

ALGORITMI KLASTER ANALIZE SA PRIMENOM**ALGORITHMS OF CLUSTER ANALYSIS WITH AN APPLICATION**Dragana Galogaža, Jelena Ivetić, *Fakultet tehničkih nauka, Novi Sad***Oblast – MATEMATIKA**

Kratak sadržaj – *Kroz ovaj rad je predstavljen pojam klaster analize i njena primena. Rad objašnjava razliku hijerarhijskih i nehijerarhijskih metoda, kao i više algoritama klaster analize. Optimizacija broja klastera, ali i određivanja samih klastera ilustrovani su kroz primenu na realnim podacima i realizovani u programskom jeziku R.*

Ključne reči: *Klaster analiza, Hijerarhijske metode, Nehijerarhijske metode, algoritam k-sredina*

Abstract – *This paper presents different algorithms of cluster analysis and its application. The difference between hierarchical and non-hierarchical methods is explained, as well as a number of cluster analysis algorithms. The procedure for finding the optimal number of clusters, followed by determination of the clusters themselves is illustrated via an application example on real data, and processed in the programming language R.*

Keywords: *Cluster analysis, Hierarchical methods, Non-hierarchical methods, k-means algorithm*

1. UVOD

Klaster analiza je jedna od metoda multivarijacione analize koja se zasniva na posmatranju i analizi više varijabli istovremeno. Koristi se za razvrstavanje podataka u grupe ili klastera tako što se slični podaci nalaze u istom klasteru. Grupe koje nastaju su disjunktni podskupovi celokupnog skupa podataka nad kojim se vrši analiza i one imaju svojstvo da se podaci koji pripadaju različitim klasterima međusobno razlikuju mnogo više od podataka koji pripadaju istom klasteru.

Cilj klaster analize je otkrivanje određene vrste prirodne strukture u skupu podataka, tj. grupisanje objekata prema njihovoj međusobnoj sličnosti u grupe [1]. Sredstva koja omogućavaju ovakvu podelu su mere sličnosti ili različitosti među podacima. Na osnovu mera sličnosti među podacima razvijene su dve osnovne metode za klasterovanje klasifikaciju, hijerarhijski i nehijerarhijski metod, koji će u nastavku biti objašnjeni. Objekti koji su međusobno slični u dovoljno visokom stepenu čine homogenu grupu tj. klaster [2].

Pojam klaster analize obuhvata mnogo različitih metoda grupisanja, te je iz tog razloga nemoguće odrediti generalno najbolji metod klasterovanja. Sa druge strane, poznavanje različitih metoda klaster analize je veoma

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bila dr Jelena Ivetić, vanr. prof.

bitno jer je tako moguće prepoznati prednosti korišćenja neke određene metode u konkretnom slučaju. Istraživač je taj koji treba da formuliše cilj istraživanja i da na osnovu njega odredi metod klaster analize koji njemu najviše odgovara [3], obzirom da isti skup objekata može imati različita smislena grupisanja u zavisnosti od svrhe grupisanja, kao i da primena neodgovarajućeg algoritma može dovesti do potpuno beskorisnih klastera. Stoga pre započinjanja klaster analize treba utvrditi jasne kriterijume o tome kako se procenjuje kvalitet pojedinog klastera i kvalitet odnosa između različitih klastera, a pored toga i šta sličnost/različitost objekata znači u određenoj oblasti primene.

2. ALGORITMI I METODE KLASTER ANALIZE

Postupak klaster analize se može podeliti u sledećih 5 koraka:

- 1) Postavljanje ciljeva klaster analize
- 2) Dizajniranje istraživanja
- 3) Kreiranje klastera
- 4) Interpretacija klastera
- 5) Validacija dobijenih klastera

Najpre je potrebno identifikovati varijable koje će biti klasterovane u skladu sa ciljevima istraživanja [4]. Pri formiranju klastera istraživač može imati za cilj da otkrije prirodne grupe među podacima, da pojednostavi podatke tj. da ih podeli u manje grupe koje će biti smislenije ili da otkrije vezu među podacima a samim tim i odnos, sličnosti i razlike među njima. Zatim je potrebno utvrditi mere sličnosti, da li je uzorak reprezentativan i da li podaci imaju neke ekstremne vrednosti [5]. Kada se dobiju rešenja klaster analize, potrebno ih je interpretirati. Ukoliko dobijemo više rešenja, potrebno je izdvojiti ono koje se najbolje slaže sa teorijskom osnovom. Postoji nekoliko načina za proveravanje tačnosti rezultata klaster analize, ali najbolji je da se primeni klaster analiza na nekom drugom uzorku, pa se zatim uporede rezultati [6].

