



**Univertyitet u Beogradu – Mašinski
fakultet**

Master akademske studije
Industrija 4.0

Istraživanje podataka
Seminarski rad

Predmetni nastavnik: Prof. dr. Nenad Mitić

Asistent: Stefan Kapunac

Srudent: Tea Tepavčević 4005/21

Školska godina: 2023/2024

Sadržaj

1. Pretprocesiranje.....	4
1.1 Deskriptivna statistika.....	4
1.2 Histogrami učestalosti numeričkih atributa.....	4
1.3 Boxplot dijagrami.....	5
1.4 Skaliranje podataka.....	6
1.5 Predprocesiranje kategoričkih objekata.....	6
1.6 Toplotna mapa.....	7
2. Klasifikacija.....	8
2.1 Random Forest Classifier.....	9
2.2 Gradient Boosting Classifier.....	10
3. Klasterovanje.....	11
2. 1 Algoritam KMeans.....	11
3.2 DBSCAN algoritam.....	12
4. Pravila pridruživanja.....	14
6. Literatura.....	16

Uvod

Ovaj projekat se fokusira na analizu podataka iz baze nekretnina, koja obuhvata raznovrsne informacije o stanovima, uključujući karakteristike poput lokacije, cene, veličine, broja soba i drugih relevantnih parametara. Kroz proces pretprocesiranja podataka, kao i primenu tehnika klasifikacije i klasterovanja, cilj nam je da istražimo i identifikujemo uzorke u podacima koji mogu pružiti uvid u dinamiku tržišta nekretnina.

Prva faza ovog projekta obuhvata prikupljanje podataka iz različitih izvora, njihovo čišćenje i pripremu za analizu. Nakon toga, sledi fazu eksplorativne analize podataka, gde ćemo istražiti osnovne karakteristike skupa podataka i identifikovati potencijalne trendove ili anomalije. Zatim, primenom tehnika klasifikacije, želimo da razvijemo modele koji mogu predvideti cene ili druge važne faktore nekretnina na osnovu dostupnih informacija.

Nakon klasifikacije, sledi i proces klasterovanja, gde ćemo grupisati slične nekretnine u određene kategorije ili klase, što može pružiti dodatni uvid u strukturu tržišta i preferencije potrošača.

Na kraju, prikazana su interesantna pravila pridruživanja na osnovu baze podataka potrošačke korpe.

1. Pretprocesiranje

Podaci su ključni za svaku analizu podataka, ali često zahtevaju određene korake pretprocesiranja kako bi se osigurala kvalitetna analiza i interpretacija rezultata. U ovom segmentu projekta, nakon sprovedene deskriptivne statistike i vizualizacije histograma i boxplot-ova za svaki atribut, fokusirali smo se na dodatne korake pretprocesiranja kako bismo pripremili podatke za dalju analizu. To uključuje skaliranje podataka, predprocesiranje kategoričkih objekata, računanje interkvartalnog raspona i granicnih vrednosti, kao i kreiranje toplotne mape.

1.1 Deskriptivna statistika

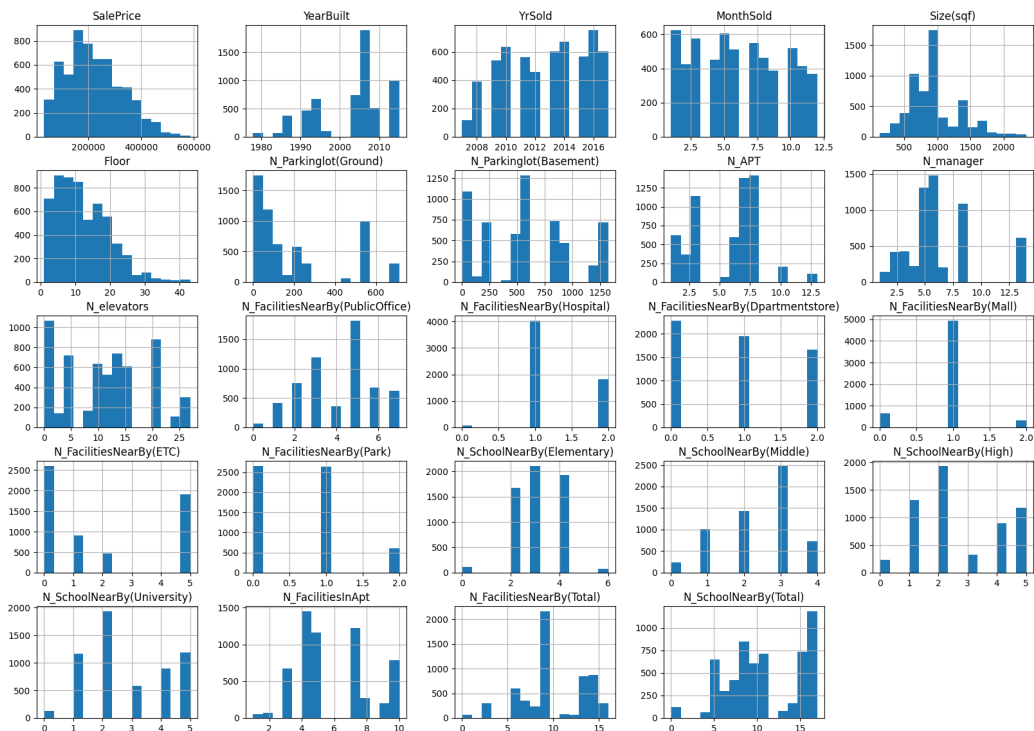
Data set se sastoji od 47 atributa i 5891 instanci. Na slici 1 je prikazana deskriptivna statistika za prvih 7 atributa.

