

Analiza prometnih tokova autoceste Zagreb-Karlovac primjenom neuronskih mreža

Slavulj, Marko

Master's thesis / Diplomski rad

2021

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

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:119:687508>

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

Download date / Datum preuzimanja: **2025-03-01**



Repository / Repozitorij:

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



SVEUČILIŠTE U ZAGREBU
FAKULTET PROMETNIH ZNANOSTI

Marko Slavulj

**ANALIZA PROMETNIH TOKOVA AUTOCESTE ZAGREB-
KARLOVAC PRIMJENOM NEURONSKIH MREŽA**

DIPLOMSKI RAD

Zagreb, 2021.

Zagreb, 30. kolovoza 2021.

Zavod: **Zavod za inteligentne transportne sustave**
Predmet: **Umjetna inteligencija**

DIPLOMSKI ZADATAK br. 6073

Pristupnik: **Marko Slavulj (0135238122)**
Studij: **Inteligentni transportni sustavi i logistika**
Smjer: **Inteligentni transportni sustavi**

Zadatak: **Analiza prometnih tokova autoceste Zagreb Karlovac primjenom neuronskih mreža**

Opis zadatka:

Analiza prometnih tokova se provodi kako bi se dobili karakteristični vremenski intervali (sati, dani, sezonalnost itd.) u godini radi simulacijskog ispitivanja raznih prometnih rješenja. Jedna od mogućnosti analize prometnih tokova jest i primjena metoda iz područja umjetne inteligencije kao što su samoorganizirajuće neuronske mreže. Pomoću njih je postojeći skup prometnih podataka moguće grupirati kako bi se odredile karakteristične klase ponašanja prometnih tokova kao podloga za umjeravanje pripadnih mikroskopskih simulacijskih modela. U radu je potrebno opisati karakteristike prometnih tokova na autocestama i postojeći skup prometnih podataka s autoceste Zagreb-Karlovac analizirati te odrediti slične grupe odnosno karakteristične vremenske intervale u godini primjenom neuronskih mreža. Na osnovu dobivenih karakterističnih grupa predložiti scenarije za izradu umjerenih simulacijskih modela.

Mentor:

Predsjednik povjerenstva za
diplomski ispit:

izv. prof. dr. sc. Edouard Ivanjko

Sveučilište u Zagrebu
Fakultet prometnih znanosti

DIPLOMSKI RAD

**ANALIZA PROMETNIH TOKOVA AUTOCESTE ZAGREB-
KARLOVAC PRIMJENOM NEURONSKIH MREŽA**

**ANALYSIS OF TRAFFIC FLOWS ON THE ZAGREB-
KARLOVAC MOTORWAY USING NEURAL NETWORKS**

Mentor: izv. prof. dr. sc. Edouard Ivanjko

Student: Marko Slavulj
JMBAG: 0135238122

Zagreb, rujan 2021.

Neizmjerno se zahvaljujem izv. prof. dr. sc. Edouard Ivanjku na mentorstvu tijekom studiranja i prilikom izrade ovog diplomskog rada. Također se zahvaljujem asistentu Mladenu Miletiću, mag. ing. traff. na iznimnom strpljenju, savjetima i pomoći prilikom izvedbe ovog rada. Zahvaljujem se i svojoj obitelji koja me podržavala tijekom cijelog trajanja studija.

SAŽETAK

Sa sve većom prometnom potražnjom na autocestama, osnovni prometni pokazatelji razine usluznosti kao što su: sigurnost, vrijeme putovanja i onečišćenje okoliša dolaze do sve većeg izražaja i u našoj zemlji. To je posljedica sve većeg obujma turista koji svoj godišnji odmor upravo provode na teritoriju Republike Hrvatske, ponajviše na obali Jadrana. To dovodi do potencijalne opasnosti na autocestama poput prometnih nezgoda i nesreća, nezadovoljstva korisnika dobivenom uslugom, itd. Republika Hrvatska ima jedne od najrazvijenijih autocesta u Europi pa i time prikuplja veliku količinu prometnih podataka. Ti prometni podatci osnova su za procjenu stanja prometnog sustava i prijedloga mogućih rješenja. Prometni podatci običnom čovjeku ne predstavljaju ništa dok se ne obrade i analiziraju. Za rješenje problema obrade velike količine prometnih podataka primijenjena je metoda K-srednjih vrijednosti (engl. K-means) koja određuje optimalni broj grupa koje sadrže dane u godini te je primijenjena samoorganizirajuća neuronska mreža koja analizom prikazuje potencijalne van-distribucijske raspodjele i grupe podataka. Također je analizirana i vizualizirana kategorizacija vozila, brzine i sezonalnost prometa.

KLJUČNE RIJEČI: k-means; neuronske mreže; prometni tok autoceste; kategorizacija vozila; samoorganizirajuća neuronska mreža (SOM).

SUMMARY

With increasing traffic on motorways, basic traffic level of service indicators, e.g.: safety, travel time and environmental pollution, are coming to the fore. This is a consequence of a growing number of tourists who spend their vacations in Croatia, mostly on the Adriatic coast. This again leads to potential risks on motorways, such as car accidents and collisions, customer dissatisfaction, etc. Republic of Croatia has one of the most developed motorway networks in Europe, and it collects a large quantity of data. These traffic data is basis for checking the conditions of the traffic system and proposing possible solutions. To an average person the traffic data does not mean anything, until it is processed and analyzed. To solve the issue of big traffic data processing, K-means clustering was applied as a method of processing traffic data. It determines an optimal number of clusters, that contain days of the year, then a self-organizing neural network is applied, which, through analysis, shows potential outlier allocations and clusters. On top of that, categories of vehicles, speed and seasonality of the traffic is analyzed and visualized.

KEY WORDS: k-means clustering; neural networks; highway traffic flow; vehicle categorization; self-organizing map (SOM).

Sadržaj

1.Uvod	1
2.Karakteristike prometnih tokova na autocestama.....	4
2.1. Podjela prometa na javnim cestama.....	4
2.2. Autoceste.....	4
2.3. Dijelovi autocesta.....	5
2.4. Karakteristike autocesta	8
2.5. Namjena i vrsta prometa	11
3.Metode za grupiranje podataka.....	13
3.1. Najčešće korišteni programski jezici za grupiranje podataka	15
3.2. Najčešći algoritmi za grupiranje podataka	16
3.3. Algoritam K-means	23
3.4. Samoorganizirajuće mape	25
4.Implementacija neuronskih mreža za grupiranje podataka.....	29
5.Dostupni prometni podaci.....	33
5.1. Dobiveni prometni podatci.....	34
5.2. Osnovna prilagodba podataka	37
5.3. Priprema podataka za obradu	39
6.Dobiveni rezultati.....	44
6.1. Analiza sezonalnosti podataka	44
6.2. Analiza vizualizacije podataka.....	44
6.3. Rezultati za određivanje optimalnog broja grupa	54
6.4. Rezultati dobiveni metodom K-means.....	62
6.5. Rezultati dobiveni metodom neuronskih mreža.....	70
6.6. Rasprava.....	81
7.Zaključak	84
LITERATURA	85
POPIS SLIKA.....	88
POPIS TABLICA	91
POPIS FORMULA.....	92
POPIS KRATICA.....	93
PRILOG 1. MATLAB PROGRAMSKI KÔD ZA METODU K-MEANS	94

1. Uvod

Trenutna prometna situacija u Republici Hrvatskoj nije na razini na kojoj bi trebala biti turistička zemlja poput nje. Najveći problem je u naplati cestarina i vođenju prometnog toka zbog sezonalnosti. Hrvatska ima jako razvijene autoceste i samim time prikuplja se velika količina podataka s njih, ali ta velika prikupljena količina podataka se ne obrađuje sustavno i rade se neke osnovne analize i objavljuje statistika na makroskopskoj razini. S toga se u ovom diplomskom radu obrađuju i analiziraju prikupljeni prometni podatci s autoceste Zagreb-Karlovac kako bi se vidjelo ponašanje prometnih tokova primjenom metode grupiranja podataka i neuronske mreže i kako bi se vidjeli potencijalni scenariji koji se rjeđe pojavljuju. Oni su bitni da ih se proanalizira kako bi se na temelju tih rezultata obrade kasnije mogli iskoristiti za izradu umjerenih simulacijskih modela za provjeru sustava upravljanja odnosno za bolje sustave (pred)putnog informiranja [1].

U izradi diplomskog rada upotrijebljeni su podatci o prometu prikupljeni na dionici autoceste A1 blizu Karlovca za oba smjera. Vršena su mjerenja kategorije vozila i kategorizacija brzina vozila u satnim rezolucijama po pojedinim danima što je sve pohranjeno u dvije Excel tablice. Kako godine prolaze, vidljivo je da je prometna potražnja sve više u porastu na autocestama, pogotovo jer je Hrvatska turistička zemlja te sa sve većom i ljepšom ponudom za turiste i intezitet prometa u ljetnim mjesecima raste. U ljetnim mjesecima autoceste su zagušene cestovnim prometom što smanjuje kvalitetu življenja, utječe na zdravlje ljudi, povećava vrijeme putovanja, doprinosi manjoj sigurnosti odvijanja prometa te je iz tog razloga potrebno pronaći potencijalno rješenje problema zagušenja. Rješenje nije samo proširiti prometnice, već se potrebno okrenuti novim tehnologijama iz domene inteligentnih transportnih sustava i time potencijalno riješiti ili barem smanjiti problem zagušenja.

Analiza i obrada prometnih podataka vrši se kako bi se dobili karakteristični vremenski intervali (vršni sati, dani, sezonalnost, itd.) u pojedinoj godini. Analiza prometnih tokova napravljena je primjenom metoda umjetne inteligencije algoritmom K-središnjih vrijednosti (engl. K-means) i samoorganizirajućih neuronskih mreža (engl. self-organizing map, SOM) radi implementacije okruženja za grupiranje podataka na osnovu sličnosti ponašanja prometnih tokova. Primjenom dobivenih grupa podataka određene su karakteristične klase scenarija ponašanja prometnih tokova na razmatranoj autocesti unutar jednog dana kao već spomenuta podloga za izradu reprezentativnih simulacijskih modela radi realističnog ispitivanja novih prometnih rješenja te za potencijalnu buduću izradu sustava upravljanja prometom primjenom karakterističnih problematičnih scenarija. Uz sve to na zadanim skupovima prometnih podataka napravljena je i obrada smanjivanja dimenzionalnosti, dohvaćanja jedinstvenih vrijednosti pojedinih atributa, izrada statistike s vizualizacijom podataka, itd. Izrađena je detaljnija analiza prometnih tokova uz prikaz bitnih parametara poput: karakterističnih vršnih perioda (sata, dana, mjeseci), zatim postotak prekoračenja brzina vožnje, promjene brzine vozila, kategorizacije vozila, itd.

Svrha diplomskog rada je obrada dobivenog skupa prometnih podataka sa ciljem grupiranja, smanjivanja dimenzionalnosti, dohvaćanja jedinstvenih vrijednosti pojedinih atributa te izrada statistike s vizualizacijom podataka. Kao rezultat dobit će se izrađena detaljnija analiza prometnih tokova gdje će biti prikazani parametri poput:

- Karakterističnih vršnih sata, dana, mjeseci;
- Postotak prekoračenja dozvoljene brzina vožnje;
- Brzine vozila;
- Kategorizacije vozila;
- Itd.

Cilj diplomskog rada je dokazati da se obradom dobivenog skupa prometnih podataka i primjenom metoda umjetne inteligencije (metoda K-Means i SOM) može pospješiti analiza velikog skupa podataka u smislu njihovog grupiranja te kreiranja grupa karakterističnog ponašanja prometnih tokova. Naslov ovog diplomskog rada je: Analiza prometnih tokova autoceste Zagreb-Karlovac primjenom neuronskih mreža. Rad je podijeljen u sedam cjelina:

1. Uvod;
2. Karakteristike prometnih tokova na autocestama;
3. Metode za grupiranje podataka;
4. Implementacija neuronskih mreža za grupiranje podataka;
5. Dostupni prometni podaci;
6. Dobiveni rezultati;
7. Zaključak.

Nakon uvodnog poglavlja u drugom poglavlju je opisana podjela cestovnog prometa i njegove karakteristike. Posebno je obrađen podnaslov autocesta gdje su opisane karakteristike autoceste, zbog čega su izgrađene, čemu služe, kakav se promet na njima odvija, razina uslužnosti, brzine, transport dobara, turistički promet, itd.

U trećem poglavlju riječ je o metodama za grupiranje podataka, najčešće korištenim programskim jezicima za grupiranje podataka, najčešćim algoritmima za grupiranje podataka, u diplomskom radu korištenom algoritmu K-means za grupiranje podataka te SOM-u. Dat je kratki opis o grupiranju podacima, programskim jezicima te njihovim karakteristikama. Posebno je opisana metoda za grupiranje K-means i SOM jer su upravo te metode korištene za obradu prometnih podataka u ovom diplomskom radu.

U četvrtom poglavlju opisan je postupak implementacije neuronskih mreža za grupiranje podataka nad identičnim skupom podataka koji je korišten i za metodu grupiranja K-means. Objasnjeno je kako se SOM može primijeniti za grupiranje podataka.

U petom poglavlju opisani su i analizirani dobiveni prometni podatci. Riječ je o organiziranosti podataka i njihovoj osnovnoj prilagodbi te je prikazano smanjenje dimenzionalnosti, dohvaćanje jedinstvenih vrijednosti pojedinih atributa i statistika s vizualizacijom podataka.

U šestom poglavlju navedeni su dobiveni rezultati. Prvo su analizirani rezultati za određivanje optimalnog broja grupa, zatim je prikazana analiza sezonalnosti podataka, potom dobiveni rezultati pomoću K-means algoritma, zatim rezultati dobiveni od SOM-a. Za kraj je ostavljena rasprava za dobivene rezultate gdje je odlučeno koji je optimalni broj grupa za podlogu kod izrade umjerenog simulacijskog modela.

Diplomski rad završava zaključkom i prijedlozima za nastavak rada na ovoj tematici.

2. Karakteristike prometnih tokova na autocestama

Promet se odvija na javnim državnim cestama različitih kopnenih površina kao što su to urbana područja (gradovi i predgrađa) i ruralna područja (manja naselja). Državna cesta je prometno-tehnički uređena građevina s podlogom i zastorom za cestovni promet. Površina je od općeg značenja za promet, kojom se svatko može koristiti na način i uz uvjete određene *Zakonom o cestama* i drugim propisima [1].

2.1. Podjela prometa na javnim cestama

Javne ceste se razvrstavaju u četiri skupine [2]:

- *Autoceste*: su javne ceste koje imaju funkciju povezivanja Republike Hrvatske u europski prometni sustav;
- *Državne ceste*: su javne ceste koje povezuju cjelokupni teritorij Republike Hrvatske i povezuju ga s mrežom glavnih europskih cesta;
- *Županijske ceste*: su javne ceste koje povezuju područje jedne ili više županija;
- *Lokalne ceste*: su javne ceste koje spadaju u mrežu županijskih cesta i povezuju područje grada i/ili općina.

2.2. Autoceste

Autocesta (slika 1) je javna cesta visoke razine usluge namijenjena isključivo motornom prometu, bez raskrižja u jednoj razini te s odvojenim kolnicima za jednosmjerni promet i sa zaustavnim trakovima te s posebno uređenim priključcima za ulazak i izlazak sa autoceste. Ima bitnu ulogu u suvremenome prostornom povezivanju i važan je čimbenik gospodarskog razvoja te turizma [3].

Autoceste u Republici Hrvatskoj

Mreža i društva

Društvo	Autoceste	Oznake izgrađenosti	Km u prometu
HRVATSKE AUTOCESTE d.o.o.		U PROMETU U IZGRADNJI U PLANIRANJU	918,50
AUTOCESTA RIJEKA-ZAGREB d.d.			187,03
BINA-ISTRA d.d.			141,00
AUTOCESTA ZAGREB-MACEJ d.o.o.			60,00
UKUPNO u prometu na dan 31.12.2019.			1.306,53

- Dijelom autoceste A1 od Zagreba do Bosiljeva upravlja ARZ d.d.
- ARZ d.d. upravlja dijelom autoceste A7
- D10 i D12 predane su na upravljanje Hrvatskim cestama d.o.o. i nisu pod naplatom



Slika 1. Autoceste u Republici Hrvatskoj, [5].

2.3. Dijelovi autocesta

Autocesta ima dva kolnička traka za promet u suprotnim smjerovima, odvojena razdjelnim pojasom, kao i zaustavni trak, voznu traku za spora vozila, zaštitni rubnik i rigol. Svaki kolnički trak ima barem dva prometna traka (pretjecajna i vozna traka), široka najmanje po 3,5 [m] (tablica 1). Zatim ima jedan zaustavni trak (slika 3) širine 2,5 [m] i voznu traku za spora vozila, označena žutom bojom na slici 4 koja se izvodi na usponima većim od 4 [%] na cestama s računskom brzinom većom od 60 [km/h] i s više od 3.000 [voz/dan], a širina trake je 3,0 do 3,25 [m]. Razdjelni pojas (slika 5) izvodi se u presjeku ceste kod koje je zbog sigurnosti prometa nužno fizički

razdvojiti dva kolnika s prometom u suprotnim smjerovima. Kod autocesta koje imaju četiri i više prometnih trakova, razdjelni pojas (slika 5) se izvodi u širini od 4,0 [m] (minimalno 3,0 [m]). Rigol (slika 6) je betonska ili kamena konstrukcija smještena duž ruba kolnika koja služi za preuzimanje padalina i njihovu odvodnju [4].



Slika 2. Autocesta, [3].



Slika 3. Zaustavni trak autoceste, [3].



Slika 4. Vozna traka za spora vozila, [3].



Slika 5. Razdjelni pojas između dva smjera autoceste, [3].



Slika 6. Rigol na autocesti, [3].

Raskrižja autoceste s drugim prometnicama su izvedena barem u dvjema razinama, a u zoni raskrižja dodaju se trakovi za uključivanje i isključivanje, tj. za ulazak na autocestu i za izlazak s nje. Pojedini dijelovi autoceste mogu u svojoj izvedbi, već prema prostornim i prometnim uvjetima (planinski teren, prolaz kroz grad, most, tunel), neznatno odstupati od navedenih značajki. Autocesta s dva prometna traka za svaki smjer vožnje široka je oko 30 [m], kolika je približno i širina autoceste Zagreb–Karlovac [4].

Tablica 1. Širine prometnih trakova s obzirom na dopuštenu brzinu kretanja vozila, [3].

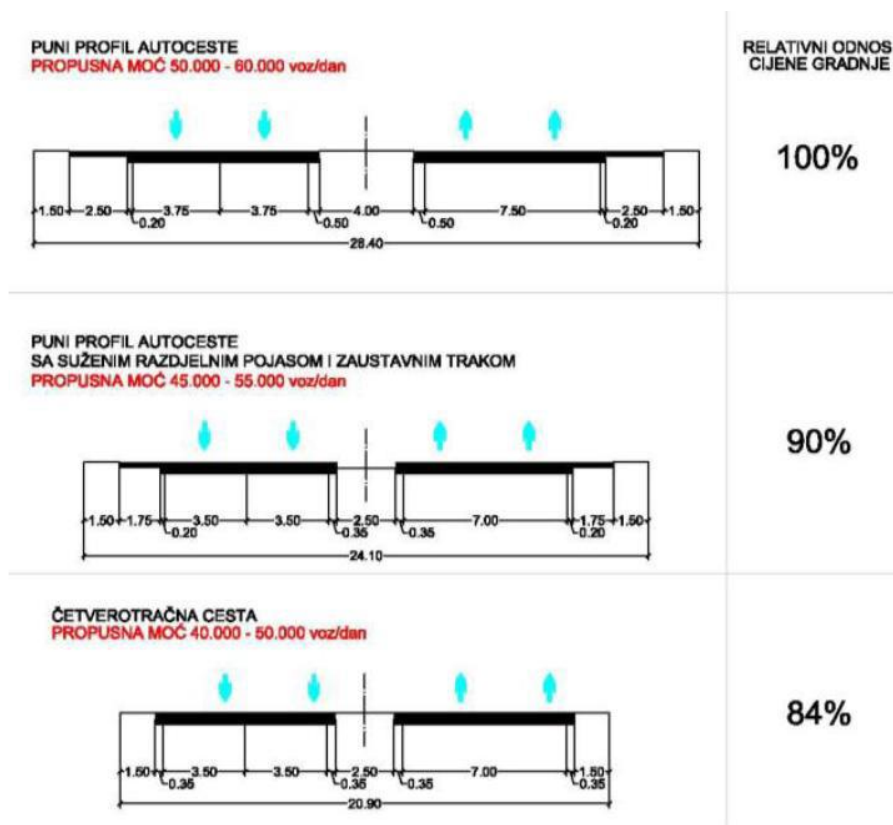
Projektna brzina [km/h]	Širina prometnog traka [m]
120	3,75
90	3,50
80	3,25
60	3,00
40	2,75 (2,50)

2.4. Karakteristike autocesta

U optimalnim uvjetima, kapacitet autoceste s po dva traka za svaki smjer vožnje iznosi oko 1.800 do 2.000 [voz/h po traku], odnosno 3.600 do 4.000 [voz/h] za dva traka. Dakle gledano u presjeku autoceste, ukupni kapacitet (oba kolnička traka) iznosi 7.200 do 8.000 [voz/h]. Kod ovog diplomskog rada dionica autoceste Zagreb-Karlovac je punog profila. Prema slici 7 vidljivo je da propusna moć takvih autocesta iznosi do 60.000 [voz/dan].

Optimalni uvjeti autoceste:

- Širina prometnog traka: 3,75 [m];
- Bočna smetnja na većoj udaljenosti od 1,80 [m];
- Dužina preglednosti najmanje 450 [m];
- Prometni tok: samo putnički automobili;
- Uzdužni nagib oko +/- 2,00 [%].



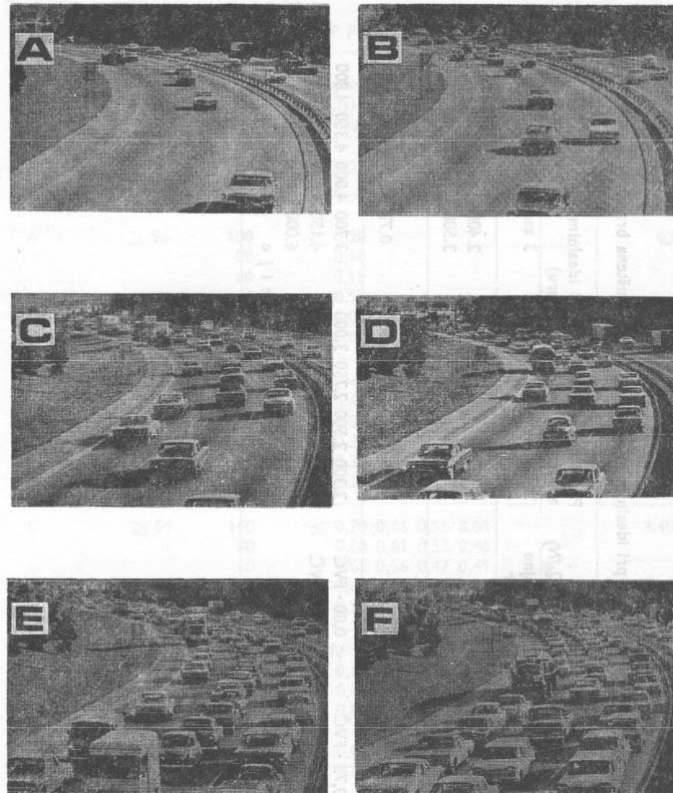
Slika 7. Propusna moć autocesta raznih profila, [7].

Razina uslužnosti odnosno razina prometne usluge LOS (engl. Level of service) predstavlja razinu operativnih uvjeta odnosno niz karakteristika koje opisuju uvjete vožnje koje se pojavljuju na određenom potezu ceste (brzina, vrijeme putovanja, prekid, sigurnost, udobnost vožnje, cijenu usluge, itd.). LOS se zasniva na [7]:

- Gustoći/prostoru;
- Slobodi odabira brzine;
- Slobodi obilaženja drugih entiteta;
- Veličini protoka;
- Sigurnosti vožnje;
- Prekidu prometa, čekanju...

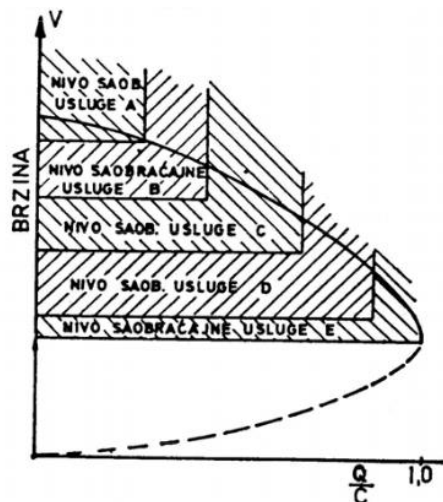
Razina uslužnosti definirana je sa 6 stupnjeva od A (najbolji) do F (najlošiji) kako slijedi [6]:

1. **Razina uslužnosti A:** Radi se o uvjetima potpunog slobodnog toka i pune slobode manevriranja. Vozači mogu birati brzinu u širokom rasponu, no ona je ograničena minimalnim vrijednostima odnosno komforom izgradnje ceste;
2. **Razina uslužnosti B:** označava stabilni prometni tok s minimalno ograničenim brzinama kao posljedicom gustoće toka;
3. **Razina uslužnosti C:** Prometni tok je i dalje stabilan uz ograničenje slobode manevriranja i izbora brzine;
4. **Razina uslužnosti D:** Prometni tokovi tendiraju k nestabilnosti, brzina toka se bitno smanjuje, komfor vožnje je nizak;
5. **Razina uslužnosti E:** Kod ove razine, brzina toka često koleba, postižu se vrijednosti protoka bliže ili identične kapacitetu ceste, mogući su i trenutni zastoji i takav prometni tok je nestabilan;
6. **Razina uslužnosti F:** Tok kod koga se iskazuju najveće gustoće prometa, tok je izrazito nestabilan s čestim zastojima, mogu se postići protoci jednaki kapacitetu prometnice, ali je moguć i potpun raspad sustava i prometni tok je prisilan.



Slika 8. Prikaz razine uslužnosti na autocesti s pripadnom količinom vozila, [6].

Na slici 8 vidljiv je slikovni prikaz razine uslužnosti na autocesti i u pravilu kod autocesta veličina protoka odgovara razini uslužnosti C.



Slika 9. razine uslužnosti u odnosu na brzinu prometnog toka i stupanj iskoristivosti kapaciteta ceste, [6].

Odnos prometnog toka i kapaciteta (stupanj zasićenja q/c) ukazuje koliki dio kapaciteta je iskorišten sadašnjim ili planiranim opterećenjem, a vrijednost se kreće između 0 i 1 (slika 9). Koristi se kao mjera dostupnosti postojećeg ili planiranog kapaciteta. Ako je stupanj zasićenja iznad vrijednosti jedan, tada kapacitet ceste ne može zadovoljiti prometnu potražnju i u tom slučaju nastaju gužve i repovi čekanja. U tablici 2 vidljiv je prikaz LOS-a prema gustoći, zauzeću i brzini prometa.

Tablica 2. Razine uslužnosti s obzirom na gustoću, zauzetost i brzinu prometnog toka, [7].

LOS	Gustoća (voz/km)	Zauzeće (%)	Brzina (Km/h)
A	0 – 7	0 – 5	≥ 97
B	7 – 12	5 – 8	≥ 92
C	12 – 19	8 – 12	≥ 87
D	19 – 26	12 – 17	≥ 74
E	26 – 42	17 – 28	≥ 48
F	42 - 62 > 62	28 – 42 > 42	≤ 48

2.5. Namjena i vrsta prometa

Autocesta je namijenjena isključivo prometovanju motornim vozilima, a pod „prometovanjem“ se misli na prijevoz ljudi i dobara. Također ima važnu ulogu u povezivanju gradova i ostalih zemalja, turizmu, gospodarstvu, brzini i sigurnosti odvijanja prometa jer je znatno

sigurnija od ostalih vrsta cesta i omogućuje protok velikog broja vozila velikim brzinama, uobičajeno većima od 120 km/h, budući da nema oštih zavoja i velikih uzdužnih nagiba. Prema vrsti prometa na autocesti, vozila koja prometuju su: osobna vozila, osobna vozila s prikolicom, motocikli, kombi vozila, kamioni, kamioni s prikolicom, tegljači i autobusi.

Gradska autocesta ima mnogo ulaznih i izlaznih mjesta, dok izvangradska autocesta uobičajeno služi prometu na srednje i velike udaljenosti, a prolazi dovoljno daleko od naselja. Uz autocestu se grade prateći objekti koji pridonose sigurnosti i udobnosti vožnje (npr. benzinske postaje, servisne radionice, moteli, odmorišta).

Kada se uporaba autoceste naplaćuje, mora postojati i paralelna nenaplatna cesta. U otvorenom naplatnom sustavu vozači se, radi naplate cestarine, zaustavljaju samo na jednome mjestu na autocesti, odnosno odsječku autoceste, a u zatvorenom naplatnom sustavu zaustavljaju se i na ulazu i na izlazu. Sve se više širi negotovinska naplata, npr. elektroničkim naplatnim karticama, beskontaktnim uređajima ili naljepnicama kako bi se naplata ubrzala te smanjio negativan utjecaj naplate na brzinu odvijanja prometa.

Širenje autocestovne mreže nastoji se ubrzati izgradnjom poluautocesta, koje su projektirane kao buduće autoceste, ali se u prvoj fazi gradi samo jedan kolnički trak kojim se odvija dvosmjerni promet. U skladu s porastom prometnih potreba, u idućoj fazi gradi se i drugi kolnički trak [4].

3. Metode za grupiranje podataka

Grupiranje se odnosi na problem analize podataka pronalaska grupa u podacima, tako da objekti koji pripadaju jednoj skupini jesu međusobno slični i u isto vrijeme postoji značajna različitost elemenata različitih skupina [26]. Proces grupiranja podataka dijeli podatke u određeni broj grupa koji međusobno imaju zajedničke karakteristike/sličnosti. Kod grupiranja podataka, cilj je pronaći i grupirati međusobno slične podatke, a sličnost definira algoritam za grupiranje na temelju ulaznih podataka.

Da bi se pojam grupiranja podataka razumio, u nastavku su opisana područja kojima pripada grupiranje podataka. Za početak, nadzirano i nenadzirano učenje su dva temeljna koncepta strojnog učenja. Grupiranje podataka spada pod nenadzirano strojno učenje. Strojno učenje je grana umjetne inteligencije koja se bavi oblikovanjem algoritma koji svoju učinkovitost poboljšavaju na temelju naučenih podataka. Za kraj, umjetna inteligencija je područje računalne znanosti koje se bavi razvojem alata (strojeva, aparata, aplikacija) koje reagiraju i uče kao ljudi. U ovo područje spadaju i pojmovi poput strojnog učenja (engl. Machine learning) i Internet stvari (engl. Internet of things) [9].

Pojam grupiranja postoji dugo vremena u civilizaciji koju svi poznajemo. Danas grupiranje označava određenu podjelu informacija u digitalnom obliku, stvari, bića, itd., dok je prije dolaska računala taj pojam imao isto značenje izuzev informacija u digitalnom obliku. S današnjim tehnologijama nailazimo na veliku količinu podataka jer su razne vrste osjetnika (senzora) svuda oko nas i samim time postoji velika količina podataka. Ali ti svi podatci su besmisleni i nečitljivi dok se ne pretvore u informacije razumljive ljudima. Tu nastupaju razni programski jezici i metode unutar samih programskih alata koji te podatke obrađuju zajedno s ljudima koji pišu kôd.

Zašto je potrebno obrađivati podatke? Zato što upravo ti podatci ljudima pružaju temelj za određenu analizu, opažanje, zaključivanje, odluke te razumijevanje i rješavanje određenih problema. Podatci koji se grupiraju u iste grupe trebaju imati slična svojstva na temelju nekih kriterija. Grupa je skupina podataka koji moraju zadovoljavati određene kriterije kako bi sam podatak pripao toj grupi. Postoji činjenica koja govori da ljudi oduvijek imaju potrebu klasificirati. Kako bi shvatili/razumjeli novi objekt ili pojavu, ljudi oduvijek pokušavaju identificirati opisne značajke tih objekta ili pojava te ih potom usporediti sa značajkama ili svojstvima već poznatih na temelju njihovih sličnosti, odnosno različitosti. Za primjer, sva drveća su grupirana u razne kategorije (po imenima, veličini, ogrjevnoj moći, krošnji, itd.). Već samo imenovanje drveća je samo po sebi grupiranje. Na slici 10 vidljivo je grupirano voće po vrsti i boji [9].



Slika 10. Grupirano voće, [10].

Kod rješavanja problema grupiranja ne postoji specifičan algoritam koji može riješiti svaki problem. Postoje različiti algoritmi koji se u velikoj mjeri razlikuju, jer je različito razumijevanje pojma što je grupa te kako je najučinkovitije pronaći. Važno je istražiti karakteristike problema da bi se odabrao najbolji algoritam grupiranja.

Primjenu grupiranja moguće je pronaći u svim područjima znanstvenih i primijenjenih istraživanja. U ekonomiji grupiranje služi u grupiranju nabavljača (za određivanje povoljnih nabava), u marketingu gdje se grupiraju kupci sa sličnim zanimanjima u svrhu efikasnog reklamiranja. Grupiranje životinja primjer je grupiranja u biologiji. U prometu grupiranje se primjenjuje u identifikaciji prometnih čepova, prometnim podacima dobivenima od strane prometne infrastrukture, itd.

Jedan konkretan problem koji se rješava grupiranjem je identificiranje spam poruka. Spamom se može doći i do krađe identiteta. Poruke koje se nalaze u spam folderu su poruke koje algoritam identificira kao spam. Algoritam koji se pokazao kao efektivan algoritam za pronalazak neželjene elektroničke pošte (engl. spam) je algoritam k-means. Algoritam gleda različite odlomke poruke (zaglavlje, naslov, tekst poruke), riječi iz odlomaka onda filtrira kroz spam filter koji određuje rang korisnosti pojedinih odlomaka i riječi. Još neki zanimljivi problemi koji se mogu riješiti algoritmom k-means su: profiliranje sumnjivaca, optimizacija trgovine isporuke (optimalan broj stanica isporuke), segmentacija korisnika u marketingu i segmentacija slika [18].

3.1. Najčešće korišteni programski jezici za grupiranje podataka

Najčešći programski jezici za razvoj aplikacija temeljenih na strojnom učenju su R i Python. Također se mogu koristiti i drugi jezici kao što su Java, C++ te skriptni jezik unutar Matlab programskog okruženja koji je i korišten u diplomskom radu. U nastavku su kratko opisani programski jezici pojedinačno.