Dve osnovne grupe metoda klaster analize su hijerarhijske i nehijerarhijske metode. Hijerarhijske metode spadaju u grupu tradicionalnih tehnika klaster analize. Sastoje se od računanja udaljenosti među podacima, od kojih je ili inicijalno svaki podatak u zasebnom klasteru, ili se svi podaci na početku nalaze u istom klasteru. Sledeći korak je stvaranje novih klastera, te ponovno izračunavanje udaljenosti elemenata novonastalih klastera i stvaranje novih klastera sve dok ne dođemo do jednog klastera ili dok se svi klasteri ne razdvoje [7]. Prvi pristup se zove tehnika spajanja, a drugi tehnika podele. Neke od najviše korišćenih metoda iz ove grupe su: metoda grupisanja na osnovu najbližih suseda, metoda grupisanja na osnovu

najdaljih suseda, metoda grupisanja na osnovu prosečnog rastojanja suseda.

Rastojanja između susednih tačaka se mogu računati preko Euklidskog rastojanja, Menhetn rastojanja, ili uopštenijeg rastojanja Minkovskog.

Nasuprot ovim tradicionalnim tehnikama se nalaze nehirerarhijske, savremene metode klaster analize. Nehierarhijski metod (ili particionalno klasterovanje) je postupak grupisanja kod kojeg jedinice mogu da prelaze iz jednog u drugi klaster.

Broj grupa je poznat unapred na osnovu algoritama za pronalaženje optimalnog broja klastera.

Kada se ustanovi broj klastera, traži se jedna početna tačka za svaku grupu [8]. U zavisnosti od izbora algoritma, ova tačka se određuje na različite načine. Neki od poznatih algoritama iz ove grupe su algoritam k-sredina, algoritam k-medoida, KLARA algoritam i PAM algoritam.

3. PRIMENA KLASTER ANALIZE

U ovom poglavlju biće detaljno prikazani rezultati klasterovanja primenom četiri algoritma, po dva hijerarhijska i nehijerarhijska.

Analizirani su realni podacima preuzeti od Republičkog zavoda za statistiku [9], a odnose se na izvoz robe iz Republike Srbije. Baza predstavlja masu izvezene robe prema 225 država u 10 kategorija (hrana i žive životinje, pića i duvan, sirove nejestive materije, itd.).

Za statističku obradu podataka korišćen je programski jezik R. Osim standardnih ugrađenih funkcija instalirane su i dve dodatne: funkcija *facextra* koja ekstrahuje i vizualizuje rezultate klaster analize i funkcija *cluster* koja pronalazi grupe u podacima.

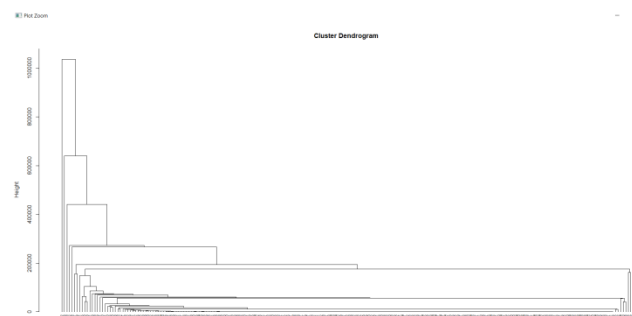
3.1. Hijerarhijska metoda spajanja – grupisanje na osnovu najbližih suseda

Za sve hijerarhijske metode, pa tako i za metodu grupisanja na osnovu najbližih suseda, koristi se komanda *hclust*.

```

Console Terminal Jobs
R 4.1.3 - C:/Users/e5650238/OneDrive - FIS/Desktop/master rad/
> setwd("C:/Users/e5650238/OneDrive - FIS/Desktop/master rad")
> tabela<-read.csv("dataforcluster.csv")
> tabela2<-na.omit(tabela)
> d<-dist(tabela2, method="euclidean")
> hclust<-hclust(d, method = "single")
> plot(hclust, cex = 0.6, hang = -10)
  
```

Slika 1. Kod programskog jezika R za pronalaženje klastera metodom najbližih suseda



Slika 2. Dobijeni klasteri metodom najbližih suseda

Posmatrajući dendrogram na Slici 2, uočavamo da su u prvoj iteraciji države pod rednim brojem 126 (Crna Gora) i 129 (Severna Makedonija) formirale jedan klaster, kao i države pod rednim brojem 50 (Češka) i 68 (Francuska) i države pod rednim brojem 87 (Hrvatska) i 89 (Mađarska).

