



ISTRAŽIVANJE PODATAKA PROJEKAT -DRUG CONSUMPTION-

RADIO:
DAVID ŽIVKOVIĆ 098/2020

PROFESOR:
PROF. DR NENAD MITIĆ

ASISTENT:
STEFAN KAPUNAC

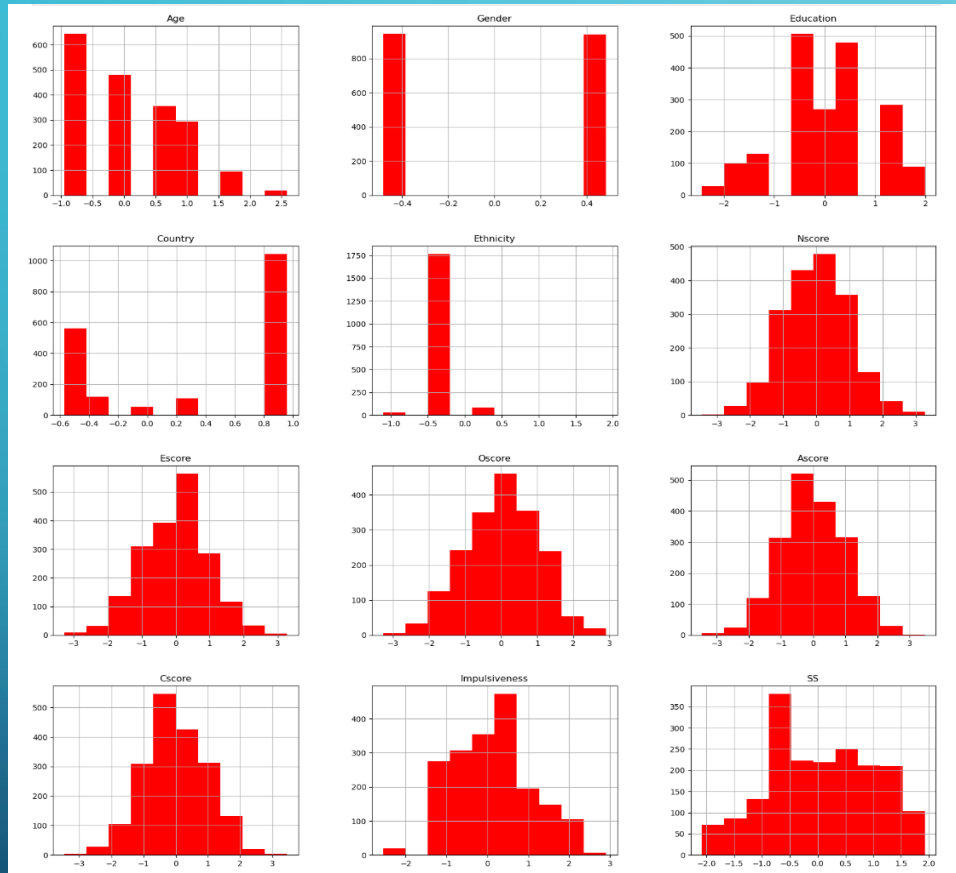
UOPŠTENO O PODACIMA

- Broj instanci je 1885, dok je broj kolona 32
- Kolone su atributi ili klase
- Atributa ima 13, dok kolona ima čak 19
- Sve vrednosti atributa su numeričkog tipa
- Atributi predstavljaju:
 - Lične podatke o instancama
 - Psihološke karakteristike o instancama
- Klase predstavljaju supstancu čija se učestalost konzumacije ispituju