	SalePrice	YearBuilt	YrSold	MonthSold	Size(sqf)	Floor	N_Parkinglot(Ground)
count	5891.000000	5.891000e+03	5.891000e+03	5.891000e+03	5.891000e+03	5891.000000	5.891000e+03
mean	0.000000	-1.929839e-17	7.719358e-17	-1.929839e-17	-9.649197e-18	0.000000	1.375011e-16
std	1.000085	1.000085e+00	1.000085e+00	1.000085e+00	1.000085e+00	1.000085	1.000085e+00
min	-1.805398	-2.833653e+00	-1.959067e+00	-1.523335e+00	-2.365988e+00	-1.525144	-1.154098e+00
25%	-0.724400	-1.131242e+00	-9.264508e-01	-9.329233e-01	-8.411830e-01	-0.815480	-1.016445e+00
50%	-0.106683	3.441812e-01	1.061656e-01	-4.730513e-02	-4.433017e-02	-0.105817	9.729003e-02
75%	0.665461	5.711693e-01	7.945764e-01	8.383130e-01	5.338375e-01	0.745779	2.349427e-01
max	2.741668	1.365628e+00	1.482987e+00	1.723931e+00	2.505000e+00	3.016702	4.689883e+00

Slika 1 Deskriptivna statistika

1.2 Histogrami učestalosti numeričkih atributa

Nakon deskriptivne statistike, kreirani su histogrami za svaki atribut kako bismo bolje razumeli raspodelu podataka. Ovi histogrami pružaju vizuelni prikaz distribucije vrednosti za svaki atribut i omogućavaju identifikaciju eventualnih anomalija ili specifičnih obrazaca u podacima(slika 2).

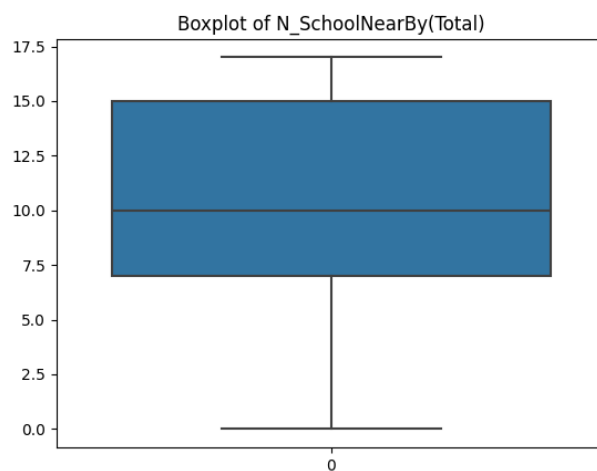
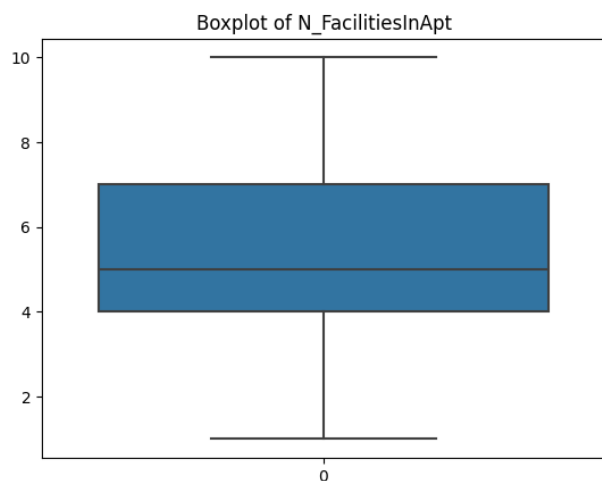


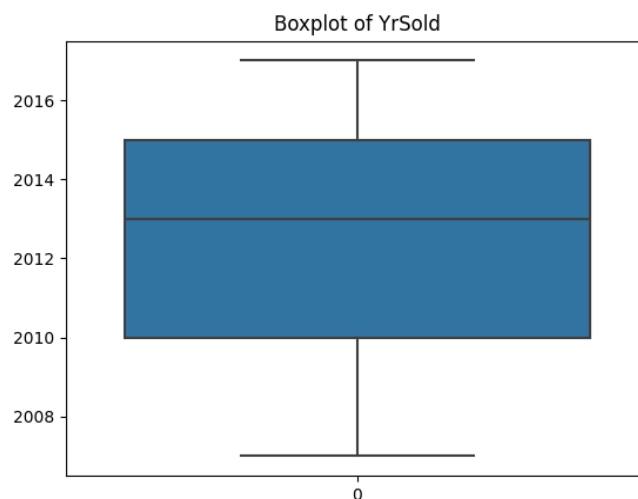
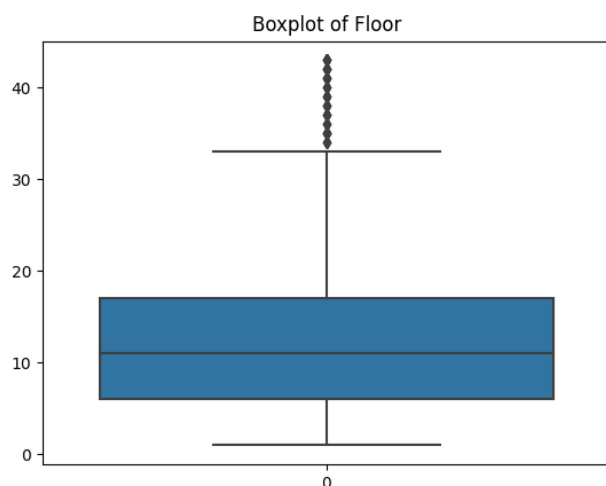
Slika 2: Histogrami učestalosti numeričkih atributa

1.3 Boxplot dijagrami

U deskriptivnoj statistici, boxplot je metoda za grafičko prikazivanje lokacije, rasprostranjenosti i asimetrije grupa numeričkih podataka kroz njihove kvartile.

Prikazana su boxplot dijagrami za 4 atributa(za kompletan skup boxplot dijagrama pogledati folder *graphs*)





1.4 Skaliranje podataka

Skaliranje podataka je važan korak koji omogućava normalizaciju različitih skala vrednosti atributa, što može poboljšati performanse algoritama za analizu podataka.

U projektu je korišćeno standardno skaliranje.