R je programski jezik i okruženje za statističke izračune i vizualizaciju. Slobodan je programski kôd što znači da se može slobodno koristiti i distribuirati te da je otvorenog kôda (engl. Open-source). R pruža širok izbor statističkih metoda za linearno i nelinearno modeliranje, klasične statističke testove, analize vremenskih serija i grupiranje podataka. Po svojoj prirodi nalazi se na razmeđi između klasičnih programskih jezika kao što su Python, Java ili C++ i statističkih alata kao što su SAS ili SPSS. Lako je proširiv s velikim izborom grafičkih tehnika. Statističari su razvili stotine specijaliziranih statističkih procedura za širok raspon uporabe putem tzv. pridodanih paketa (engl. contributed packages) koji su slobodno dostupni i integrirani direktno sa sustavom R [10]. R se danas nametnuo kao jedan od vodećih analitičkih programskih jezika s kojim se uz pomoć pratećih paketa na vrlo učinkovit način mogu provesti složene analize podatkovnih skupova te stvoriti izvještaj popraćen kompleksnim vizualizacijama i izračunima.

Python je programski jezik opće namjene, interpretirani, interaktivni, objektno orijentirani programski jezik visokog nivoa. Odlikuje ga visokokvalitetna struktura podataka koja ga u kombinaciji s dinamičkim pisanjem i vezivanjem čini veoma atraktivnim za brz razvoj različitih aplikacija. Python je besplatan programski alat, a podržava ga većina operativnih sustava, kao što su Mac, Windows ili Linux. Rasprostranjenost Pythona je široka, a ovo su samo neke od industrija u kojima je ovaj popularni programski jezik uspješno primijenjen: desktop grafički razvoj aplikacija (uključujući i igrice), matematička i naučna analiza podataka, razvoj interneta (Python danas koristi neki od najvećih internet domena na svijetu poput Google-a, Facebook-a, Reddit-a, Dropbox-a, YouTube-a, Instagram-a, Pinterest-a, itd.) [11].

Unutar IT zajednice česte su kritike Pythona na račun njegove sporosti. Pošto je Python interpreterski jezik, programi napisani u njemu vrše se malo sporije za usporedbu od kompajlerskih jezika, kao što su C, C++ i slični. Međutim, unatoč toj brzinskoj manjkavosti, u industriji se Python poprilično koristi (ponajviše kao back end programski jezik). Python se često uspoređuje s Javom. Oboje su interpreterski jezici, i oboje imaju gotovo nikakvu podršku za višejezgreno izvođenje programa, pošto i Python i Java koriste samo jednu procesorsku jezgru. Java se kao jezik češće primjenjuje u izradi mobilnih aplikacija i interaktivnog web sadržaja, dok je Python gospodar PC svijeta. Što se tiče brzine izvođenja programa, Java i Python su približno jednaki [12].

Matlab (Matrix Laboratory) je visoko učinkovit (engl. high-performance) programski jezik namijenjen za tehničke proračune. Objedinjava računanje, vizualizaciju i programiranje u lako uporabljivoj okolini u kojoj se problem i rješenje definirani poznatom matematičkom notacijom. Uobičajena je uporaba Matlab-a za:

- Matematiku i izračune;
- Razvoj algoritama;
- Modeliranje, simulaciju, analizu;
- Analizu i obradu podataka, vizualizaciju;
- Znanstvenu i inženjersku grafiku;
- Razvoj aplikacija, uključujući i izgradnju (engl. graphical user interface, GUI), interaktivno korisničko grafičko sučelje.

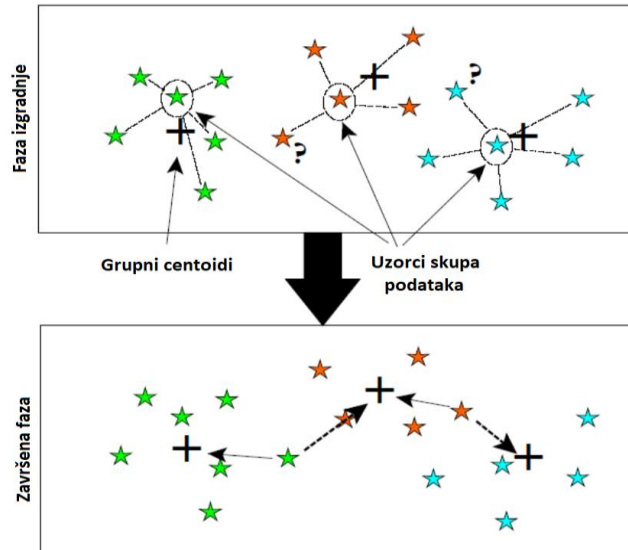
Matlab je i okružje i programski jezik. Jedna od jačih strana Matlab-a je činjenica da njegov programski jezik omogućava izgradnju vlastitih alata za višekratnu uporabu. Svaki korisnik može sam kreirati vlastite funkcije i programe (poznate kao M-datoteke) u kôdu MATLAB-a. Skup specijaliziranih M-datoteka za rad na određenoj klasi problema naziva se Toolbox. S Matlab-om dolazi niz Toolbox-ova koji su i više od kolekcije korisnih funkcija. Oni predstavljaju rezultate istraživanja vrhunskih stručnjaka iz područja upravljanja, obrade signala, identifikacije procesa, i drugih. Dakle uz pomoć MATLAB-a može se razviti nove ili adaptirati postojeće Toolbox-ove za rješavanje određenih problema [13].

3.2. Najčešći algoritmi za grupiranje podataka

Sada kada se priča o grupiranju podataka, postoje razni algoritmi koji pomažu pri grupiranju podataka kada nije oku jasno vidljivo koji podatak spada pod koju grupu. Neki od algoritma za grupiranje su [19]:

- a) Particijsko grupiranje;
 - b) Hijerarhijsko grupiranje;
 - c) Grupiranje zasnovano na gustoći podataka;
 - d) Grupiranje zasnovano na mreži podataka.
- a) Particijsko grupiranje;

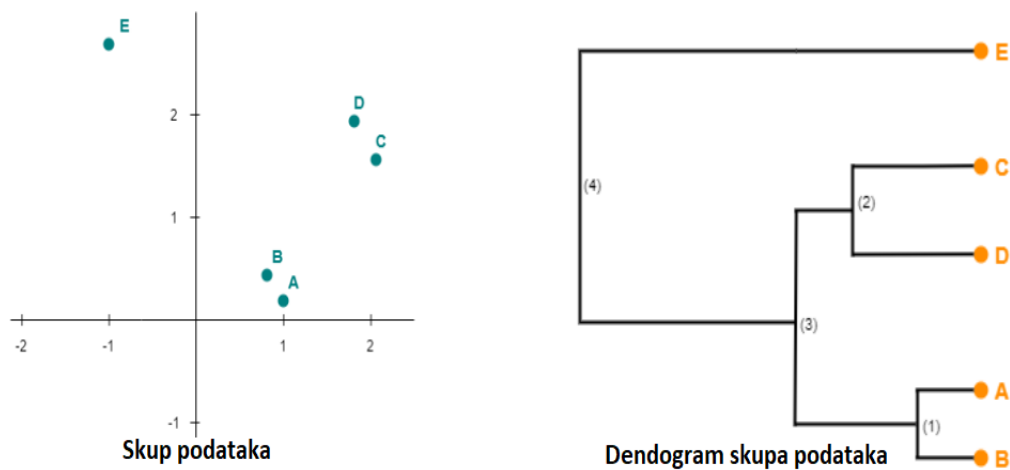
Particijsko grupiranje je podjela skupa podataka u grupe koji sadrže slične podatke. Grupe predstavlja centroid (centar grupe), odnosno podatak koji sažima opis svih podataka u toj grupi. Definicija centroida grupe ovisi o vrsti podataka koja se proučava. Na primjer, ako postoji zadan skup realnih brojeva tada će njihov centroid biti aritmetička sredina danog skupa. Pod particijsko grupiranje spadaju algoritmi: K-means, K-metoda, neizrazitih C-srednjih vrijednosti i algoritam maksimizacije očekivanja [19]. Na slici 11 vidljive su faze particijskog grupiranja kao i položaji centroida, uzoraka skupa podataka i grupa.



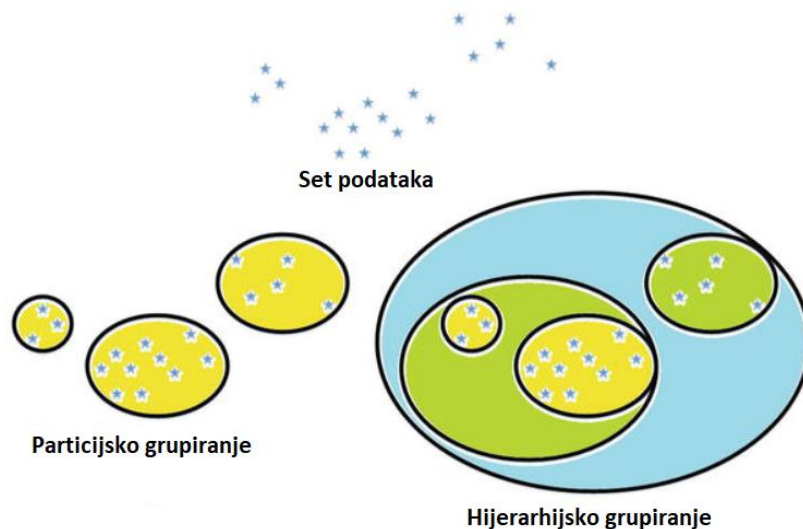
Slika 11. Particijsko grupiranje, [20].

b) Hijerarhijsko grupiranje:

Za razliku od particijskog grupiranja, hijerarhijsko grupiranje rezultira hijerarhijom grupa od uzoraka (slika 12). Hijerarhija grupa prikazuje se dendrogramom. Dendrogram je stablo u kojem listovi odgovaraju podacima (točke A-E), a vodoravne linije (1-4) odgovaraju povezivanjima na određenoj udaljenosti. Ovakav prikaz grupiranja je zanimljiv jer se može presjeći na bilo kojoj udaljenosti i dobiti grupe koji bi se dobili particijskim grupiranjem na toj udaljenosti. Hijerarhijsko grupiranje provodi se pomoću funkcije udaljenosti ili mjere sličnosti, s ciljem pronalaska grupa podataka koji su najbliži jedni drugima. Postoje dvije vrste hijerarhijskog grupiranja: aglomerativno i divizno [19].



Slika 12. Skup podataka i dendrogram tog skupa, [19].

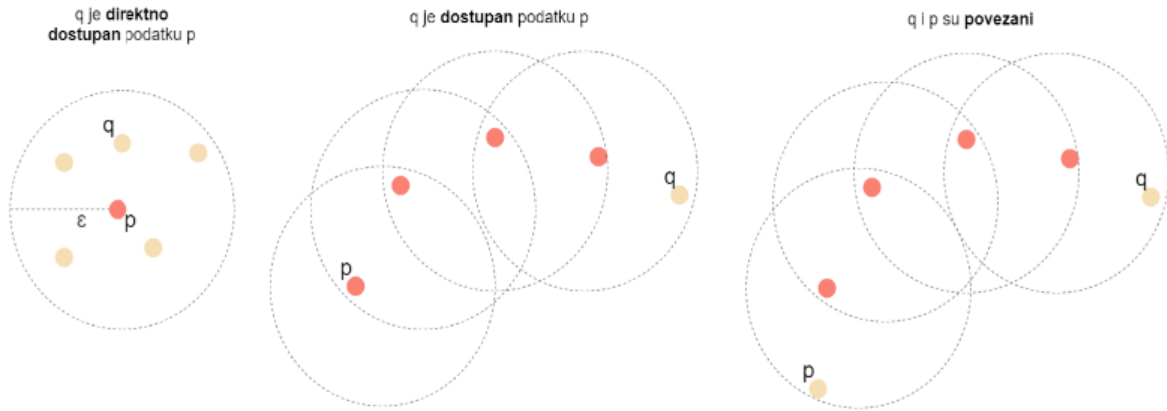


Slika 13. Tehnike grupiranja, [21].

Slika 13 prikazuje hijerarhijsko i particijsko grupiranje uz neobrađeni skup podataka. U hijerarhijskom grupiranju, sve podgrupe su definirane u drugim podgrupama. U hijerarhijskom grupiranju, sirovi se podaci prvo grupiraju u neke podgrupe (oblik sa tri skupine). Nakon tog postupka, podgrupe su hijerarhijski definirane u dvije zelene grupe. Posljednji postupak uključuje sve te grupe definirane u skupu unija [21].

c) Grupiranje zasnovano na gustoći podataka;

Algoritmi za grupiranje temeljeni na gustoći podataka su vrsta tehnike grupiranja koja u osnovi primjenjuje kriterij lokalne grupe. Grupe se promatraju kao regije u prostoru u kojem podaci tvore gusto područje i razdvojeni su regijama s malom gustoćom podataka. Guste regije podataka ponekad mogu stvoriti neke proizvoljne oblike i podaci mogu biti nasumično raspoređeni unutar te regije. Stoga, algoritmi temeljeni na gustoći podataka mogu lako prepoznati grupe bilo kakvih oblika, ali se oslanjaju na uvjetu da podaci unutar određene grupe čine stisnutu regiju. Na primjer, za rudarenje podataka je traženje van-distribucijskih vrijednosti važnije od pronalaska običnih slučajeva. Jedni od algoritama zasnovani na gustoći podataka su: DBSCAN algoritam (engl. Density-based spatial clustering of application with noise), GDBSCAN algoritam (engl. Generalized density-based spatial clustering of application with noise) i OPTICS algoritam (engl. Ordering points to identify the clustering structure) [19].



Slika 14. Direktna dostupnost, dostupnost i povezanost podataka, [19].

Slika 14 prikazuje grupiranje podataka DBSCAN algoritmom. S obzirom na dani skup podataka, algoritam grupira podatke koji su bliski jedni drugima tako da van-distribucijske vrijednosti ostanu području s malom gustoćom podataka. Uspješno pronalazi proizvoljne oblike sve dok grupe stvaraju gustu regiju. DBSCAN algoritam je temeljen na konceptu dostupnosti temeljenoj na gustoći, koju definiramo na sljedeći način:

Podatak q je **direktno dostupan** (engl. direct density-reachable, DDR) podatku p ako je p udaljen od q za najviše ϵ te ako postoji dovoljan broj podataka oko p tako da može biti stvorena grupa oko p i q . Vidljivo je da relacija direktne dostupnosti nije simetrična, tj. ako je p direktno dostupan podatku q , ne mora značiti da je q direktno dostupan podatku p .

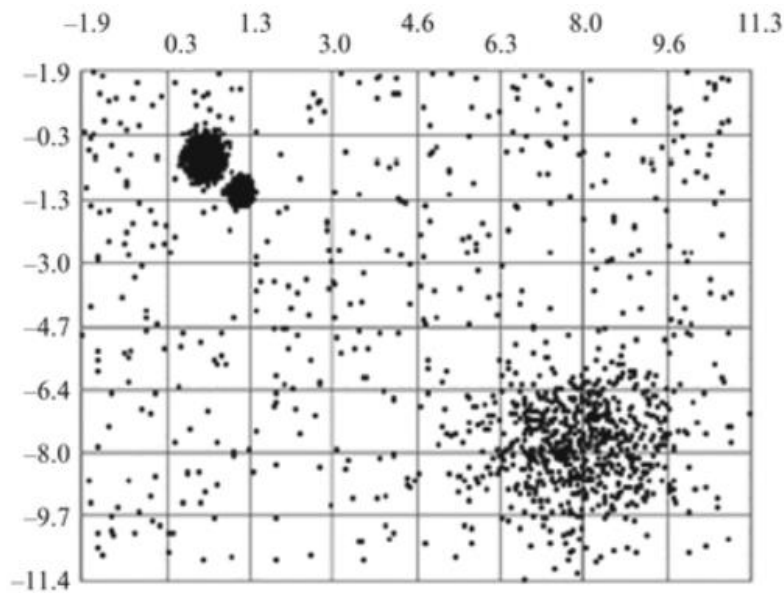
Podatak q je **dostupan** (engl. density reachable, DR) podatku p ako postoji niz podataka p_1, \dots, p_n , $p_1 = p$, $p_n = q$, gdje je svaki p_{i+1} direktno dostupan (DDR) podatku p_i . Također, relacija nije simetrična jer podatak p može biti dostupan podatku q , ali q može ležati na rubu grupe i tada neće imati dovoljan broj susjeda da se broji kao pravi element grupe.

Uz to, definiramo i **povezanost** baziranu na gustoći (engl. density-connectedness, DC) kao: dva podatka p i q su povezana ako postoji niz podataka o_1, o_2, \dots, o_n takvi da su o_1 i p dostupni, o_2 i o_1 dostupni, o_3 i o_2 dostupni, ..., te su o_n i q dostupni.

Grupa pronađena DBSCAN algoritmom mora zadovoljavati dva uvjeta: svi podatci unutar određene grupe moraju biti međusobno povezani te ako je podatak povezan s drugim podatkom u grupi, tada je on uključen u strukturu grupe [19].

d) Grupiranje zasnovano na mreži podataka.

Slično kao grupiranje zasnovano na gustoći, grupiranje zasnovano na mreži podataka često se koristi za određivanje grupa u velikim višedimenzionalnim prostorima (slika 15). Ponovno se grupe promatraju kao guste regije. Vremenska složenost većinskih tehnika grupiranja linearno ovisi o veličini danog skupa. Glavna prednost grupiranja zasnovanog na mreži je sposobnost rada s velikim skupovima podataka. Osnovna razlika grupiranja zasnovanog na mreži od onog zasnovanog na gustoći podataka je što ovi algoritmi ne rade sa skupovnim podatcima već s okolnim prostorom. Često korišten algoritam u ovoj tehnici grupiranja je STING (engl. Statistical information grid). On se uglavnom koristi za grupiranje prostornih baza podataka [19].



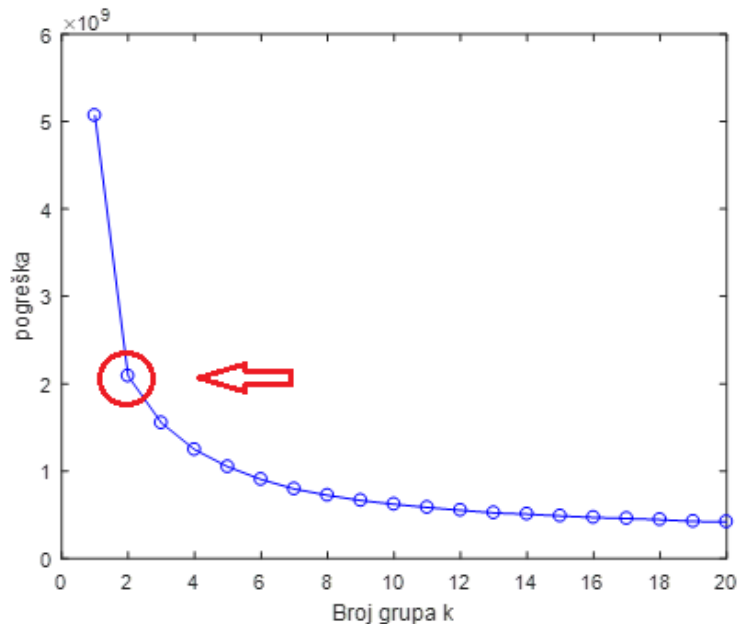
Slika 15. Mreža podataka, [19].

Iduća faza nakon grupiranja podataka određenim algoritmom je problem kako odrediti optimalni broj centroida/grupa (parametar k). U ovom diplomskom radu je slučaj takav da je broj k nepoznat u početku implementacije algoritma. Zato postoje metode koje se koriste za odabir optimalnog broja grupa k , a one su:

- a) Analiza metodom lakta (engl. Elbow method);
- b) Analiza metodom Siluete (engl. Silhouette analysis);
- c) Analiza metodom Calinski-Harabasz (engl. T.Calinski and J.Harabasz analysis, CH);
- d) Analiza metodom Davies-Bouldin (engl. Davies and Bouldin analysis, DB).

a) Analiza metodom lakta:

Analiza metodom lakta prikazuje optimalan broj centroida/grupa k tako što pokreće K-means algoritam više puta i prikazuje kvadratno odstupanje (grešku) za svaki broj k . U laktu dijagrama (označeno crvenom bojom) treba se nalaziti optimalan broj grupa za K-means algoritam. Ako se pretpostavi da je broj grupa $k = n$, onda je pogreška za $k = n - 1$ puno veća nego za $k = n$ [14]. U slučaju kao na slici 16, optimalan broj grupa k iznosi 2.



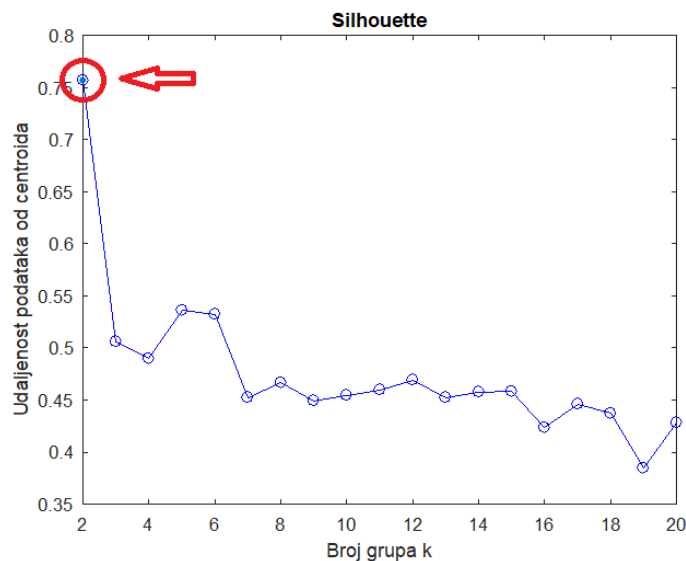
Slika 16. Analiza metodom lakta

Ako se broj grupa povećava tada vrijednost funkcije cilja opada. Zato se za optimalni broj grupa uzima onaj broj grupa za koji vrijednost funkcije cilja naglo opada. Kada se to prikaže na grafu ovisnosti funkcije cilja o broju grupa (slika 16), vidljiv je lom linije grafa poput lakta. Od tud naziv za metodu lakta. Ponekad taj kriterij odabira broja grupa nije dovoljan, pa je potrebno koristiti ostale spomenute metode/analize za pronalazak optimalnog broja grupa.

b) Analiza metodom siluete:

Analiza Siluete je metoda koja se također koristi za određivanje optimalnog broja grupa k . Izračunava udaljenost svake točke od dodijeljenog centroida i izračunava udaljenosti svake točke od susjednih centroida. Vrijednosti siluete mogu se kretati u intervalu $[-1, 1]$. Vrijednost -1 označava da je točka/podatak vrlo udaljena od centroida kojemu je dodijeljena i zaključuje se da je loše klasificirana. Ako je vrijednost Siluete 0 , to znači da se točka nalazi na granici između susjednih centroida. Vrijednost 1 označava točku/podatak vrlo blizu dodijeljenom centroidu i to

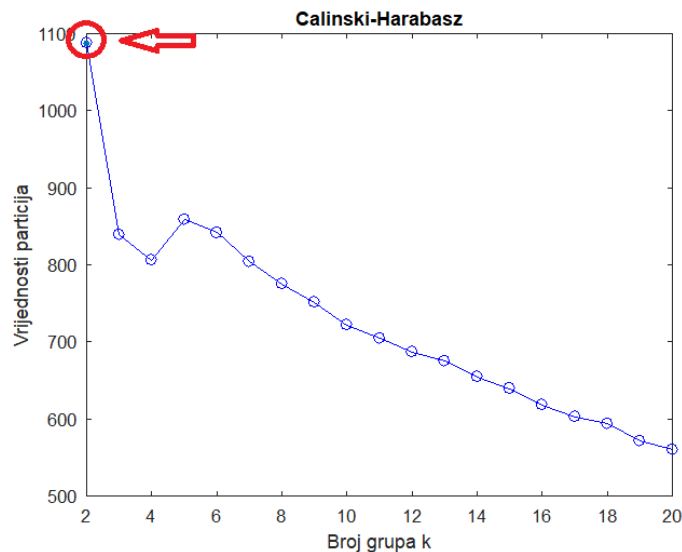
za ovu metodu predstavlja idealan slučaj grupiranja [14]. Prema slici 17 optimalan broj grupa k iznosu 2 jer gledamo maksimum krivulje (označeno crvenom bojom).



Slika 17. Analiza Metodom siluete.

c) Analiza metodom Calinski-Harabasz:

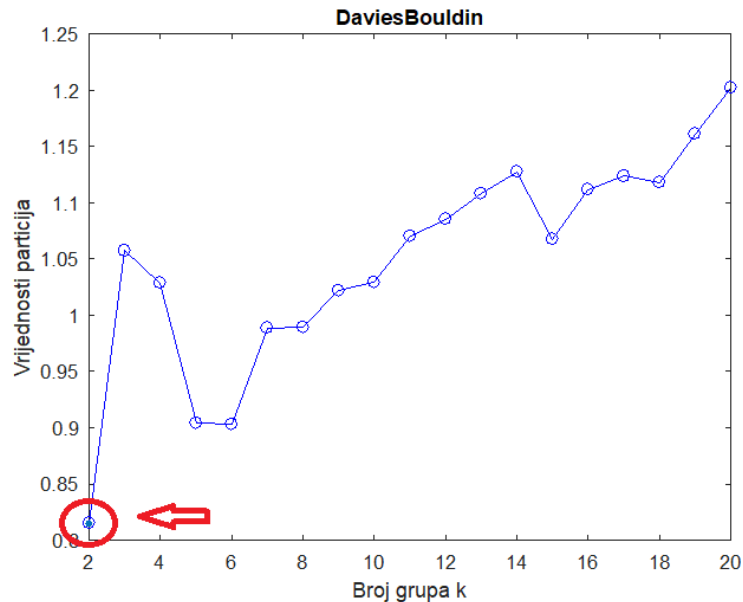
Analiza metodom Calinski-Harabasz (CH) je definirana tako da interno kompaktnija particija čije su grupe dobro međusobno razdvojene imaju veću CH vrijednost. Što znači da se particija s najvećom CH vrijednošću smatra optimalnom [18]. Prema slici 18 optimalan broj grupa k iznosu 2 jer se gleda maksimum krivulje (označeno crvenom bojom).



Slika 18. Analiza metodom Calinski-Harabasz.

d) Analiza metodom Davies-Bouldin.

Analiza metodom Davies-Bouldin (DB) je definirana tako da interno kompaktnija particija čije grupe su međusobno bolje razdvojene ima manju DB vrijednost. Što znači da se particija s najmanjom DB vrijednošću smatra optimalnom [18]. Prema slici 19 optimalan broj grupa k iznosi 2 jer se gleda minimum krivulje (označeno crvenom bojom).



Slika 19. Analiza metodom Davies-Bouldin.

Najjednostavniji način za odrediti optimalni broj grupa k je vizualizacijom rezultata podataka ako su jednodimenzionalni ili dvodimenzionalni jer se tada mogu prikazati u ravnini. U slučaju ovog diplomskog rada, vizualni odabir optimalnog broja grupa k nije moguć jer se podaci ne mogu lijepo prikazati, a i vizualno određeni parametar k je subjektivnog osjećaja.

U ovom diplomskom radu za analizu grupiranja podataka korišten je algoritam K-means i SOM koji su detaljnije opisani u nastavku.

3.3. Algoritam K-means

Algoritam k-means je najpoznatiji i najjednostavniji algoritam za grupiranje podataka koji je svrstan u algoritme nenadziranog učenja. Tipičan je algoritam grupiranja temeljen na udaljenosti, a udaljenost je vrednovani pokazatelj sličnosti. Ako je udaljenost dva objekta mala, njihova sličnost je veća [24]. Koristi se u slučajevima kada se obrađuje ne označeni skup podataka kojega je potrebno podijeliti u fiksni broj grupa u kojima podaci posjeduju slične karakteristike. Glavni cilj je definirati k centroida za svaku grupu, gdje su grupe modelirane euklidskom udaljenošću od središta. Prema slici 20, prvi korak algoritma je odabir broja grupa i inicijalno

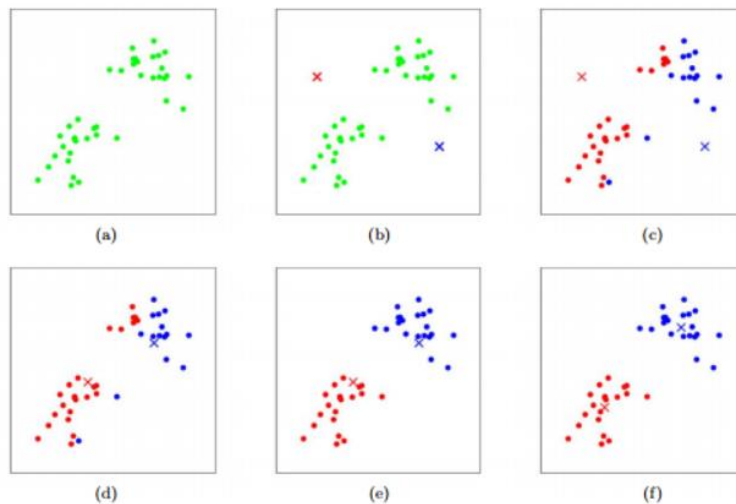
pozicioniranje centroida. Nakon pozicioniranja centroida, računa se udaljenost između određenih centroida i svakog podatka. Centroidi se pomiču s obzirom na aritmetičku sredinu udaljenosti od podataka i ovaj postupak se iterativno ponavlja sve dok nema promjena u poziciji centroida [14].

Algoritam pronalazi najmanju udaljenost između centroida i podataka, stoga se ciljna funkcija može izraziti kao (1):

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2, \quad (1)$$

gdje je:

- $\|x_i^{(j)} - c_j\|^2$ – udaljenost između $x_i^{(j)}$ i c_j ,
- $x_i^{(j)}$ – podatak,
- c_j – centroid grupe.



Slika 20. K-means algoritam kroz iteracije, [19].

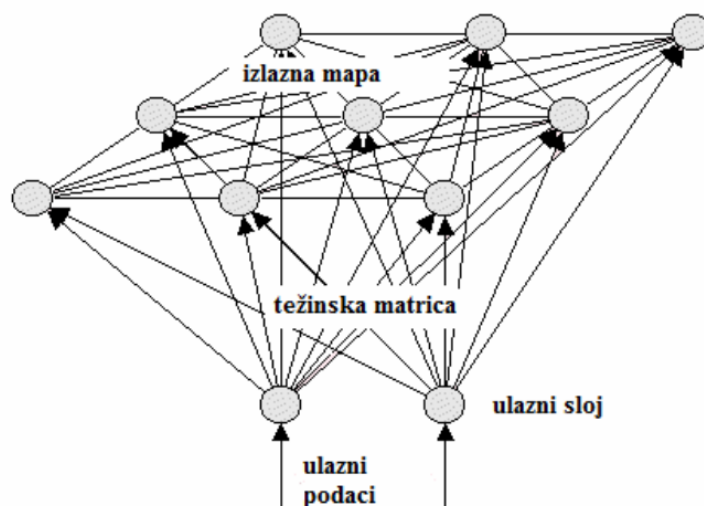
K-Means ima prednost što je prilično brz, jer sve što radi je računanje udaljenosti između točaka i središta grupa; vrlo malo izračuna. Također počinje nasumičnim odabirom centara grupe i stoga može dati različite rezultate grupiranja na različitim izvođenjima algoritma. Stoga se rezultati ne mogu ponoviti i nedostaje im dosljednost. K-means algoritam se koristi naravno i u području prometa. Neki od primjera su:

- Grupiranje prometnih puteva i raskrižja prema brzini i repovima čekanja (LOS);
- Grupiranje podataka dobivenih sa autocesta zbog boljeg sustava upravljanja (pretputnog i putnog informiranja korisnika), vizualizacije prometnih parametara, procjene stanja prometnog sustava i prijedloga mogućih rješenja, itd;

- Analiza učestalosti prometnih skupina (biciklisti, automobili, taksiji) na pojedinim prometnim putevima.

3.4. Samoorganizirajuće mape

Samoorganizirajuće mape (engl. Self-organizing map, SOM) su vrste umjetnih neuronskih mreža (engl. Artificial neural network, ANN) i zapravo se mogu definirati kao dvoslojne neuronske mreže. Naziv „Samoorganizirajuća“ jer nije potreban nadzor, a „Mapa“ jer pokušava mapirati svoje težine u skladu s datim ulaznim podacima [25]. ANN je model bioloških neuronskih mreža ljudskog živčanog sustava kojega je 1980-ih uveo Finski profesor Teuvo Kohonen i ANN se ponekad naziva Kohonenova mreža. ANN se uvježbava pomoću nenadziranog učenja za proizvodnju nisko-dimenzionalni (obično dvodimenzionalni), diskretizirani prikaz ulaznog prostora uzoraka treninga, nazvan mapa, i stoga je metoda za smanjenje dimenzionalnosti. To je zapravo, paralelan spoj međupovezanih (svaki sa svakim) neurona koji su organizirani u slojevima. Također, treba napomenuti da svaka veza među neuronima ulazno-izlaznog sloja ima svoju pripadnu težinu što je vidljivo na slici 21 gdje je prikazan koncept samo-organizirajućih mapa [15], [16].



Slika 21. Koncept modela samoorganizirajućih mapa, [15].

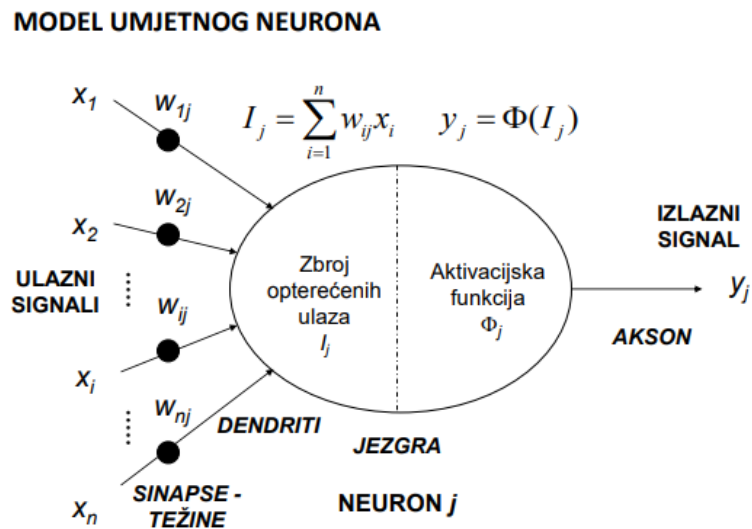
Razni osjetilni podražaji kod ljudi su neurološki mapirani u mozgu tako da prostorni odnos među podražajima određenih dijelova mozga odgovara prostornom odnosu među neuronima organiziranim u dvodimenzionalnu mapu. Tako lokacija točaka na mapi pokazuje relativnu sličnost točaka u višedimenzionalnom prostoru, tj. što su dva podataka sličnija, to će bliže biti prikazani na dvodimenzionalnoj samoorganizirajućoj mapi (slika 24). U skladu s time, primarna funkcija SOM-a je prikazati višedimenzionalne podatke s ulaza u mrežu, na dvodimenzionalnu mapu, pritom zadržavajući odnose među podacima [15].

