Optimization of 3G Mobile Network Design Using a Hybrid Search Strategy*, Journal of communications and networks, 2005.
- [10] D. Varoutas, D. Katsianis, Th. Sphicopoulos, K. Stordahl, I. Welling: *On the Economics of 3G Mobile Virtual Network Operators*, Wireless Personal Communications, 2006.
- [11] Suk Yu Hui, Kai Hau Yeung: *Challenges in the Migration to 4G Mobile Systems*, City University of Hong Kong, 2003.
- [12] Bao-Shuh P. Lin, Wen-Hsiang Tsai, C.C. Wu, P.H. Hsu, J.Y. Huang, Tsai-Hwa Liu: *The Design of Cloud-based 4G/LTE for Mobile Augmented Reality with Smart Mobile Devices*, National Chiao Tung University, 2013.
- [13] Pol Blasco, Deniz Gunduz: *Learning-Based Optimization of Cache Content in a Small Cell Base Station*, Imperial College London, UK, 2014.
- [14] Juan Carrasquilla, Roger G. Melko: *Machine learning phases of matter*, University of Waterloo, Ontario, Canada, 2016.

- [15] Alpaydin, Ethem: *Introduction to Machine Learning*, MIT Press, 2010.
- [16] Mohri Mehryar, Rostamizadeh Afshin, Talwalkar Ameet: *Foundations of Machine Learning*, The MIT Press, 2012.
- [17] Kanungo Tapas, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Silverman Ruth, Angela Y. Wu: *An efficient k-means clustering algorithm: Analysis and implementation*, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002.
- [18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery: *Section 16.1. Gaussian Mixture Models and k-Means Clustering, Numerical Recipes: The Art of Scientific Computing(3rd ed.)*, New York (NY): Cambridge University Press., 2007.
- [19] Coates Adam, Lee Honglak, Andrew Y. Ng: *An analysis of single-layer networks in unsupervised feature learning*, International Conference on Artificial Intelligence and Statistics, 2011.
- [20] Cepy Slamet, Ali Rahman, Muhammad Ali Ramdhani, Wahyudin Darmalaksana: *Clustering the Verses of the Holy Qur'an using K-Means Algorithm*, Department of Informatics Engineering, UIN Sunan Gunung Djati Bandung, Indonesia, 2016.
- [21] M. P. Bhatia, D. Khurana: *Analysis of Initial Centers for K-Means Clustering Algorithm*, International Journal Computer Application, 2013.
- [22] Vijayarani S., Ilamathi M. J., Nithya, M.: *Preprocessing Techniques for Text Mining: An Overview*, International Journal Computer Science and Communication Network, 2015.
- [23] Marco Capó, Aritz Pérez, Jose A. Lozano: *An efficient approximation to the K-means clustering for Massive Data*, University of the Basque Country UPV/EHU, 2016.
- [24] Yu, Guoshen: *Solving Inverse Problems with Piecewise Linear Estimators: From Gaussian Mixture Models to Structured Sparsity*, IEEE Transactions on Image Processing., 2012.
- [25] Permuter H., Francos, J., Jermyn I.H.: *Gaussian mixture models of texture and colour for image database retrieval*, IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003.
- [26] B. G. Lindsay B. G.: *Mixture Models: Theory, Geometry, and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Hayward: Institute of Mathematical Statistics., 2015.