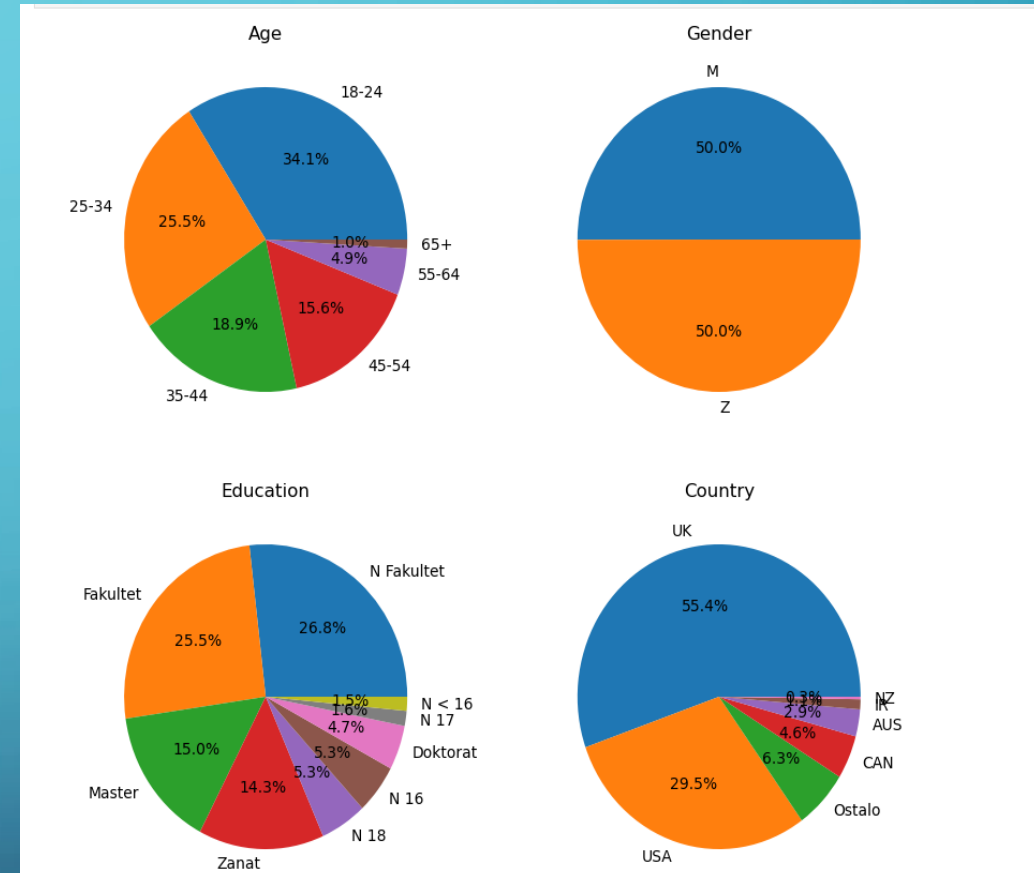
UOPŠTENO O PODACIMA

- Lični atributi: age, gender, country, education, ...
- Psihološke karakteristike: impulsivnost, neurotičnost, ...
- Supstance koje se ispituju variraju od kofeina i nikotina, preko slabih opijata, pa sve do jakih opijata
- Klase koje se ispituju imaju vrednosti CL0-CL6
- CL0 ukazuje na to da nije nikad korišćeno
- CL6 ukazuje na svakodnevno učestalost