Premda su umjetne neuronske mreže po svojim sposobnostima obrade podataka daleko ispod mogućnosti obrade bioloških neuronskih mreža, one ipak pokazuju sposobnosti [17]:

- Učenja na temelju iskustva - prilagođavanje okolini promjenom ponašanja;
- Poopćavanja - neosjetljivost na male promjene ulaznih podataka;
- Odjeljivanja bitnog od nebitnog - apstrakcija;
- Prepoznavanja uzoraka - grupiranje, predviđanje, identifikacija, optimizacija, prilagodljivo vođenje.

Elementi umjetnog neurona vide se na slici 22 [17]:

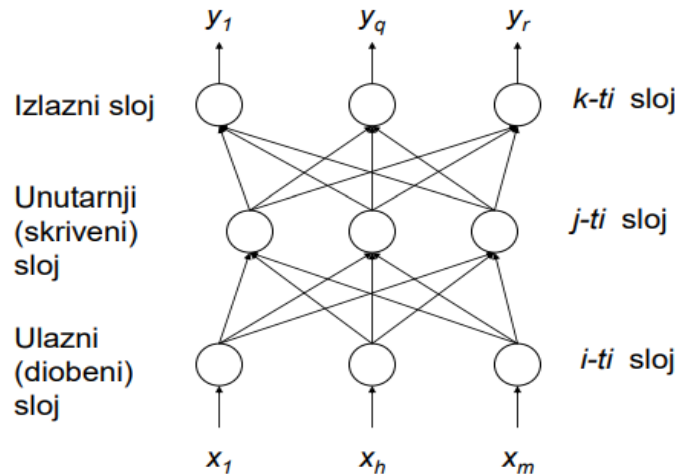
- 1) Ulazni signali - veličine kontinuiranih vrijednost;
- 2) Težine - prilagodljive veličine i pozitivne ili negativne vrijednosti;
- 3) Jezgra - zbrajanje opterećenih ulaznih signala i aktivacijska funkcija (prijenosna funkcija) neurona;
- 4) Izlazni signal.



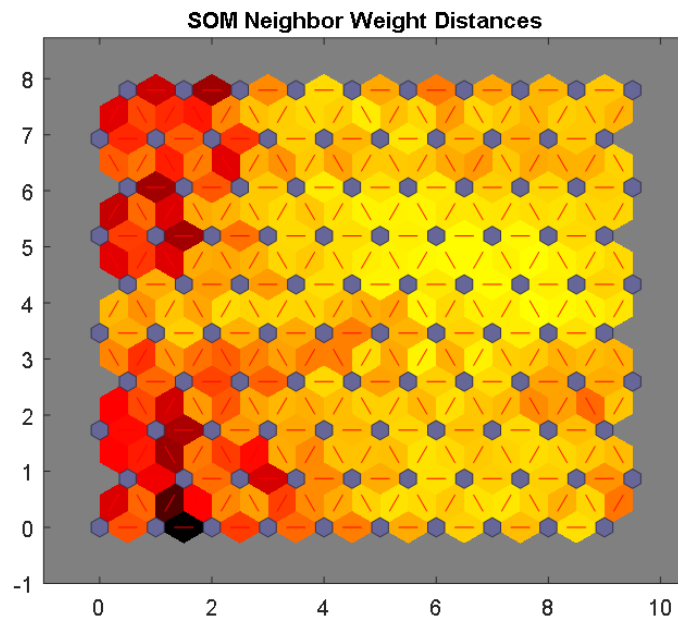
Slika 22. Elementi umjetnog neurona, [17].

Arhitektura neuronske mreže (slika 23) [17]:

- Ulazni sloj - Sučelje za predaju podataka iz okoline neuronskoj mreži;
- Unutarnji (skriveni) sloj ili slojevi - Nisu povezani s okolinom i obrađuju signalne;
- Izlazni sloj - Predaje rezultate obrade neuronske mreže okolini.



Slika 23. Arhitektura neuronske mreže, [17].



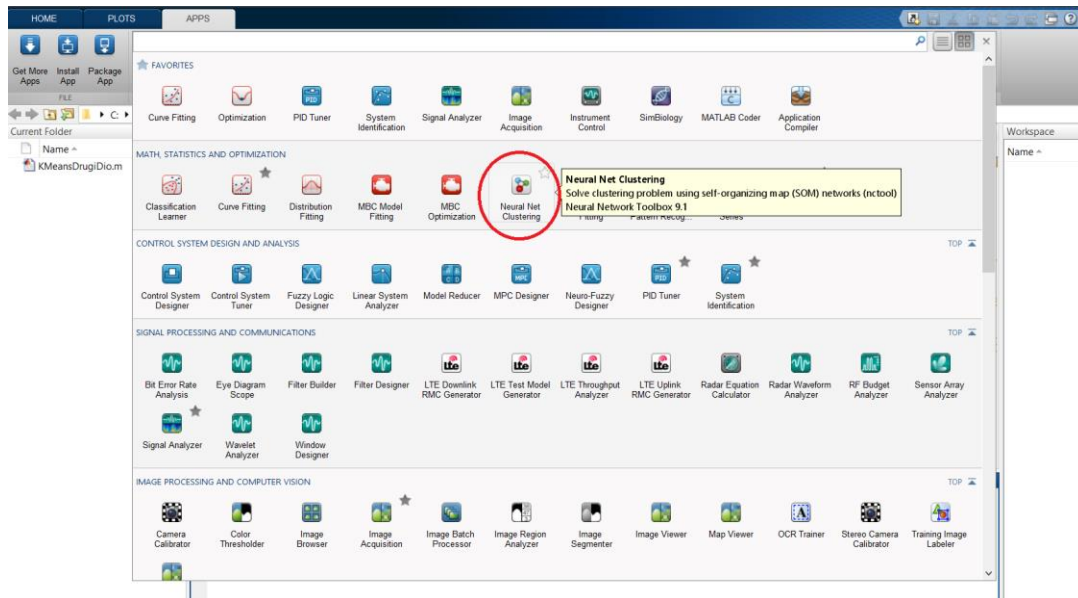
Slika 24. Dvodimenzionalna samoorganizirajuća mapa.

Kao i većina umjetnih neuronskih mreža, i SOM-ovi djeluju u dva načina: trening i mapiranje. "Trening" gradi kartu koristeći primjere unosa (natjecateljski postupak, također nazvan kvantizacija vektora), dok "mapiranje" automatski grupira novi ulazni vektor. Vidljivi dio samoorganizirajuće mape (slika 24) je prostor mape koji se sastoji od komponenta koje se nazivaju čvorovi ili neuroni. Prostor mape unaprijed je definiran, obično kao konačno dvodimenzionalno područje gdje su čvorovi poredani u pravilnu šesterokutnu ili pravokutnu mrežu. Svaki čvor povezan je s vektorom "težine", koji je položaj u ulaznom prostoru, odnosno

ima istu dimenziju kao svaki ulazni vektor. Dok čvorovi u prostoru karte ostaju fiksni, trening se sastoji u pomicanju vektora težine prema ulaznim podacima (smanjenje metrike udaljenosti) bez kvarenja topologije inducirane iz prostora karte. Prema tome, samoorganizirajuća karta opisuje preslikavanje iz ulaznog prostora veće dimenzije u prostor karte niže dimenzije. Jednom osposobljena, karta može klasificirati vektor iz ulaznog prostora pronalaženjem čvora s vektorom težine (najbliže metričke udaljenosti) najbližem vektoru ulaznog prostora [16].

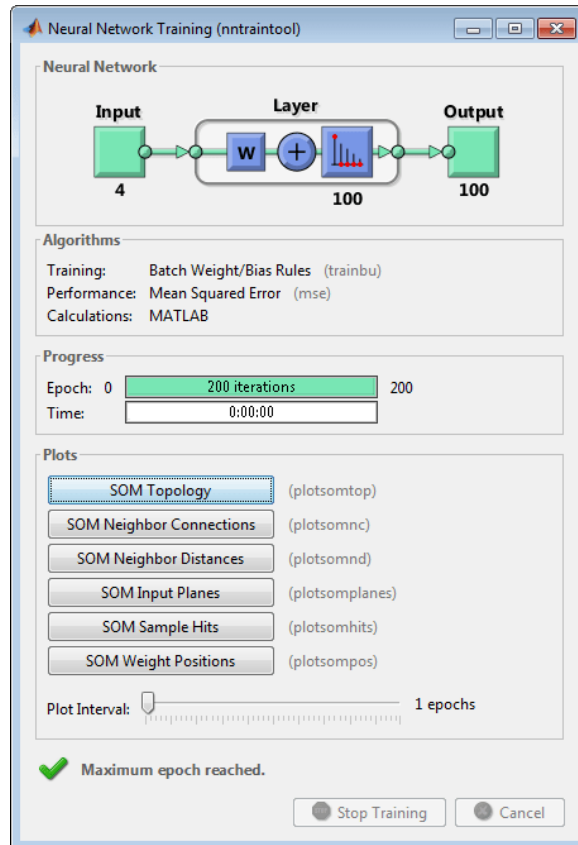
4. Implementacija neuronskih mreža za grupiranje podataka

U Matlab programskom alatu postoji aplikacija grupiranje neuronskih mreža (engl. Neural Net Clustering, NCTOOL) koja je prikazana slikom 25. NCTOOL rješava problem grupiranja pomoću SOM-a i te podatke grupira prema sličnosti. Ta aplikacija pomaže pri odabiru podataka, definiranju mrežne arhitekture i osposobljavanju mreže. Mogu se odabrati vlastiti podatci iz radnog prostora Matlab ili upotrijebiti jedan od primjera skupova podataka. Nakon obuke mreže, omogućeno je analiziranje rezultata pomoću različitih alata za vizualizaciju.



Slika 25. Aplikacija Neural Net Clustering.

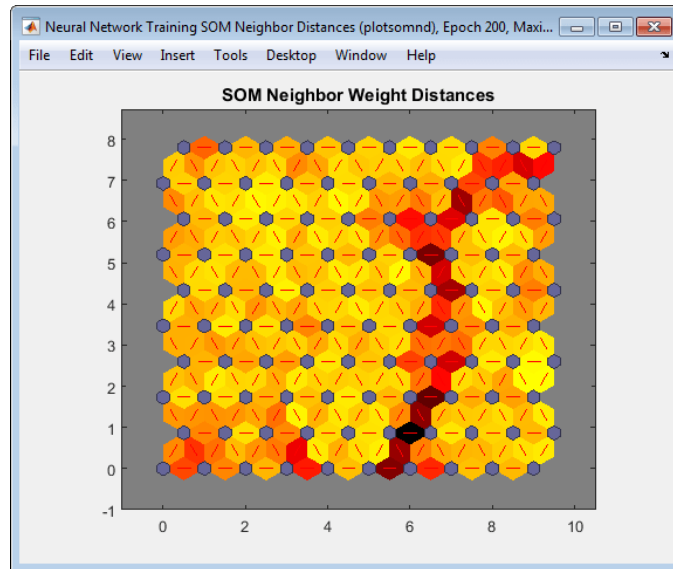
SOM se sastoji od konkurentnog sloja koji može klasificirati skup vektora s bilo kojim dimenzijama u onoliko grupa koliko sloj ima neurona. Neuroni su raspoređeni u 2D topologiji, što omogućuje sloju da formira prikaz distribucije i dvodimenzionalnu aproksimaciju topologije skupa podataka. Mreža se trenira pomoću SOM paketnog algoritma (trainbu, Learnsomb).



Slika 26. Prikaz vrsta vizualizacija podataka.

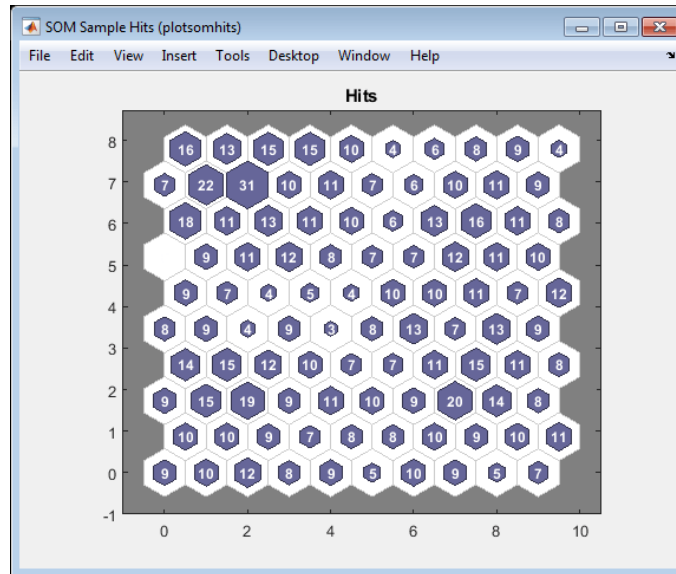
Kada se NCTOOL za grupiranje podataka pravilno podesi za trening i mapiranje, onda dobiveni rezultati bivaju prikazani sljedećim prikazima:

- 1) prikaz udaljenosti neurona/grupa opisuje razlike između neurona. Prikaz se vrši odabirom na SOM Neighbor Distances prema slici 26. Svaki neuron prikazuje jednu grupu. Na slici 27 sivi šesterokuti predstavljaju neurone. Crvene tanke linije povezuju susjedne neurone. Boje u regijama koje sadrže te tanke crvene crte označavaju udaljenosti između neurona. Tamnije boje predstavljaju veće udaljenosti, a svjetlije boje predstavljaju manje udaljenosti. Neuroni koji imaju malu udaljenost između sebe predstavljaju slične podatke. SOM algoritam raspoređuje članove (neurone) u dvodimenzionalnu mrežu stavljajući slične članove bliže jedni drugima, gdje više sličnih članova (skupina) tvori jednu grupu. Član se sastoji od ulaznih podataka koji su međusobno slični. Hijerarhijski gledano, ulazni podatci koji su međusobno slični tvore jedan član, a određena skupina tih sličnih članova čini grupu. Topologija mreže je dimenzije 10 x 10, pa postoji 100 neurona. Traka tamnih segmenata prelazi iz područja donjeg središta u gornje desno područje i čini granicu dviju grupa.



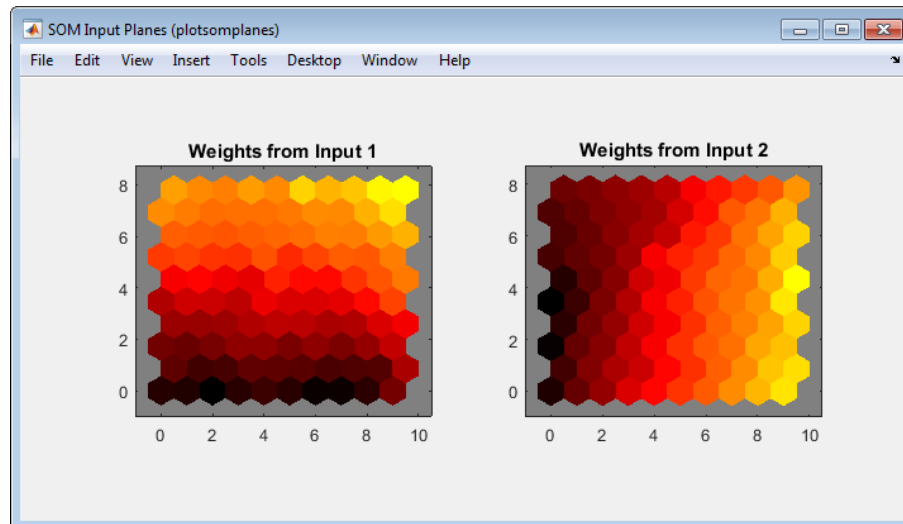
Slika 27. Prikaz udaljenosti između neurona, [22].

- 2) prikaz broja sličnih ulaznih pogodaka koji predstavlja svaki neuron, a naposljetku skupina tih neurona tvori jednu grupu. Prikaz se vrši odabirom na SOM Sample Hits prema slici 26. Za SOM trening, vektor težine povezan sa svakim neuronom pomiče se i postaje središte skupine ulaznih vektora. Osim toga, neuroni koji su susjedni u topologiji trebali bi se također približavati jedan drugom u ulaznom prostoru. Zadana topologija SOM-a je šesterokutna. Ova slika prikazuje lokacije neurona u topologiji i pokazuje koliko je podataka o treningu povezano sa svakim od neurona (centri grupa). Topologija mreže je dimenzije 10 x 10, pa postoji 100 neurona. Maksimalan broj pogodaka povezanih s bilo kojim neuronom prikazuje broj u svakom članu/neuronu i u ovom slučaju on iznosi 31. Dakle, u toj skupini postoji broj ulaznih vektora koliko piše na neuronu, vidljivo na slici 28 i on iznosi 31.



Slika 28. Prikaz broja sličnih ulaznih podataka po neuronima, [22].

- 3) prikaz težinskih ravnina gdje broj težinskih ravnina ovisi o ulaznim podacima. Prikaz se vrši odabirom na SOM Weight Planes prema slici 26. Slika 29 prikazuje težinsku ravninu za svaki element ulaznog vektora (u ovom slučaju dva). Oni su vizualizacije težina koje povezuju svaki ulaz sa svakim od neurona. Tamnije boje predstavljaju veće težine. Ako su prikazi povezivanja dva ulaza bili vrlo slični, tada su ulazi jako povezani. U ovom slučaju, ulaz 1 ima veze koje se jako razlikuju od onih na ulazu 2.



Slika 29. Prikaz težinskih ravnina, [22].

5. Dostupni prometni podaci

Brojenjem prometa obuhvaćena je mreža državnih te značajan dio županijskih i lokalnih cesta, a dodatno se prikupljaju, obrađuju i prikazuju prometni podaci s autocesta, drugih cesta i objekata s naplatom uporabe te državnih trajektnih linija odnosno s preko 950 brojačkih mjesta. Dostupni prometni podaci prikupljeni su s državne autoceste A1 kod Karlovca. Prometni podaci su prikupljeni na način neprekidnog automatskog brojanja [8].

Sustavno brojanje prometa se odvija na:

- Državnim cestama;
- Cestovnim građevinama s naplatnim postajama;
- Županijskim cestama;
- Lokalnim cestama;
- Trajektnim pristaništima.

Za dosadašnje prikupljanje podataka o količini i strukturi prometa, korištena su tri osnovna načina brojenja [23]:

- Neprekidno automatsko brojenje brojilima na državnim, županijskim i lokalnim cestama;
- Povremeno automatsko brojenje brojilima na državnim, županijskim i lokalnim cestama;
- Naplatno brojenje na autocestama, drugim cestovnim građevinama s naplatom prolaska vozila i u trajektnim pristaništima.

Republika Hrvatska ima poprilično dobru količinu prometnih podataka u bazi podataka jer se brojanje prometa izvodi sustavno još od ranih sedamdesetih godina prošlog stoljeća. Time je stvoren bogat fond podataka nužnih za racionalno gospodarenje cestama, što podrazumijeva istraživanje, planiranje, projektiranje, izgradnju, rekonstrukciju i održavanje cesta. S tim prikupljenim prometnim podacima *Hrvatske Autoceste d.o.o.* izrađuju prosječni godišnji dnevni promet (PGDP), prosječni ljetni dnevni promet (PLDP), utvrđivanje osnovnih promjena prometa (porast ili pad prometa u odnosu na prethodnu godinu) te izračunan prosječnog godišnjeg postotka promjene PLDP-a [23].

5.1. Dobiveni prometni podatci

Prometni podatci nad kojima je vršena obrada podataka su datoteke s nastavkom .xlsx. Za osnovnu prilagodbu/organiziranje dobivenih prometnih podataka korišten je programski alat Excel iz programskog paketa Microsoft Office. Dobivene su dvije tablice, gdje su podatci organizirani po satnim rezolucijama za tri godine (2018, 2019, 2020) posebno za voznu i pretjecajnu prometnu traku i oba smjera (Zagreb-Karlovac i Karlovac-Zagreb). Rezolucija prikupljenih podataka je svakih sat vremena, dakle 24 stupaca u tablici predstavlja prometni tok jednoga dana po satnim rezolucijama. Tako uređen vektor (redak) od 24 podataka jedan je od niza vektora i oni predstavljaju ulaz za grupiranje.

Prva tablica naziva „PoBrzinama“ (slika 30) sadrži podatke o ukupnom prometnom toku (Q_{uk}) osobnih i teretnih vozila s obzirom na voznu i pretjecajnu traku, te broj vozila unutar određene kategorije brzine, također za pretjecajnu i voznu traku za oba smjera prometnog toka. Ukupno ima 30 atributa/stupaca i 23.812 redaka/zapisa za 3 godine.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Dan	Sat	Q_uk_pretjecajna	Qov_pretjecajna	Qtv_pretjecajna	Vov_pretjecajna	Vtv_pretjecajna	0-40 km/h_pretjecajna	41-50 km/h_pretjecajna	51-60 km/h_pretjecajna	61-70 km/h_pretjecajna	71-80 km/h_p	
2	0	2018-01-01 00:00:00	0	5	5	0	165	0	0	0	0	0	
3	1	2018-01-01 01:00:00	1	6	6	0	155,67	0	0	0	0	0	
4	2	2018-01-01 02:00:00	2	15	15	0	155,47	0	0	0	0	0	
5	3	2018-01-01 03:00:00	3	7	7	0	153,29	0	0	0	0	0	
6	4	2018-01-01 04:00:00	4	10	10	0	144,9	0	0	0	0	0	
7	5	2018-01-01 05:00:00	5	1	1	0	173	0	0	0	0	0	
8	6	2018-01-01 06:00:00	6	5	5	0	153	0	0	0	0	0	
9	7	2018-01-01 07:00:00	7	7	7	0	163,57	0	0	0	0	0	
10	8	2018-01-01 08:00:00	8	13	12	1	157,33	104	0	0	0	0	
11	9	2018-01-01 09:00:00	9	13	13	0	144,31	0	0	0	0	0	
12	10	2018-01-01 10:00:00	10	55	55	0	149,2	0	0	0	0	0	
13	11	2018-01-01 11:00:00	11	117	117	0	153,4	0	0	0	0	0	
14	12	2018-01-01 12:00:00	12	182	182	0	152,47	0	0	0	0	0	
15	13	2018-01-01 13:00:00	13	189	189	0	153,88	0	0	0	0	0	
16	14	2018-01-01 14:00:00	14	200	200	0	153,6	0	0	0	0	0	
17	15	2018-01-01 15:00:00	15	235	234	1	137,19	106	0	0	0	0	
18	16	2018-01-01 16:00:00	16	240	237	3	123,28	97,33	0	0	0	0	
19	17	2018-01-01 17:00:00	17	165	165	0	125,15	0	0	0	0	0	
20	18	2018-01-01 18:00:00	18	154	149	5	128,26	108,8	0	0	0	0	
21	19	2018-01-01 19:00:00	19	85	84	1	136,88	106	0	0	0	0	
22	20	2018-01-01 20:00:00	20	64	64	0	143,56	0	0	0	0	0	
23	21	2018-01-01 21:00:00	21	43	43	0	139,74	0	0	0	0	0	
24	22	2018-01-01 22:00:00	22	25	25	0	136,32	0	0	0	0	0	
25	23	2018-01-01 23:00:00	23	18	17	1	138,06	100	0	0	0	0	
26	24	2018-01-02 00:00:00	0	14	14	0	129	0	0	0	0	0	
27	25	2018-01-02 01:00:00	1	10	10	0	139,4	0	0	0	0	0	
28	26	2018-01-02 02:00:00	2	5	4	1	132	104	0	0	0	0	
29	27	2018-01-02 03:00:00	3	7	6	1	158,17	105	0	0	0	0	

Slika 30. Tablica „PoBrzinama“.

Tablica 3. Opis atributa tablice „PoBrzinama“.

Atribut	Opis
Dan	yyyy-mm-dd hh:mm:ss - format prikaza snimljenog podatka
Sat	0 : 24 - sat u danu kada je snimljen podatak
Q_{uk} pretjecajna i vozna	Ukupni prometni tok svih vozila za pretjecajnu i voznu traku
Q_{ov} pretjecajna i vozna	Prometni tok osobnih vozila za pretjecajnu i voznu traku
Q_{tv} pretjecajna i vozna	Prometni tok teretnih vozila za pretjecajnu i voznu traku
V_{ov} pretjecajna i vozna	Brzina osobnih vozila za pretjecajnu i voznu traku
V_{tv} pretjecajna i vozna	Brzina teretnih vozila za pretjecajnu i voznu traku
0-40 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 0-40 [km/h] za pretjecajnu i voznu traku
41-50 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 41-50 [km/h] za pretjecajnu i voznu traku
51-60 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 51-60 [km/h] za pretjecajnu i voznu traku
61-70 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 61-70 [km/h] za pretjecajnu i voznu traku
71-80 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 71-80 [km/h] za pretjecajnu i voznu traku
81-90 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 81-90 [km/h] za pretjecajnu i voznu traku
91-100 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 91-100 [km/h] za pretjecajnu i voznu traku
101-130 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom 101-130 [km/h] za pretjecajnu i voznu traku
> 130 [km/h] pretjecajna i vozna	Broj vozila koji se voze brzinom >130 [km/h] za pretjecajnu i voznu traku

Druga tablica naziva „PoKategorijama“ (slika 31) sadrži podatke o grupiranim vozilima s obzirom na voznu i pretjecajnu traku te smjeru odvijanja prometa. Te grupe vozila (atributi) su prikazani tablicom 31.

	A	B	C	D	E	F	G	H	I	J
		Dan	Sat	Neklasificirano_pretjecajna	Motocikli_pretjecajna	Osobna vozila_pretjecajna	Kombij_pretjecajna	Osobna vozila s prikolicom_pretjecajna	Kamioni_pretjecajna	Kamioni s prikolicom_pretjecajna
2	0	2018-01-01 00:00:00	0	0	0	0	5	0	0	0
3	1	2018-01-01 01:00:00	1	2	2	0	4	0	0	0
4	2	2018-01-01 02:00:00	2	0	0	0	14	1	0	0
5	3	2018-01-01 03:00:00	3	1	0	0	6	0	0	0
6	4	2018-01-01 04:00:00	4	4	0	0	6	0	0	0
7	5	2018-01-01 05:00:00	5	0	0	0	1	0	0	0
8	6	2018-01-01 06:00:00	6	0	0	0	5	0	0	0
9	7	2018-01-01 07:00:00	7	1	0	0	6	0	0	0
10	8	2018-01-01 08:00:00	8	3	0	0	9	0	1	0
11	9	2018-01-01 09:00:00	9	3	0	0	10	0	0	0
12	10	2018-01-01 10:00:00	10	4	0	0	50	1	0	0
13	11	2018-01-01 11:00:00	11	8	0	0	109	0	0	0
14	12	2018-01-01 12:00:00	12	4	0	0	176	2	0	0
15	13	2018-01-01 13:00:00	13	15	0	0	173	1	0	0
16	14	2018-01-01 14:00:00	14	9	0	0	189	2	0	0
17	15	2018-01-01 15:00:00	15	10	0	0	216	8	1	0
18	16	2018-01-01 16:00:00	16	9	0	0	196	32	1	0
19	17	2018-01-01 17:00:00	17	11	0	0	141	13	0	0
20	18	2018-01-01 18:00:00	18	11	0	0	128	10	1	1
21	19	2018-01-01 19:00:00	19	9	0	0	68	7	0	1
22	20	2018-01-01 20:00:00	20	7	0	0	55	2	0	0
23	21	2018-01-01 21:00:00	21	4	0	0	38	1	0	0
24	22	2018-01-01 22:00:00	22	2	0	0	21	2	0	0
25	23	2018-01-01 23:00:00	23	1	0	0	15	1	0	0
26	24	2018-01-02 00:00:00	0	2	0	0	10	2	0	0
27	25	2018-01-02 01:00:00	1	2	0	0	7	1	0	0
28	26	2018-01-02 02:00:00	2	0	0	0	4	0	0	0
29	27	2018-01-02 03:00:00	3	0	0	0	6	0	0	0

Slika 31. Tablica „PoKategorijama“.

Tablica 4. Opis atributa tablice „PoKategorijama“.

Atributi	Opis
Dan	yyyy-mm-dd hh:mm:ss - format prikaza snimljenog podatka
Sat	0 : 24 - sat u danu kada je snimljen podatak
Neklasificirano – pretjecajna i vozna	Vozila koja nisu mogla biti klasificirana za pretjecajnu i voznu traku
Motocikli – pretjecajna i vozna	Broj motocikala na pretjecajnoj i voznoj prometnoj traci
Osobna vozila – pretjecajna i vozna	Broj osobnih vozila na pretjecajnoj i voznoj prometnoj traci
Kombiji – pretjecajna i vozna	Broj kombija na pretjecajnoj i voznoj prometnoj traci
Osobna vozila s prikolicom – pretjecajna i vozna	Broj osobnih vozila sa prikolicom na pretjecajnoj i voznoj prometnoj traci
Kamioni – pretjecajna i vozna	Broj kamiona na pretjecajnoj i voznoj prometnoj traci
Kamioni s prikolicom – pretjecajna i vozna	Broj kamiona sa prikolicom na pretjecajnoj i voznoj prometnoj traci
Tegljači – pretjecajna i vozna	Broj tegljača na pretjecajnoj i voznoj prometnoj traci
Autobusi – pretjecajna i vozna	Broj autobusa na pretjecajnoj i voznoj prometnoj traci
Ukupno – pretjecajna i vozna	Ukupni broj svih vozila na pretjecajnoj i voznoj prometnoj traci

5.2. Osnovna prilagodba podataka

Naravno kao kod svakog skupa podataka potrebno je izvršiti osnovnu prilagodbu podataka iz razloga što je svaki postupak rješavanja određenog problema individualan za sebe. Zato što je potrebno izvršiti osnovnu prilagodbu podataka prema samom problemu kojem se pristupa. U prvom skupu podataka (tablica „PoBrzinama“) nije bilo podatka o ukupnom prometnom toku obje prometne trake (jednog smjera), već ukupni prometni tok pretjecajne i vozne trake. Ta dva atributa podatka su zbrojena i dobiven je podatak o ukupnom prometnom toku za jedan smjer prometnog toka. To je rađeno upravo iz razloga dobivanja karakteristika po danu, gdje će kasnije slični dani imati sličnu karakteristiku. Drugi skup podataka (tablica „PoKategorijama“) nije podvrgnut osnovnim prilagodbama podataka zbog dobre organizacije podataka unutar tablice nad kojima se vrši vizualizacija tih istih podataka.

Sada kada su postojani svi podatci za nastavak obrade podataka, uklonjeni su atributi (Q_{uk} , Q_{ov} , Q_{tv}) za obje kolničke trake (vozna i pretjecajna traka) i oba smjera (slika 32 – atributi označeni crvenom bojom). Razlog tomu je dodavanje nova tri atributa naziva: $Q_{uk} - ukupno$, $Q_{ov} - ukupno$, $Q_{tv} - ukupno$ (slika 33 – atributi označeni crvenom bojom). To je potrebno radi smanjenja dimenzionalnosti podataka unutar tablice i načinu vizualiziranja jer se ta vizualizacija prikazuje posebno za prometne tokove i posebno za prometne trake.

Uklonjeni su i atributi (V_{ov} , V_{tv}) za obje kolničke trake i oba smjera prometnih tokova (slika 32 - atributi označeni crvenom bojom), a razlog tomu je dodavanje nova dva atributa naziva: $V_{ov} - max$ i $V_{tv} - max$ (slika 33 – atributi označeni crvenom bojom). To je potrebno iz istog razloga koji je naveden u tekstu gore. Također su još dodani atributi: dan u tjednu, dan u mjesecu i dan u godini zbog kasnije vizualizacije podataka po danima, tjednima, mjesecima i godinama (slika 33 - označeno plavom bojom).

D	E	F	G	H
Q _{uk_pretjecajna}	Q _{ov_pretjecajna}	Q _{tv_pretjecajna}	V _{ov_pretjecajna}	V _{tv_pretjecajna}
5	5	0	165	
6	6	0	155,67	
15	15	0	155,47	
7	7	0	153,29	
10	10	0	144,9	
1	1	0	173	

R	S	T	U	V
Q _{uk.1_vozna}	Q _{ov.1_vozna}	Q _{tv.1_vozna}	V _{ov.1_vozna}	V _{tv.1_vozna}
32	30	2	123,07	94,5
95	93	2	124,28	137,5
94	93	1	118,67	123
80	79	1	124,33	87
60	59	1	120,32	138
37	34	3	119,29	108,33

Slika 32. Izbrisani podatci/atributi.

V	W	X	Y	Z	AA	AB	AC	
Q _{uk}	Q _{ov} - ukupno	Q _{tv} - ukupno	V _{ov} - max	V _{tv} - max	Dan u tjednu	Tjedan u godini	Dan u mjesecu	
3	37	35	2	165	94,5	Pon	1	1
2	101	99	2	155,67	137,5	Pon	1	1
1	109	108	1	155,47	123	Pon	1	1
1	87	86	1	153,29	87	Pon	1	1
7	70	69	1	144,9	138	Pon	1	1
1	38	35	3	173	108,33	Pon	1	1
3	69	64	5	153	93,4	Pon	1	1
0	77	71	6	163,57	98,67	Pon	1	1
4	118	99	19	157,33	104	Pon	1	1
5	167	154	13	144,31	105,31	Pon	1	1
5	305	292	13	149,2	104,77	Pon	1	1
5	490	475	15	153,4	110,2	Pon	1	1
3	655	640	15	152,47	112,67	Pon	1	1
4	676	653	23	153,88	108,35	Pon	1	1
2	654	638	16	153,6	109,13	Pon	1	1
3	689	658	31	137,19	106	Pon	1	1
4	652	615	37	123,28	97,33	Pon	1	1
7	547	498	49	125,15	93,29	Pon	1	1
2	497	450	47	128,26	108,8	Pon	1	1
3	341	304	37	136,88	106	Pon	1	1
3	278	251	27	143,56	98,93	Pon	1	1
1	190	159	31	139,74	91,39	Pon	1	1
3	147	111	36	136,32	92	Pon	1	1
4	107	85	22	138,06	100	Pon	1	1
3	76	55	21	129	93,86	Uto	1	2
5	70	54	16	139,4	94,25	Uto	1	2
3	46	34	12	132	104	Uto	1	2
2	55	28	27	158,17	105	Uto	1	2

Slika 33. Svi dodani atributi.

Tablica 5. Opis novododatih atributa tablice „PoBrzinama“.