Primitimo da se do poslednje iteracija država pod rednim brojem 172 (Rumunija) izdvajala, tj. ostala sama u klasteru, i tek na kraju algoritma se spojila sa ostalim klasterima.

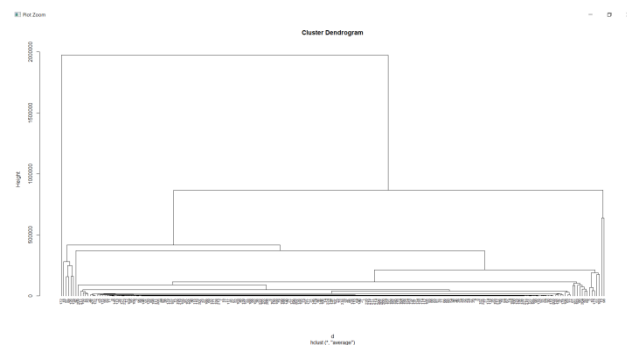
3.2. Hijerarhijska metoda spajanja – grupisanje na osnovu najdaljih suseda

Metoda grupisanja na osnovu najdaljih suseda takođe koristi komandu *hclust*. Slike 3 i 4 prikazuju kod u jeziku R i dobijene rezultate u formi dendograma.

```

Console Terminal Jobs
R 4.1.3 - C:/Users/e5650238/OneDrive - FIS/Desktop/master rad/
> setwd("C:/Users/e5650238/OneDrive - FIS/Desktop/master rad")
> tabela<-read.csv("dataforcluster.csv")
> tabela2<-na.omit(tabela)
> d<-dist(tabela2, method="euclidean")
> hclust<-hclust(d, method = "average")
> plot(hclust, cex = 0.6, hang = -10)
  
```

Slika 3. Kod programskog jezika R za pronalaženje klastera metodom najdaljih suseda



Slika 4. Dobijeni klasteri metodom najdaljih suseda

Posmatrajući Sliku 4, može se primetiti da su u prvoj iteraciji države pod rednim brojem 13 (Bosna i Hercegovina) i 98 (Italija) formirale jedan klaster. Kao i sledeći parovi država: 50 (Češka) i 68 (Francuska), 87 (Hrvatska) i 89 (Mađarska), 126 (Crna Gora) i 129 (Severna Makedonija). Primitimo da se takođe do poslednje iteracije izdvajala država 172 (Rumunija), i tek na kraju algoritma se spojila sa ostalim državama.

3.3. Algoritam K- sredina

Algoritam k-sredina je jedan od najpopularnijih nehijerarhijskih metoda klaster analize. Da bi odredili klaster, potrebno je da unapred znamo optimalan broj klastera.

Za potrebe master rada korišćene su metoda lakta, metoda siluete i GEP statistika u cilju pronalaženja optimalnog broja klastera. U ovom primeru prikazana je metoda lakta.

Slika 5 prikazuje funkcije potrebne za pronalaženje optimalnog broja klastera: funkcije koje zadaju naredbu putanje gde se nalazi tabela podataka, učitavanje tabele, a zatim funkcija koja izračunava optimalan broj klastera za metodu k-sredina uz pomoć metode lakta.

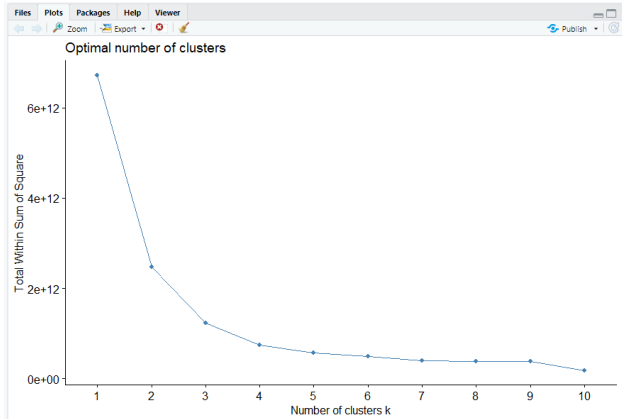
```

R 4.1.3 - C:/Users/e5650238/OneDrive - FIS/Desktop/master rad/
> setwd("C:/Users/e5650238/OneDrive - FIS/Desktop/master rad")
> tabela<-read.csv("DataforCluster.csv")
> fviz_nbclust(tabela,Funcluster = kmeans, method = "wss")
> |

```

Slika 5. Kod programskog jezika R za pronalaženje optimalnog broja klastera metodom lakta za algoritam k-sredina

Rezultat je grafikon prikazan na Slici 6. Vidi se da do pregiba dolazi u tački 3 na x-osi, te je 3 optimalan broj klastera korišćen za metodu k-sredina.