UOPŠTENO O PODACIMA

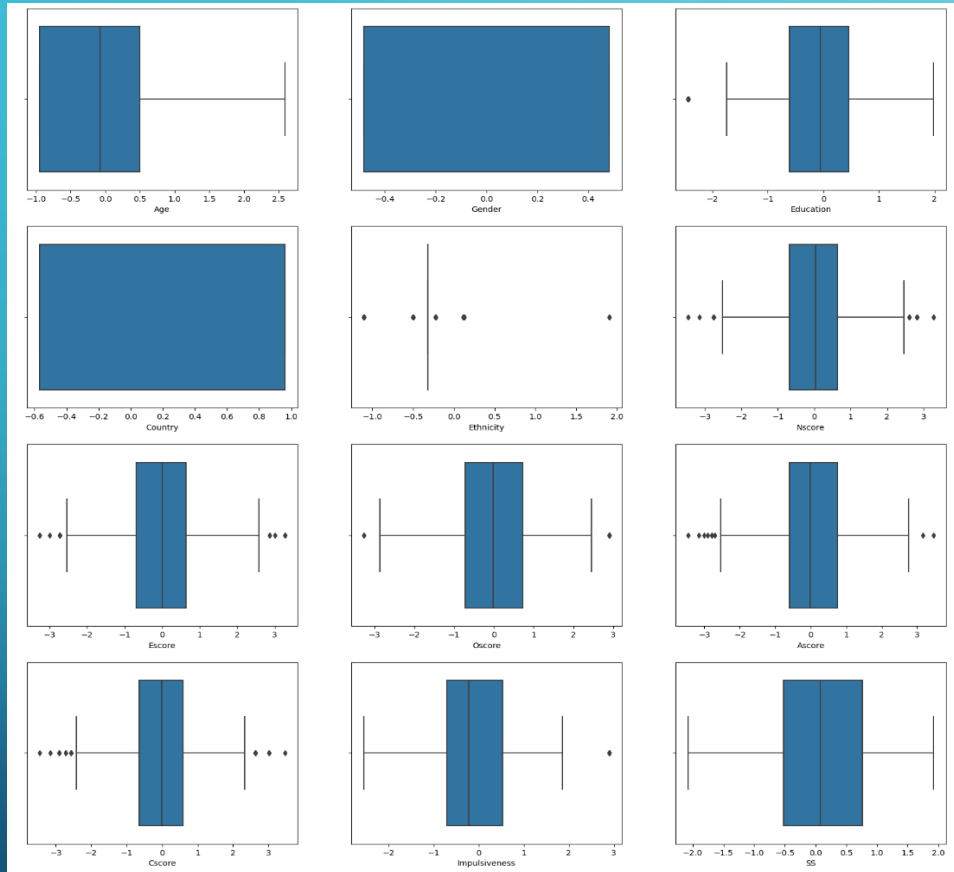


Raspodela atributa

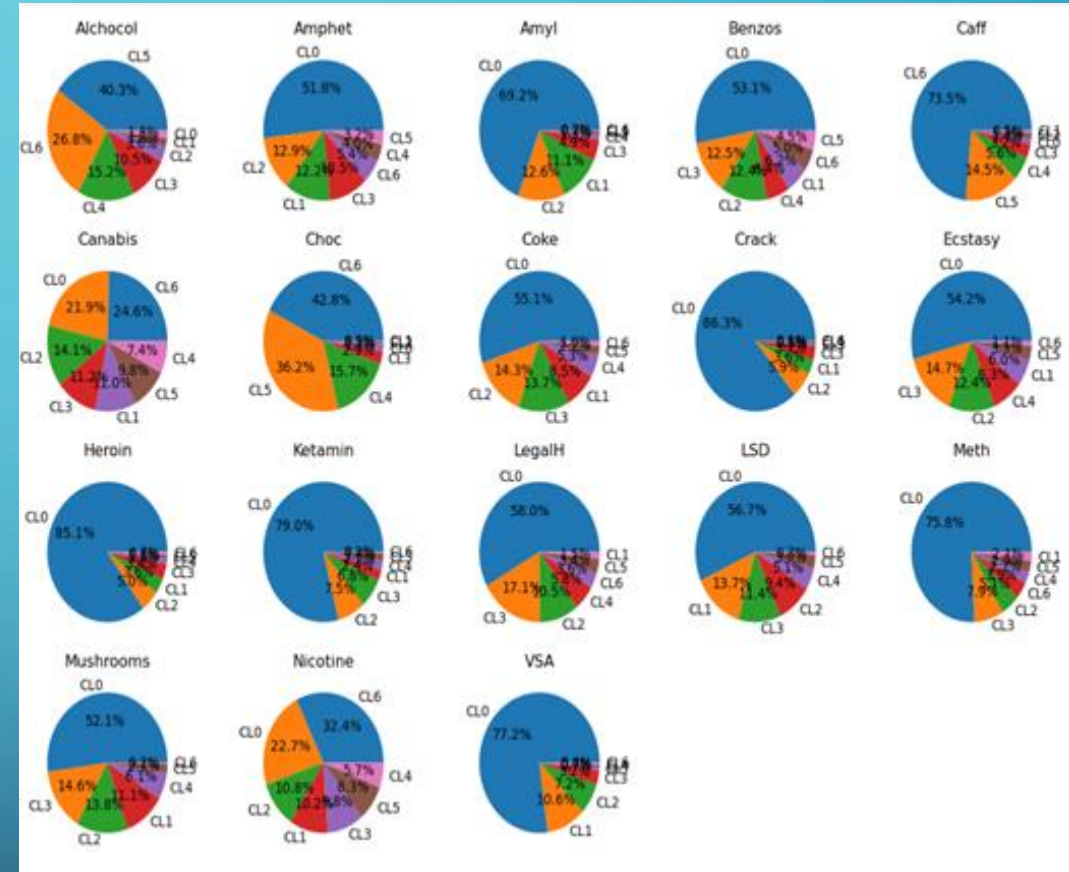


Balansiranost atributa

UOPŠTENO O PODACIMA



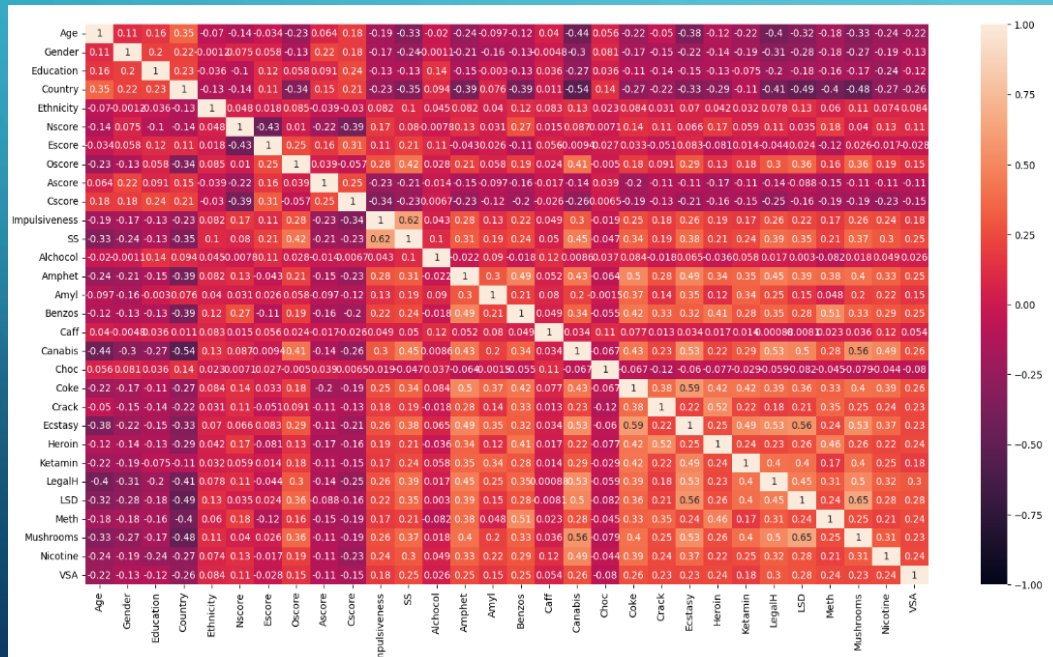