Atributi	Opis
Q_{uk}	Ukupni prometni tok svih vozila za prometni tok jednog i drugog smjera
Q_{ov} - ukupno	Ukupni prometni tok osobnih vozila za jedan i drugi smjer
Q_{tv} - ukupno	Ukupni prometni tok teretnih vozila za jedan i drugi smjer
V_{ov} - max	Maksimalna brzina osobnih vozila za jedan i drugi smjer
V_{tv} - max	Maksimalna brzina teretnih vozila za jedan i drugi smjer
Dan u tjednu	Pon, uto, sri, čet, pet, sub, ned
Tjedan u godini	Broj koji prikazuje o kojem tjednu u godini se radi
Dan u mjesecu	Broj koji prikazuje o kojem danu u mjesecu se radi

5.3. Priprema podataka za obradu

Najviše vremena potrošeno je na osnovnu prilagodbu i pripremu za obradu dobivenih prometnih podataka. Priprema prilagođenih podataka za daljnju obradu podrazumijeva filtriranje dobivenih skupova podataka, odnosno uklanjanje nekih van-distribucijskih podataka (engl. outlier). Van-distribucijski podatak (slika 34) je objekt koji značajno odstupa od ostatka zbirke objekata. Prije pripreme podataka za obradu normalno je opažanje tijekom faze analize podataka da taj podatak leži daleko od drugih vrijednosti. Nije normalno da se podatak nalazi daleko od dobro strukturiranih podataka.

0	3	12	22	146	85	338	280	58	136,59	137
0	3	10	23	104	53	241	186	55	137,84	141,75
0	0	10	13	70	43	157	117	40	149,24	112,13
498	4941	45398	22684	120791	43750	381572	301204	80368	132,39	115,7
0	1	11	9	45	32	107	75	32	145,22	111,56
0	0	4	7	23	19	57	45	12	148,25	118,33
0	0	2	6	23	10	41	30	11	121,9	108,45
0	1	5	5	19	16	50	37	13	167,25	107,69
0	0	4	8	20	14	50	39	11	139	121
0	2	4	8	52	18	93	80	13	135,56	104,46

Slika 34. Van-distribucijski podatci.

Također u dobivenom skupu podataka nije samo problem u van-distribucijskim podacima. Velik problem bili su nedostatak zapisa na dnevnoj razini, čak i mjesečnoj. Prikaz takvih problema dan je slikom 35 gdje je prikazano nepostojanje zapisa više sati. Isto tako na slici 36 prikazan je izostanak zapisa cijeloga mjeseca rujna. Ovo su jedni od nedostajućih zapisa, bilo ih znatno više. Ti nedostajući zapisi produkt su vjerojatno kvara na sustavu.

The screenshot shows an Excel spreadsheet with the following data structure:

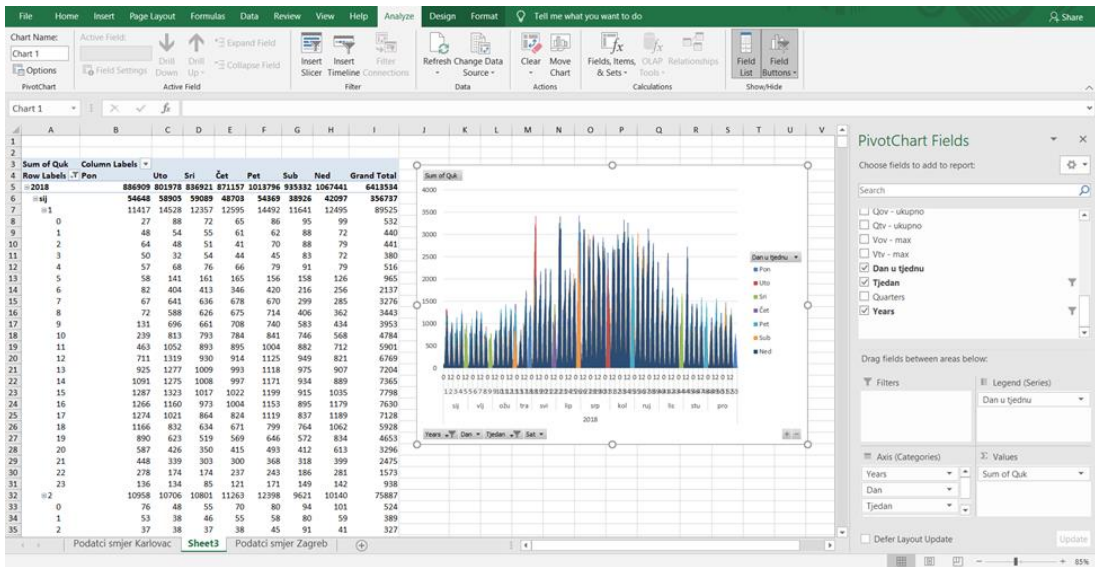
	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
3371	3	6	151	86	236	82	1114	845	269	0	0	Pon	21	21	
3372	0	6	139	74	240	102	1095	849	246	0	0	Pon	21	21	
3373	0	4	116	107	245	81	1052	816	236	0	0	Pon	21	21	
3374	0	6	134	93	285	92	1173	919	254	0	0	Pon	21	21	
3375	0	1	47	27	112	31	423	342	81	0	0	Pon	21	21	
3376							0	0	0	0	0	Pon	21	21	
3377							0	0	0	0	0	Pon	21	21	
3378							0	0	0	0	0	Pon	21	21	
3379							0	0	0	0	0	Pon	21	21	
3380							0	0	0	0	0	Pon	21	21	
3381							0	0	0	0	0	Pon	21	21	
3382							0	0	0	0	0	Pon	21	21	
3383							0	0	0	0	0	Pon	21	21	
3384							0	0	0	0	0	Pon	21	21	
3385							0	0	0	0	0	Pon	21	21	
3386							0	0	0	0	0	Pon	21	21	
3387							0	0	0	0	0	Pon	21	21	
3388							0	0	0	0	0	Pon	21	21	
3389							0	0	0	0	0	Pon	21	21	
3390							0	0	0	0	0	Pon	21	21	
3391							0	0	0	0	0	Pon	21	21	
3392							0	0	0	0	0	Pon	21	21	
3393							0	0	0	0	0	Pon	21	21	
3394							0	0	0	0	0	Pon	21	21	
3395							0	0	0	0	0	Pon	21	21	
3396							0	0	0	0	0	Pon	21	21	
3397							0	0	0	0	0	Pon	21	21	
3398							0	0	0	0	0	Pon	21	21	
3399							0	0	0	0	0	Pon	21	21	

Slika 35. Nedostatak zapisa više sati.

	A	B	C	D	E	F	G	H	I	J	K	L	
189	15.kol		509	525	574	773	1250	1880	2442	2636	2753	2910	2734
190	16.kol		328	322	449	588	918	1484	1943	2046	2309	2462	2152
191	17.kol		615	710	887	1138	1667	2510	2715	2691	2652	2945	3002
192	18.kol		312	391	472	679	1004	1600	2070	2327	2432	2503	2794
193	19.kol		331	392	481	665	1070	1749	1942	1591	2288	3132	2827
194	20.kol		265	249	285	410	631	1157	1738	1928	2061	2120	2049
195	21.kol		236	232	239	366	555	921	1452	1689	1828	1847	1836
196	22.kol		459	267	275	360	594	888	1436	1678	1867	1945	1907
197	23.kol		241	292	313	462	587	1096	1591	1727	1691	1969	1992
198	24.kol		530	526	777	1005	1237	1888	2538	2791	2978	2682	3245
199	25.kol		243	274	343	426	697	913	1364	1735	1797	2188	2287
200	26.kol		275	305	353	454	718	1203	1715	1827	1991	2311	2258
201	27.kol		191	191	214	240	374	622	1156	1441	1609	1546	1468
202	28.kol		180	190	175	280	377	670	1249	1409	1594	1569	1600
203	29.kol		191	215	208	282	464	773	1370	1535	1828	1698	1654
204	30.kol		229	216	198	284	402	758	1209	1444	1691	1669	1761
205	31.kol		395	428	549	721	965	1448	2132	2401	2455	2810	2619
206	1.lis		84	90	80	122	178	385	831	1038	1233	1032	882
207	2.lis		129	92	72	103	154	371	816	1002	1159	1139	930
208	3.lis		158	176	112	128	195	459	777	1077	1178	1165	964
209	4.lis		108	95	93	111	185	447	808	1099	1334	1272	1273
210	5.lis		191	155	118	131	224	468	847	1200	1689	1977	2141
211	6.lis		118	77	79	87	92	144	279	446	714	996	1180
212	7.lis		99	66	62	82	161	414	648	868	883	887	1011

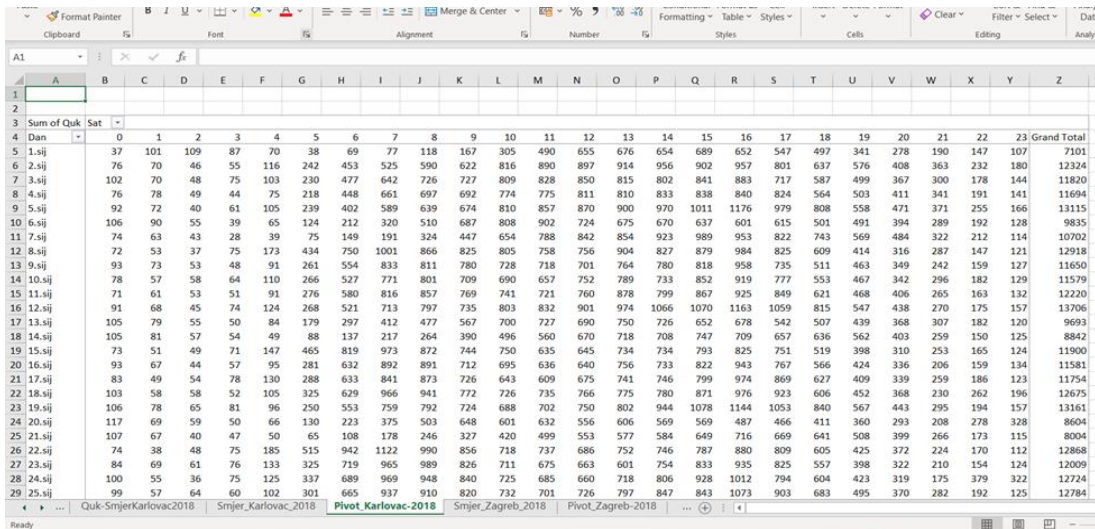
Slika 36. Nedostatak zapisa mjeseca rujna.

Sada kada je opisana osnovna prilagodba prvog skupa podataka i kada su završene pripreme obrade podataka, dovedena je u red nova Excel tablica gdje su podatci organizirani kao vektor s 24 podatka i on čini jedan dan. Svaki od tih 24 podatka predstavlja ukupni prometni tok (Q_{uk}) po satu za jedan smjer. Takvih tablica ima **četiri**: za 6, 7, 8, 9 mjesec (ljetna sezona) za jedan i drugi smjer, te za 2018., 2019., te 2020. godinu također za jedan i drugi smjer. U nastavku rada bit će objašnjeno zašto četiri pivot tablice. Takve tablice s tako organiziranim podacima su dobivene putem pivot tablice (slika 37) kao Excel-ovog alata. Pivot tablica moćan je alat za izračunavanje, sažimanje i analizu podataka koji omogućuje vidljivost usporedba i trendova u podacima.



Slika 37. Pivot tablica 1.

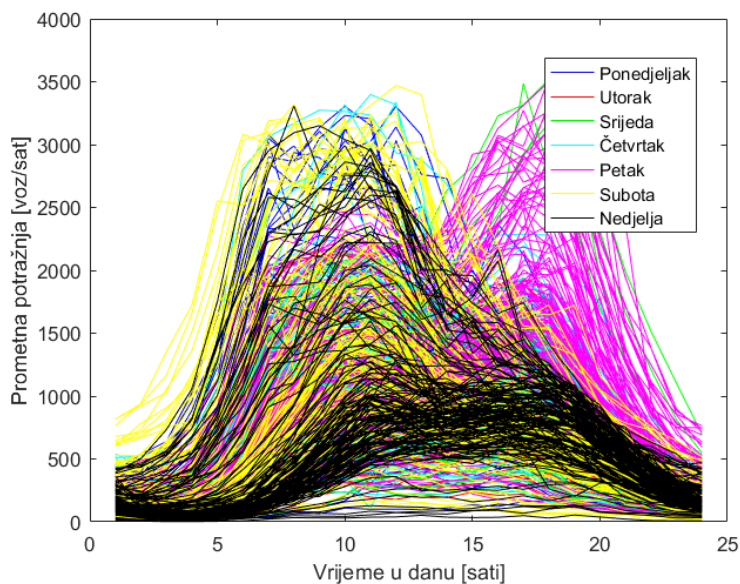
Izrađene pivot tablice sadrže podatke ukupnog prometnog (Q_{uk}) toka po satnim rezolucijama za svaki dan u godini. Prikaz pivot tablice vidljiv je na slici 38. Vertikalni stupci tablice predstavljaju atribute svakog sata u danu dok horizontalni redci predstavljaju dane u godini. Razlika tih četiri tablica je u veličini broja zapisa, odnosno u pitanju su različita vremena prikupljanja zapisa kao i smjera prometnih tokova.



Slika 38. Pivot tablica 2.

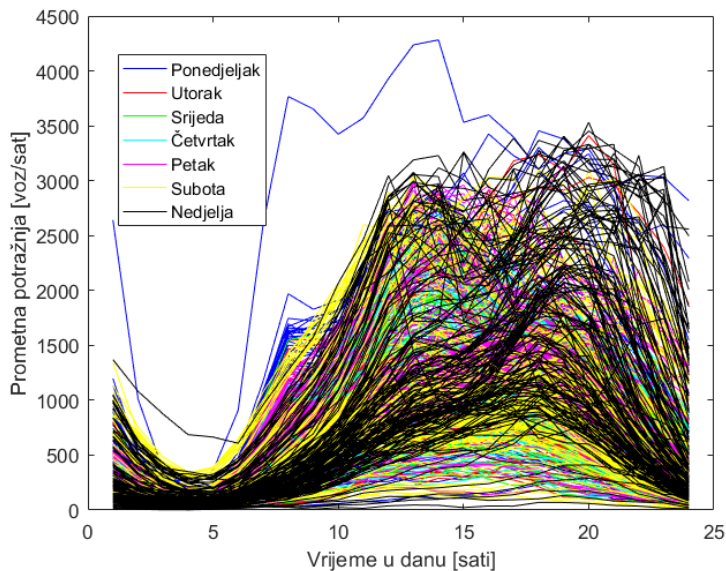
Nakon što je skup podataka prilagođen za grupiranje, prikazana su četiri korištena skupa podataka za grupiranje. Ta četiri skupa podataka sastoje se od krivulja koje označuju broj protekli vozila u satnoj rezoluciji svakoga dana u tjednu. Skupovi podataka korišteni za grupiranje:

- 1) Ukupni prometni tok smjer Karlovac (sve tri godine);



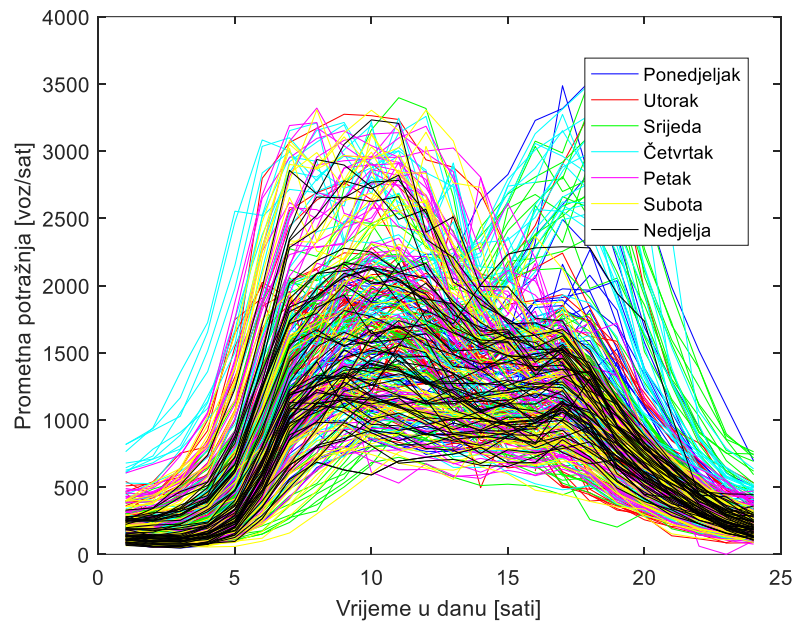
Slika 39 Prikaz sirovih podataka prometnog toka.

- 2) Ukupni prometni tok smjer Zagreb (sve tri godine);



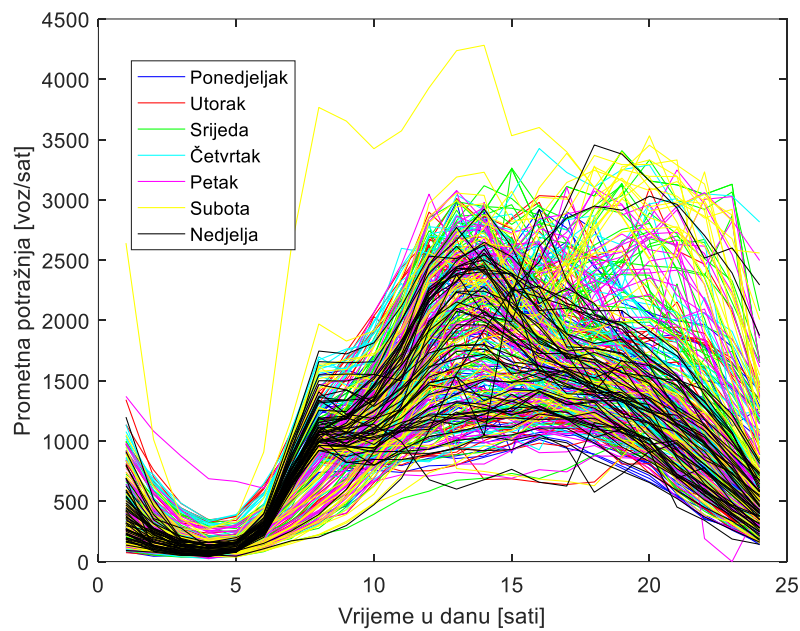
Slika 40. Prikaz sirovih podataka prometnog toka.

3) Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec);



Slika 41. Prikaz sirovih podataka prometnog toka.

4) Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).



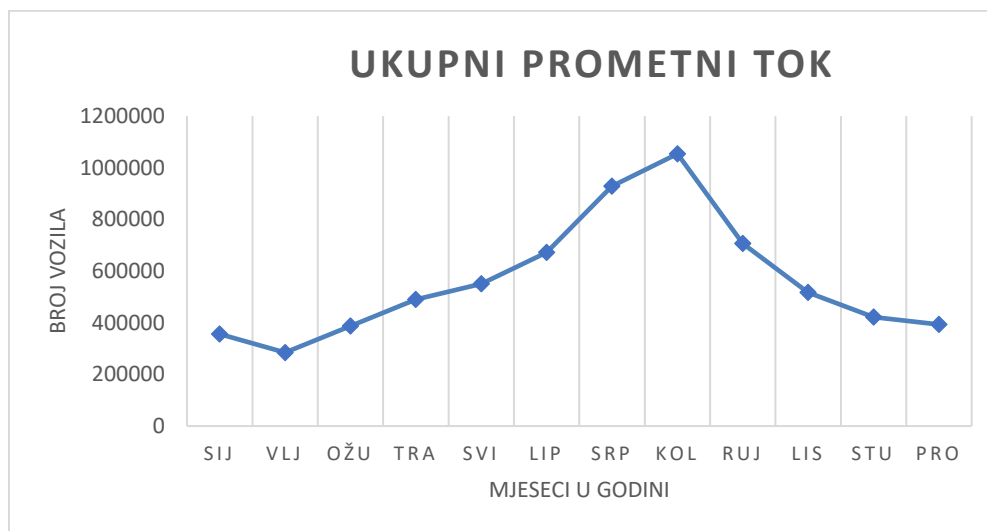
Slika 42. Prikaz sirovih podataka prometnog toka.

6. Dobiveni rezultati

U ovom poglavlju iznesena su svi dobiveni rezultati. Rezultati nisu u skladu s očekivanjima jer prikupljeni prometni podatci od strane Hrvatskih autocesta u dosta velikoj mjeri nisu ispravna. Skup podataka maksimalno je prilagođen i obrađen kako bi se vizualizirali čim reprezentativniji prometni podaci.

6.1. Analiza sezonalnosti podataka

Sezonalnost je pojam koji označava utjecaj određenog godišnjeg razdoblja na cijelu godinu. Konkretno za ovaj rad, sezonalnost je pojam koji opisuje utjecaj ljetnih mjeseci (sezone) na odvijanje prometnog toka tijekom cijele godine. U Republici Hrvatskoj sezonalnost znatno utječe na prometne tokove državnih autocesta. Tijekom ljetnih mjeseci zbog velikog broja priljeva turista koji dolaze na godišnji odmor, vidljiva je jasna razlika u količini prometa u ljetnim mjesecima i svim ostalim (slika 43).



Slika 43. Utjecaj sezonalnosti na prometni tok.

6.2. Analiza vizualizacije podataka

Ljudi oduvijek vizualiziraju podatke. Prikupljeni su podaci od zemljopisnih karti do grafikona i dijagrama i naposljetku slagani tako da bolje i detaljnije prenesu značenje podataka. Isto tako analizirani su dobiveni prometni podaci i na kraju vizualizirani. Vizualizirani su jer vizualizacija omogućuje brzo i učinkovito tumačenje podataka za razliku od ostalih prikaza podataka poput tabličnog ili tekstualnog prikaza.

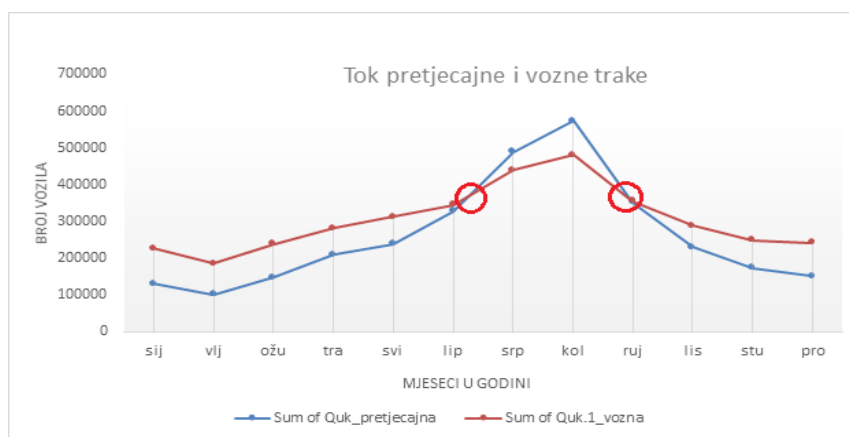
S ovoliko velikim skupom podataka moguće je vizualizirati jako velik broj informacija ali zbog čitljivosti i svrhe ovog diplomskog rada vizualizirani su neki najvažniji prometni parametri. U nastavku bit će prikazani:

- 1) Odnos pretjecajne i vozne trake;
- 2) Kategorizacija brzina;
- 3) Prikaz brzina;
- 4) Kategorizacija vozila pretjecajne i vozne trake;
- 5) Prekoračenja brzine;
- 6) Utjecaj potpunog zatvaranja zbog korona situacije na odvijanje prometa.

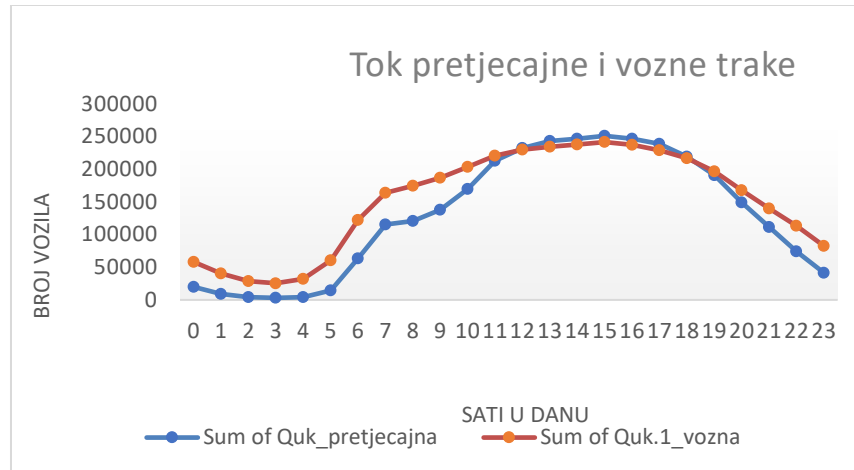
1) Odnos pretjecajne i vozne trake:

Prikaz slike 44 pruža uvid u odnos pretjecajne i vozne trake kroz 2018. godinu i temeljen je na podacima za osobna vozila u smjeru Zagreba. Prikaz se odnosi na 2018. godinu zato što u svim godinama nedostaju zapisi na mjesečnoj razini. Crvena linija je prikaz vozne trake, a plava pretjecajne trake. Slika 45 prikaz je prometnog toka pretjecajne vozne trake u danu za smjer Zagreb 2018. godine.

Vidljivo je kako je u ljetnim mjesecima pretjecajna traka više opterećena od vozne trake. Na temelju toga može se zaključiti kako je u tim mjesecima promet gust i stupanj zasićenja prometnica je blizu vrijednosti 1, što znači da je broj vozila blizu ukupnog kapaciteta prometnice kojom se prometuje. Nadalje, crvenim kružićima su označena dva sjecišta linija gdje se izjednačava tok vozne i pretjecajne trake. Prvo sjecište se nalazi početkom lipnja, a drugo krajem kolovoza. Vidljivo je kako su sjecišta od prilike na istoj razini prometnog toka, oko 350.000 [voz/mjesecc]. S prometne strane gledano, do prometnog toka (od prilike 350.000 [voz/mjesecc]) voznom trakom se odvija jači intenzitet prometa, dok nakon toga pretjecajnom trakom i tako u krug.



Slika 44. Odnos pretjecajne i vozne trake u godini.



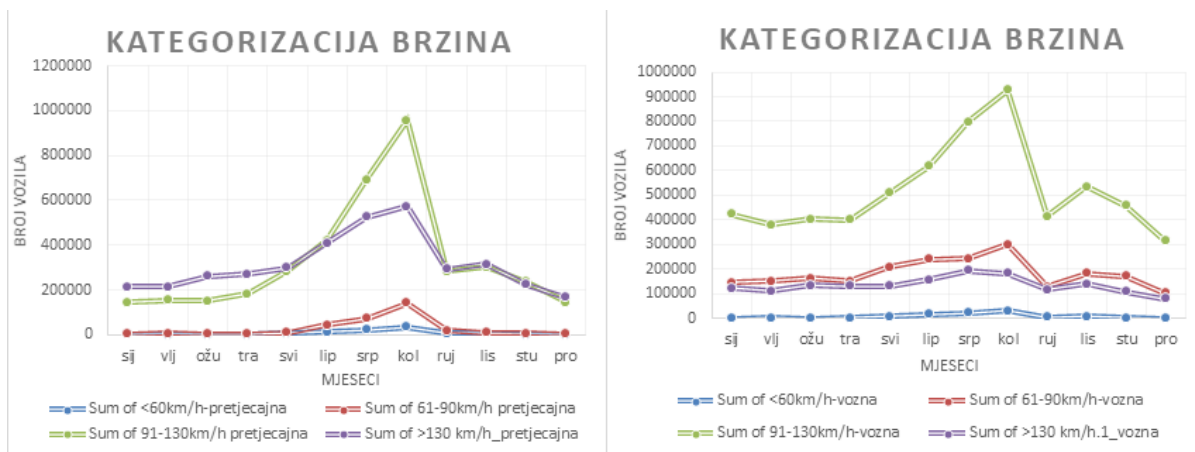
Slika 45. Odnos pretjecajne i vozne trake u danu.

2) Kategorizacija brzina:

Iz prikaza slike 46 daju se pročitati zanimljivi prometni parametri. Radi se o prikazima broja vozila po kategorijama brzina. Četiri su kategorije brzina: 0-60 [km/h], 61-90 [km/h], 91-130 [km/h] i >130 [km/h]. Svaka točka na krivuljama slike 46 predstavlja broj vozila određene kategorije brzine za svaki mjesec u godini i za oba prometna traka. Prikaz se odnosi na smjer Zagreb, a smjer Karlovac je istog prikaza pa je svejedno o kojem smjeru se radi za ovaj konkretan primjer.

Plava linija predstavlja brzinu do 60 [km/h]. Za mjesec srpanj i kolovoz vidljiv je blagi porast broja vozila s tom kategorijom brzine. To znači da je u tim mjesecima došlo do zastoja ili otežanog odvijanja prometa na obje prometne trke. Crvena linija predstavlja brzinu između 61-90 [km/h]. Za početak vidljivo je kako u ljetnim mjesecima postoji porast broja vozila sa tom kategorijom brzine. Jasno je vidljivo kako je na pretjecajnoj traci porast broja vozila zabilježen isključivo u ljetnim mjesecima, dok je na voznoj traci stalno prisutan približan broj vozila takvih brzina što je dokaz da se odvijanje prometa u tom razdoblju odvijalo otežano. Zelena linija predstavlja brzinu između 91-130 [km/h]. Sam vrhunac broja vozila s tom kategorijom brzina desio se u mjesecu kolovozu na obje prometne trake. Na pretjecajnoj traci broj vozila s tom brzinom je dvostruko manja do mjeseca svibnja. Pretjecajna traka od mjeseca svibnja ima nešto kasniji i jači linearni porast u odnosu na voznu traku što je i očekivano zbog toga što se odvijanje prometa vrši prvenstveno voznom trakom, a onda i pretjecajnom kada je promet gust. Ljubičasta linija predstavlja brzinu veću od 130 [km/h]. Na pretjecajnoj prometnoj traci do mjeseca svibnja trostruko je veći broj vozila sa prekoračenjem brzine gdje je nakon toga linearni porast do mjeseca kolovoza i nakon toga se vraća u iznos kao i na početku godine do mjeseca svibnja. Zanimljivo je opažanje da je broj vozila za kategorije brzina približno istog iznosa izuzev mjeseci: lipanj, srpanj i kolovoz. Na voznoj prometnoj traci broj vozila sa prekoračenjem brzine je u najvećoj mjeri skoro pa konstantan i znatno manji nego što je to slučaj na pretjecajnoj traci.

Za kraj usporedba brzina pretjecajne i vozne trake opisuje kako voznom trakom prometuje kontinuiraniji promet s manjim oscilacijama brzina. Drugim riječima sezonalnost manje pogađa voznu traku u odnosu na pretjecajnu što je i za očekivati jer autocesta ne služi samo turistima za vrijeme ljetne sezone već se njome odvija transport ljudi i dobara tijekom cijele godine. A kada se promet odvija u potpuno slobodno toku, većina vozila vozi se voznom prometnom trakom kao je i pravilo.



Slika 46. Kategorije brzina za obje prometne trake

3) Prikaz brzina:

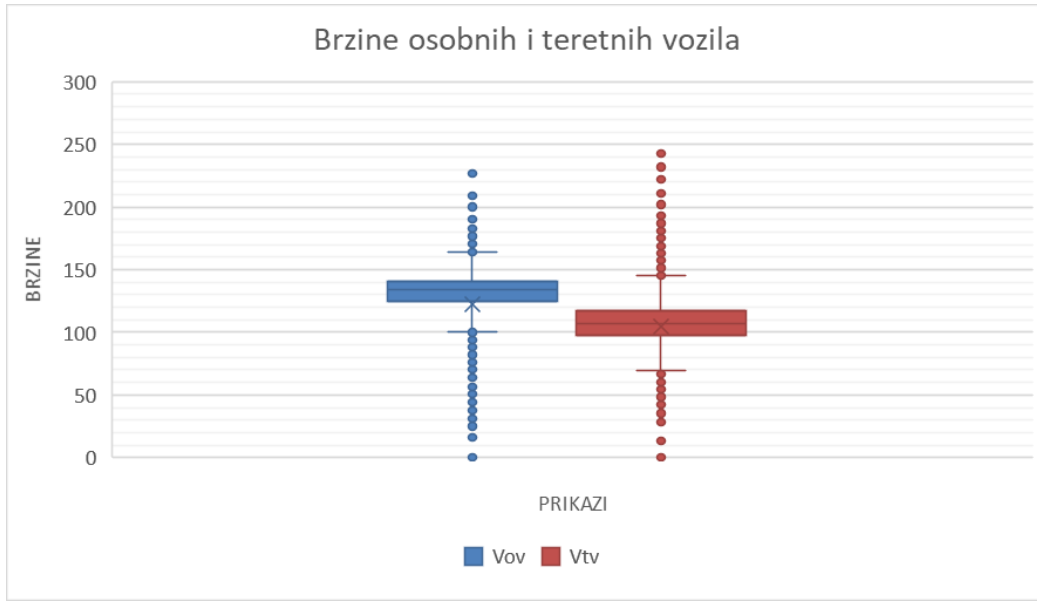
Prikaz brzina prikazan je je vizualni prikaz podataka koji prikazuje distribuciju brzina osobnih i teretnih vozila. Također prikazuje minimalne i maksimalne vrijednosti, smještaj većine podataka, median i srednju vrijednost. Median je statistički pojam koji određuje sredinu distribucije. Pola članova skupa (distribucije) nalazi se iznad mediana, a pola ispod. Manje je osjetljiv na ekstremne vrijednosti od aritmetičke sredine što ga čini posebno pogodnim za nepravilne asimetrične distribucije (skupove).

Slika 47 prikaz je brzina osobnih i teretnih vozila za smjer Karlovac. Slika pruža prikaz dvaju BoxPlot-a gdje plavi označava brzinu osobnih vozila u satnoj rezoluciji sa svim ostalim karakteristikama, a crveni brzinu teretnih vozila također sa svim ostalim karakteristikama kroz sve 3 godine. Karakteristike BoxPlot-ova:

- a) Plavi BoxPlot/Brzina osobnih vozila:
- Min: 100 [km/h]
 - Max: 164 [km/h]
 - Median: 134 [km/h]
 - Srednja vrijednost: 122 [km/h]

b) Crveni Boxplot/Brzina teških vozila:

- Min: 70 [km/h]
- Max: 145[km/h]
- Median: 107 [km/h]
- Srednja vrijednost: 104 [km/h]



Slika 47. Brzine osobnih i teretnih vozila.