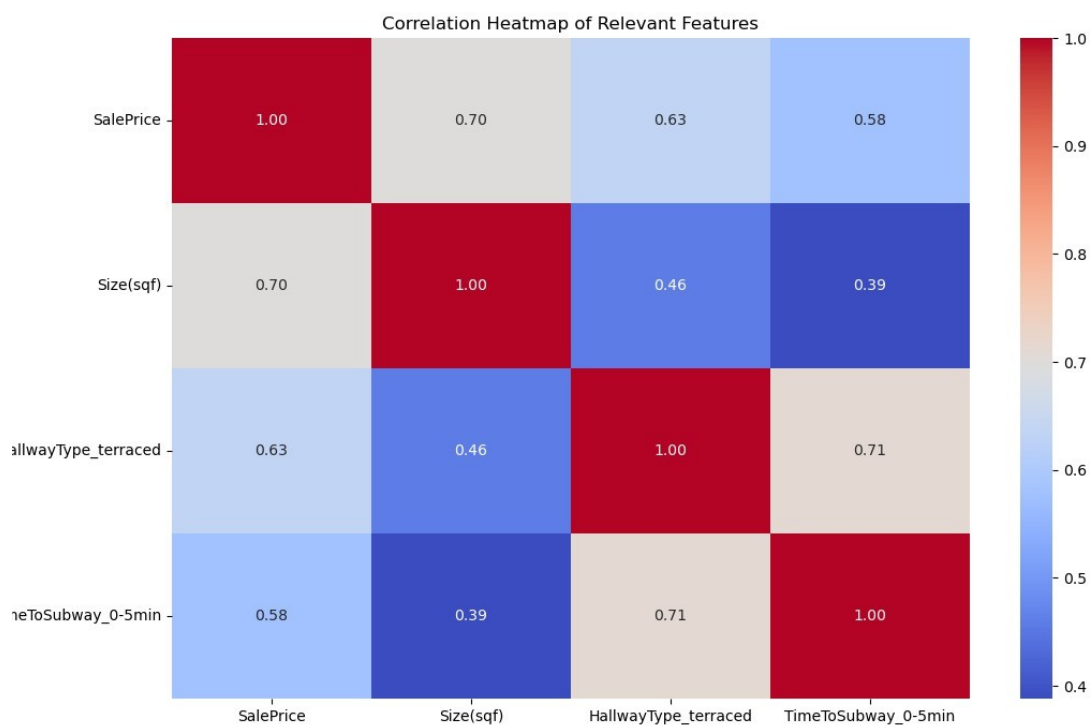
1.5 Predprocesiranje kategoričkih objekata

Kategorički atributi su često prisutni u datasetovima i zahtevaju posebnu obradu pre nego što se mogu koristiti u analizi. Jedan od uobičajenih koraka u pretprocesiranju je kodiranje kategoričkih atributa, što omogućava algoritmima da bolje razumeju i obrade ove informacije.

Korišćena je funkcija ***pd.get_dummies()*** iz biblioteke *Pandas* da bi se kodirali kategorički atributi u binarne (*dummy*) promenljive. Svaki kategorički atribut sa više kategorija se transformiše u niz binarnih atributa, gde svaki od njih označava prisustvo određene kategorije.

1.6 Toplotna mapa

Izračunata je matrica korelacije između svih atributa u datasetu koristeći Pearson-ov koeficijent korelacije. Zatim se odabiru relevantni atributi koji imaju jaku korelaciju sa ciljnom promenljivom ("SalePrice") većom od 0.5 (ili manjom od -0.5).



Slika 3: Toplotna mapa

2. Klasifikacija

Klasifikacija je tehnika mašinskog učenja koja se koristi za predviđanje pripadnosti instanci određenim klasama ili kategorijama na osnovu njihovih karakteristika ili atributa. Cilj klasifikacije je naučiti model koji može automatski klasifikovati nove instance na osnovu obrazaca naučenih iz podataka za obuku.

U ovom delu projekta, prvo je ciljna promenljiva (***SalePrice***) podeljena na 4 kategorije (***Q1***, ***Q2***, ***Q3***, ***Q4***) kako bismo mogla da se izvrši klasifikacija i prikaže raspodela po klasama (slika 4). Nakon toga, pronađene su relevantni atributi koristeći Pirsonov koeficijent korelacije.



Slika 4

Podaci su podeljeni na trening i test skupove. Korišćen su modeli ***Random Forest Classifier*** i ***Gradient Boosting Classifier***.

2.1 Random Forest Classifier

Random Forest Classifier je popularan algoritam mašinskog učenja koji se koristi za klasifikaciju i regresiju. On je ansambl metoda, što znači da kombinuje više modela za postizanje boljih performansi u poređenju sa pojedinačnim modelima. Glavna ideja iza *Random Forest* algoritma je kreiranje velikog broja stabala odlučivanja tokom treninga i kombinovanje njihovih predikcija kako bi se donela konačna odluka. Često se koristi zbog svoje sposobnosti za rad sa visokodimenzionalnim podacima, tolerancije na preprilagođavanje i stabilnosti. Takođe, može se koristiti za identifikaciju važnih atributa u datasetu, što je korisno za analizu podataka.

u Projektu je korišćen algoritam **GridSearchCV** kako bi se pronašli najbolji parametri modela (broj satabala, maksimalna dubina stabla i minimalni broj uzoraka potreban za deljenje čvora) kroz unakrsnu validaciju.

Dobijeni model je zatim evaluiran na test skupu, a rezultati su sačuvani i prikazani kroz metrike kao što su preciznost, odziv i F1 mera.

Na kraju, rezultati su sačuvani u tekstualnoj datoteci, a takođe su kreirane i toplotne mape rezultata za poređenje performansi modela.

Random Forest Classifier Metrike:				
	precision	recall	f1-score	support
Q1	0.79	0.71	0.74	350
Q2	0.62	0.83	0.71	496
Q3	0.66	0.38	0.48	278
Q4	0.61	0.45	0.52	55
accuracy			0.67	1179
macro avg	0.67	0.59	0.61	1179
weighted avg	0.68	0.67	0.66	1179

Slika 5: Classification report - Random Forest Classifier

2.2 Gradient Boosting Classifier

Gradient Boosting Classifier je još jedan popularan algoritam mašinskog učenja koji se koristi za klasifikaciju i regresiju. Ovaj algoritam radi tako što gradi niz slabih modela (obično stabala odlučivanja) iterativno, pri čemu svaki sledeći model pokušava da ispravi greške prethodnog modela. Na taj način, svaki naredni model se fokusira na instance koje su pogrešno klasifikovane ili imaju veliku grešku u predikciji. On je poznat po svojoj sposobnosti da pruži veoma dobre rezultate kada su podaci visoke dimenzionalnosti ili kada postoji kompleksna zavisnost između atributa i ciljne promenljive. Takođe, ovaj algoritam je otporan na overfitting i može efikasno raditi sa različitim tipovima podataka.