## 7. PRILOZI

### I. Programski kôd za određivanje raspona vremena

```
import sqlite3
import numpy as np
import scipy
import matplotlib.pyplot as plt
from statsmodels.distributions.empirical_distribution import ECDF
import math

def geometricMean(arr, n):

    # declare sum variable and
    # initialize it to 1.
    sum = 0;

    # Compute the sum of all
    # the elements in the array.
    for i in range(n):
        sum = sum + math.log(arr[i]);

    # compute geometric mean
    # through formula antilog
    # (((log(1) + log(2) + . .
    # ... + log(n))/n)
    # and return the value to
    # main function.
    sum = sum / n;

    return math.exp(sum);

con=sqlite3.connect('C:/Users/data.db')
cur = con.cursor()

cgi_group_query = '''
    SELECT TIME_IN_CELL_s
    FROM ctr_mobility
    ...

cur.execute(cgi_group_query)

data = []
for values in cur.fetchall():
    if values[0] != '':# and float(values[0]) < 26 :
        data.append(float(values[0]))

for q in [50, 90, 95, 100]:
    print ("{}% percentile: {}".format (q, np.percentile(data, q)))

num_bins = 10

counts, bin_edges = np.histogram (data, bins=num_bins, normed=True)
```

```

cdf = np.cumsum (counts)
#plt.plot (bin_edges[1:], cdf/cdf[-1])

#plt.hist(data, bins=15)

ecdf = ECDF(data)
# get cumulative probability for values
print('P(x<2): %.3f' % ecdf(2))
print('P(x<3): %.3f' % ecdf(3))
print('P(x<5): %.3f' % ecdf(5))
print('P(x<10): %.3f' % ecdf(10))
print('P(x<15): %.3f' % ecdf(15))
print('P(x<20): %.3f' % ecdf(20))
print('P(x<40): %.3f' % ecdf(40))
print('P(x<60): %.3f' % ecdf(60))
print('P(x<80): %.3f' % ecdf(80))
print('P(x<110): %.3f' % ecdf(110))
# plot the cdf
plt.plot(ecdf.x, ecdf.y)

gmean = geometricMean(data, len(data))
print("Geometric mean: ", gmean)

plt.show()

```

## II. Programski kôd za razvrstavanje korisnika unutar ćelija u zadanom vremenskom rasponu

```

import sqlite3
import numpy as np
import matplotlib.pyplot as plt

con=sqlite3.connect('C:/Users/data.db')
cur = con.cursor()

cgi_group_query = '''
        SELECT TIME_IN_CELL_s
        FROM ctr_mobility JOIN ENM_CGI ON
ENM_CGI.CGI=ctr_mobility.SOURCE_CGI
        where X1 = 'False' and X2 = 'False'
        ...

cur.execute(cgi_group_query)

buckets = ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-
20', '20-25', '25+']
#buckets = ['0-2', '2-3', '3-4', '4-5', '5-10', '10-11', '11-12', '12-15', '15-
20', '20-25', '25+']

```

```

data = [0] * len(buckets)
for values in cur.fetchall():
    if values[0] != '':
        value = float(values[0])
        bucket_index = 0
        for bucket in buckets:
            if '+' in bucket:
                vals = bucket.split('+')
                max = float(vals[0])
                if value >= max:
                    data[bucket_index] += 1
            else:
                vals = bucket.split('-')
                min = float(vals[0])
                max = float(vals[1])
                if value >= min and value < max:
                    data[bucket_index] += 1
                    break
            else:
                bucket_index += 1

x_pos = [i for i, _ in enumerate(buckets)]
plt.bar(x_pos, data, color='green')

plt.xlabel("Ranges")
plt.ylabel("Number of samples")
plt.title("Series distribution")

plt.xticks(x_pos, buckets)

plt.show()

```

### III. Programski kôd za provedbu K-means metode

```

import sqlite3
import numpy as np
from sklearn.cluster import KMeans

con=sqlite3.connect('C:/Users/data.db')
cur = con.cursor()

cgi_group_query = '''
    SELECT SOURCE_CGI, group_concat(TIME_IN_CELL_s)
    FROM ctr_mobility
    GROUP BY SOURCE_CGI
    '''

cur.execute(cgi_group_query)

```

```
buckets = ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-25', '25+']

data = []
cells = []
for cgi, values in cur.fetchall():
    time_in_cell_for_cgi = values.split(',')

    feature_vector = [0] * len(buckets)
    for current_time in time_in_cell_for_cgi:
        if current_time == '':
            continue

        working_time = float(current_time)

        bucket_index = 0
        for bucket in buckets:
            if '+' in bucket:
                vals = bucket.split('+')
                max = float(vals[0])
                if working_time >= max:
                    feature_vector[bucket_index] += 1
            else:
                vals = bucket.split('-')
                min = float(vals[0])
                max = float(vals[1])
                if working_time >= min and working_time < max:
                    feature_vector[bucket_index] += 1
                    break
                else:
                    bucket_index += 1

        data.append(feature_vector)
        cells.append(cgi)

X = np.array(data)
kmeans = KMeans(n_clusters=4, random_state=23).fit(X)

byCluster = dict()
for clustindex in kmeans.labels_:
    byCluster.update({clustindex: []})

for i in range(len(data)):
    byCluster[kmeans.labels_[i]].append(cells[i])
    # print (cells[i], ' je u clusteru: ', kmeans.labels_[i])

for clust in byCluster.keys():
    print("Clust: ", clust)
    for cells in byCluster[clust]:
        print(cells)
```