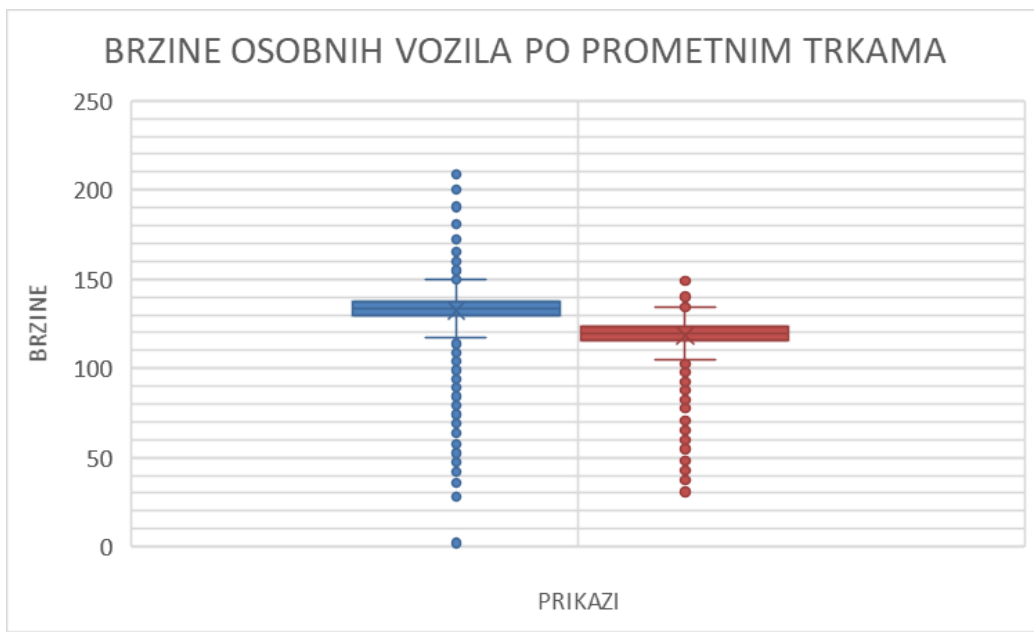
Slika 48 je prikaz brzina osobnih vozila po pretjecajnoj i voznoj traci za smjer Karlovac. Slika pruža prikaz dvaju BoxPlota gdje plavi označava brzinu osobnih vozila pretjecajne trake u satnoj rezoluciji sa svim ostalim karakteristikama, a crveni brzinu osobnih vozila vozne trake također sa svim ostalim karakteristikama kroz sve 3 godine. Karakteristike BoxPlot-ova:

a) Plavi BoxPlot/Brzina osobnih vozila pretjecajne trake:

- Min: 119 [km/h];
- Max: 150 [km/h];
- Median: 135[km/h];
- Srednja vrijednost: 133 [km/h].

b) Crveni Boxplot/Brzina osobnih vozila vozne trake:

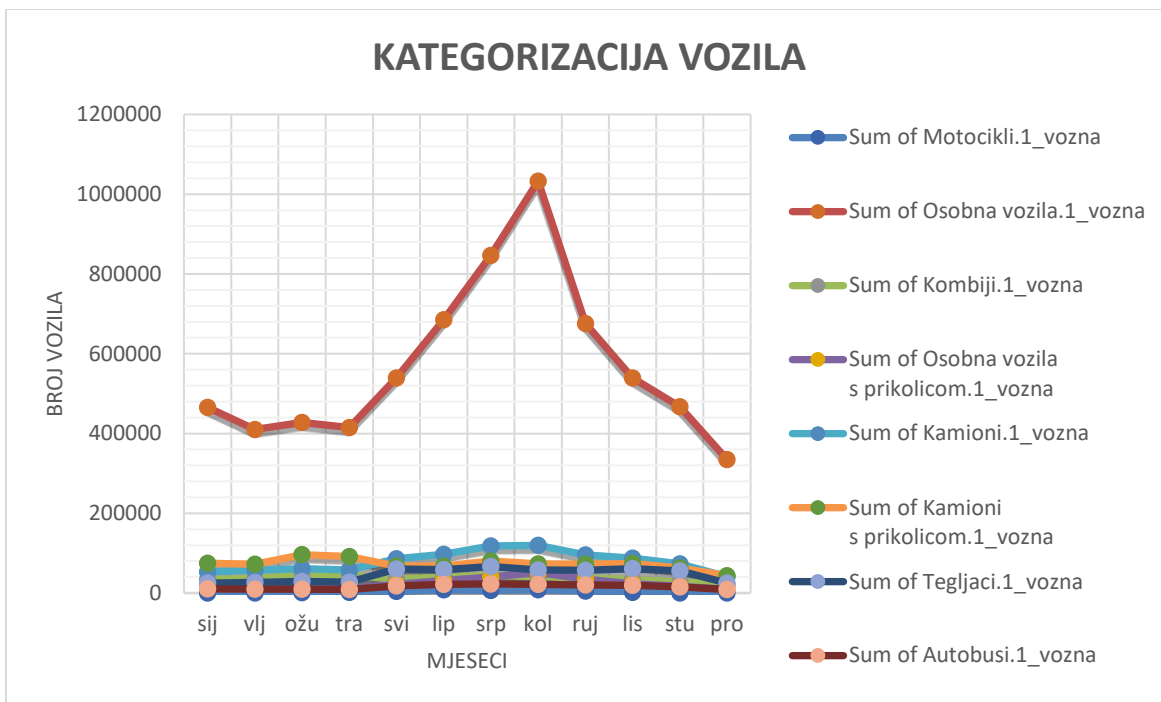
- Min: 105 [km/h];
- Max: 135[km/h];
- Median: 119 [km/h];
- Srednja vrijednost: 117 [km/h].



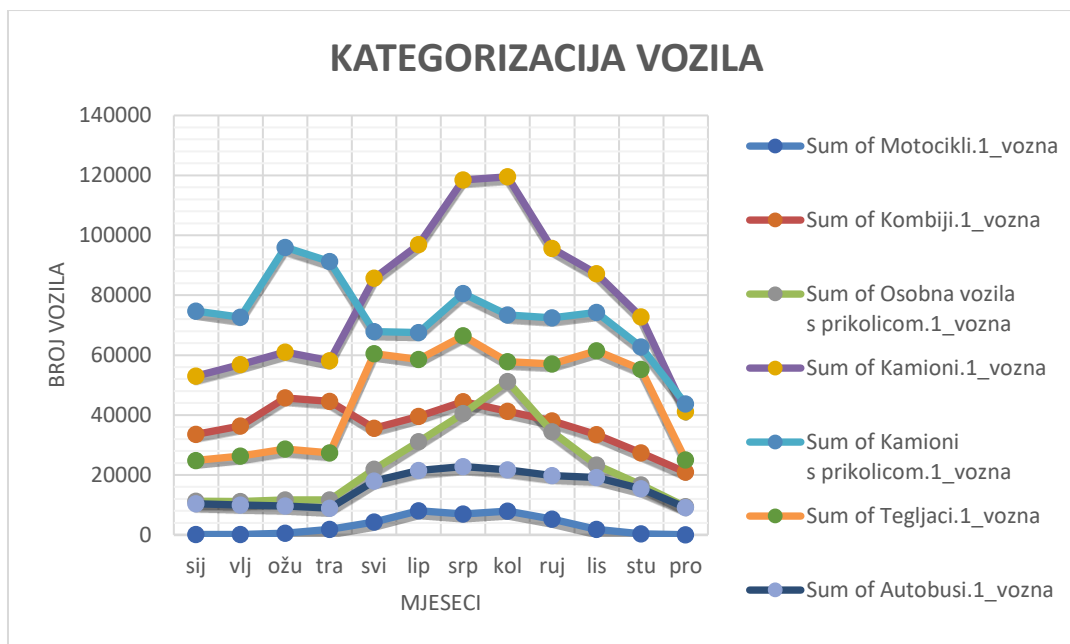
Slika 48. Brzine osobnih vozila po prometnim trakama.

4) Kategorizacija vozila pretjecajne i vozne trake;

Kategorizacija vozila opisuje broj vozila po kategorijama u zadanom vremenskom razdoblju. Slike 49 i 50 prikazuju prosječan broj vozila po kategorijama za voznu traku u smjeru Zagreba za sve tri godine. Na slici 49 vidljiv je značajno veći broj osobnih vozila u odnosu na sve ostale što je i očekivano. Na slici 50 je isti prikaz kao na slici 49 ali izuzev osobnih vozila zato što ostala vozila nisu vidljiva u prikazu jer su u manjini. Također je zanimljiv prikaz osobnih vozila s prikolicom jer jasno pokazuje ovisnost o ljetnoj sezoni kada ljudi vuku plovila na prikolicama i obične auto prikolice.



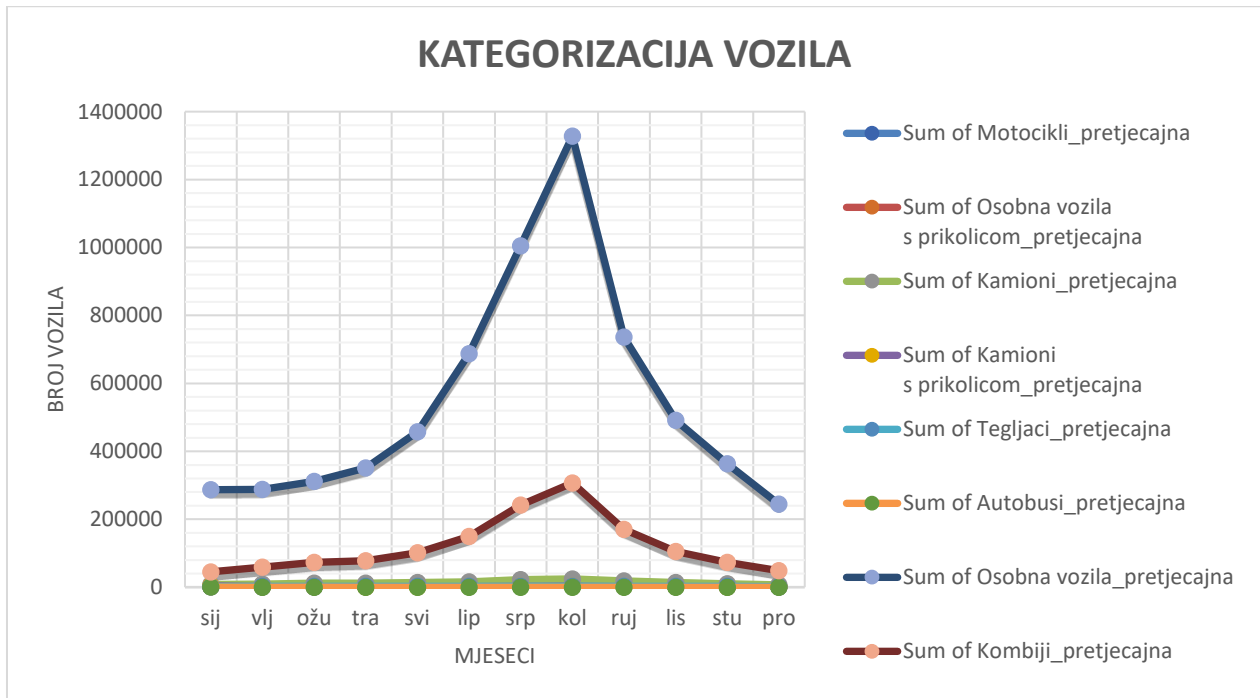
Slika 49. Kategorije vozila vozne trake za sve tri godine.



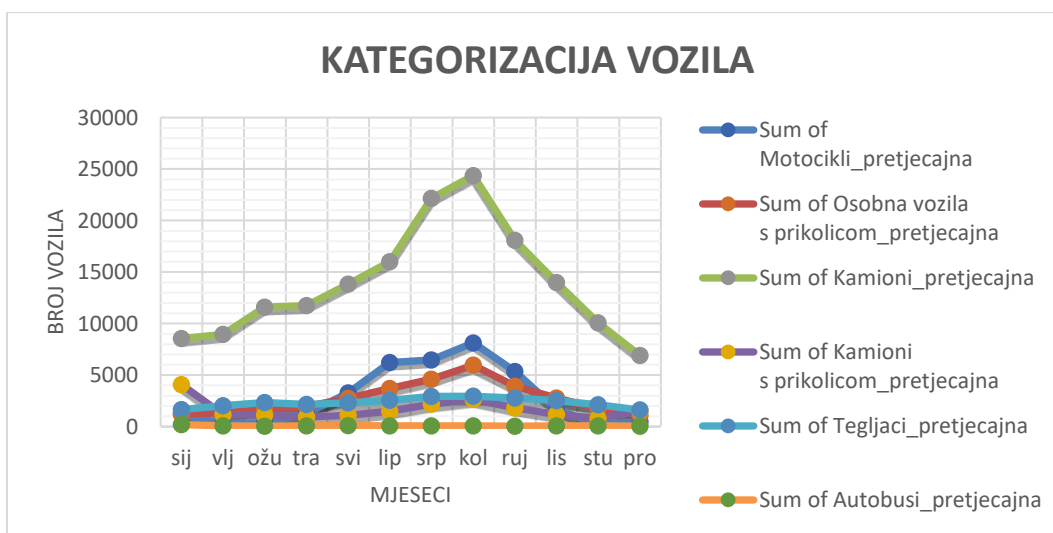
Slika 50. Kategorije vozila vozne trake za sve tri godine bez osobnih vozila.

Slike 51 i 52 prikazuju prosječan broj vozila po kategorijama za pretjecajnu traku u smjeru Zagreba za sve tri godine. Na slici 51 također je vidljiv je značajno veći broj osobnih vozila u odnosu na sve ostale što je i očekivano. Slika 52 prikazuje istiprikaz kao i slika 51 ali izuzev

osobnih i kombi vozila zato što ostala vozila nisu vidljiva u prikazu jer su u manjini. Međutim ovdje je situacija dosta drugačija nego kod prikaza vozne trake zbog dva razloga. Prvi razlog je što je ovdje znatno veći broj kombi vozila što je očekivano jer pretjecajnom trakom prometuju vozila koja postižu nešto veće brzine i koji zbog sigurnosnih razloga to smiju. Drugi razlog je što je broj ostalih vozila znatno manji broj ostalih vozila. Nešto veće brzine na pretjecajnoj traci odvijaju se zbog pretjecanja sporijih vozila u voznoj traci.



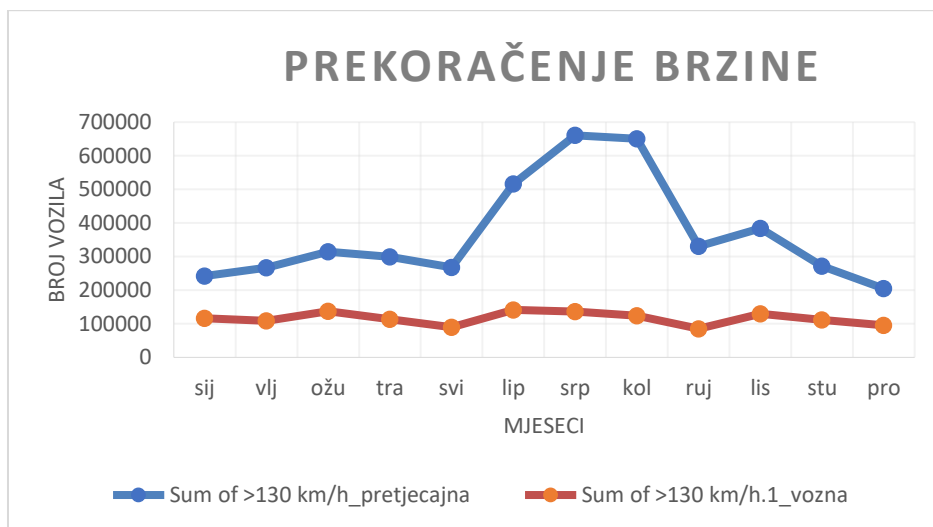
Slika 51. Kategorije vozila pretjecajne trake.



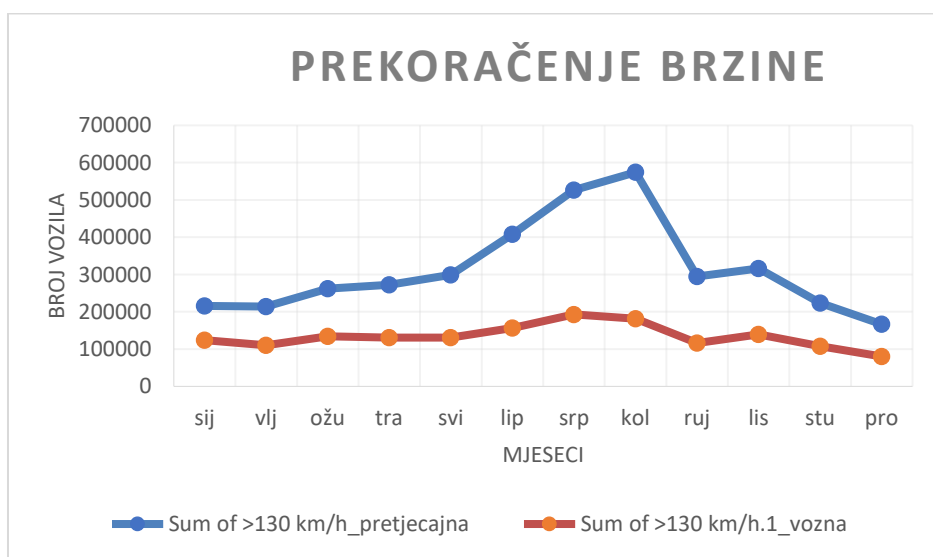
Slika 52. Kategorije vozila pretjecajne trake izuzev osobnih i kombi vozila.

5) Prekoračenje brzine:

Na autocesti A1 od kuda su i prikupljeni podatci, maksimalna dopuštena brzina vožnje iznosi 130 [km/h]. Od uvijek postoji skupina ljudi koji su skloni nepoštivanju prometnih propisa kao u ovom slučaju maksimalne propisane brzine i samim time smanjuju sigurnost odvijanja prometnih tokova. Upravo takav prikaz dan je na slici 53 i 54 čije dvije krivulje opisuju ponašanje broja vozila koja prekoračuju dozvoljenu brzinu prometovanja za jedan i drugi smjer. Prikazi su temeljeni na podacima od svih triju godina. Vidljiv je isti trend ponašanja broja prekoračenja za oba smjera.

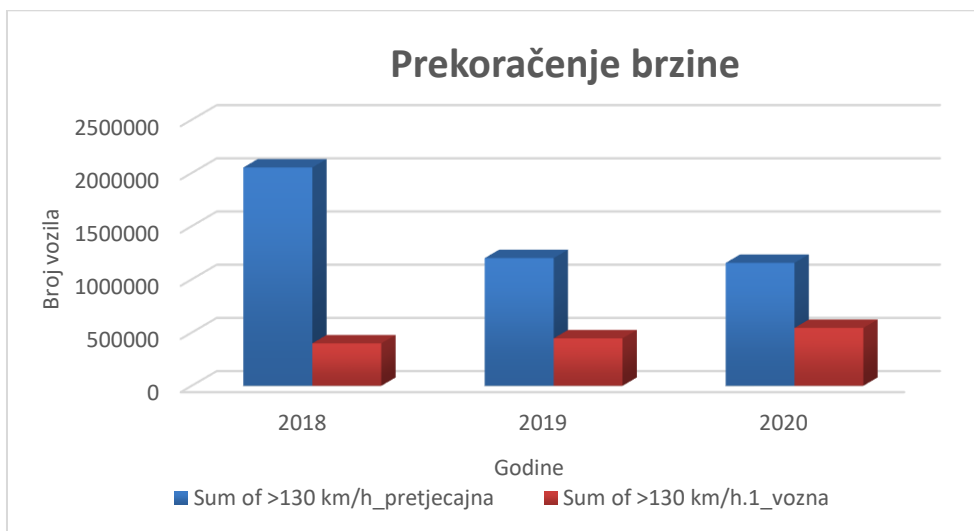


Slika 53. Prikaz prekoračenja brzina na godišnjoj razini, smjer Karlovac za sve godine.



Slika 54. Prikaz prekoračenja brzina na godišnjoj razini, smjer Zagreb za sve godine.

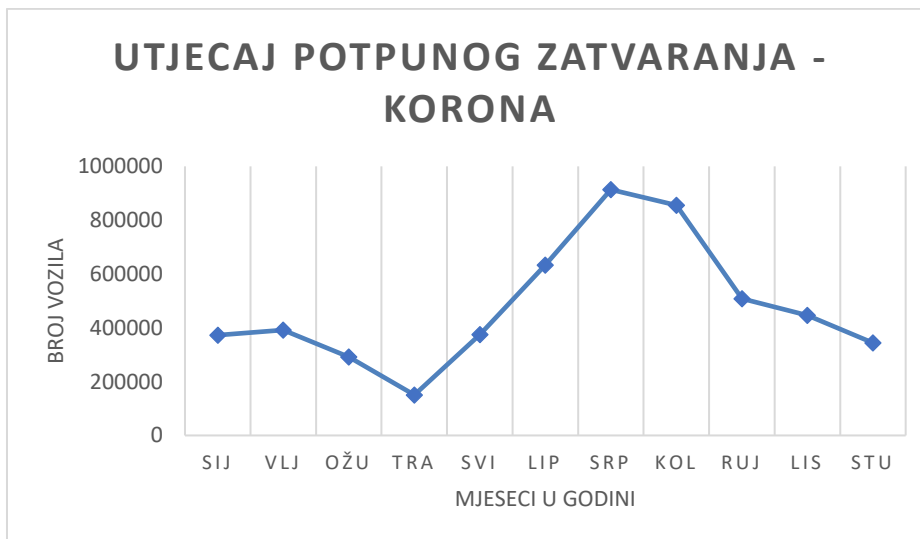
Prikaz na slici 55 daje uvid u trend ponašanja broja prekoračenja dozvoljene brzine prometovanja. S godinama broj vozila s prekoračenjem na pretjecajnoj traci opada, dok je na voznoj traci situacija obrnuta, ali u manjem intenzitetu.



Slika 55. Prikaz prekoračenja brzina po godinama, smjer Karlovac.

6) Utjecaj potpunog zatvaranja zbog korona situacije na odvijanje prometa.

Kada se slika 56 uspoređi s prije prikazanom slikom 43, jasno se vidi utjecaj potpunog zatvaranja na prometni tok autoceste. Mjesec ožujak i travanj najbolje opisuju kako se potpuno zatvaranje zbog korona situacije manifestiralo na prometni tok autoceste. Također se jasno vidi raniji završetak sezone u odnosu na godinu ranije (slika 43).



Slika 56. Prometni tok u godini potpunog zatvaranja godine 2020.

6.3. Rezultati za određivanje optimalnog broja grupa

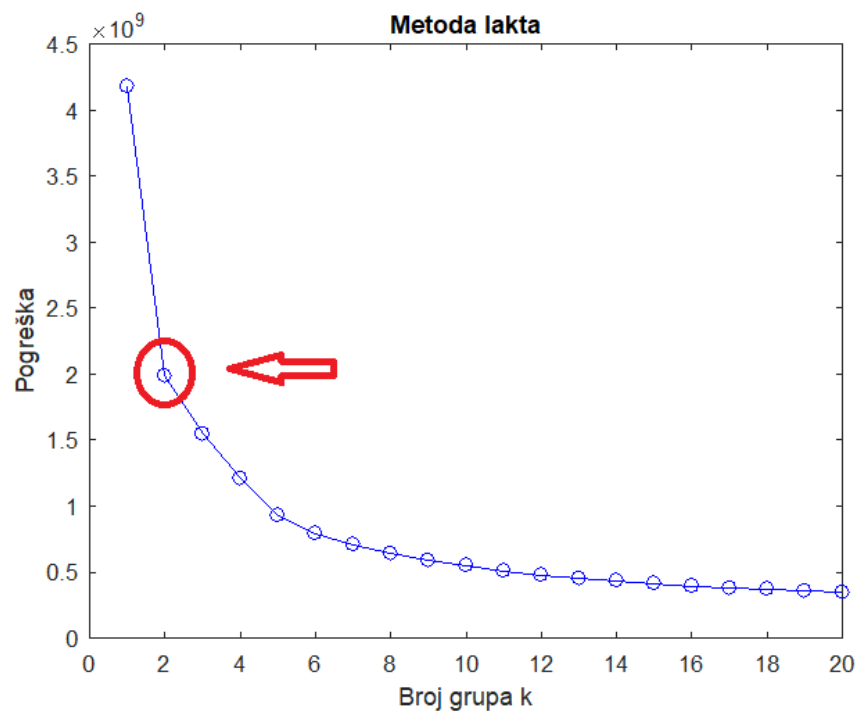
U ovom radu optimalni broj grupa određuju četiri metode, a one su:

- Metoda lakta (engl. Elbow method);
- Analiza Siluete (engl. Silhouette analysis);
- Analiza Calinski-Harabasz (engl. T.Calinski and J.Harabasz analysis, CH);
- Analiza Davies-Bouldin (engl. Davies and Bouldin analysis, DB).

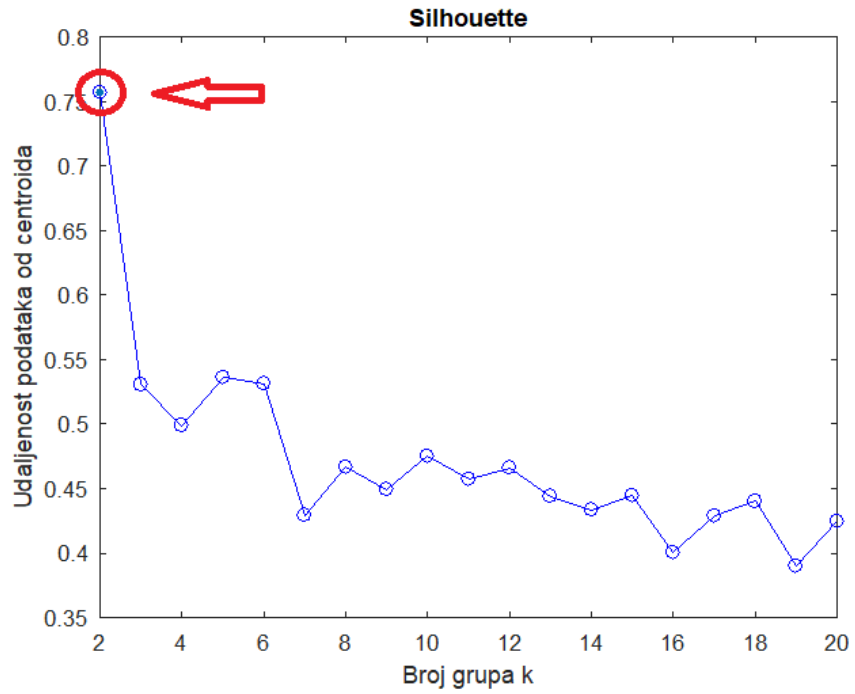
U nastavku teksta prikazan je svaki skup podataka posebno kojega opisuju sve četiri metode. Korišteni skupovi podataka za grupiranje su:

- I. Ukupni prometni tok smjer Karlovac (sve tri godine);
- II. Ukupni prometni tok smjer Zagreb (sve tri godine);
- III. Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec);
- IV. Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).

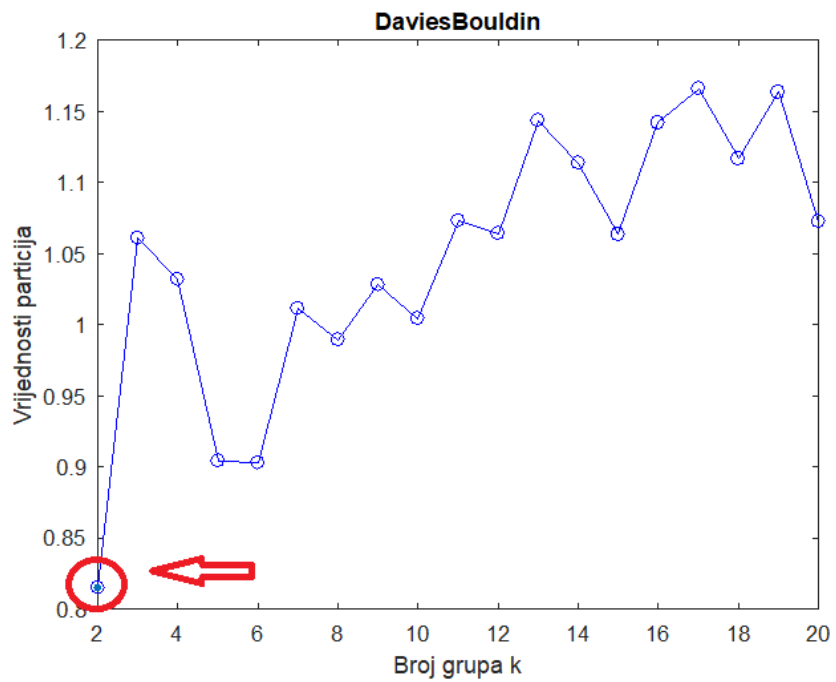
- I. Ukupni prometni tok smjer Karlovac (sve tri godine);



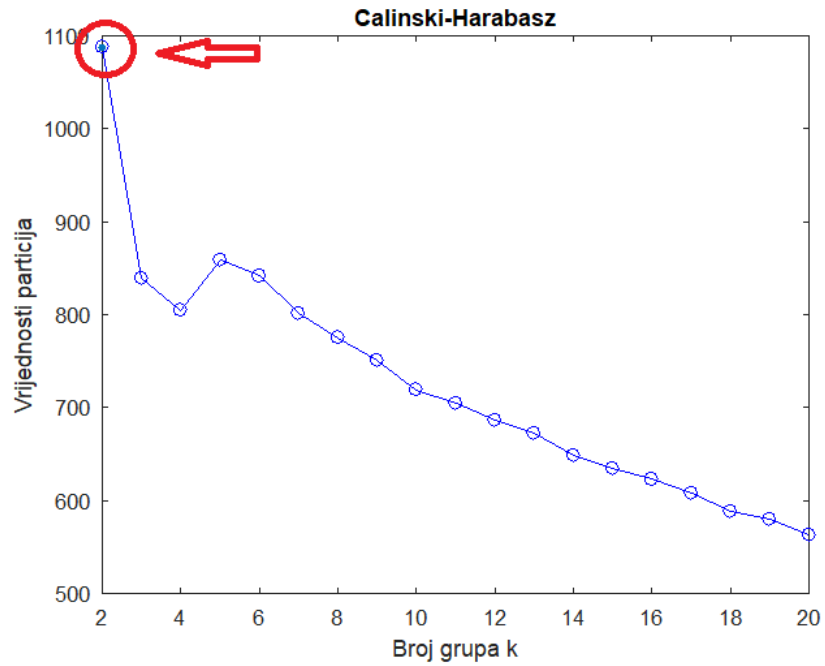
Slika 57. Metoda lakta i broj grupa = 2.



Slika 58. Metoda siluete i broj grupa = 2.

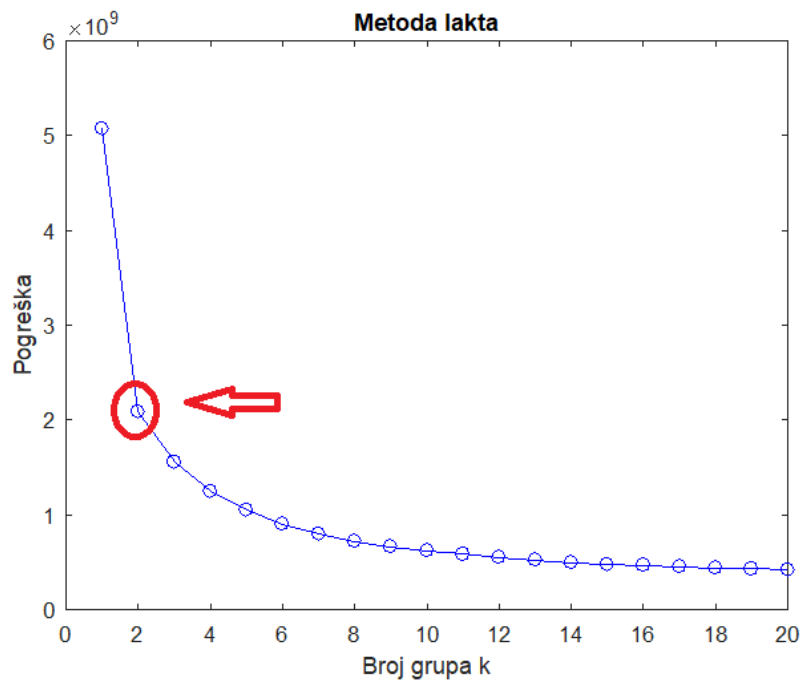


Slika 59. DB metoda i broj grupa = 2.

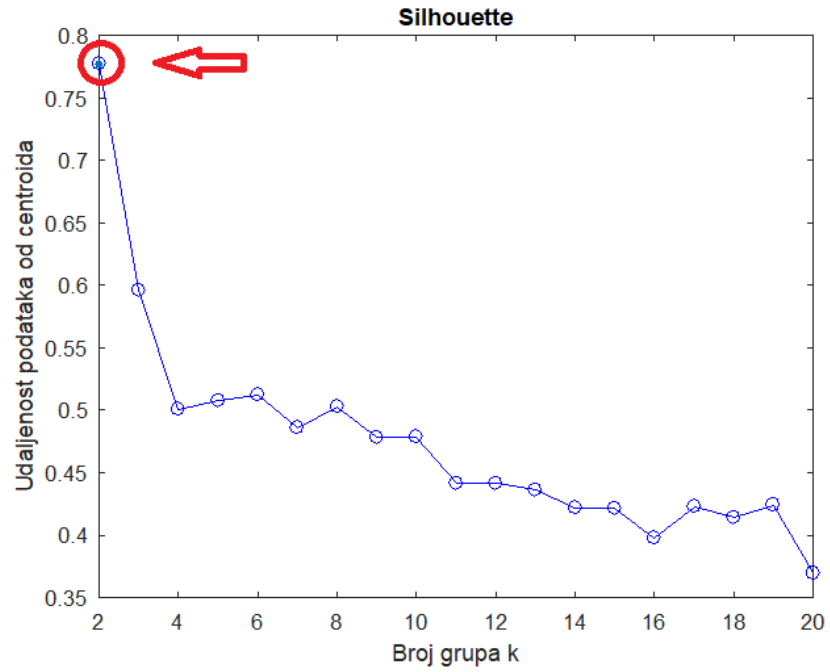


Slika 60. CH metoda i broj grupa = 2.