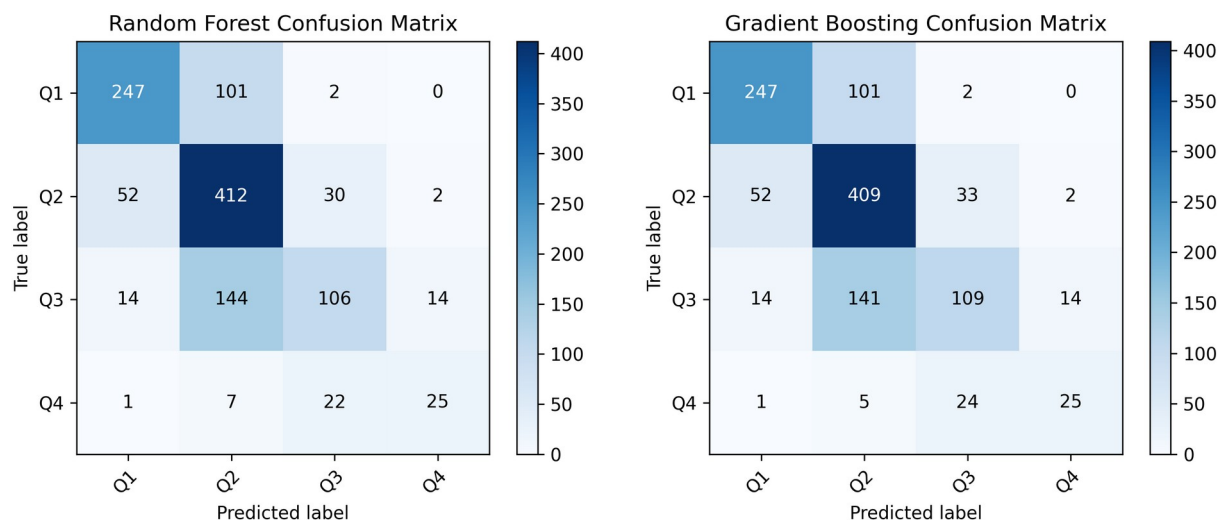
Korišćen je **GridSearchCV** za traženje najboljih parametara modela (broj estimatora, *learning rate*, i maksimalna dubina stabla) kroz unakrsnu validaciju.

Kao i kod *Random Forest Classifier* algoritma, model je evaluiran i kreirana je tpolutna mapa.

Gradient Boosting Classifier Metrike:					
	precision	recall	f1-score	support	
Q1	0.79	0.71	0.74	350	
Q2	0.62	0.82	0.71	496	
Q3	0.65	0.39	0.49	278	
Q4	0.61	0.45	0.52	55	
accuracy			0.67	1179	
macro avg	0.67	0.59	0.62	1179	
weighted avg	0.68	0.67	0.66	1179	