Elementi van granica



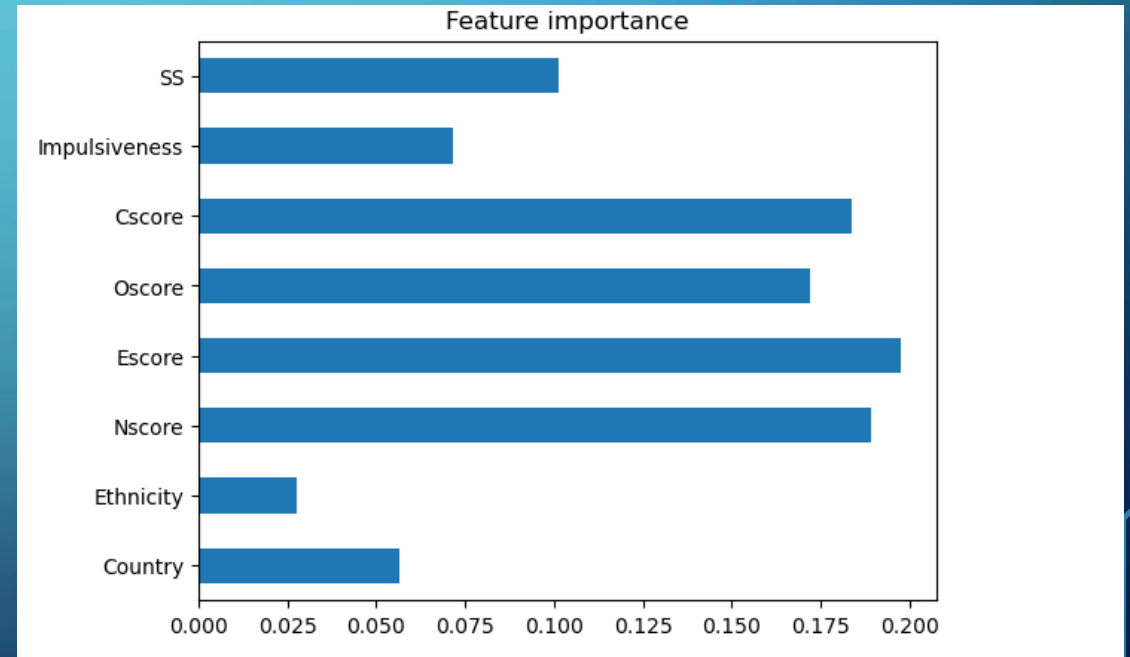
Balansiranost klasa

PRETPROCESIRANJE

- Nema nedostajućih vrednosti
- Postoji minimalno elemenata van granice
- Pojedini atributi imaju loš stepen korelacije