## IV. Programski kôd za provedbu GMM metode i usporedbu K-means i GMM metode

```
# jutarnji dataset
con=sqlite3.connect('C:\Users\data.db')

cur = con.cursor()

cgi_group_query = '''
    SELECT SOURCE_CGI, group_concat(TIME_IN_CELL_s)
    FROM ctr_mobility
    GROUP BY SOURCE_CGI
    '''

cur.execute(cgi_group_query)

buckets = ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-25', '25+']

data = []
cells = []
for cgi, values in cur.fetchall():
    time_in_cell_for_cgi = values.split(',')

    feature_vector = [0] * len(buckets)
    for current_time in time_in_cell_for_cgi:
        if current_time == '':
            continue

        working_time = float(current_time)

        bucket_index = 0
        for bucket in buckets:
            if '+' in bucket:
                vals = bucket.split('+')
                max = float(vals[0])
                if working_time >= max:
                    feature_vector[bucket_index] += 1
            else:
                vals = bucket.split('-')
                min = float(vals[0])
                max = float(vals[1])
                if working_time >= min and working_time < max:
                    feature_vector[bucket_index] += 1
                    break
                else:
                    bucket_index += 1

        data.append(feature_vector)
        cells.append(cgi)

X = np.array(data)

# metoda-1
kmeans=KMeans(n_clusters=4, init='k-means++', n_init=20, max_iter=500,
tol=0.00001).fit(X)
clusters_kmeans = kmeans.labels_
```

```
byCluster_kmeans_1 = dict()
for clustindex in clusters_kmeans:
    byCluster_kmeans_1.update({clustindex: []})

for i in range(len(data)):
    byCluster_kmeans_1[clusters_kmeans[i]].append(cells[i])

# popodnevni dataset
con=sqlite3.connect('C:\Users\Data\ctr_mobility.db')

cur = con.cursor()

cgi_group_query = '''
    SELECT SOURCE_CGI, group_concat(TIME_IN_CELL_s)
    FROM ctr_mobility
    GROUP BY SOURCE_CGI
    '''

cur.execute(cgi_group_query)

buckets = ['0-1', '1-2', '2-3', '3-4', '4-5', '5-7', '7-10', '10-15', '15-20', '20-
25', '25+']

data = []
cells = []
for cgi, values in cur.fetchall():
    time_in_cell_for_cgi = values.split(',')

    feature_vector = [0] * len(buckets)
    for current_time in time_in_cell_for_cgi:
        if current_time == '':
            continue
        working_time = float(current_time)

        bucket_index = 0
        for bucket in buckets:
            if '+' in bucket:
                vals = bucket.split('+')
                max = float(vals[0])
                if working_time >= max:
                    feature_vector[bucket_index] += 1
            else:
                vals = bucket.split('-')
                min = float(vals[0])
                max = float(vals[1])
                if working_time >= min and working_time < max:
                    feature_vector[bucket_index] += 1
                    break
            else:
                bucket_index += 1

        data.append(feature_vector)
        cells.append(cgi)

X = np.array(data)
```



```
# metoda-1
kmeans = KMeans(n_clusters=4, init='k-means++', n_init=20,
max_iter=500tol=0.00001).fit(X)
clusters_kmeans = kmeans.labels_

byCluster_kmeans_2 = dict()
for clustindex in clusters_kmeans:
    byCluster_kmeans_2.update({clustindex:[]})
V.
for i in range(len(data)):
    byCluster_kmeans_2[clusters_kmeans[i]].append(cells[i])

# print header
for gmmCluster in byCluster_kmeans_1.keys():
    print("UJUTRO-"+str(gmmCluster), sep=' \t ', end=' \t ', flush=True)

for kmeansCluster in byCluster_kmeans_2.keys():
    kmeansCells = set(byCluster_kmeans_2[kmeansCluster])
    print("\n POPODNE-"+str(kmeansCluster), sep='\t', end=' \t ', flush=True)
    for gmmCluster in byCluster_kmeans_1.keys():
        gmmCells = set(byCluster_kmeans_1[gmmCluster])

        commonElements = len(kmeansCells.intersection(gmmCells))
        diff = len(list(kmeansCells - gmmCells))
        diff += len(list(gmmCells - kmeansCells))
        print(commonElements, sep=' \t ', end=' \t ', flush=True)
```