Slika 6: Classification report - Gradient Boosting Classifier

Toplotne mape za *Random Forest Classifier* i *Gradient Boosting Classifier*



Slika 7: Toplotne mape

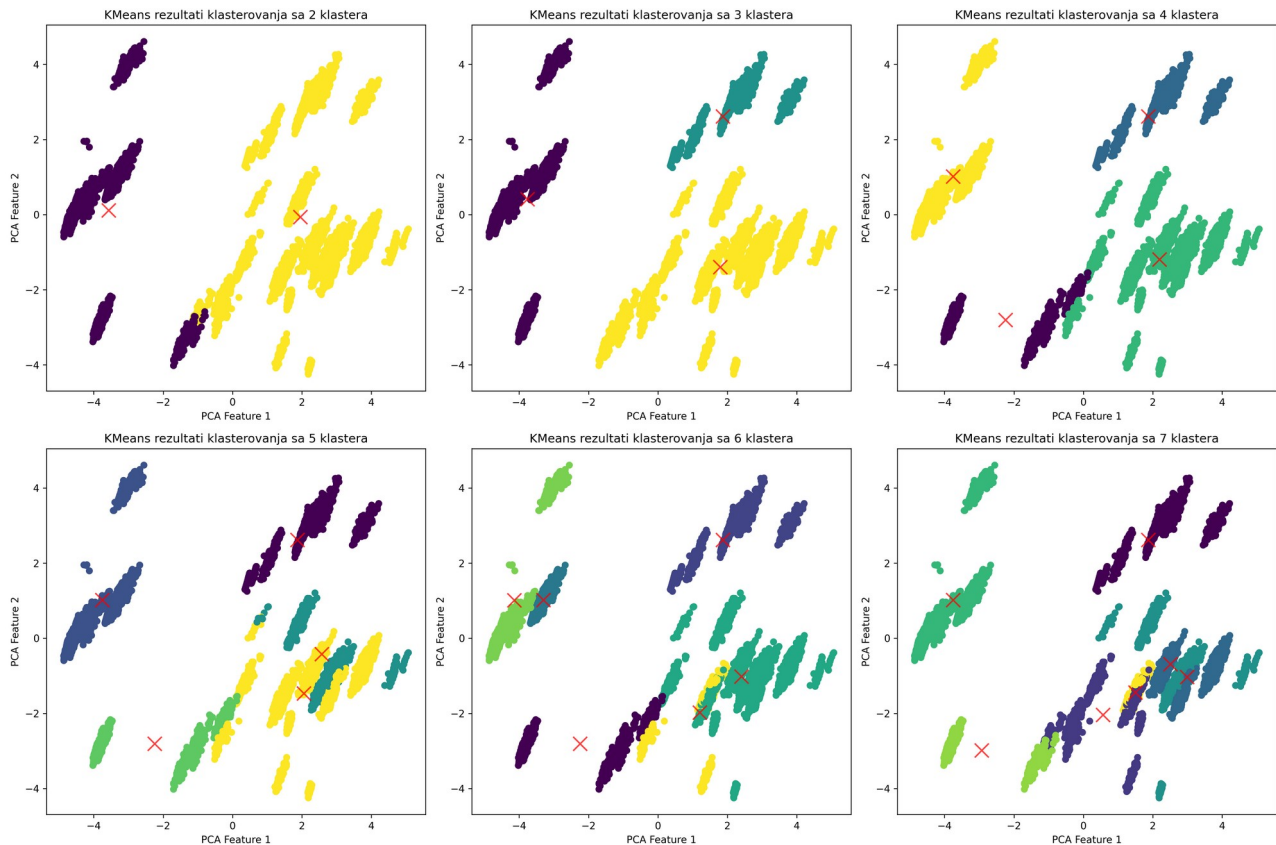
3. Klasterovanje

Klasterovanje je tehnika mašinskog učenja koja se koristi za grupisanje sličnih instanci podataka u skupove nazvane klasteri. Cilj klasterovanja je otkrivanje prirodnih grupa ili struktura unutar podataka, bez prethodnog poznavanja oznaka klasa. Ova tehnika je široko korišćena u različitim oblastima, uključujući analizu podataka, obradu slika, biomedicinsko istraživanje, i mnoge druge.

2. 1 Algoritam KMeans

Jedan od najčešće korišćenih algoritama za klasterovanje je *KMeans*. Ovaj algoritam funkcioniše tako što grupiše instance podataka u predefinisani broj klastera na osnovu sličnosti među njima. *KMeans* radi iterativno i pokušava da minimizuje sumu kvadrata udaljenosti između tačaka i centara klastera.

U projektu je korišćena variabla **num_clusters** sa opsegom [2, 8] koja označava boj klastera. Prikazani su dijagrami koji predstavljaju KMeans rezultate klasterovanja sa različitim vrenostima broja klastera (slika 8).



Slika 8: KMeans rezultati klasterovanja sa različitim vrenostima broja klastera

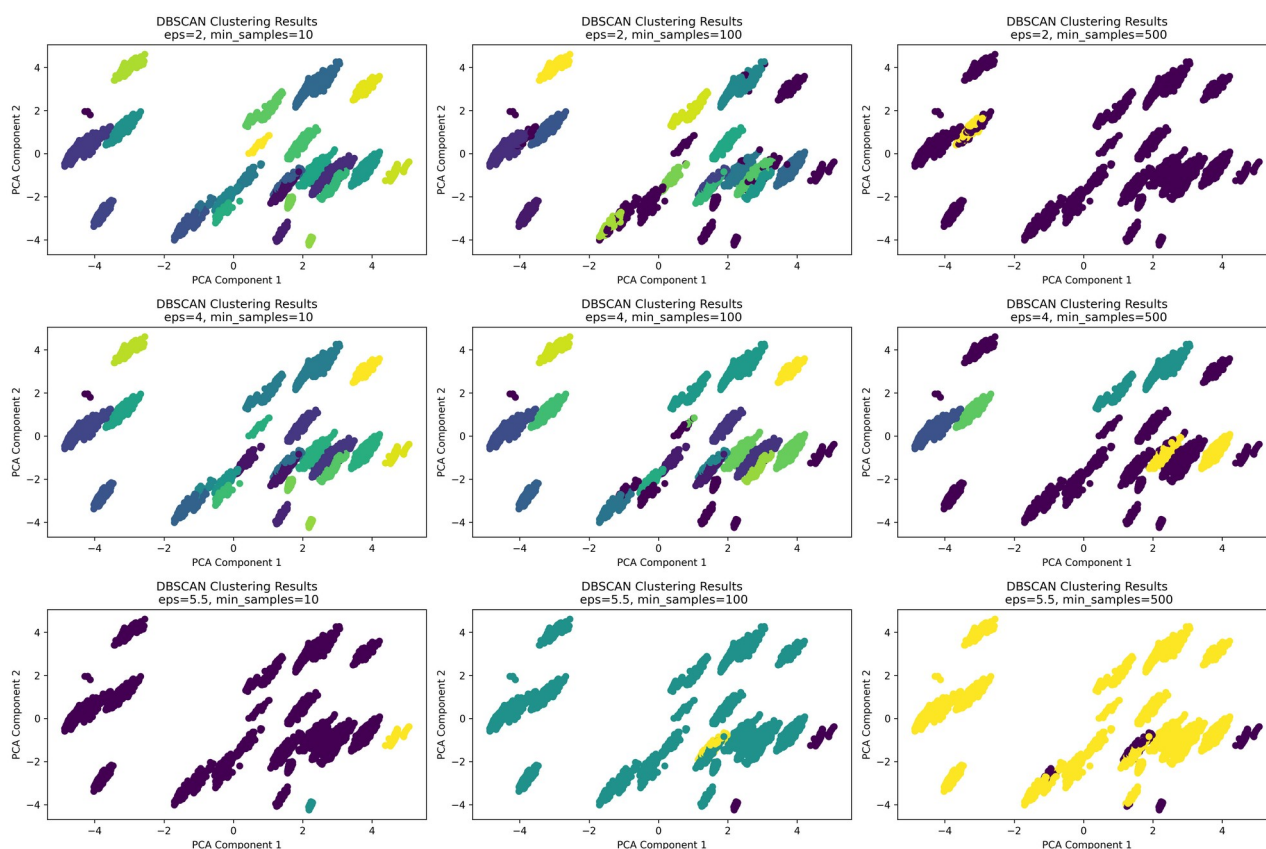
3.2 DBSCAN algoritam

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) je algoritam za klasterovanje koji se koristi za identifikaciju gustoće zasnovanih klastera u prostornim podacima. Ovaj algoritam ne zahteva pretpostavku o obliku klastera i može efikasno raditi sa podacima visoke gustine.

Ovaj algoritam koristi dva parametra za definisanje klastera:

- ***eps***: Ovo je parametar koji definiše maksimalnu udaljenost između dve tačke kako bi se smatrala deo istog klastera. Tačke unutar ovog *eps* okruženja smatraju se susedima.
- ***min_samples***: Ovo je minimalni broj tačaka koji treba da bude unutar *eps* okruženja određene tačke da bi ta tačka bila označena kao deo klastera. Manji broj *min_samples* rezultira uključivanjem više tačaka u klaster, dok veći broj *min_samples* čini algoritam manje podložnim buci.