Matrica korelacije



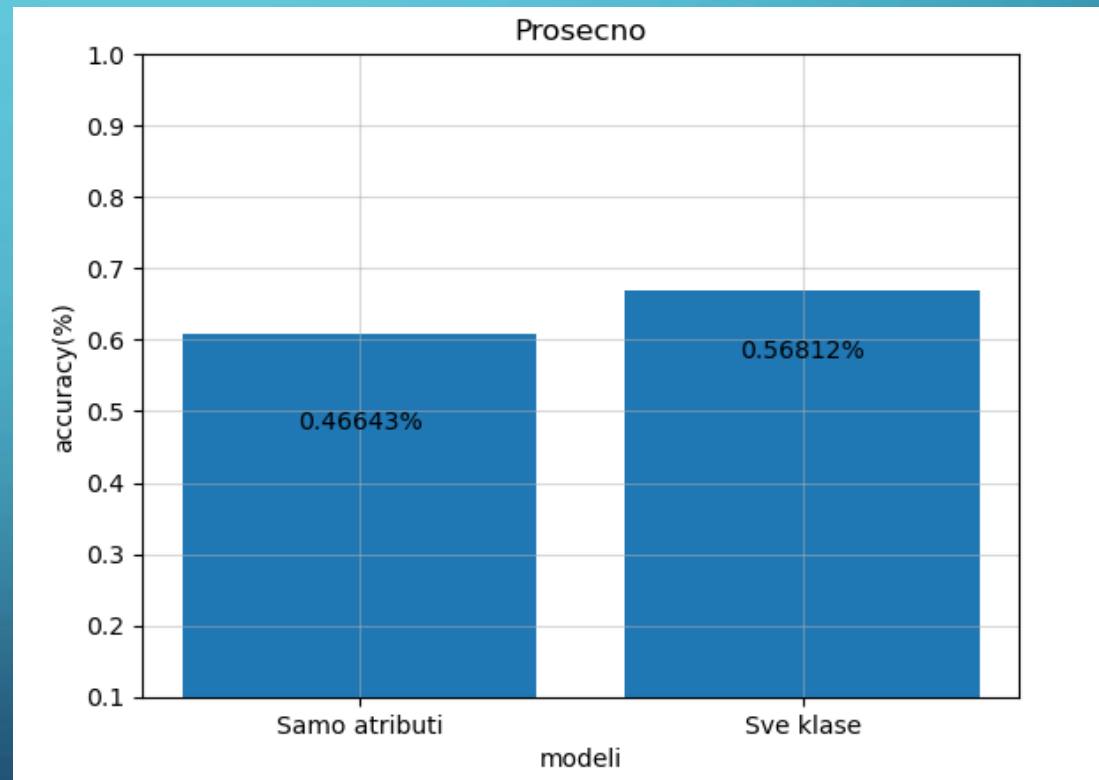
Značajnost atributa

KLASIFIKACIJA

- Modeli korišćeni pri klasifikaciji:
 - Stabla odlučivanja
 - KNN
 - Neuronske mreže (MLPClassifier)
 - SVM (SVC)
 - Logistička regresija
 - Ansambl (RandomForestClassifier)
- Radi se klasifikacija nad svakom klasom