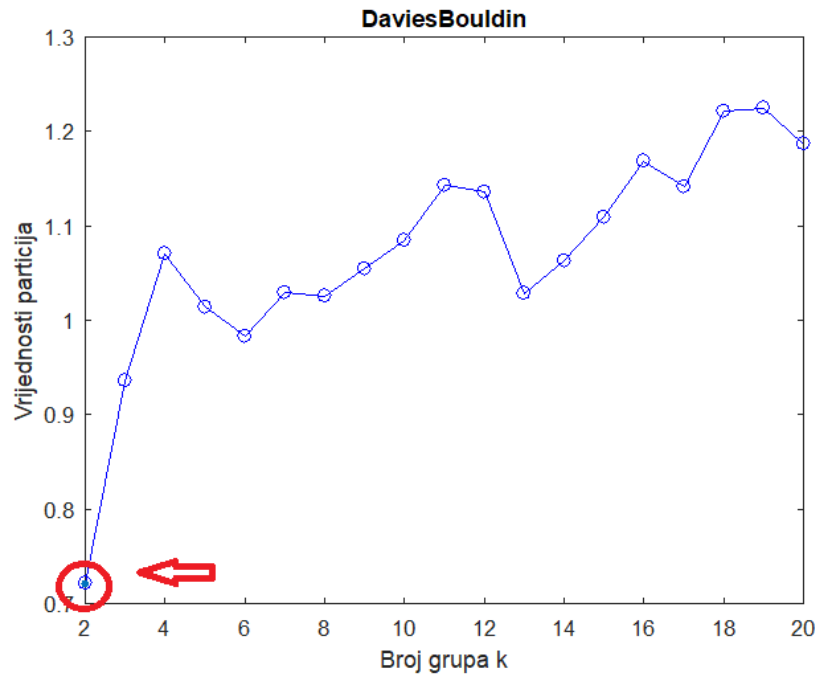
II. Ukupni prometni tok smjer Zagreb (sve tri godine);



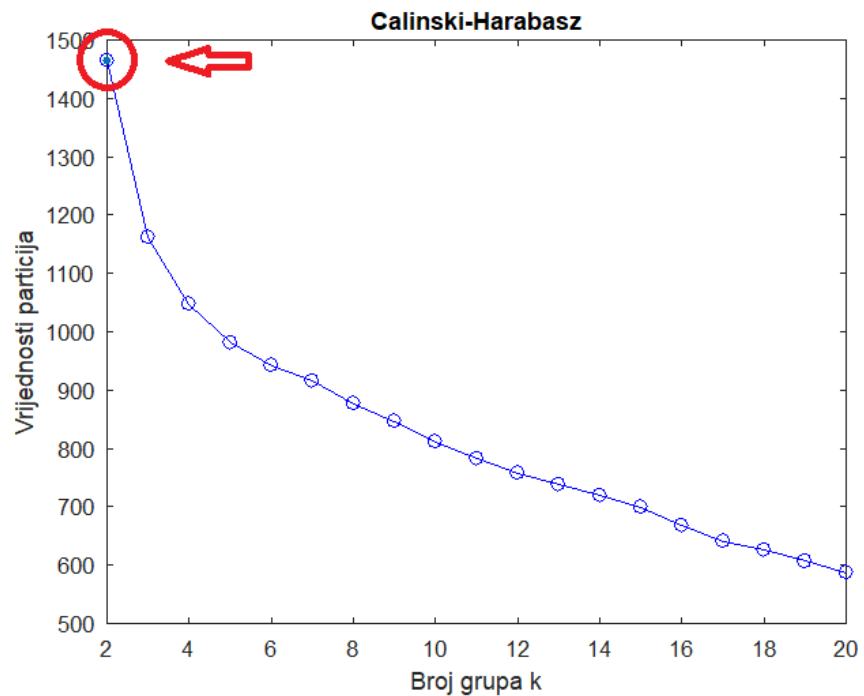
Slika 61. Metoda lakta i broj grupa = 2.



Slika 62. Metoda siluete i broj grupa = 2.

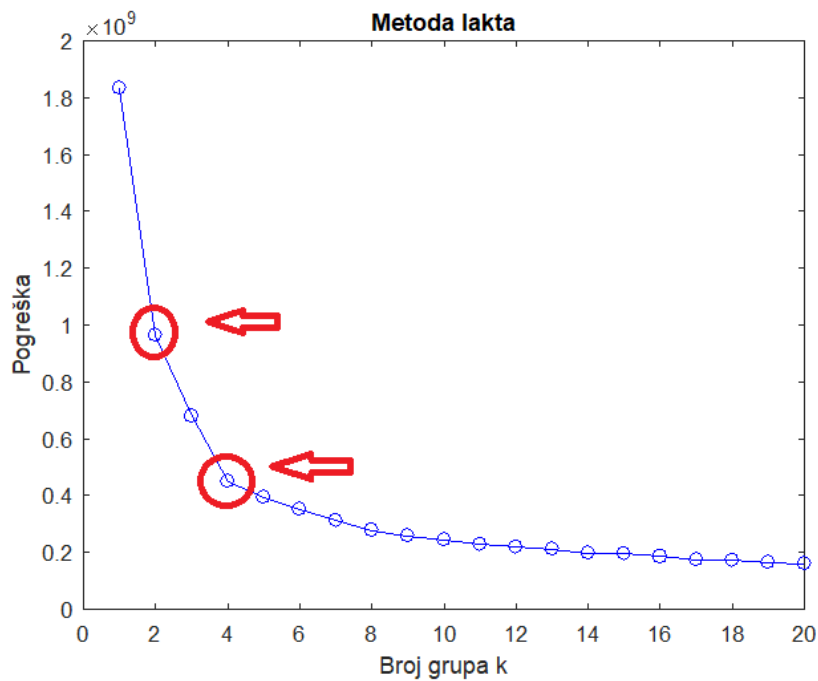


Slika 63. DB metoda i broj grupa = 2.

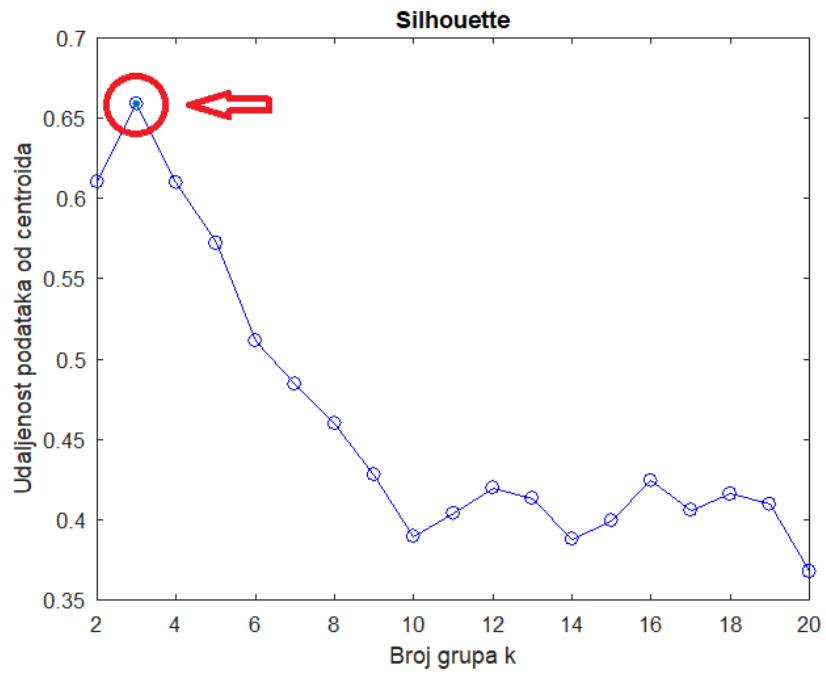


Slika 64. CH metoda i broj grupa = 2.

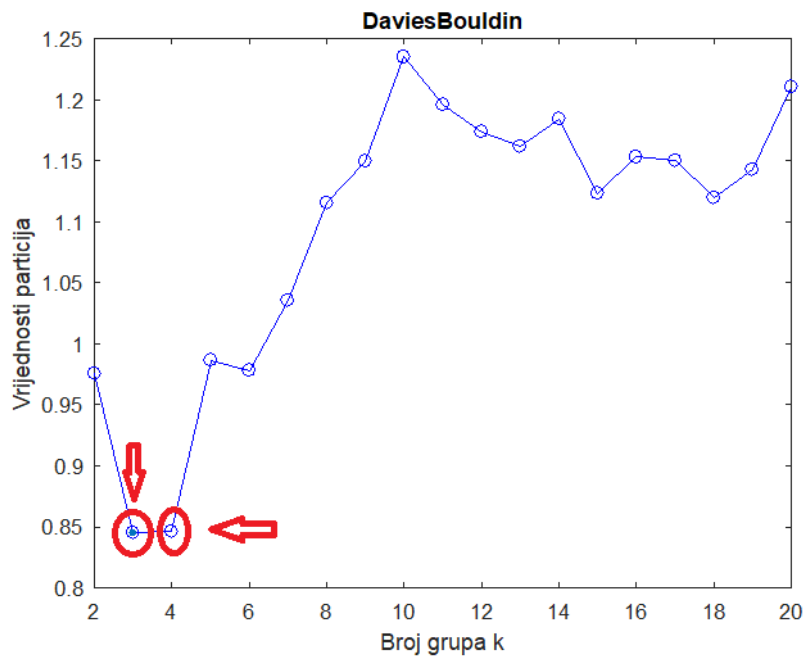
III. Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec);



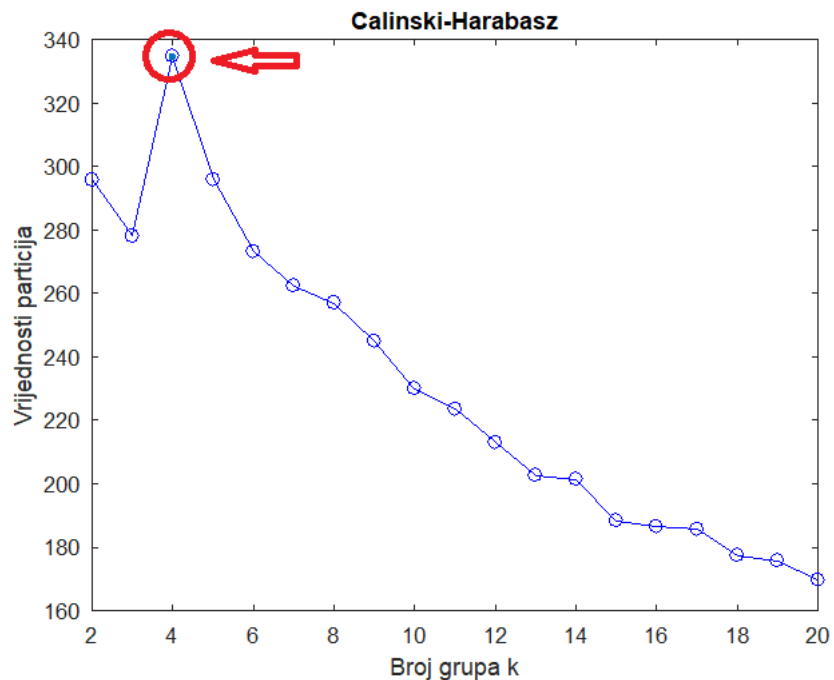
Slika 65. Metoda lakta i broj grupa = 2, može i 4.



Slika 66. Metoda siluete i broj grupa = 3.

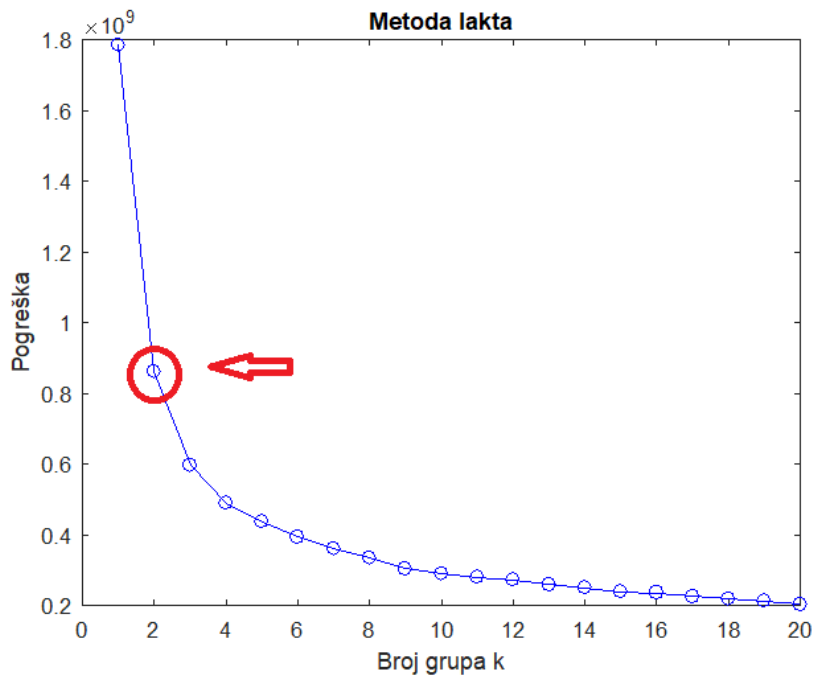


Slika 67. DB metoda i broj grupa = 3, može i 4.

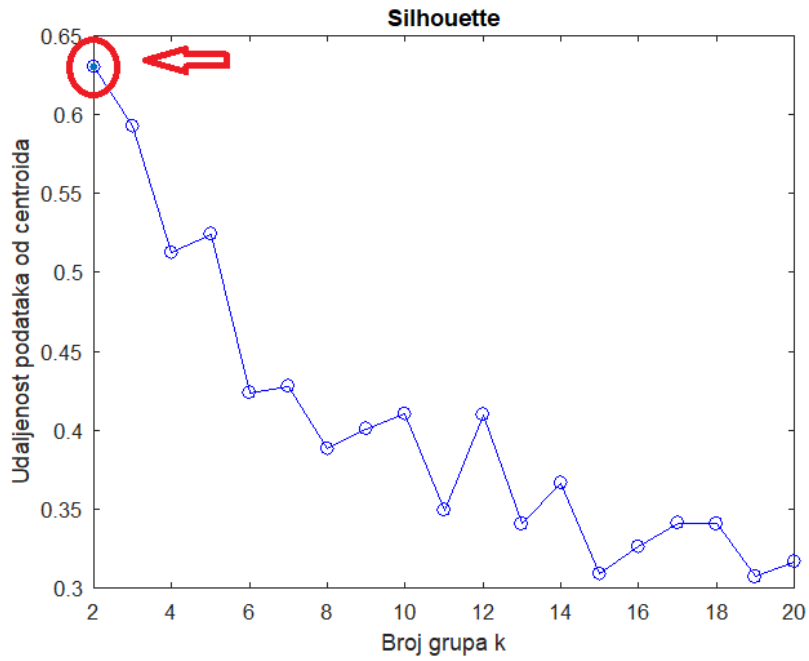


Slika 68. CH metoda i broj grupa = 4.

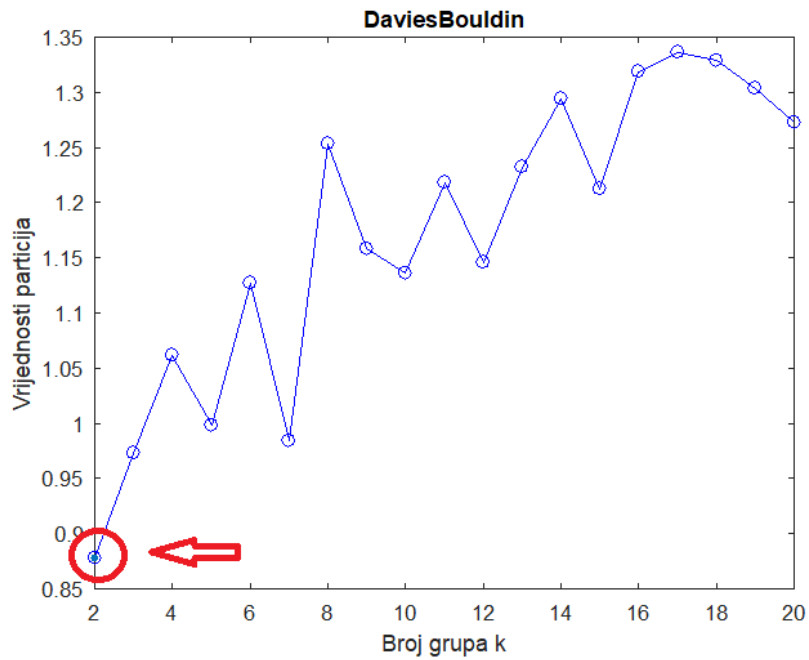
IV. Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).



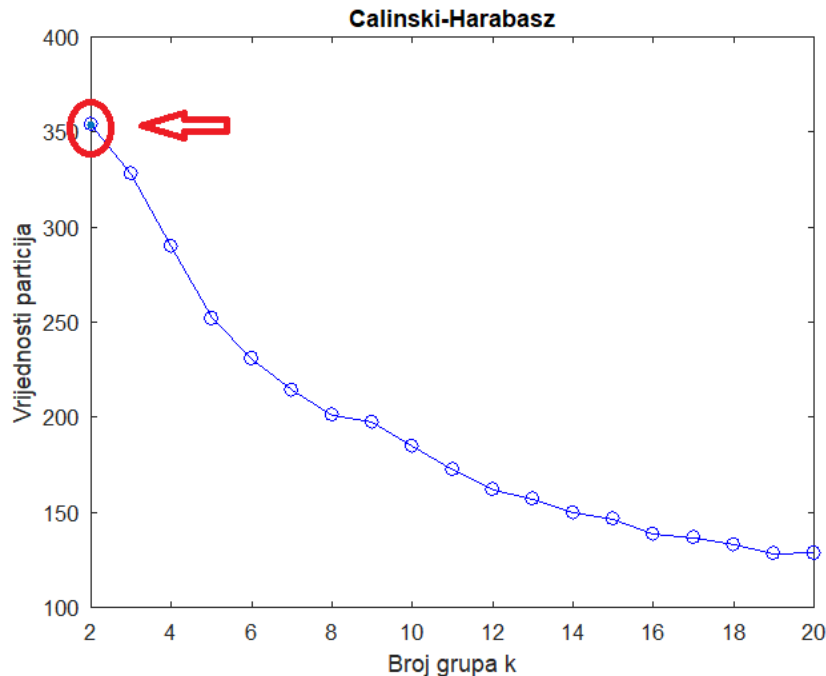
Slika 69. Metoda lakta i broj grupa = 2.



Slika 70. Metoda siluete i broj grupa = 2.



Slika 71. DB metoda i broj grupa = 2.



Slika 72. CH metoda i broj grupa = 2.

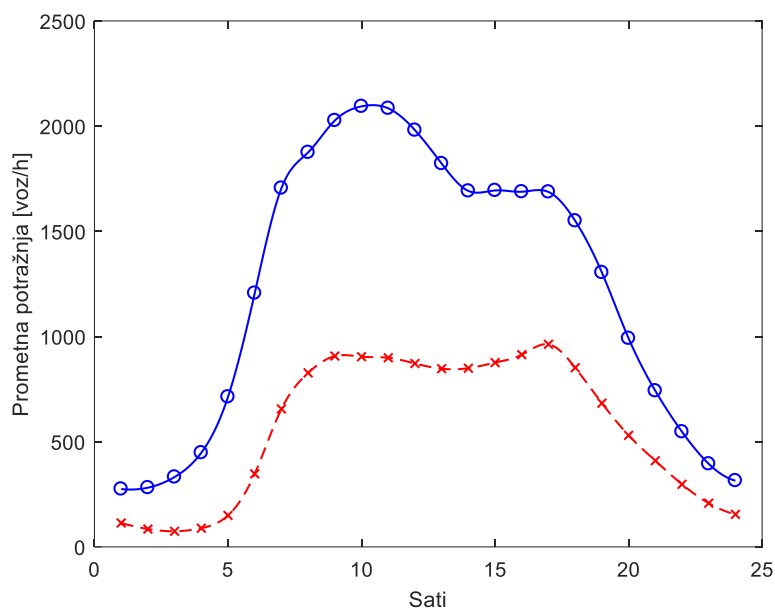
6.4. Rezultati dobiveni metodom K-means

Algoritam K-means primijenjen je na više skupova podataka zbog sezonalnosti. Kao rezultat daje odnos ukupnog prometnog toka u ovisnosti o satima u danu za određeni broj grupa. Korišteni skupovi podataka za grupiranje su:

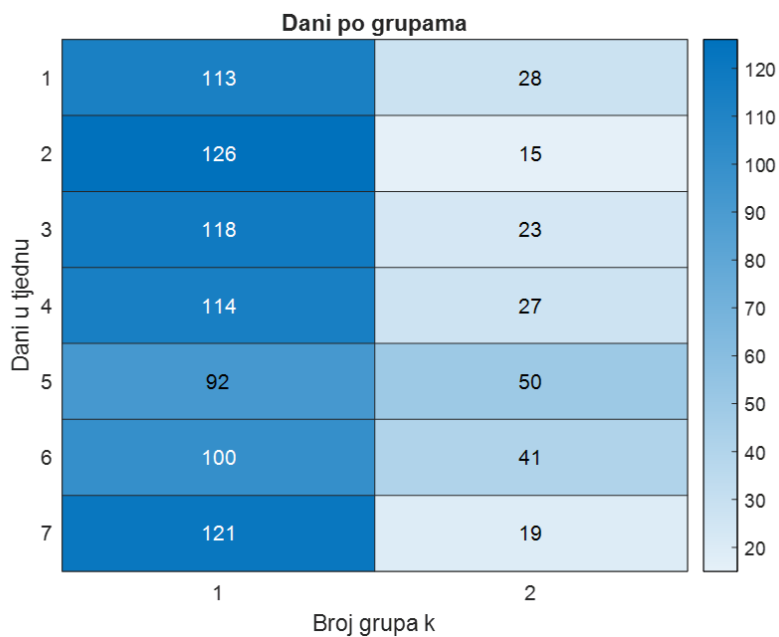
- 1) Ukupni prometni tok smjer Karlovac (sve tri godine);
- 2) Ukupni prometni tok smjer Zagreb (sve tri godine);
- 3) Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec);
- 4) Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).

1) Ukupni prometni tok smjer Karlovac (sve tri godine);

Prikaz slike 73 odnosi se na ukupni prometni tok smjer Karlovac za sve tri godine. Optimalni broj grupa k za taj skup podataka iznosi 2. Slika 74 prikaz je broja grupa za taj skup podataka. Na x-osi prikazan je broj grupa, a na y-osi dani u tjednu (1-7) gdje broj jedan označava ponedjeljak pa tako redom. Svaka grupa u sebi ima dane u tjednu, gdje je u svakom danu pojedine grupe prikazan broj dana koji pripadaju tom danu u tjednu. Isto vrijedi i za ostale takve prikaze.



Slika 73. Optimalni broj grupa $k = 2$.



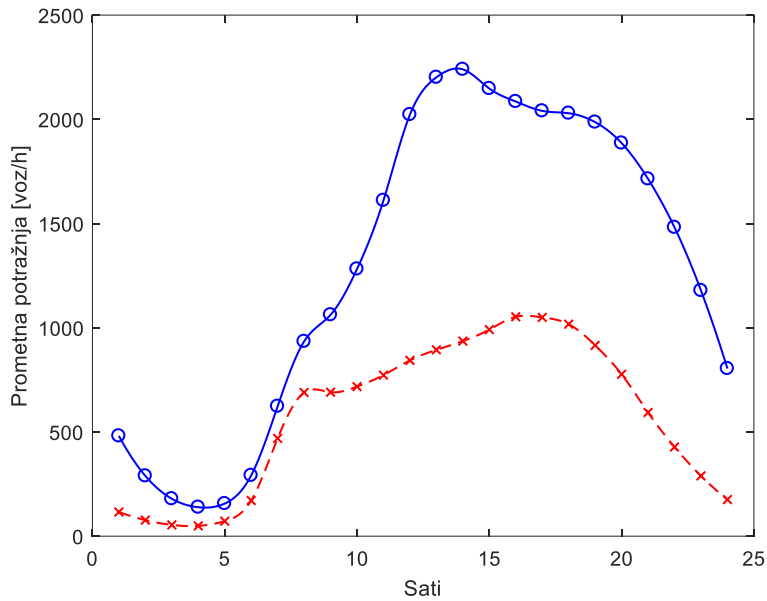
Slika 74. Pripadnost podataka grupama.

Tablica 6. Pripadnost dana pojedinim grupama prema slici 74.

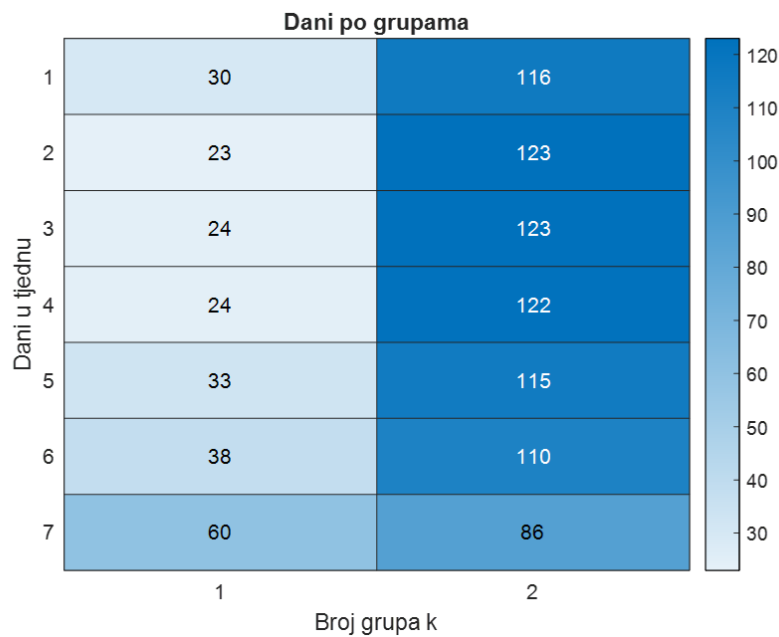
Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Pon, uto, sri, čet, ned
Grupa 2	Pet, sub

2) Ukupni prometni tok smjer Zagreb (sve tri godine):

Prikaz slike 75 odnosi se na ukupni prometni tok smjer Zagreb za sve tri godine. Optimalni broj grupa k za taj skup podataka iznosi 2 i prikazan je slikom 75.



Slika 75. Optimalni broj grupa $k = 2$.



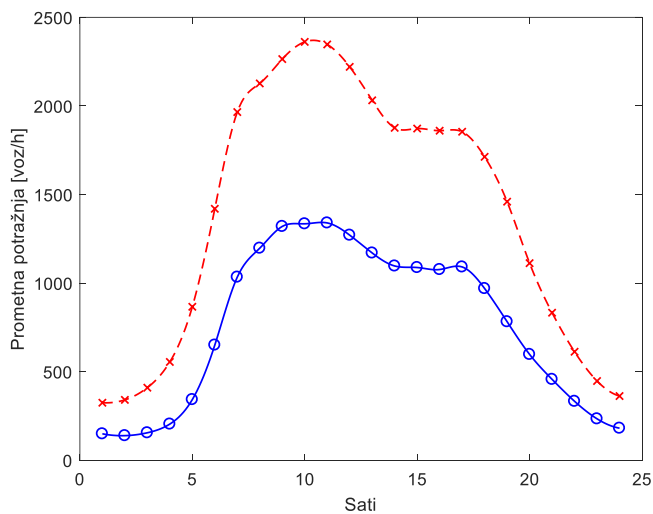
Slika 76. Pripadnost podataka grupama.

Tablica 7. Pripadnost dana pojedinim grupama prema slici 76.

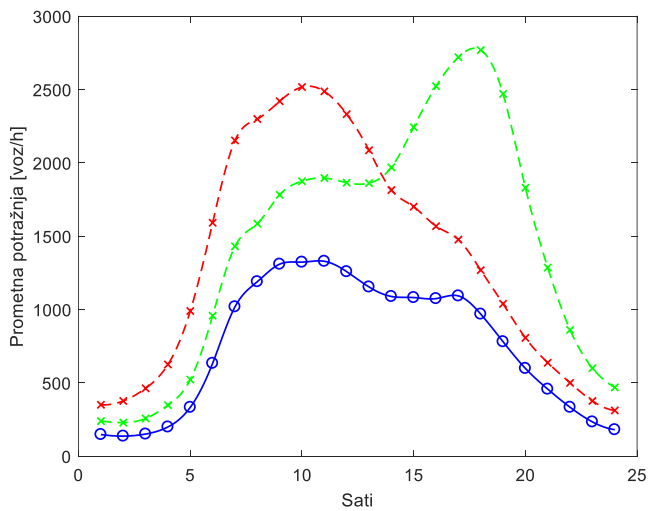
Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	ned
Grupa 2	Pon, uto, sri, čet, pet, sub

3) Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec):

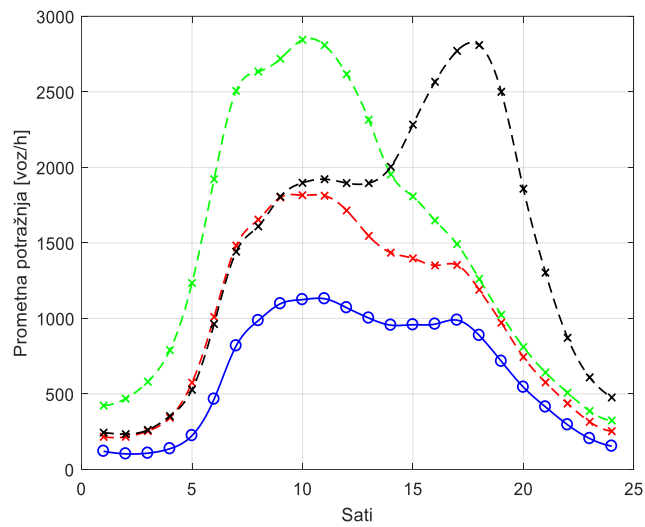
Prikaz slika 77, 78, 79 odnosi se na ukupni prometni tok smjer Karlovac za 6, 7, 8, 9 mjesec. Optimalni broj grupa k za taj skup podataka iznosi 2, 3 ili 4.



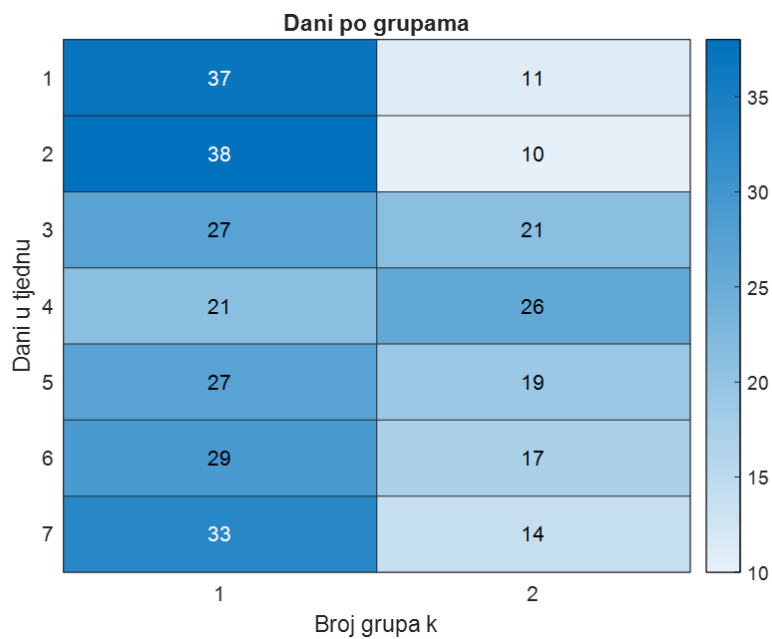
Slika 77. Optimalni broj grupa $k = 2$



Slika 78. Optimalni broj grupa $k = 3$.



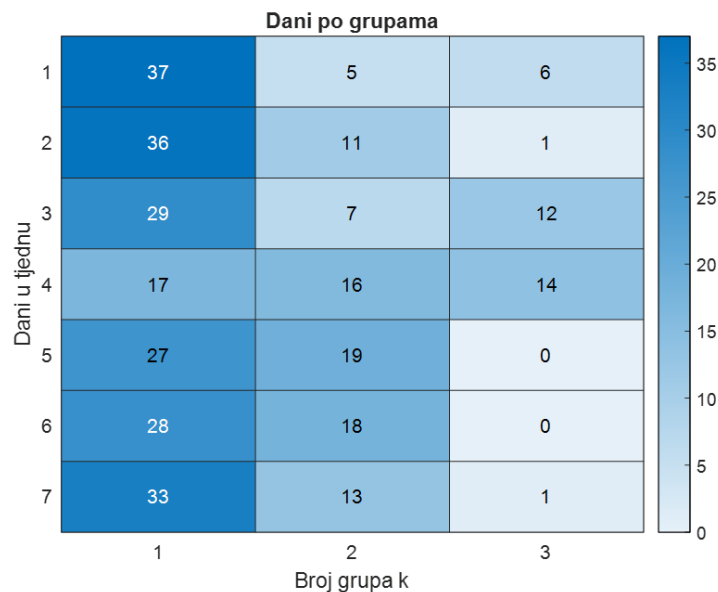
Slika 79. Optimalni broj grupa $k = 4$.



Slika 80. Pripadnost podataka grupama.

Tablica 8. Pripadnost dana pojedinim grupama prema slici 80.

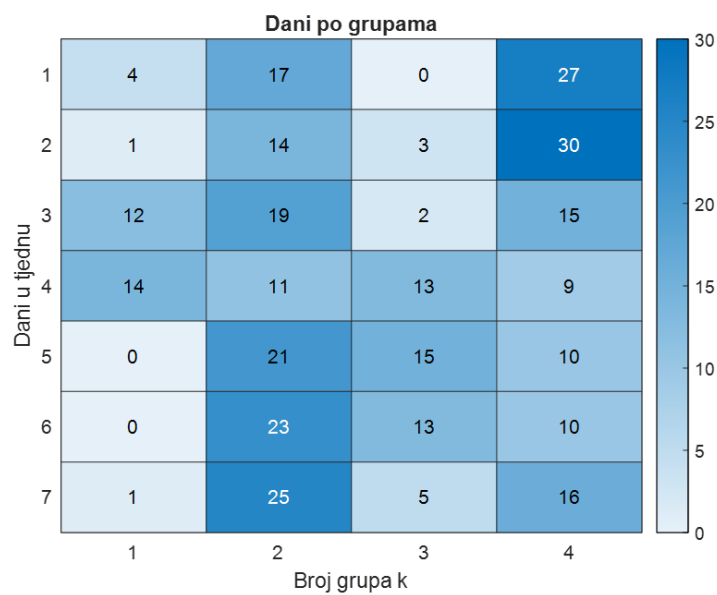
Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Pon, uto, ned
Grupa 2	Sri, čet, pet, sub



Slika 81. Pripadnost podataka grupama.

Tablica 9. Pripadnost dana pojedinim grupama prema slici 81.

Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Pon, uto, ned
Grupa 2	Pet, sub
Grupa 3	Sri, čet



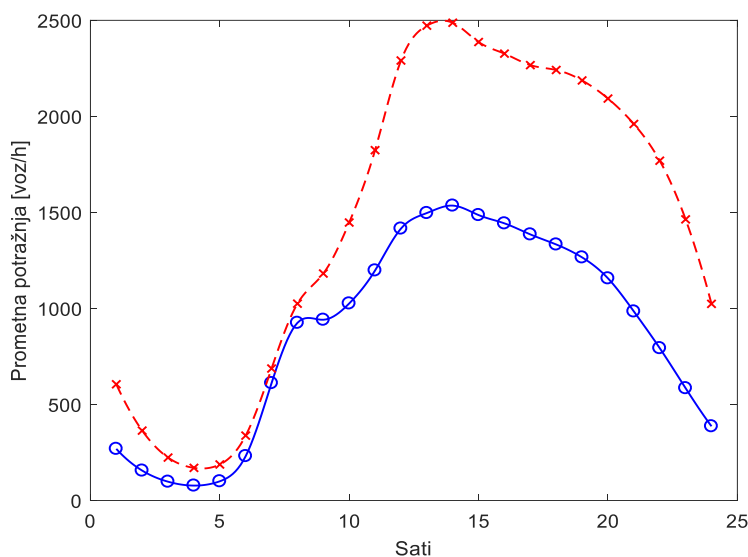
Slika 82. Pripadnost podataka grupama.

Tablica 10. Pripadnost dana pojedinim grupama prema slici 82.

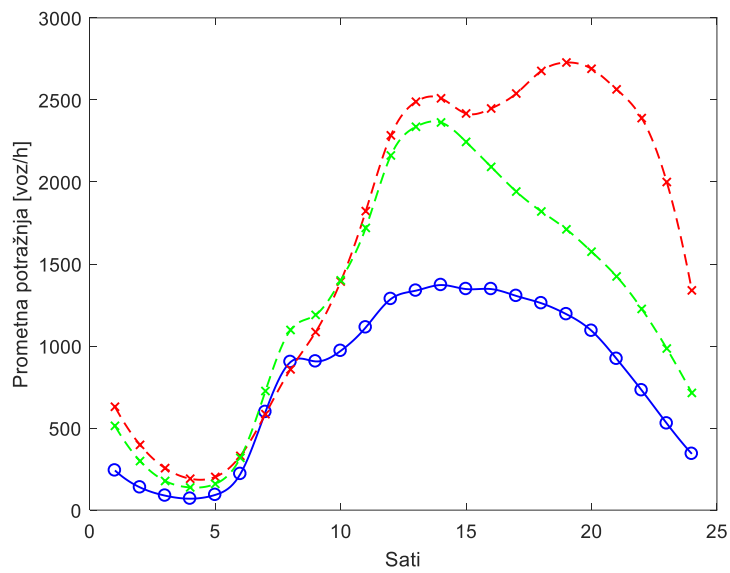
Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Sri, čet
Grupa 2	ned
Grupa 3	Pet, sub
Grupa 4	Pon, uto

4) Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).

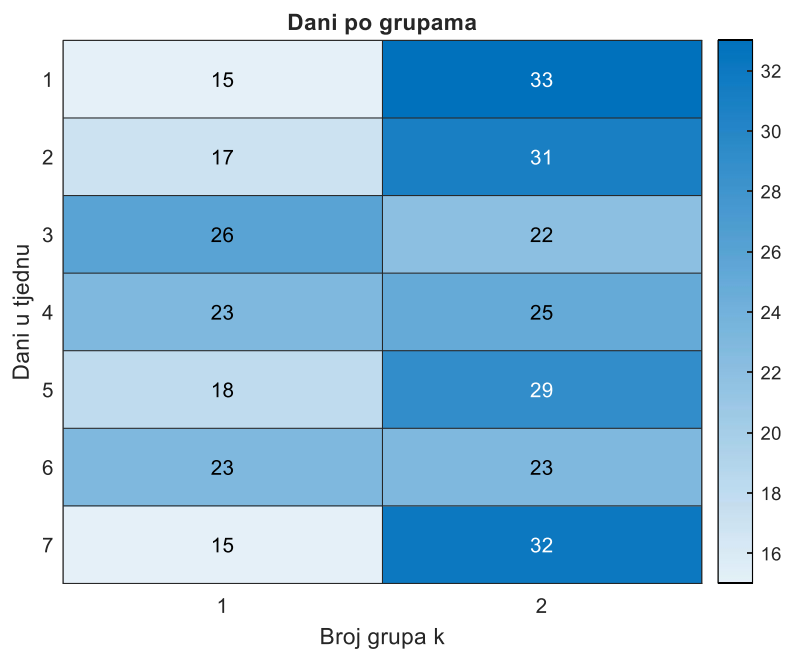
Prikaz slika 83 i 84 odnosi se na ukupni prometni tok smjer Zagreb za 6, 7, 8, 9 mjesec. Optimalni broj grupa k za taj skup podataka iznosi 2, ali je još prikazano rješenje s 3 grupe koje nije bilo potrebno i prikazan je slikama 83 i 84. Rješenje sa optimalnim brojem grupa 3 (slika 86) je prikazano isključivo zbog dokaza kako metode za određivanje broja grupa dobro određuju parametar k . Jer je vidljivo kako nema mogućnosti jasne raspodjele dana po grupama ko što je to slučaj za rješenja za parametar k .



Slika 83. Optimalni broj grupa $k = 2$.



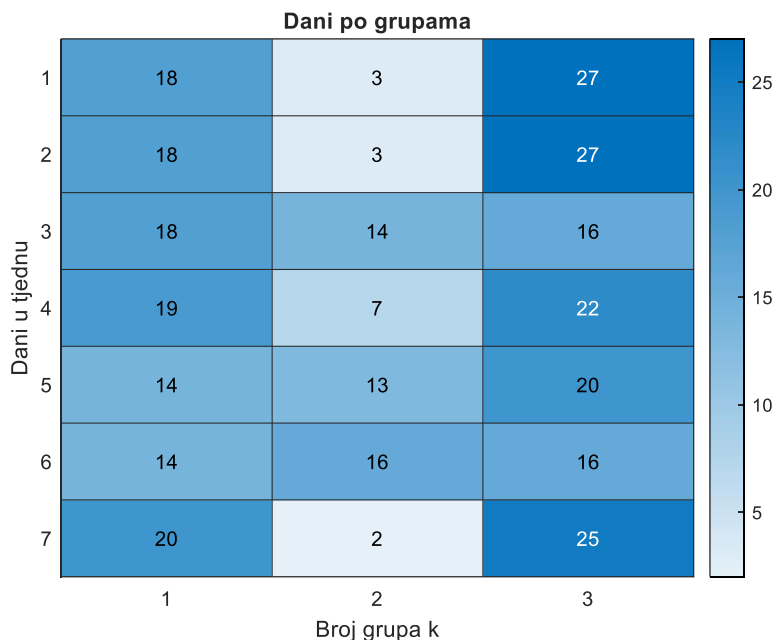
Slika 84. Optimalni broj grupa $k = 3$



Slika 85. Pripadnost podataka grupama.

Tablica 11. Pripadnost dana pojedinim grupama prema slici 85.

Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Sri, čet, sub
Grupa 2	Pon, uto, pet, ned



Slika 86. Pripadnost podataka grupama.

Tablica 12. Pripadnost dana pojedinim grupama prema slici 86.

Grupe	Dani koji pripadaju pojedinoj grupi
Grupa 1	Čet, ned
Grupa 2	Sri, pet, sub
Grupa 3	Pon, uto

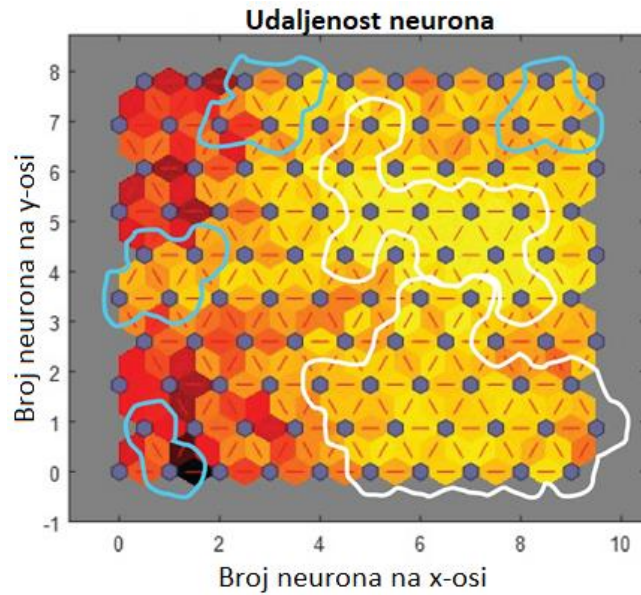
6.5. Rezultati dobiveni metodom neuronskih mreža

Dobiveni rezultati nisu u skladu sa očekivanim prvenstveno zbog sezonalnosti, a potom i velikim djelom zbog nedostajućih zapisa. Metoda neuronskih mreža primijenjena je na četiri skupa podataka zbog sezonalnosti. Kao rezultati neuronskih mreža prikazani su sljedeći prikazi: prikaz udaljenosti neurona/grupa, prikaz sličnosti ulaznih podataka i prikaz težinskih ravnina. Svaki od tih prikaza nudi raspored neurona/grupe. U nastavku će biti opisan svaki prikaz rezultata. Kao i u pod poglavlju prije, korišteni su sljedeći skupovi podataka za grupiranje:

- 1) Ukupni prometni tok smjer Karlovac (sve tri godine);
- 2) Ukupni prometni tok smjer Zagreb (sve tri godine);
- 3) Ukupni prometni tok smjer Karlovac (6, 7, 8, 9 mjesec);
- 4) Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).

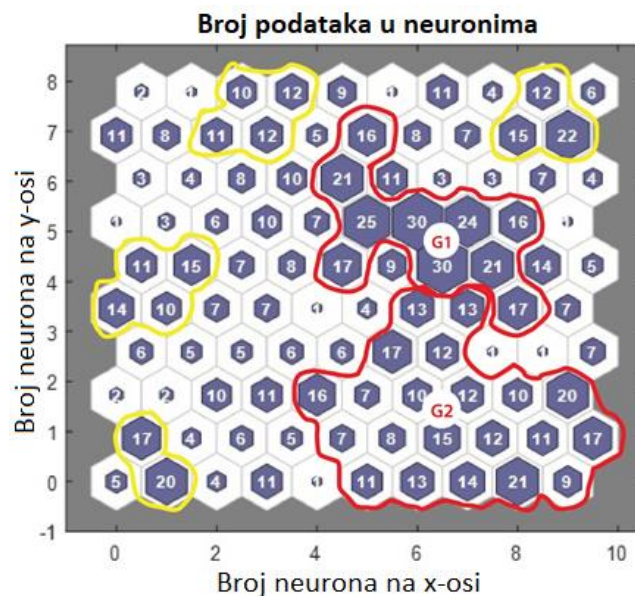
1) Ukupni prometni tok smjer Karlovac (sve tri godine);

Optimalni broj grupa za ovaj skup podataka prema K-means algoritmu iznosi 2.



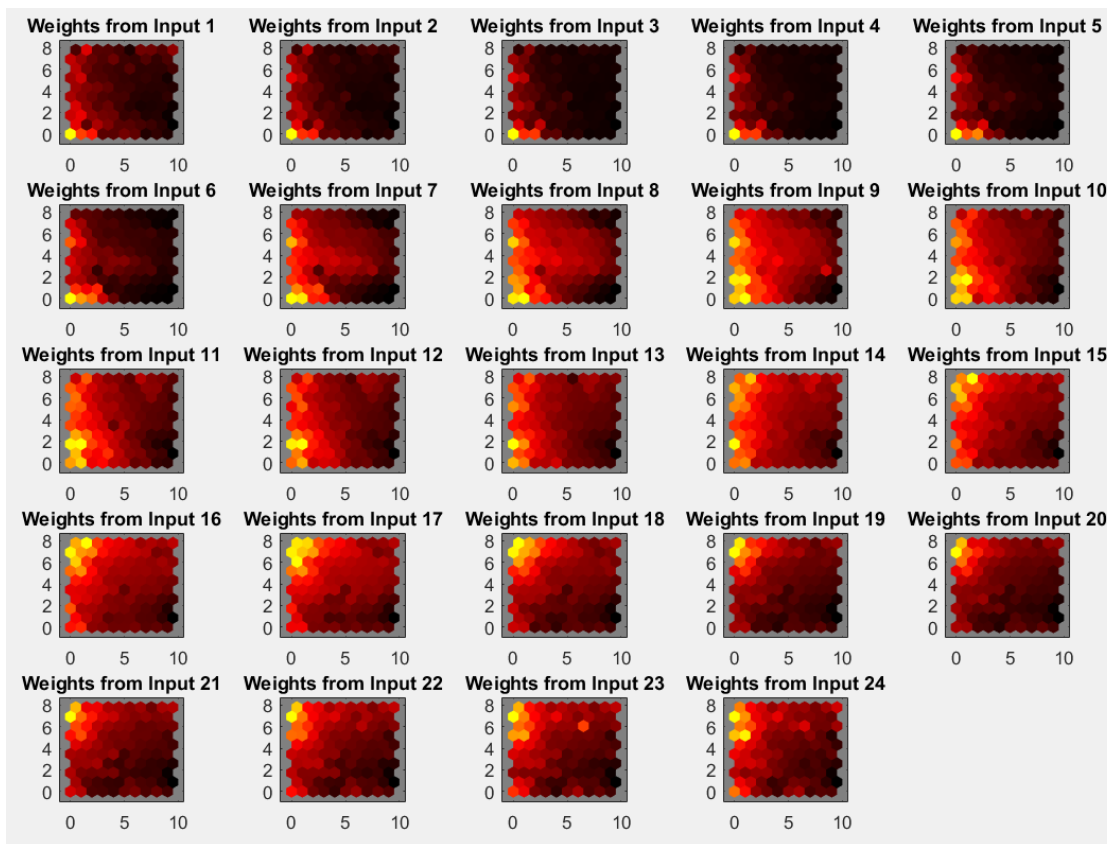
Slika 87. Dvodimenzionalna samoorganizirajuća mapa.

Prikaz slike 87 ne nudi jasno razdvajanje grupa ali u nekoj manjoj mjeri ipak su razdvojene. Dolje i gore lijevo gdje prevladava tamnija crvena boja jasno je da je među neuronima velika razlika i tu ne dolazi do grupiranja. Označena područja bijelom bojom su dvije grupe dok područja označena plavom bojom predstavljaju potencijalne male grupe.



Slika 88. Prikaz broja sličnih ulaznih podataka po neuronima.

Slika 88 pruža uvid u broj dana koji pripada svakom neuronu/grupi (plavi šesterokut). Kao što je gore spomenuto, prilično dobro se vide dvije veće grupe označene crvenom bojom s još manjim grupama oko koje su označene žutom bojom.



Slika 89. Prikaz težinskih ravnina.

Slika 89 prikaz je težinskih ravnina po satima u danu, a njen prikaz podataka je na slici 39. Pokazuje kako pojedina težina utječe na cjelokupnu sliku. Žuti neuroni su ujedno i najaktiviraniji, dok su crni neaktivni. Žuti neuroni također opisuju visoke vrijednosti prometnog toka.

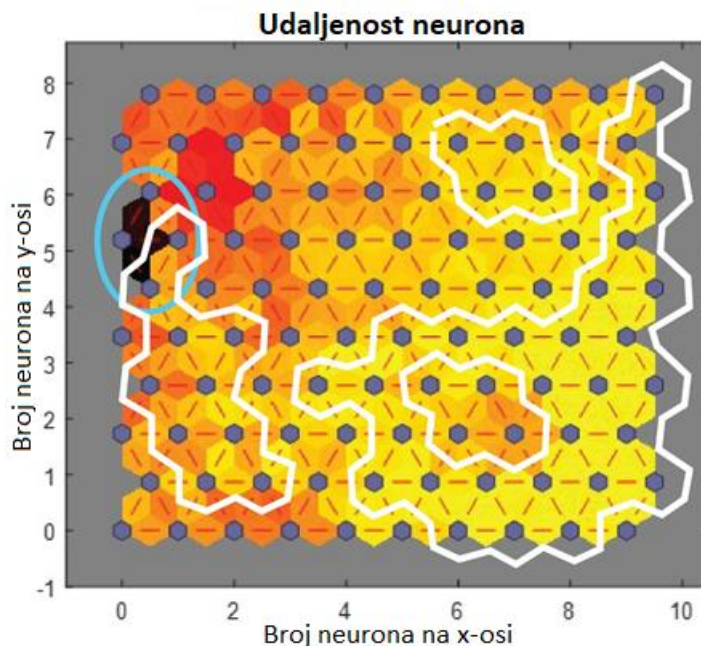
Žute grupe neurona predstavljaju izrazitu osjetljivost na ponašanje podataka. Tako u prvom prikazu slike 89 dolje lijevo postoji žuta grupa neurona koji predstavlja izrazitu osjetljivost na ponašanje podataka u ponoći. To znači da kod prikaza prometnog toka u danu postoji krivulja koja u ponoć ima visoku vrijednost. Što govori da neuroni koji su aktivni u donjem lijevom kutu su grupe koje po noći imaju visoke vrijednosti.

Gledano za središnji dio dana (16h, 17h) vidljivo je da su aktivirane grupe gore lijevo. To znači da je taj dio skup grupa kada postoji visok poslijepodnevni vršni sat. Općenito kada se grupirani aktivni neuroni nalaze u gornjem lijevom kutu predstavlja grupe kada je jako izražen popodnevni vršni sat. Dok kada je izražen jutarnji vršni sat (9h, 10h) ti grupirani aktivni neuroni su smješteni dolje lijevo.

Na dalje, moguć je slučaj za isti broj grupa aktivnih neurona za različite sate, a vidljiv je na prvih pet prikaza (od ponoći do 5 sati ujutro). To znači da je ponašanje grupa u tom vremenskom razdoblju jako slično kao i prometni tok. Zadnja tri prikaza (od 22h do ponoći) također su istog broja aktivnih neurona, dok odmah nakon ponoći to nije slučaj.

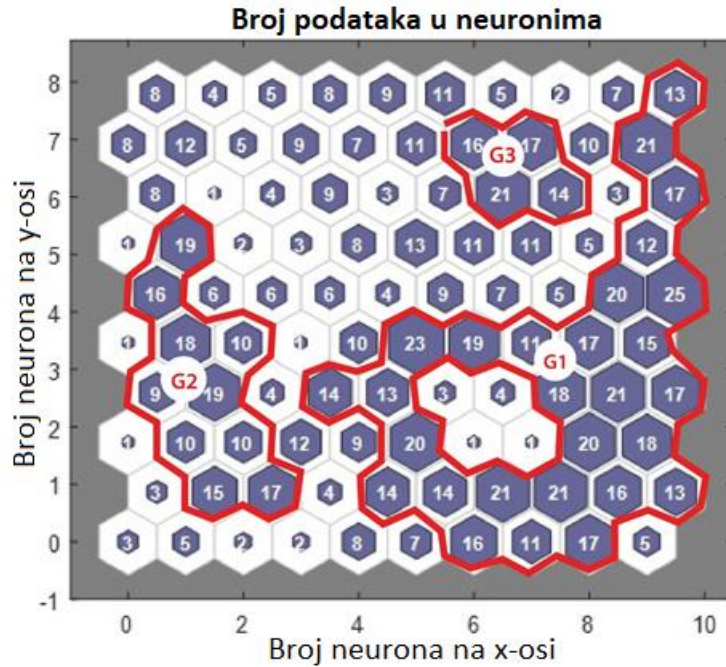
2) Ukupni prometni tok smjer Zagreb (sve tri godine):

Optimalni broj grupa prema K-means algoritmu iznosi 2.



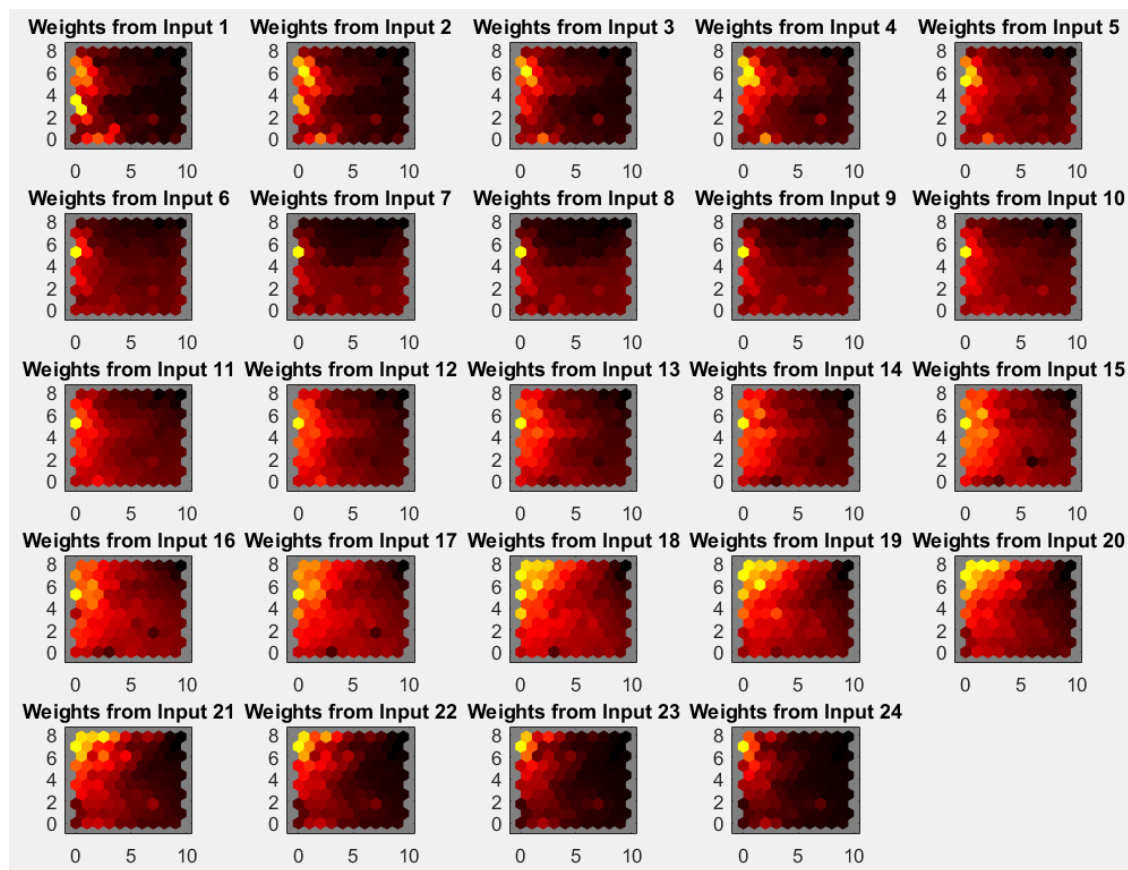
Slika 90. Dvdimenzionalni SOM.

Prikaz slike 90 ne nudi jasno razdvajanje grupa ali u nekoj manjoj mjeri ipak su razdvojene. Ono što je jako dobro vidljivo iz slike je izuzeti neuron (zaokružen plavom linijom) gdje je njegova udaljenost od ostalih jako velika. Izuzet je zbog jedne plave krivulje koja predstavlja prometni tok za svaki sat u tom danu (slika 40). Taj podatak je toliko različit od ostalih da ga je SOM izdvojio, dok algoritam K-means to ne vidi osim ako bi postojao velik broj zapisa s tim približnim iznosom. Dakle SOM prikazuje svaku specifičnu situaciju koja je jako različita od drugih, dok to nije slučaj kod algoritma K-means. Takvi slučajevi su jako interesantni i važni za učenje nekog sustava upravljanja. Kod takvih slučajeva važno je prvo provjeriti da se ne radi o zapisu koji je van-distribucijski. Jer ako je, potrebno ga je izbrisati iz skupa podataka. Ako nije onda je to specifičan podatak koji nešto govori i definitivno ga treba uzeti u obzir kod bilo kakvog upravljanja ili vizualizacije jer on može predstavljati određeni incident, loše vremenske uvjete i slično. Ostala područja označena bijelom bojom predstavljaju tri grupe.



Slika 91. Prikaz broja sličnih ulaznih podataka po neuronima.

Kao što je gore spomenuto, na slici 91 prilično dobro se vide jedna velika grupa, druga nešto manja i treća kao najmanja. Također na lijevoj strani postoje neuroni s vrijednostima jednog podatka. Potencijalno, to bi mogli biti podaci koji imaju poseban značaj i kao takve važno ih je istražiti.



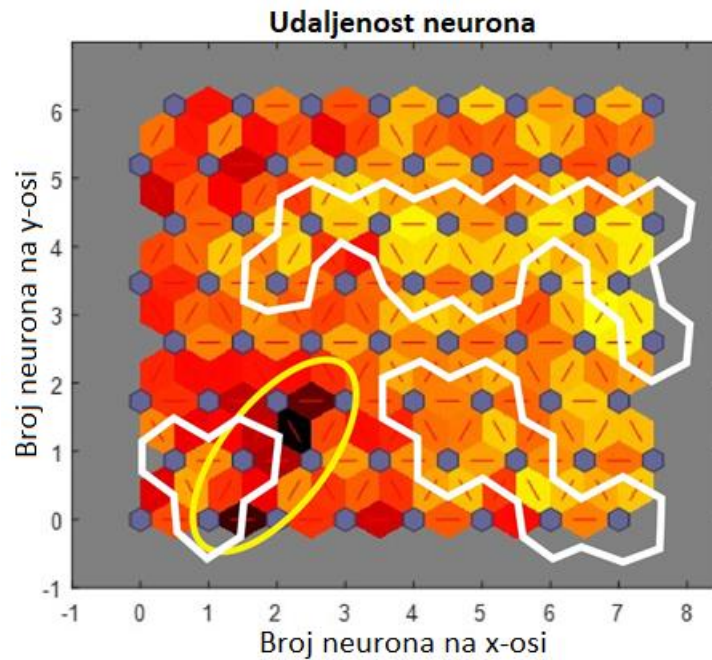
Slika 92. Prikaz težinskih ravnina.

Prikaz slike 92 vidljivo se razlikuje od prikaza težinskih ravnina (slika 89). Interpretacija težinskih ravnina nije ista u ovim slučajevima jer podaci koji se analiziraju trenutno nisu isti. Drugi razlog je da se početna rešetka SOM-a postavlja nasumično i tada može doći do različitog pozicioniranja. Dakle važno je težinske ravnine promatrati odvojeno i pratiti razlike između svakog od tih 24 prikaza kako bi se došlo do nekog zaključka.

Ono što se može usporediti s prikazom slike 89 je velika aktivacija neurona u rano jutarnjim satima (od ponoći do 6 sati u jutro), dok tamo to nije bio slučaj. Isto tako od 6 sati ujutro pa do 10 h prije podne su jako slabe aktivacije, što također nije slučaj na prikazu slike 89. Situacija je takva jer se radi o smjeru prema Zagrebu i slikom 40 se može vidjeti kako do podne nema vršnih opterećenja, odnosno da krivulje znatno sporije rastu, tj. odvijanje prometnog toka krenulo znatno kasnije u odnosu na suprotan smjer. Razlog tomu je opet sezonalnost jer tada velikim djelom prometnog toka čine ljudi koji se vraćaju kućama s godišnjih odmora. Ljudi na put kreću oko podnevnih sati jer se većinom u jutarnjim satima ljudi još kupaju, spremaju za put, napuštaju apartmane i slično.

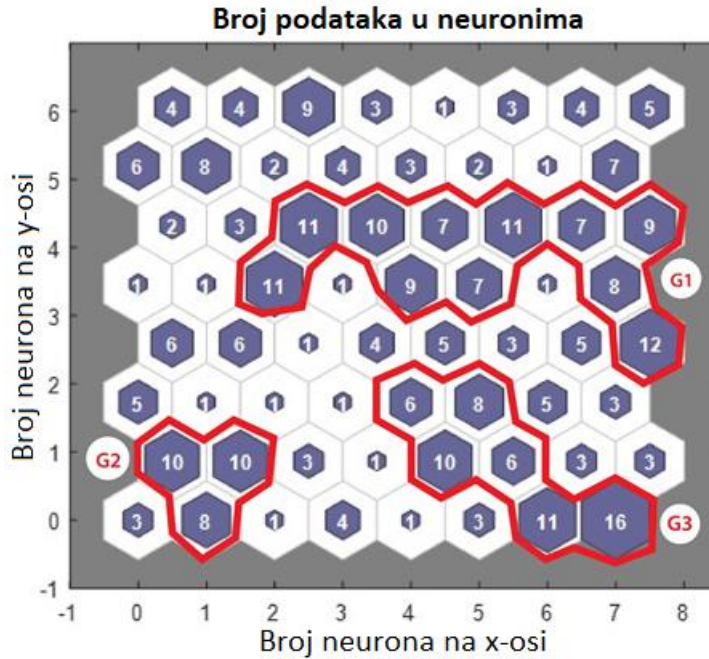
3) Ukupni prometni tok smjer Karlovac (6. 7. 8. 9 mjesec):

Optimalni broj grupa prema K-means algoritmu iznosi 2, 3 i 4.



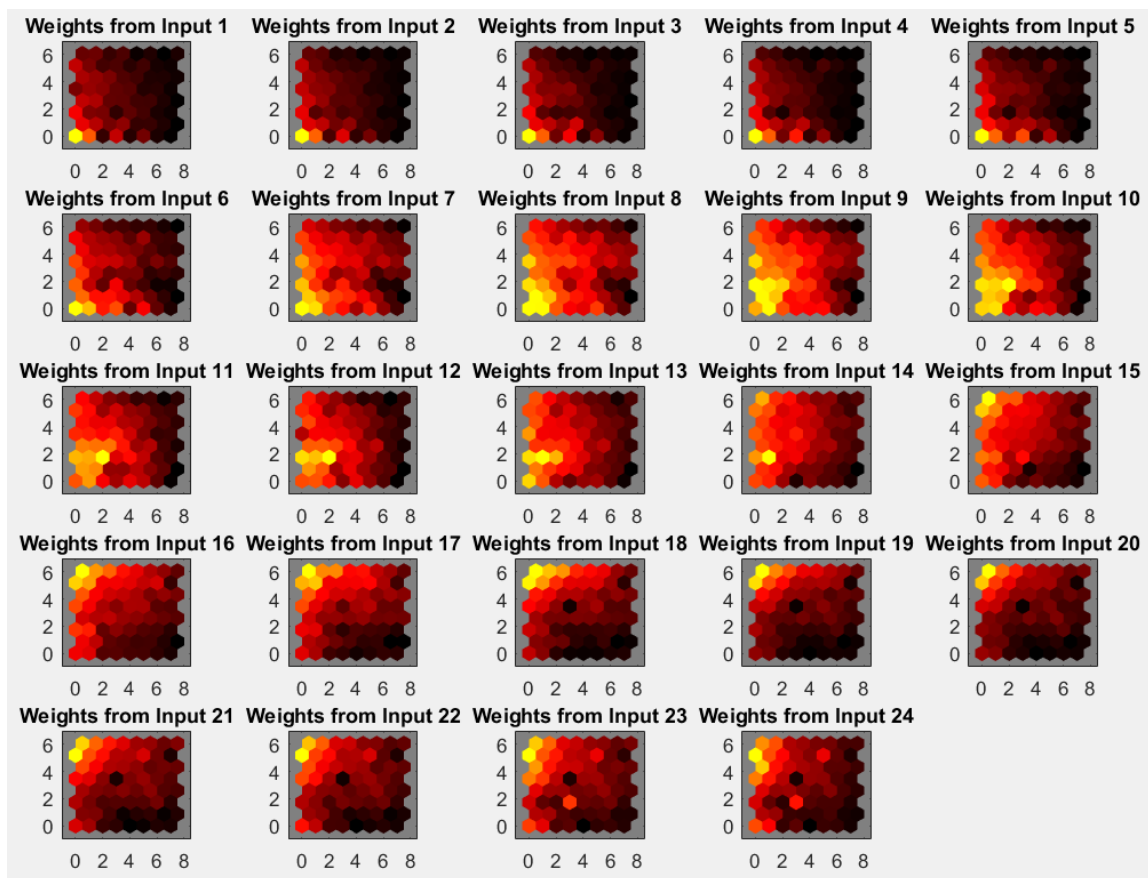
Slika 93. Dvodomensjonalna samoorganizirajuća mapa.

Prikaz slike 93 također ne nudi jasno razdvajanje grupa ali u nekoj manjoj mjeri ipak se vide neke granice grupa. Žutom bojom zaokružene tamne udaljenosti između 6 neurona govori da je velika razlika u tim podacima i tu ne dolazi do grupiranja. Potencijalno to su možda rijetke situacije u prometu poput već spomenutih incidenata, lošeg vremena i slično. Jasni odabir optimalnih broja grupa na ovom prikazu nije vidljiv. S toga ove tri označene grupe na slici 93 nisu i konačne. Može biti i manje i više grupa dato kao rješenje. Područja označena bijelom bojom označuju te tri spomenute grupe.



Slika 94. Prikaz broja sličnih ulaznih podataka po neuronima.

U prikazu slike 94 nije jasno određen prijelaz s jedne grupe na drugu. Kod takvih rezultata potrebno je tražiti bilo kakvu razliku u broju podataka po neuronima kako bi se uspjele stvoriti kakve takve grupe. Kao i na ovom prikazu, često takve grupe nisu u obliku nekih krugova i slično, već jako često budu u obliku ravnih i krivudavih linija kao u ovom slučaju.

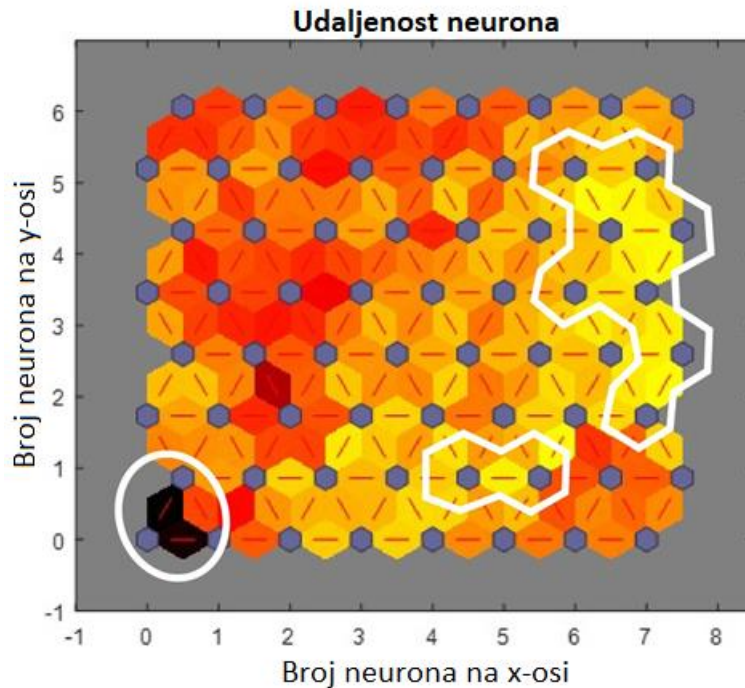


Slika 95. Prikaz težinskih ravnina.