U projektu su korišćene variable ***eps_values*** i ***min_samples_values*** kako bi se prikazal rezultati različitih za različite vrednosti ovih parametara (slika 9).



Slika 9: DBSCAN

4. Pravila pridruživanja

Pravila pridruživanja su ključni koncepti u analizi podataka u području trgovine, marketinga i poslovne inteligencije. Ova tehnika omogućuje otkrivanje značajnih veza između različitih stavki koje se često kupuju zajedno. Kroz analizu podataka o transakcijama, pravila pridruživanja otkrivaju asocijacije između stavki koje se često pojavljuju zajedno, što omogućuje trgovcima da poboljšaju strategije prodaje, prilagode ponudu proizvoda i optimizuju raspored trgovine radi poboljšanja iskustva kupaca i povećanja prodaje.

U projektu je korišćena druga baza podataka jer prethodni dataset nije bio pogodan za pravila pridruživanja. Korišćen je data set *Market Basket Analysis*.

U nastavku se nalaze dobijena pravila pridruživanja sa različitim kriterijumima (*antecedent support, consequent support, support, confidence i lift*).

	A	B	C	D	E	F	G	H	I
1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	convic
2	{('LUNCH BAG BLACK SKULL')}	{('LUNCH BAG RED RETROSPOT')}	0.0581636892397175	0.071135115173337	0.02917	0.5015873	7.05119	0.02504	1.86
3	{('JUMBO BAG PINK POLKADOT')}	{('JUMBO BAG RED RETROSPOT')}	0.0559017679915063	0.0952776623736325	0.03781	0.67630058	7.09821	0.03248	2.79
4	{('JUMBO STORAGE BAG SUKI')}	{('JUMBO BAG RED RETROSPOT')}	0.0543784332733232	0.0952776623736325	0.03328	0.61205433	6.4239	0.0281	2.33
5	{('JUMBO SHOPPER VINTAGE RED PAISLEY')}	{('JUMBO BAG RED RETROSPOT')}	0.0538244933758021	0.0952776623736325	0.03121	0.57975986	6.08495	0.02608	2.15
6	{('LUNCH BAG SUKI DESIGN')}	{('LUNCH BAG RED RETROSPOT')}	0.0497622674606472	0.071135115173337	0.02557	0.51391466	7.22449	0.02203	1.91
7	{('LUNCH BAG PINK POLKADOT')}	{('LUNCH BAG RED RETROSPOT')}	0.0495314591700134	0.071135115173337	0.02737	0.5526561	7.7691	0.02385	2.0
8	{('RED RETROSPOT CHARLOTTE BAG')}	{('CHARLOTTE BAG PINK POLKADOT')}	0.0474541845543092	0.0340673036975488	0.024	0.50583658	14.8482	0.02239	1.95
9	{('ALARM CLOCK BAKELIKE RED')}	{('ALARM CLOCK BAKELIKE GREEN')}	0.0474080228961824	0.0446383234085768	0.02922	0.61635833	13.8078	0.0271	2.49
10	{('ROSES REGENCY TEACUP AND SAUCER')}	{('GREEN REGENCY TEACUP AND SAUCER')}	0.0467617596824078	0.0450076166735909	0.03374	0.72161895	16.0333	0.03164	3.43
11	{('ROSES REGENCY TEACUP AND SAUCER')}	{('PINK REGENCY TEACUP AND SAUCER')}	0.0467617596824078	0.0339749803812953	0.02631	0.56268509	16.5617	0.02472	2.20
12	{('ROSES REGENCY TEACUP AND SAUCER')}	{('GREEN REGENCY TEACUP AND SAUCER')}	0.0467617596824078	0.0279278031666897	0.02377	0.50839092	18.2038	0.02247	1.97
13	{('LUNCH BAG WOODLAND')}	{('LUNCH BAG RED RETROSPOT')}	0.046346304759267	0.071135115173337	0.024	0.51792829	7.28091	0.02071	1.92
14	{('GREEN REGENCY TEACUP AND SAUCER')}	{('PINK REGENCY TEACUP AND SAUCER')}	0.0450076166735909	0.0339749803812953	0.02793	0.62051282	18.2638	0.0264	2.54
15	{('GREEN REGENCY TEACUP AND SAUCER')}	{('ROSES REGENCY TEACUP AND SAUCER')}	0.0450076166735909	0.0467617596824078	0.03374	0.74974359	16.0333	0.03164	3.80

	A	B	C	D	E	F	G	H	I
1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	convic
2	{('RED HANGING HEART T-LIGHT HOLDER')}	{('WHITE HANGING HEART T-LIGHT HOLDER')}	0.0332825555093939	0.101647971195125	0.02211	0.66435506	6.53584	0.01873	2.6
3	{('CANDLEHOLDER PINK HANGING HEART')}	{('WHITE HANGING HEART T-LIGHT HOLDER')}	0.0176799150625491	0.101647971195125	0.01246	0.70496084	6.93532	0.01067	3.04
4	{('JUMBO BAG PINK POLKADOT')}	{('JUMBO BAG RED RETROSPOT')}	0.0559017679915063	0.0952776623736325	0.03781	0.67630058	7.09821	0.03248	2.79
5	{('JUMBO STORAGE BAG SUKI')}	{('JUMBO BAG RED RETROSPOT')}	0.0543784332733232	0.0952776623736325	0.03328	0.61205433	6.4239	0.0281	2.33
6	{('JUMBO SHOPPER VINTAGE RED PAISLEY')}	{('JUMBO BAG RED RETROSPOT')}	0.0538244933758021	0.0952776623736325	0.03121	0.57975986	6.08495	0.02608	2.15
7	{('JUMBO BAG APPLES')}	{('JUMBO BAG RED RETROSPOT')}	0.0444998384341965	0.0952776623736325	0.02497	0.56120332	5.89019	0.02073	2.06
8	{('JUMBO BAG BAROQUE BLACK WHITE')}	{('JUMBO BAG RED RETROSPOT')}	0.0426995337672529	0.0952776623736325	0.02677	0.62702703	6.58105	0.02271	2.4
9	{('JUMBO BAG PINK VINTAGE PAISLEY')}	{('JUMBO BAG RED RETROSPOT')}	0.039422056040253	0.0952776623736325	0.02363	0.59953162	6.29247	0.01988	2.25
10	{('JUMBO BAG WOODLAND ANIMALS')}	{('JUMBO BAG RED RETROSPOT')}	0.039422056040253	0.0952776623736325	0.0246	0.62412178	6.55056	0.02085	2.40
11	{('JUMBO BAG STRAWBERRY')}	{('JUMBO BAG RED RETROSPOT')}	0.0376217513733093	0.0952776623736325	0.02442	0.64907975	6.81251	0.02084	2.57
12	{('RECYCLING BAG RETROSPOT')}	{('JUMBO BAG RED RETROSPOT')}	0.0342981119881826	0.0952776623736325	0.01911	0.55720054	5.84818	0.01584	2.04
13	{('DOTCOM POSTAGE')}	{('JUMBO BAG RED RETROSPOT')}	0.032682453953746	0.0952776623736325	0.02239	0.68502825	7.18981	0.01927	2.87
14	{('JUMBO STORAGE BAG SKULLS')}	{('JUMBO BAG RED RETROSPOT')}	0.0324516456631122	0.0952776623736325	0.01906	0.58748222	6.166	0.01597	2.19
15	{('JUMBO BAG VINTAGE DOILY')}	{('JUMBO BAG RED RETROSPOT')}	0.0321746757143517	0.0952776623736325	0.01616	0.50215208	5.27041	0.01309	1.81