STABLA ODLUČIVANJA

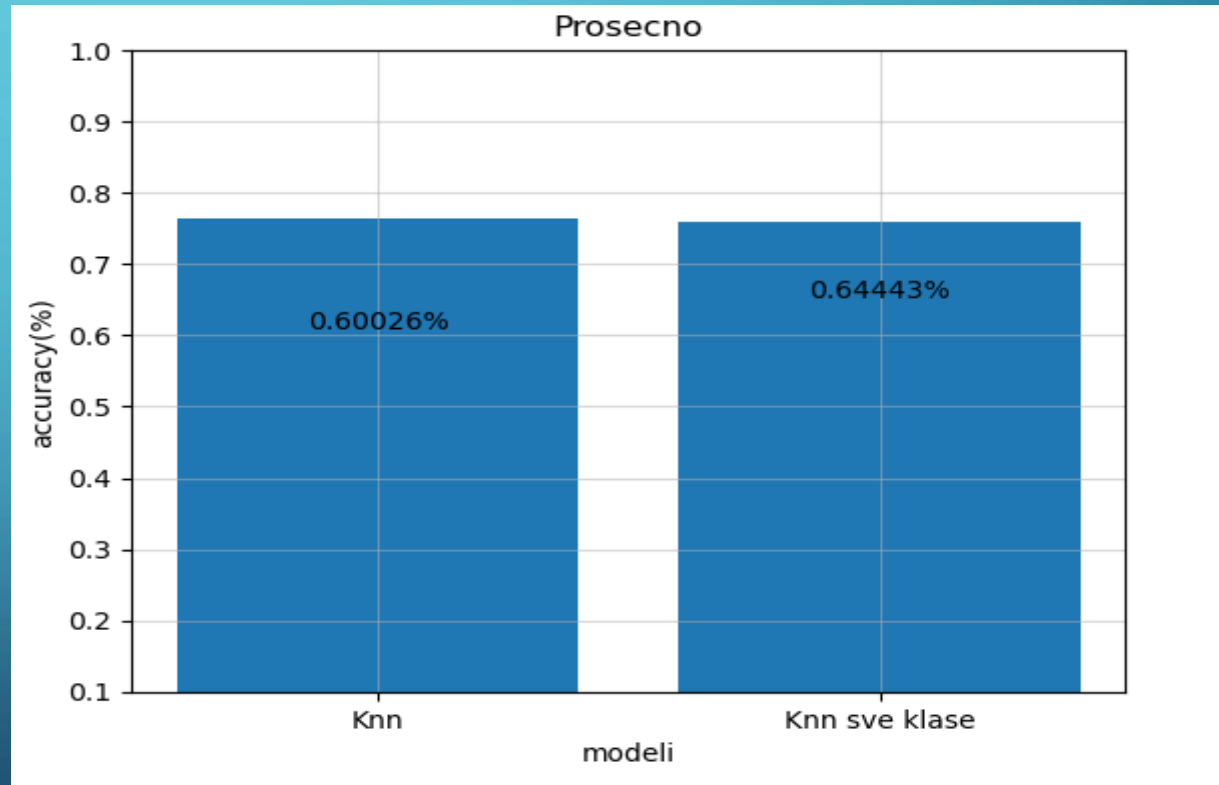
- Izvršeno balansiranje klasa uzorkovanjem
- Rađena klasifikacija i kada se koriste svi atributi
- Hiperparametri:
 - Kriterijum atributa: gini
 - Najveća dubina: 21
 - Najveći broj atributa: 6



Rezultati

KNN

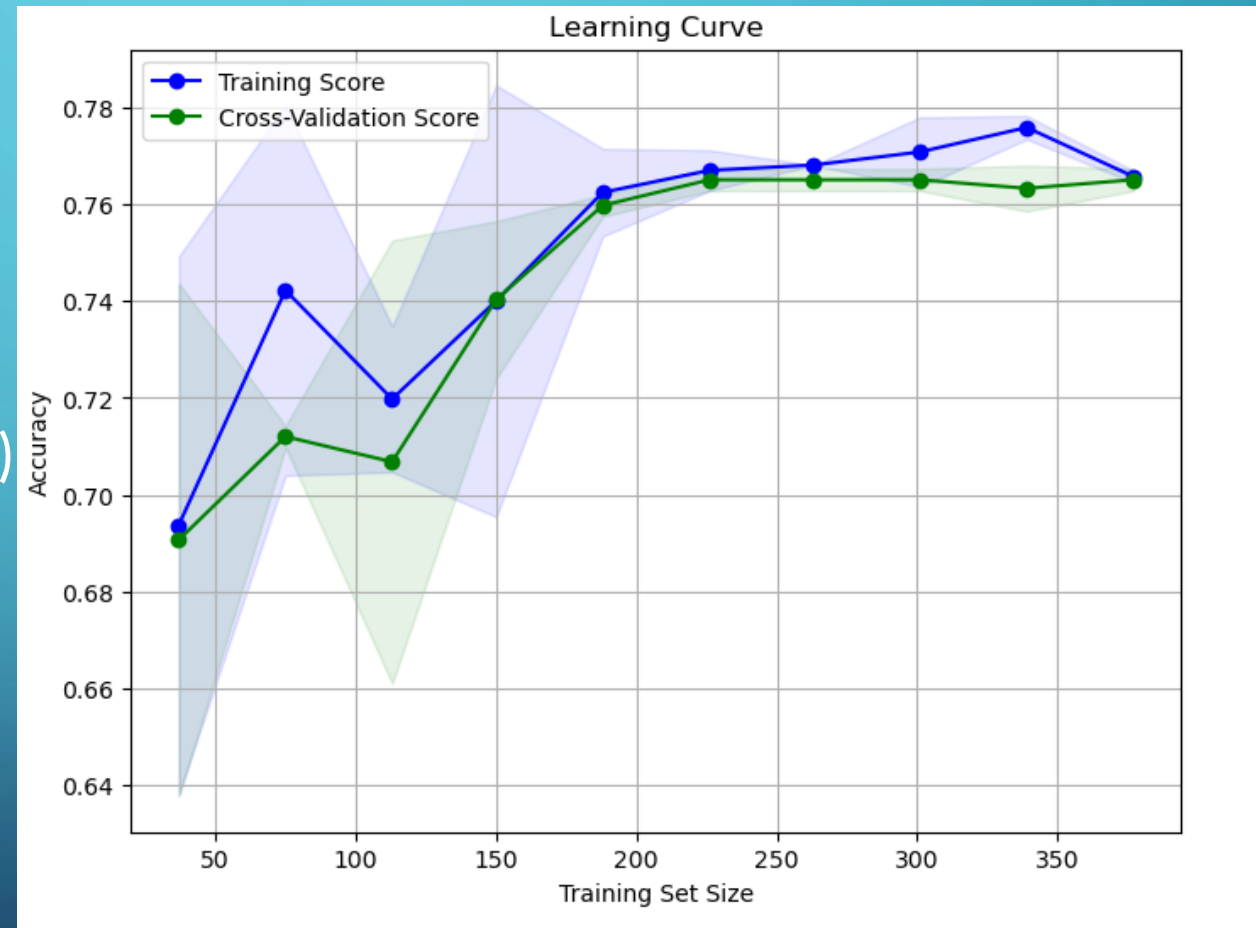
- Zbog prirode algoritma neophodno je skaliranje podataka
- Brzo se prilagodi sa porastom broja suseda
- Hiperparametri:
 - Algoritam: auto
 - Broj suseda: 10
 - Težina: uniform