Slika 6. Grafikon metode lakta za algoritam k-sredina

Nakon što je izračunat optimalan broj klastera, može se preći na sam postupak algoritma k-sredina. Koristi se funkcija *kmeans* koja raspoređuje podatke u zadati broj klastera (u ovom slučaju 3).

U funkciji *kmeans* treba zadati i početni broj tačaka, *nstart*, jer je rezultat metode k-sredina osetljiv na nasumične početne podele tačaka.

Ovo znači da će programski jezik R pokušati *n* različitih nasumičnih podela, a zatim odabrati najbolje rezultate koji odgovaraju podacima sa najnižom varijacijom unutar klastera.

Smatra se da što je veći broj *nstart*, algoritam k-sredina u programskom jeziku R postaje bolji. Ipak, preporučena vrednost je 25 ili 50, kako bi se dobio najstabilniji rezultat.

```

R 4.1.3 - C:/Users/e5650238/OneDrive - FIS/Desktop/master rad/
> setwd("C:/Users/e5650238/OneDrive - FIS/Desktop/master rad")
> tabela<-read.csv("DataforCluster.csv")
> tabela2<-na.omit(tabela)
> fviz_nbclust(tabela2, Funcluster = kmeans, method = "wss")
> view(tabela2)
> kmeans(tabela2, 3, nstart = 50)
k-means clustering with 3 clusters of sizes 7, 1, 217

Cluster means:
  NO      NO.1      NO.2      NO.3      NO.4      NO.5      NO.6      NO.7      NO.8      NO.9
1 348257.124 65231.1429 160780.600  94883.571 20954.9000 151832.157 251314.043 24961.657 20946.1429 50.98571
2 1906302.800  8649.9000 149638.900 144195.3000 2329.4000 360494.500 400579.000 17467.400 46123.1000 14.20000
3  5343.056   800.8366  3246.965   546.7783  462.9399  3250.271  6269.432  1726.714  934.6333 43.40737

Clustering vector:
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3  3
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
 3  3  3  3  3  3  3  3  3  3  3  3  3  3  3

within cluster sum of squares by cluster:
[1] 7437202328
[2] 48601193490
(between_SS / total_SS = 81.7 %)

available components:
[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss" "size"
[8] "iter" "ifault"
|

```

Slika 7. Dobijeni klasteri pomoću algoritma k-sredina

Rezultati metode k-sredina daju 3 klastera različitih veličina. Prvi klaster se sastoji od 7 država, drugi od jedne države a treći od 217 država:

- Klaster 1: 13 (Bosna i Hercegovina), 18 (Bugarska), 87 (Hrvatska), 89 (Mađarska), 98 (Italija), 126 (Crna Gora) i 129 (Severna Makedonija).
- Klaster 2: 172 (Rumunija).
- Klaster 3: Sve ostale države.

Upoređujući ovu metodu sa prethodno obrađenim, može se primetiti konzistentnost analiza u pogledu izdvojenog položaja Rumunije, koja se nalazi u zasebnom klasteru.

3.4. Algoritam K metodis – PAM algoritam

PAM algoritam je jedan od popularnih metoda nehijerarhijske klaster analize. Da bi odredili klaster, potrebno je da unapred znamo optimalan broj klastera.