	A	B	C	D	E	F	G	H	I
1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	onv
2	{{'JUMBO BAG PINK POLKADOT'}})	{{'JUMBO BAG RED RETROSPOT'}})	0.0559017679915063	0.0952776623736325	0.03781	0.67630058	7.09821	0.03248	2.79
3	{{'GREEN REGENCY TEACUP AND SAUCER'}})	{{'ROSES REGENCY TEACUP AND SAUCER'}})	0.0450076166735909	0.0467617596824078	0.03374	0.74974359	16.0333	0.03164	3.80
4	{{'ROSES REGENCY TEACUP AND SAUCER'}})	{{'GREEN REGENCY TEACUP AND SAUCER'}})	0.0467617596824078	0.0450076166735909	0.03374	0.72161895	16.0333	0.03164	3.43
5	{{'JUMBO STORAGE BAG SUKI'}})	{{'JUMBO BAG RED RETROSPOT'}})	0.0543784332733232	0.0952776623736325	0.03328	0.61205433	6.4239	0.0281	2.33
6	{{'JUMBO SHOPPER VINTAGE RED PAISLEY'}})	{{'JUMBO BAG RED RETROSPOT'}})	0.0538244933758021	0.0952776623736325	0.03121	0.57975986	6.08495	0.02608	2.15
7	{{'ALARM CLOCK BAKELIKE GREEN'}})	{{'ALARM CLOCK BAKELIKE RED'}})	0.0446383234085768	0.0474080228961824	0.02922	0.65460186	13.8078	0.0271	2.75
8	{{'ALARM CLOCK BAKELIKE RED'}})	{{'ALARM CLOCK BAKELIKE GREEN'}})	0.0474080228961824	0.0446383234085768	0.02922	0.61635833	13.8078	0.0271	2.49
9	{{'LUNCH BAG BLACK SKULL'}})	{{'LUNCH BAG RED RETROSPOT'}})	0.0581636892397175	0.071135115173337	0.02917	0.5015873	7.05119	0.02504	1.86
10	{{'PINK REGENCY TEACUP AND SAUCER'}})	{{'GREEN REGENCY TEACUP AND SAUCER'}})	0.0339749803812953	0.0450076166735909	0.02793	0.82201087	18.2638	0.0264	5.36
11	{{'GREEN REGENCY TEACUP AND SAUCER'}})	{{'PINK REGENCY TEACUP AND SAUCER'}})	0.0450076166735909	0.0339749803812953	0.02793	0.62051282	18.2638	0.0264	2.54
12	{{'LUNCH BAG PINK POLKADOT'}})	{{'LUNCH BAG RED RETROSPOT'}})	0.0495314591700134	0.071135115173337	0.02737	0.5526561	7.7691	0.02385	2.0
13	{{'JUMBO BAG BAROQUE BLACK WHITE'}})	{{'JUMBO BAG RED RETROSPOT'}})	0.0426995337672529	0.0952776623736325	0.02677	0.62702703	6.58105	0.02271	2.4
14	{{'PINK REGENCY TEACUP AND SAUCER'}})	{{'ROSES REGENCY TEACUP AND SAUCER'}})	0.0339749803812953	0.0467617596824078	0.02631	0.77445652	16.5617	0.02472	4.22
15	{{'ROSES REGENCY TEACUP AND SAUCER'}})	{{'PINK REGENCY TEACUP AND SAUCER'}})	0.0467617596824078	0.0339749803812953	0.02631	0.56268509	16.5617	0.02472	2.20