Rezultati

NEURONSKE MREŽE

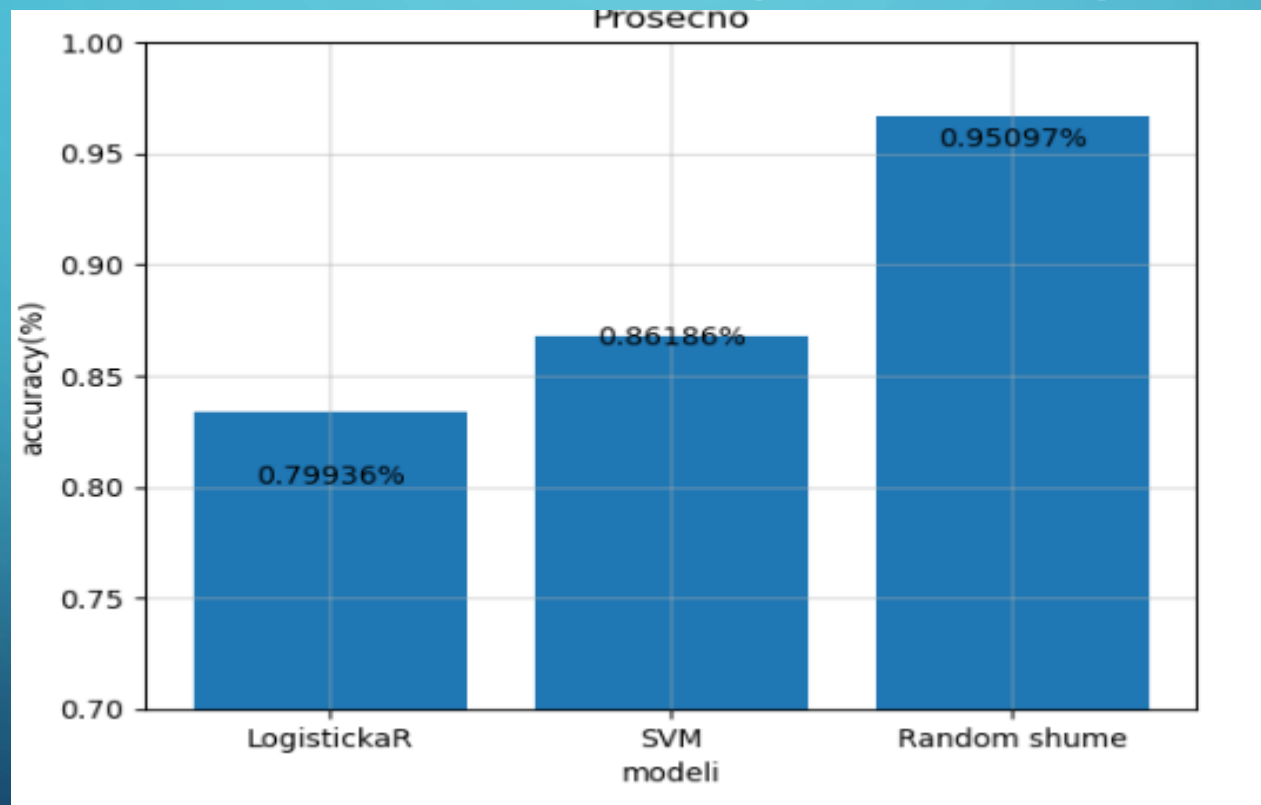
- Zbog prirode algoritma neophodno je skaliranje podataka
- Hiperparametri:
 - Fja aktivacije: tanh
 - Brzina učenja: constant
 - Skriveni slojevi: (50, 30, 20)
 - Najviše iteracija: 20