Računanje optimalnog broja klastera je u ovom primeru određeno pomoću metode GEP statistike i predstavljeno je na slikama 8 i 9.

```

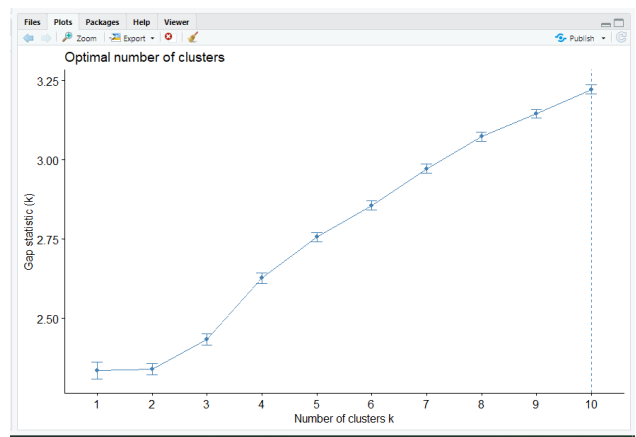
> gap_stat <- clusgap(tabela2, pam, nstart = 25,
+                    k.max = 10, B = 50)
+ clustering k = 1,2,..., k.max (= 10): ... done
+ Bootstrapping, b = 1,2,..., B (= 50) [one "." per sample]:
+ ..... 50
> print(gap_stat, method = "firstmax")
Clustering Gap statistic ["clusgap"] from call:
clusgap(x = tabela2, FUNcluster = pam, k.max = 10, B = 50, nstart = 25)
B=50 simulated reference sets, k = 1..10; spaceHO="scaledPCA"
--> Number of clusters (method "firstmax"): 10
      logw      e_logw      gap      se_sim
[1.] 15.25208 17.58831 2.336223 0.02651488
[2.] 14.79739 17.13762 2.340229 0.01736329
[3.] 14.52434 16.95785 2.433509 0.01706230
[4.] 14.24181 16.86856 2.626753 0.01588005
[5.] 14.05962 16.81579 2.756173 0.01467274
[6.] 13.91474 16.77082 2.856083 0.01382337
[7.] 13.75748 16.72932 2.971837 0.01427100
[8.] 13.61828 16.69126 3.072979 0.01419550
[9.] 13.51172 16.65624 3.144517 0.01319518
[10.] 13.40208 16.62426 3.222182 0.01416416
> fviz_gap_stat(gap_stat)
> |

```

Slika 8. Kod programskog jezika R za pronalaženje optimalnog broja klastera metodom GEP statistike

Rezultat predhodno napisane funkcije predstavlja grafik GEP statistike, prikazan na Slici 9.

GEP metoda daje optimalan broj klastera 10 koji će biti korišćen za PAM algoritam.



Slika 9. Grafikon metode GEP statistike

Nakon što je izračunat optimalan broj klastera, može se preći na sam postupak PAM algoritma. Zadati broj klastera za PAM algoritam jeste 10.

Rezultati algoritma daju sledeće klasterne različitih veličina:

- 1) Sve ostale države
- 2) 5 (Albanija), 10 (Austrija), 51 (Nemačka), 173 (Rusija)

- 3) 13 (Bosna i Hercegovina)
- 4) 18 (Bugarska)
- 5) 42 (Kina)
- 6) 50 (Češka), 61 (Španija), 68 (Francuska), 80 (Grčka), 163 (Poljska), 181 (Slovenija), 183 (Slovačka), 204 (Turska), 211 (SAD)
- 7) 87 (Hrvatska), 89 (Mađarska)
- 8) 98 (Italija)
- 9) 126 (Crna Gora), 129 (Severna Makedonija)
- 10) 172 (Rumunija)

```

C:\Users\655238\OneDrive - FIS\Desktop\master_rad\
> setwd("C:/Users/655238/OneDrive - FIS/Desktop/master_rad")
> table1a<-read.csv("data/cluster.csv")
> table1a2<-na.omit(table1a)
> pam(table1a2, 10)
Medoids:
  10  29  13  42  50  61  68  80  87  98  126  129  172
  10  198811.7 2294.9 31908.5 3535.5 11387.5 42739.0 99310.1 11255.2 13860.8 8.3
  13  492291.0 163483.3 175064.0 413006.9 150311.1 203822.3 311560.7 23002.6 39428.0 48.4
  18  18 109173.7 41376.3 134431.6 76817.8 27394.6 123363.4 398829.2 8967.7 8756.0 230.3
  42  42 2498.1 467.8 37392.5 0.0 1828.7 7207.4 37394.0 1749.9 158.1 0.0
  204 204 26813.4 1611.9 59872.0 1504.2 5822.3 17437.8 71362.9 4032.7 3567.6 71.0
  89 89 219964.7 16453.3 313072.7 170000.7 7473.5 233516.6 204886.5 35026.6 11234.5 4.5
  98 98 928227.1 14275.9 65173.6 673.9 14818.5 97500.5 263955.6 49404.6 33018.8 7.1
  129 129 298406.5 56824.8 176463.7 56940.0 29036.0 113538.0 219127.0 22864.9 14571.0 17.2
  172 172 1506302.8 86429.8 149638.9 444195.3 2320.4 300484.5 400379.0 17467.4 46123.5 14.2
Clustering vector:
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
  1  1  1  1  2  1  1  1  1  1  2  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
  1  1  1  1  1  1  1  1  1  1  1  1  5  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
  6  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
  6  1  6  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
  211 212 213 214 215 216 217 218 219 220 221 222 223 224 225
  6  1  1  1  1  1  1  1  1  1  1  1  1  1  1
Objective Function:
  build swap
  8628.437 8127.512
Available components:
  [1] "medoids" "id.med" "clustering" "objective" "isolation" "clusinfo" "silinfo" "dtss" "call"
  [10] "data"
  |