	A	B	C	D	E	F	G	H	I
1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	onv
2	{{'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE GREEN'}})	0.0114942528735632	0.0163873886349998	0.01089	0.94779116	57.8366	0.01071	18.1
3	{{'WOODEN TREE CHRISTMAS SCANDINAVIAN', 'WOODEN TREE CHRISTMAS SCANDINAVIAN'}})	{{'WOODEN STAR CHRISTMAS SCANDINAVIAN'}})	0.0110787979504224	0.0228500207727462	0.01034	0.93333333	40.8461	0.01009	14.6
4	{{'HERB MARKER THYME'}})	{{'HERB MARKER ROSEMARY'}})	0.0107556663435351	0.0108018280016618	0.01002	0.93133047	86.2197	0.0099	14.4
5	{{'HERB MARKER ROSEMARY'}})	{{'HERB MARKER THYME'}})	0.0108018280016618	0.0107556663435351	0.01002	0.92735043	86.2197	0.0099	13.6
6	{{'STRAWBERRY CHARLOTTE BAG', 'WOODEN STAR CHRISTMAS SCANDINAVIAN'}})	{{'RED RETROSPOT CHARLOTTE BAG'}})	0.0120943544292111	0.0474541845543092	0.01112	0.91984733	19.3839	0.01055	11.8
7	{{'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE ROSES'}})	0.0118635461385773	0.0192032497807321	0.01089	0.91828794	47.8194	0.01067	12.0
8	{{'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE GREEN'}})	0.0130175875917463	0.0163873886349998	0.01186	0.91134752	55.6127	0.01165	11.0
9	{{'REGENCY CAKESTAND 3 TIER', 'ROSES REGENCY TEACUP AND SAUCER'}})	{{'GREEN REGENCY TEACUP AND SAUCER'}})	0.0142639523611688	0.0450076166735909	0.01293	0.90614887	20.1332	0.01228	10.1
10	{{'STRAWBERRY CHARLOTTE BAG', 'CHARLOTTE BAG', 'RED RETROSPOT CHARLOTTE BAG'}})	{{'RED RETROSPOT CHARLOTTE BAG'}})	0.0124636476942252	0.0474541845543092	0.01126	0.9037037	19.0437	0.01067	9.89
11	{{'ROSES REGENCY TEACUP AND SAUCER', 'GREEN REGENCY TEACUP AND SAUCER'}})	{{'GREEN REGENCY TEACUP AND SAUCER'}})	0.0263121451322531	0.0450076166735909	0.02377	0.90350877	20.0746	0.02259	9.89
12	{{'SET/20 RED RETROSPOT PAPER NAPKINS', 'SET/6 RED SPOTTY PAPER PLATES'}})	{{'SET/6 RED SPOTTY PAPER PLATES'}})	0.0133407191986336	0.0238655772515349	0.01191	0.89273356	37.4067	0.01159	9.10
13	{{'JUMBO BAG RED RETROSPOT', 'SUKI SHAMPOO', 'DOTCOM POSTAGE'}})	{{'DOTCOM POSTAGE'}})	0.0129714259336195	0.032682453953746	0.01149	0.886121	27.113	0.01107	8.49
14	{{'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE ROSES'}})	0.0130175875917463	0.0192032497807321	0.01149	0.88297872	45.9807	0.01124	8.38
15	{{'REGENCY CAKESTAND 3 TIER', 'PINK REGENCY TEACUP AND SAUCER'}})	{{'GREEN REGENCY TEACUP AND SAUCER'}})	0.0170798135069012	0.0450076166735909	0.01486	0.87027027	19.3361	0.0141	7.3

	A	B	C	D	E	F	G	H	I
1	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	onv
2	{{'HERB MARKER THYME'}})	{{'HERB MARKER ROSEMARY'}})	0.0107556663435351	0.0108018280016618	0.01002	0.93133047	86.2197	0.0099	14.4
3	{{'HERB MARKER ROSEMARY'}})	{{'HERB MARKER THYME'}})	0.0108018280016618	0.0107556663435351	0.01002	0.92735043	86.2197	0.0099	13.6
4	{{'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE ROSES', 'REGENCY TEA PLATE PINK'}})	0.0130175875917463	0.0136638508055209	0.01089	0.83687943	61.2477	0.01072	6.04
5	{{'REGENCY TEA PLATE ROSES', 'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE PINK'}})	0.0136638508055209	0.0130175875917463	0.01089	0.7972973	61.2477	0.01072	4.86
6	{{'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN'}})	{{'REGENCY TEA PLATE GREEN'}})	0.0114942528735632	0.0163873886349998	0.01089	0.94779116	57.8366	0.01071	18.1
7	{{'REGENCY TEA PLATE GREEN'}})	{{'REGENCY TEA PLATE PINK', 'REGENCY TEA PLATE GREEN'}})	0.0163873886349998	0.0114942528735632	0.01089	0.66478873	57.8366	0.01071	2.9
8	{{'REGENCY TEA PLATE PINK'}})	{{'REGENCY TEA PLATE GREEN'}})	0.0130175875917463	0.0163873886349998	0.01186	0.91134752	55.6127	0.01165	11.0
9	{{'REGENCY TEA PLATE GREEN'}})	{{'REGENCY TEA PLATE PINK'}})	0.0163873886349998	0.0130175875917463	0.01186	0.72394366	55.6127	0.01165	3.57
10	{{'REGENCY SUGAR BOWL GREEN'}})	{{'REGENCY MILK JUG PINK'}})	0.0135715274892674	0.0138946590961547	0.01016	0.74829932	53.8552	0.00997	3.91
11	{{'REGENCY MILK JUG PINK'}})	{{'REGENCY SUGAR BOWL GREEN'}})	0.0138946590961547	0.0135715274892674	0.01016	0.73089701	53.8552	0.00997	3.66
12	{{'SET OF 3 WOODEN TREE DECORATIONS', 'SET OF 3 WOODEN STOCKING DECORATIONS'}})	{{'SET OF 3 WOODEN STOCKING DECORATIONS'}})	0.0132022342242533	0.0143562756774223	0.01002	0.75874126	52.8508	0.00983	4.08
13	{{'SET OF 3 WOODEN STOCKING DECORATIONS', 'SET OF 3 WOODEN TREE DECORATIONS'}})	{{'SET OF 3 WOODEN TREE DECORATIONS'}})	0.0143562756774223	0.0132022342242533	0.01002	0.6977492	52.8508	0.00983	3.26
14	{{'POPPY'S PLAYHOUSE LIVINGROOM'}})	{{'POPPY'S PLAYHOUSE BEDROOM', 'POPPY'S PLAYHOUSE LIVINGROOM'}})	0.0147717306005632	0.014033144070535	0.01006	0.68125	48.5458	0.00986	3.09
15	{{'POPPY'S PLAYHOUSE BEDROOM', 'POPPY'S PLAYHOUSE LIVINGROOM'}})	{{'POPPY'S PLAYHOUSE LIVINGROOM'}})	0.014033144070535	0.0147717306005632	0.01006	0.71710526	48.5458	0.00986	3.48

6. Literatura

Baza podataka 1 : <https://www.kaggle.com/datasets/gunhee/koreahousedata>

Baza podataka 2 : <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis?resource=download>

Istrazivanje podataka za Industriju 4.0

https://poincare.matf.bg.ac.rs/~nenad/ip_ind_4.0.html

Istraživanje podataka 1 <https://matf-istrazivanje-podataka-1.github.io/practices/>