Prikaz slike 95 dosta je sličan prikazu težinskih ravnina (slika 89). Dakle od ponoći pa do 6 sati ujutro nema puno aktivnih neurona što u prometnom smislu znači da je intenzitet prometa mali. Od 6 sati ujutro pa do 13 sati vidljiva je velika aktivacija neurona što predstavlja velik intenzitet prometa. Od 13 do 16 sati je znatno manja aktivacija neurona u odnosu na prijepodne i poslije podne od 16 sati. Zato što ljudi kada kreću na godišnje odmore, kreću rano ujutro tako da su čim prije na destinaciji pa oni koji krenu oko podneva bolje prođu što se tiče gužvi (slika 41). Zaključci koji su doneseni su takvi zbog skupa podataka ljetne sezone. Dakle nije isto radi li se o smjeru Zagreb ili Karlovac i podacima uzetim za sve tri godine ili samo iz perioda ljetne sezone.

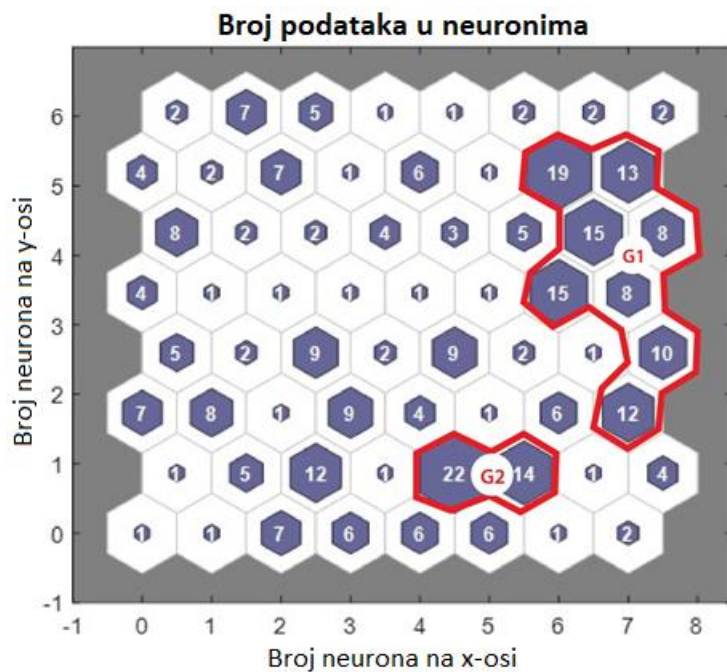
4) Ukupni prometni tok smjer Zagreb (6, 7, 8, 9 mjesec).

Optimalni broj grupa prema K-means algoritmu iznosi 2.



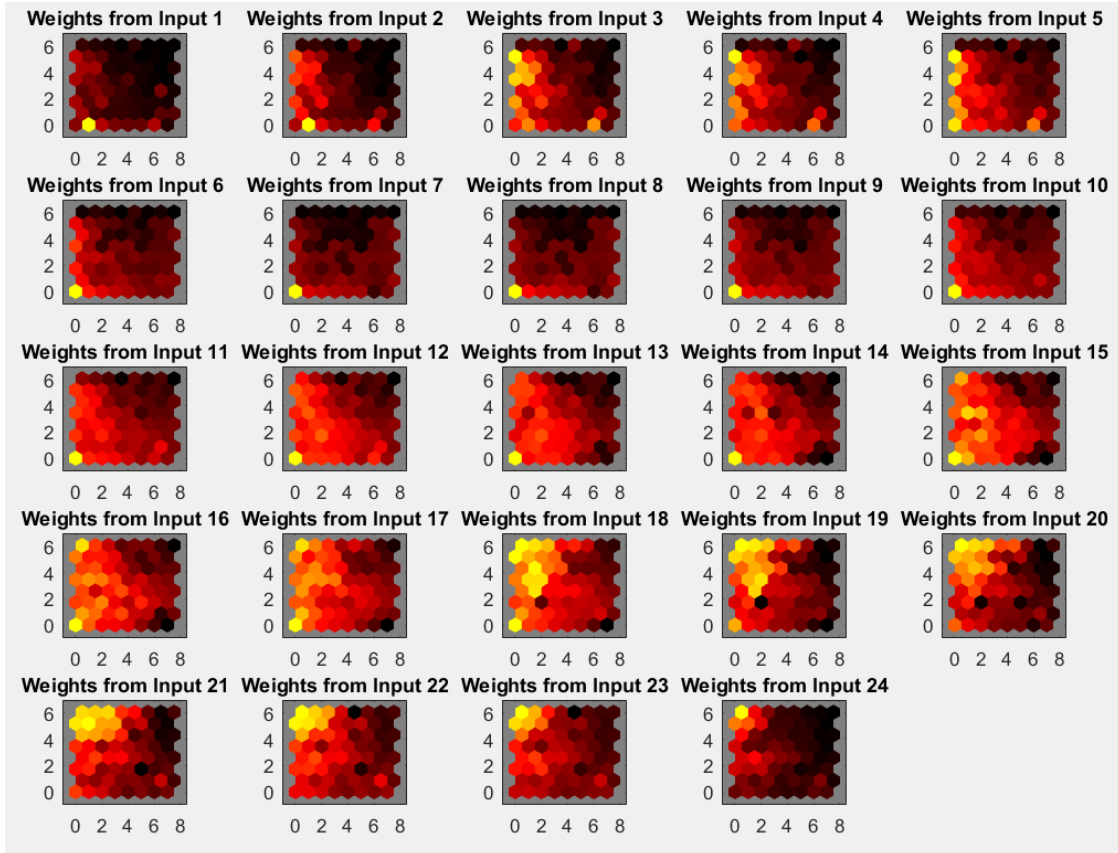
Slika 96. Dvodimenzionalna samoorganizirajuća mapa.

Prikaz slike 96 prilično dobro opisuje razdvajanje grupa u odnosu na smjer Karlovac za sezonu. Ono što je jako dobro vidljivo iz slike je izuzeti neuron (zaokružen bijelom linijom) gdje je njegova udaljenost od ostala dva susjedna neurona jako velika. Taj izuzeti neuron (van-distribucijski podatak) je zbog jedne žute krivulje koja predstavlja prometni tok za svaki sat u tom danu (slika 42) kao i u slušaju smjer Zagreb za sve 3 godine. Na slici desno, područja zaokružena bijelom linijom su područja dviju grupa.



Slika 97. Prikaz broja sličnih ulaznih podataka po neuronima.

U prikazu slike 97 prilično dobro se vide dvije grupe. Također postoje neuroni s vrijednostima jednog podatka. Potencijalno, to bi mogli biti podaci koji imaju poseban značaj i kao takve važno ih je istražiti.



Slika 98. Prikaz težinskih ravnina.

Na slici 98 u jutarnjem periodu do 6 sati vidljiva je jaka izraženost van-distribucijskih podataka koji je gore izuzet. Zbog tih van-distribucijskih podataka nema neke značajne promjene do 15h jer na slici 42, ta krivulja znatno odskaje od ostalih podataka u skupu podataka. Kasnije kada se ona pripojila ostalim podacima, vidi se jaka aktivacija neurona sve do kasno u večer.

6.6. Rasprava

Kroz obradu i analizu podataka došlo se do zaključka da na prometni tok najveći utjecaj ima sezonalnost i korona situacija. Nakon što su prikazani dobiveni rezultati, dan je kratak osvrt na njih.

Kod vizualizacije sirovih podataka za oba prometna smjera prisutno je različito ponašanje odvijanja prometnih tokova. Za smjer prema Zagrebu maksimalni prometni tok prisutan je oko podne (12 sati) i traje sve do ponoći. Tomu je tako zbog utjecaja sezonalnosti, jer tada se ljudi vraćaju s godišnjih odmora. Za smjer prema Karlovcu maksimalni prometni tok prisutan je već oko 6 sati prije podne i traje do podneva (12 sati). Od podneva do 17 sati prometni tok znatno opada gdje od 17 sati ponovno raste na intenzitet prijepodnevnog.

Odnos pretjecajne i vozne trake mijenja se isključivo zbog sezonalnosti. Kada prometni sustav nije opterećen, odvijanje prometa u većoj mjeri odvija se po voznim prometnim trakama. Za slučaj kada dolazi do preopterećenja sustava prometa (u ljetnim mjesecima), tada pretjecajnom trakom biva veće odvijanje prometnih tokova. Također zbog sezonalnosti se mijenja i broj vozila po kategorijama brzine. Dakle znatno je veći broj vozila s brzinom većom od 90 [km/h], a broj vozila s manjim brzinama se ne mijenjaju, osim u slučajevima prometnih zastoja gdje je tada jedina brzina odvijanja prometnih tokova nula ili jako mala.

Kod kategorije vozila na voznoj prometnoj traci najveći udio odnosi se na osobna vozila što je i očekivano, a potom na kamione, kamione s prikolicom, itd. Na pretjecajnoj traci najveći udio također se odnosi na osobna vozila ali iza njih su kombi vozila, kamioni, itd. Prekoračenja brzine prisutna su tijekom cijele godine ali usred sezone taj broj prekoračenja brzine drastično raste. Gledano po godinama, broj prekoračenja brzine na pretjecajnoj traci u blagom je opadanju, dok je za voznu traku u blagom porastu. Korona situacija (period potpunog zatvaranja) u velikoj mjeri se odrazila na prometni sustav. U godini 2020. treći, četvrti i peti mjesec imaju znatno manji broj vozila nego što je to bilo godinama prije.

Tablicom 13 vidljiv je prikaz dobivenih rješenja u ovisnosti o algoritmu za grupiranje podataka. U prvom i zadnjem retku tablice 13, broj grupa k jednak je za oba algoritma, dok u drugom i trećem retku to nije slučaj. Jedan od razloga tomu su skupovi podataka nad kojima se odvijalo grupiranje. Ti prometni skupovi podataka nemaju jasnu granicu između grupa, odnosno jako velik broj zapisa u podatcima ne čini grupe. Drugi razlog je taj da SOM prikazuje svaki specifični podatak koji je različit od drugih, dok algoritam K-means to ne radi. Također algoritam K-means nema svojstvo učenja na temelju iskustva kao što je to slučaj kod SOM-a.

Tablica 13. Optimalni broja grupa k algoritma K-means i SOM-a.

Smjer	K-means	SOM
Smjer Karlovac za 3 god.	2	2
Smjer Zagreb za 3 god.	2	3
Smjer Karlovac za sezonu	2,3,4	3
Smjer Zagreb za sezonu	2	2

Sva ova zapažanja kao i dobivena rješenja mogu se u budućnosti (kao nadgradnja ovog rada) upotrijebiti za provjeru i podešavanja sustava upravljanja odnosno za bolje sustave (pred)putnog informiranja. Takvi sustavi temeljili bi se Upravo kod sustava upravljanja autocestom je najvažnije dobiti karakteristične dane u godini - grupe (npr. Vikendi u zimi kada se ide na skijanje, dani u ljetnim mjesecima kada se putuje na godišnji odmor i vraća s istog i vršni sati). Najvažniji dio diplomskog rada upravo je bila primjena neuronskih mreža za grupiranje postojećih podataka s autoceste Zagreb-Karlovac radi određivanje karakterističnih vremenskih intervala kao podloga za naknadnu izradbu simulacijskih scenarija odnosno umjerenih mikroskopskih simulacijskih modela. Ti bi se scenariji koristili za provjeru različitih prometnih

rješenja odnosno sustava upravljanja kao što su upravljanje priljevnim tokovima, promjenjivim ograničenjem brzine, davanjem preporuka vozačima i dinamička naplata cestarine (u danima veće prijevozne potražnje, naplata je veća i obratno).

7. Zaključak

U praktičnom dijelu rada dobiveni prometni podaci su organizirani, vizualizirani i obrađeni kako bi se dobio temelj za određenu analizu, opažanje, zaključivanje, odluke te razumijevanje i rješavanje određenih problema. Gledano kronološkim redoslijedom, prvo je uređen dobiveni prometni skup podataka programskim alatom Excel. Zatim je izrađena vizualizacija prometnih podataka kako bi se dobio uvid u stanje prometnog sustava. Naposljetku je u novoj tablici organiziran skup podataka nad kojim se odvijalo grupiranje algoritmom K-means i SOM. Za kraj su rezultati grupiranja skupa podataka opisani, uspoređeni i komentirani u svrhu potencijalne buduće nadogradnje ovog diplomskog rada.

Posljedicom utjecaja sezonalnosti i nedostatkom velike količine prometnih zapisa, potrebno je bilo grupiranje izvršiti na dva skupa podataka za oba prometna smjera. Budući da sezonalnost traje tri mjeseca, a godina 12 mjeseci, rezultati grupiranja nisu isti za cjelogodišnji skup podataka i sezonski skup podataka za tri godine. Dobiveni rezultati metodom K-means i SOM prilično su slični. Razlike dobivenih rezultata su u načinu grupiranja podataka i u toliko se rezultati razlikuju. Prometni skupovi podataka opisani su s dvije, tri i četiri grupe. Time je prikazana raspodjela dana u tjednu po grupama, gdje svaki od dana u tjednu koji pripadaju istoj grupi imaju slične karakteristike. Rezultati su u skladu s očekivanjima i uz ne tako savršeni skup podataka zadovoljavajući. Uslijed sezonalnosti i ne savršenog skupa podataka u prikazima rješenja vidljivo je kako za većinu prikaza rezultata nema jasne granice među grupama. Pojedini dani su jednako ili slično raspoređeni po svim grupama. Idealan slučaj bio bi kada svakoj grupi pripadaju jedinstveni dani, odnosno kada dani u tjednu ne bi bili raspoređeni po svim grupama.

Gledano po smjerovima, dobiveni rezultati su u skladu s prometnom strukom. U smjeru Karlovca (prema moru) prvu grupu čine ponedjeljak, utorak, srijeda, četvrtak i nedjelja, a drugu grupu čine ostali dani u tjednu. Dobiveni rezultati za smjer Zagreb (smjer od mora prema Zagrebu) se razlikuju u odnosu na suprotni smjer. Broj grupa je ostao isti ali raspodjela dana po grupama nije, što znači da se promet odvija različito prema smjerovima. U grupu jedan pripada nedjelja, a u drugu grupu pripadaju svi ostali dani u tjednu.

Time je stvorena podloga za izradu reprezentativnih simulacijskih modela zbog realističnog ispitivanja novih prometnih rješenja i za potencijalnu buduću izradu sustava upravljanja prometom primjenom karakterističnih problematičnih scenarija.

LITERATURA

- [1] Narodne novine. Zakon o cestama(NN84/11,2022/2013, 54/2013, 148/2013,92/2014, 110/2019). Preuzeto sa: <https://narodne-novine.nn.hr> [Pristupljeno: kolovoz 2021.].
- [2] Legac I. Cestovne prometnice: Javne ceste. Zagreb: Fakultet prometnih znanosti; 2006.
- [3] Blašković Zavada, J.: Osnove prometne infrastrukture - Osnove infrastrukture cestovnog prometa, Fakultet prometnih znanosti, Sveučilište u Zagrebu, Zagreb, 2015.
- [4] Hrvatska enciklopedija - autocesta. Preuzeto sa: <https://www.enciklopedija.hr/Natuknica.aspx?ID=4714> [Pristupljeno: kolovoz 2021.].
- [5] Hrvatska udruga koncesionara za autoceste s naplatom cestarine (HUKA), Nacionalno izvješće, 2019.
- [6] Brozović I. Prometno i prostorno planiranje II. Dio. Rijeka: Veleučilište u rijeci, 2009.
- [7] Ljupko, Š.: Osnove prometnog inženjerstva - Tokovi na mreži, autorizirana predavanja, Fakultet prometnih znanosti, Sveučilište u Zagrebu, Zagreb, 2015.
- [8] Hrvatske ceste d.o.o., Brojanje prometa na cestama Republike Hrvatske, Zagreb, 2018.
- [9] Dalbello Bašić B., Šnajder J.: Strojno učenje – Grupiranje, autorizirana predavanja, Fakultet elektrotehnike i računarstva, Sveučilište u Zagrebu, Zagreb, 2012.
- [10] Radović, A.: Upoznavanje sa sintaksom jezik R i njegova primjena u osnovnoj statističkoj i grafičkoj analizi podataka, Sveučilišni računski centar, Sveučilište u Zagrebu, Zagreb 2015.
- [11] ITAcademy, Python – programski jezik. Preuzeto sa: <https://www.it-akademija.com/python-kurs-programski-jezik> [pristupljeno: kolovoz 2021.].
- [12] Wikipedija, Python. Preuzeto sa: [https://hr.wikipedia.org/wiki/Python_\(programski_jezik\)](https://hr.wikipedia.org/wiki/Python_(programski_jezik)) [Pristupljeno: kolovoz 2021.].
- [13] Vrdoljak, M.: Uvod u Matlab, Fakultet strojarstva i brodogradnje, Sveučilište u Zagrebu, Zagreb, 2007.

- [14] Tišljarić, L.: Analiza repova čekanja i razine uslužnosti urbanih prometnica korištenjem algoritama strojnog učenja i Nosql baza podataka, Diplomski rad, Fakultet prometnih znanosti, Sveučilište u Zagrebu, Zagreb, 2018.
- [15] Jurković, S. Segmentacijska analiza poslovnih klijenata banaka pomoću samo-organizirajućih mapa. 2010;10(08): 3-5.
- [16] Wikipedija, SOM. Preuzeto sa: https://hr2.wiki/wiki/Self-organizing_map [Pristupljeno: kolovoz 2021.].
- [17] Edouard, I.: Umjetna inteligencija – Umjetne neuronske mreže, autorizirana predavanja, Fakultet prometnih znanosti, Sveučilište u Zagrebu, Zagreb, 2008.
- [18] Buršić, S.: Automatsko grupiranje podataka algoritmom K-means, Završni rad, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek, Sveučilište Josipa Jurja Strossmayera u Osijeku, Osijek, 2019.
- [19] Habijanić, A.: Grupiranje podataka, Diplomski rad, Sveučilišni diplomski studij matematike i računarstva, Sveučilište J. J. Strossmayera u Osijeku Odjel za Matematiku, Osijek, 2020.
- [20] ResearchGate, Partition clustering. Preuzeto sa: https://www.researchgate.net/figure/Partitioning-clustering_fig3_330417189. [Pristupljeno: kolovoz 2021.].
- [21] Kutbay, U. Partitional Clustering. Electrical and Electronics Engineering Department, Gazi University Engineering. Ankara, 2018. DOI: 10.5772/intechopen.75836.
- [22] Mathworks, Cluster Data with a Self-Organizing Map. Preuzeto sa: <https://ch.mathworks.com/help/deeplearning/gs/cluster-data-with-a-self-organizing-map.html> [Pristupljeno: kolovoz 2021.].
- [23] Možić, M., Kopic, D. Brojanje prometa na cestama Republike Hrvatske godine 2018 – Prosječni ljetni dnevni promet. Zagreb, 2018.
- [24] Yutao Li, Wengang Jiang. Research on short-term Traffic flow Prediction Based on Big Data Environment. College of Electronic Information Jiangsu university of science and Technology. Zangzhou, 2019. DOI: 10.1109/CAC48633.2019.8996158.

- [25] Kumar, K. Self-Organizing Map (SOM) Neural Networks for Air Space Sectoring. Department of computer science, Gurukul Kangri University. Haridwar, 2014. DOI:10.1109/CICN.2014.230.
- [26] Clustering using Flower Pollination Algorithm and Calinski-Harabasz Index. Faculty of Physics and Applied Computer Science, AGH University of Science and Technology. Kraków, 2018. DOI: 10.1109/ICDSBA.2018.00025.

POPIS SLIKA

Slika 1. Autoceste u Republici Hrvatskoj, [5].	5
Slika 2. Autocesta, [3].	6
Slika 3. Zaustavni trak autoceste, [3].	6
Slika 4. Vozna traka za spora vozila, [3].	7
Slika 5. Razdjelni pojas između dva smjera autoceste, [3].	7
Slika 6. Rigol na autocesti, [3].	8
Slika 7. Propusna moć autocesta raznih profila, [7].	9
Slika 8. Prikaz razine uslužnosti na autocesti s pripadnom količinom vozila, [6].	10
Slika 9. razine uslužnosti u odnosu na brzinu prometnog toka i stupanj iskoristivosti kapaciteta ceste, [6].	11
Slika 10. Grupirano voće, [10].	14
Slika 11. Particijsko grupiranje, [20].	17
Slika 12. Skup podataka i dendogram tog skupa, [19].	17
Slika 13. Tehnike grupiranja, [21].	18
Slika 14. Direktna dostupnost, dostupnost i povezanost podataka, [19].	19
Slika 15. Mreža podataka, [19].	20
Slika 16. Analiza metodom lakta.	21
Slika 17. Analiza Metodom siluete.	22
Slika 18. Analiza metodom Calinski-Harabasz.	22
Slika 19. Analiza metodom Davies-Bouldin.	23
Slika 20. K-means algoritam kroz iteracije, [19].	24
Slika 21. Koncept modela samoorganizirajućih mapa, [15].	25
Slika 22. Elementi umjetnog neurona, [17].	26
Slika 23. Arhitektura neuronske mreže, [17].	27
Slika 24. Dvodimenzionalna samoorganizirajuća mapa.	27
Slika 25. Aplikacija Neural Net Clustering.	29
Slika 26. Prikaz vrsta vizualizacija podataka.	30
Slika 27. Prikaz udaljenosti između neurona, [22].	31
Slika 28. Prikaz broja sličnih ulaznih podataka po neuronima, [22].	32
Slika 29. Prikaz težinskih ravnina, [22].	32
Slika 30. Tablica „PoBrzinama“.	34
Slika 31. Tablica „PoKategorijama“.	36
Slika 32. Izbrisani podatci/atributi.	37
Slika 33. Svi dodani atributi.	38
Slika 34. Van-distribucijski podatci.	39
Slika 35. Nedostatak zapisa više sati.	39
Slika 36. Nedostatak zapisa mjeseca rujna.	40
Slika 37. Pivot tablica 1.	41
Slika 38. Pivot tablica 2.	41
Slika 39. Prikaz sirovih podataka prometnog toka.	42
Slika 40. Prikaz sirovih podataka prometnog toka.	42
Slika 41. Prikaz sirovih podataka prometnog toka.	43

Slika 42. Prikaz sirovih podataka prometnog toka.	43
Slika 43. Utjecaj sezonalnosti na prometni tok.....	44
Slika 44. Odnos pretjecajne i vozne trake u godini.....	45
Slika 45. Odnos pretjecajne i vozne trake u danu.....	46
Slika 46. Kategorije brzina za obje prometne trake.....	47
Slika 47. Brzine osobnih i teretnih vozila.	48
Slika 48. Brzine osobnih vozila po prometnim trakama.	49
Slika 49. Kategorije vozila vozne trake za sve tri godine.	50
Slika 50. Kategorije vozila vozne trake za sve tri godine bez osobnih vozila.....	50
Slika 51. Kategorije vozila pretjecajne trake.....	51
Slika 52. Kategorije vozila pretjecajne trake izuzev osobnih i kombi vozila.	51
Slika 53. Prikaz prekoračenja brzina na godišnjoj razini, smjer Karlovac za sve godine.....	52
Slika 54. Prikaz prekoračenja brzina na godišnjoj razini, smjer Zagreb za sve godine.	52
Slika 55. Prikaz prekoračenja brzina po godinama, smjer Karlovac.	53
Slika 56. Prometni tok u godini potpunog zatvaranja godine 2020.	53
Slika 57. Metoda lakta i broj grupa = 2.	54
Slika 58. Metoda siluete i boj grupa = 2.....	55
Slika 59. DB metoda i broj grupa = 2.....	55
Slika 60. CH metoda i broj grupa = 2.....	56
Slika 61. Metoda lakta i broj grupa = 2.	56
Slika 62. Metoda siluete i broj grupa = 2.	57
Slika 63. DB metoda i broj grupa = 2.....	57
Slika 64. CH metoda i broj grupa = 2.....	58
Slika 65. Metoda lakta i broj grupa = 2, može i 4.....	58
Slika 66. Metoda siluete i broj grupa = 3.	59
Slika 67. DB metoda i broj grupa = 3, može i 4.	59
Slika 68. CH metoda i broj grupa = 4.....	60
Slika 69. Metoda lakta i broj grupa = 2.	60
Slika 70. Metoda siluete i broj grupa = 2.	61
Slika 71. DB metoda i broj grupa = 2.....	61
Slika 72. CH metoda i broj grupa = 2.....	62
Slika 73. Optimalni broj grupa $k = 2$	63
Slika 74. Pripadnost podataka grupama.	63
Slika 75. Optimalni broj grupa $k = 2$	64
Slika 76. Pripadnost podataka grupama.	64
Slika 77. Optimalni broj grupa $k = 2$	65
Slika 78. Optimalni broj grupa $k = 3$	65
Slika 79. Optimalni broj grupa $k = 4$	66
Slika 80. Pripadnost podataka grupama.	66
Slika 81. Pripadnost podataka grupama.	67
Slika 82. Pripadnost podataka grupama.	67
Slika 83. Optimalni broj grupa $k = 2$	68
Slika 84. Optimalni broj grupa $k = 3$	69
Slika 85. Pripadnost podataka grupama.	69

Slika 86. Pripadnost podataka grupama.	70
Slika 87. Dvodimenzionalna samoorganizirajuća mapa.	71
Slika 88. Prikaz broja sličnih ulaznih podataka po neuronima.....	71
Slika 89. Prikaz težinskih ravnina.	72
Slika 90. Dvodimenzionalni SOM.	73
Slika 91. Prikaz broja sličnih ulaznih podataka po neuronima.....	74
Slika 92. Prikaz težinskih ravnina.	75
Slika 93. Dvodimenzionalna samoorganizirajuća mapa.	76
Slika 94. Prikaz broja sličnih ulaznih podataka po neuronima.....	77
Slika 95. Prikaz težinskih ravnina.	78
Slika 96. Dvodimenzionalna samoorganizirajuća mapa.	79
Slika 97. Prikaz broja sličnih ulaznih podataka po neuronima.....	80
Slika 98. Prikaz težinskih ravnina.	81

POPIS TABLICA

Tablica 1. Širine prometnih trakova s obzirom na dopuštenu brzinu kretanja vozila, [3].....	8
Tablica 2. Razine uslužnosti s obzirom na gustoću, zauzetost i brzinu prometnog toka, [7].	11
Tablica 3. Opis atributa tablice „PoBrzinama“.....	35
Tablica 4. Opis atributa tablice „PoKategorijama“.....	36
Tablica 5. Opis novododatih atributa tablice „PoBrzinama“.....	38
Tablica 6. Pripadnost dana pojedinim grupama prema slici 74.....	63
Tablica 7. Pripadnost dana pojedinim grupama prema slici 76.....	65
Tablica 8. Pripadnost dana pojedinim grupama prema slici 80.....	66
Tablica 9. Pripadnost dana pojedinim grupama prema slici 81.....	67
Tablica 10. Pripadnost dana pojedinim grupama prema slici 82.....	68
Tablica 11. Pripadnost dana pojedinim grupama prema slici 85.....	69
Tablica 12. Pripadnost dana pojedinim grupama prema slici 86.....	70
Tablica 13. Optimalni broja grupa k algoritma K-means i SOM-a.....	82

POPIS FORMULA

(1) K-means algoritam

POPIS KRATICA

SOM	Samorganizirajuća mreža (engl. Self-Organizing Map)
LOS	Razina uslužnosti (engl. Level Of Service)
GUI	Grafičko korisničko sučelje (engl. Graphical User Interface)
DBSCAN	Prostorno grupiranje aplikacije na temelju gustoće s ekstremima (engl. Density-Based Spatial Clustering of Application with Noise)
GDBSCAN	Općenito prostorno grupiranje aplikacija zasnovanih na gustoći s ekstremima (engl. Generalized Density-Based Spatial Clustering of Application with Noise)
OPTICS	Redosljed točaka za identificiranje strukture grupa (engl. Ordering Points to Identify the Clustering Structure)
DDR	Izravno dostignuta gustoća (engl. Direct Density-Reachable)
DR	Dostignuta gustoća (engl. Density Reachable)
DC	Povezanost gustoće (engl. Density-Connectedness)
CH	Metoda Calinski-Harabasz (engl. T.Calinski and J.Harabasz analysis)
DB	Metoda Davies-Bouldin (engl. Davies and Bouldin analysis)
ANN	Umjetna neuronska mreža (engl. Artificial Neural network)
NCTOOL	Aplikacija grupiranja neuronskih mreža (engl. Neural Net Clustering)
PGDP	Prosječni Godišnji Dnevni Promet
PLDP	Prosječni Ljetni Dnevni Promet

PRILOG 1. MATLAB PROGRAMSKI KÔD ZA METODU K-MEANS

```
%% Unos i prikaz sirovih podataka
A = Skup podataka za obradu;
initAnalysis1 = A;
figure()

[num, ~] = size(initAnalysis1);
days1 = initAnalysis1';

ponedjeljak = days1(:,1:7:end);
utorak = days1(:,2:7:end);
srijeda = days1(:,3:7:end);
cetvrtak = days1(:,4:7:end);
petak = days1(:,5:7:end);
subota = days1(:,6:7:end);
nedjelja = days1(:,7:7:end);
plot(ponedjeljak(:,1),'b');
hold on;
plot(utorak(:,1),'r');
plot(srijeda(:,1),'g');
plot(cetvrtak(:,1),'c');
plot(petak(:,1),'m');
plot(subota(:,1),'y');
plot(nedjelja(:,1),'k');
plot(ponedjeljak,'b');
plot(utorak,'r');
plot(srijeda,'g');
plot(cetvrtak,'c');
plot(petak,'m');
plot(subota,'y');
plot(nedjelja,'k');
legend('Ponedjeljak','Utorak','Srijeda','Çetvrtak','Petak','Subota','Nedjelja','Location','best');
xlabel('Vrijeme u danu [sati]')
ylabel('Prometna potražnja [voz/sat]')

%% Organiziranje skupa podataka po danima u tjednu
z = 1;
DayNumber = 1;
for k = 1:1096 %% 366 zapisa za 4 mjeseca i 1096 zapisa za 3 godine.
    DayNumber(k) = z;
```

```

z=z+1;
if (z==8)
    z=1;
end
end

%% K-means za svaki skup podataka
error1 = [];
clust = [];
cenroids = [];
for j = 1:20
    [idx,C,sumd,D] = kmeans(initAnalysis1,j,'MaxIter',1000, 'emptyaction','singleton',
'replicate',20);
    clust(:,j) = idx;

    error1(j) = sum(sumd);

evaluationMatrix=zeros(7,j);
    for i = 1:size(idx)
        x = DayNumber(i);
        y = idx(i);
        if(isnan(y))
            else
                evaluationMatrix(x,y) = evaluationMatrix(x,y) + 1;
            end
        end
    end
figure();
heatmap(evaluationMatrix)
title('Dani po grupama')
xlabel('Broj grupa k')
ylabel('Dani u tjednu')
end

%% Metode za pronalazak optimalnog broja grupa k za određen skup podataka
eva1 = evalclusters(initAnalysis1,clust,'CalinskiHarabasz') % CH traži maksimum krivulje
figure()
plot(eva1)
title('Calinski-Harabasz')
xlabel('Broj grupa k')
ylabel('Vrijednosti particija')

```

```

eva2 = evalclusters(initAnalysis1,clust,'DaviesBouldin') % DB traži minimum krivulje
figure()
plot(eva2)
title ('DaviesBouldin')
xlabel('Broj grupa k')
ylabel('Vrijednosti particija')

eva3 = evalclusters(initAnalysis1,clust,'silhouette') % Silueta traži maksimum krivulje
figure()
plot(eva3)
title ('Silhouette')
xlabel('Broj grupa k')
ylabel('Udaljenost podataka od centroida')
figure()

plot(error1,'b-o','DisplayName','error1') % Metoda lakta
title ('Metoda lakta')
xlabel('Broj grupa k')
ylabel('Pogreška')

%% K-means rješenja - prikaz grupiranih podataka (u nastavku primjer za 4 grupe)
[idx,C,sumd,D] = kmeans(initAnalysis1,4,'MaxIter',1000, 'emptyaction','singleton',
'replicate',20);

s = spline(1:24,C(1,:),1:0.1:24);
figure()
plot(C(1,:), 'ob','LineWidth',1);
hold on
plot(1:0.1:24,s,'b-','LineWidth',1)

plot(C(2,:), 'xr','LineWidth',1);
s = spline(1:24,C(2,:),1:0.1:24);
plot(1:0.1:24,s,'r--','LineWidth',1)
xlabel('Sati');
ylabel('Prometna potražnja [voz/h]');

plot(C(3,:), 'xg','LineWidth',1);
s = spline(1:24,C(3,:),1:0.1:24);
plot(1:0.1:24,s,'g--','LineWidth',1)

```

```
xlabel('Sati');  
ylabel('Prometna potražnja [voz/h]');
```

```
plot(C(4,:), 'xk', 'LineWidth', 1);  
s = spline(1:24, C(4,:), 1:0.1:24);  
plot(1:0.1:24, s, 'k--', 'LineWidth', 1)  
xlabel('Sati');  
ylabel('Prometna potražnja [voz/h]');
```

```
grid on
```




Sveučilište u Zagrebu
Fakultet prometnih znanosti
10000 Zagreb
Vukelićeva 4

IZJAVA O AKADEMSKOJ ČESTITOSTI I SUGLASNOST

Izjavljujem i svojim potpisom potvrđujem kako je ovaj _____ diplomski rad
isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na
objavlenu literaturu što pokazuju korištene bilješke i bibliografija.

Izjavljujem kako nijedan dio rada nije napisan na nedozvoljen način, niti je prepisan iz
necitiranog rada, te nijedan dio rada ne krši bilo čija autorska prava.

Izjavljujem također, kako nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj
visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Svojim potpisom potvrđujem i dajem suglasnost za javnu objavu _____ diplomskog rada
pod naslovom **ANALIZA PROMETNIH TOKOVA AUTOCESTE ZAGREB-KARLOVAC**

PRIMJENOM NEURONSKIH MREŽA

na internetskim stranicama i repozitoriju Fakulteta prometnih znanosti, Digitalnom akademskom
repozitoriju (DAR) pri Nacionalnoj i sveučilišnoj knjižnici u Zagrebu.

U Zagrebu, 10.9.2021 _____

Student/ica:



(potpis)