```

Slika 10. Klasteri dobijeni PAM algoritmom

Upoređujući ovu metodu sa prethodno obrađenim, možemo primetiti da su se izdvojile neke do sada nepomenute države, kao što su 5 (Albanija), 10 (Austrija), 173 (Rusija), 42 (Kina), 50 (Češka), 61 (Španija), 68 (Francuska), 80 (Grčka), 163 (Poljska), 181 (Slovenija), 183 (Slovačka), 204 (Turska), 211 (SAD). Međutim i prilikom ove klaster analize, Rumunija je ostala sama u klasteru.

4. ZAKLJUČAK

U ovom radu obrađeni su neki od algoritama hijerarhijskih i nehijerarhijskih metoda klaster analize, odabrani među većim brojem metoda obrađenim u okviru master rada. Kao što možemo primetiti u obrađenim primerima, četiri različita algoritma su dala veoma slične rezultate. U svakom od ovih algoritama u poseban klaster se izdvojila Rumunija. Ostale države su se takođe izdvajale u sličnim parovima u svim obrađenim algoritmima, te su se tako veoma često u istom klasteru nalazile Crna Gora i Severna Makedonija, Hrvatska i Mađarska, Slovenija i Poljska, Bosna i Hercegovina i Italija, kao i Češka i Francuska. Rezultati ovog istraživanja pokazali su nam da su algoritmi klaster analize dobro definisani, i da nećemo pogrešiti ako odaberemo bilo koji od njih. Interpretacija klastera ostaje van okvira ovog rada, s obzirom na to da zahteva razumevanje spoljno trgovinskih odnosa Srbije.

5. LITERATURA

[1] M. R. Anderberg (1973). Cluster Analysis for Applications London: Academic Press (print article) 51-59

[2] Brian S. Everitt , Sabine Landau, Morven Leese, Daniel Stahl(2001). Cluster Analysis London, UK, King’s College: John Wiley & Sons, Ltd (print article) 15-111

[3] Sławomir T. Wierzchoń, Mieczysław A. Kłopotek (2018). Modern Algorithms of Cluser Analysis Warsaw,Poland: Springer International Publishing (print article) 9-67

[4] Bruce L. Golden, S. Raghavan, Edward A. Wasil (2005). The next wave in computing, optimization and decision technologies New York: Springer International Publishing (print article) 213-249

[5] Ljiljana Sekulić (2002). "Primena klasteranalize u istraživanju demografske starosti stanovništva" Republika Srbija: Institut društvenih nauka 154-170

[6] Leonard Kaufman, Peter J. Rousseeuw (1990). Finding Groups in Data: An Introduction to Cluster Analysis New York: Wiley (print article) 67-125

[7] Bradley Boehmke (10th January 2023). "Hierarchical Cluster Analysis" (online blog) https://uc-r.github.io/hc_clustering

[8] JánosAbonyi , Balázs Feil (2005). Cluster Analysis for Data Mining and System Identification Basel, Switzerland: Birkhäuser Basel (print article) 129-146

[9] Republički zavod za statistiku (10th February 2023). <https://opendata.stat.gov.rs/odata/> (Public website)

Kratka biografija:



Dragana Galogaža rođena je u Novom Sadu 1997. god. Master rad na Fakultetu tehničkih nauka iz oblasti Matematika – Primenjena matematika odbranila je 2023.god.

kontakt: galogaza97@gmail.com



Jelena Ivetić je od 2003. zaposlena na Fakultetu tehničkih nauka, trenutno je u zvanju vanrednog profesora. Predaje statističke predmete, a oblasti naučnog rada su joj primenjena statistika i verovatnoća, i logika u računarstvu.

kontakt: jelenaivetic@uns.ac.rs

Zahvalnica:

Jelena Ivetić je finansijski podržana od strane projekta Departmana za opšte discipline u tehnici, FTN - UNS, „Primena informaciono komunikacionih tehnologija u nastavi opštih disciplina“.