Kriva učenja

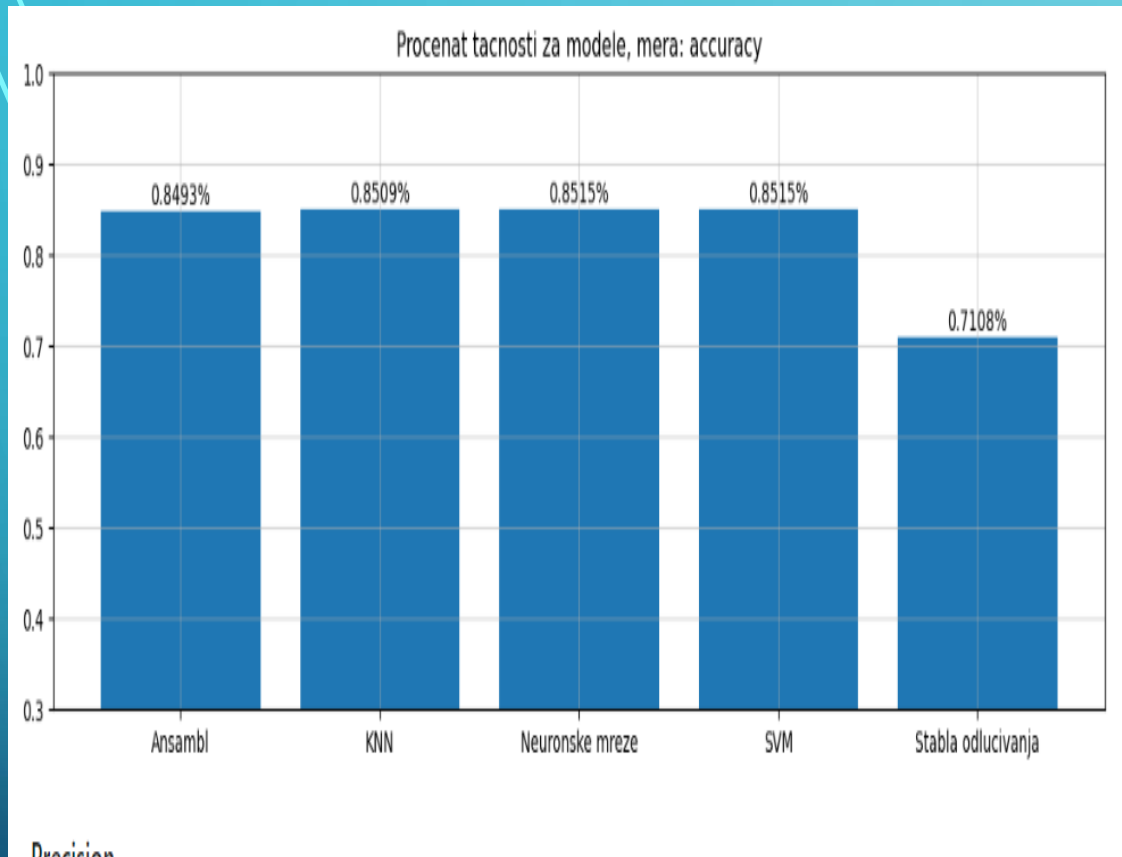
LOGISTIČKA REGRESIJA; SVM; ANSAMBL

- Logistička regresija radi kao neuronske mreže sa 1 neuronom
- SVM pokušava da nađe idealnu jednačinu prikaza podataka

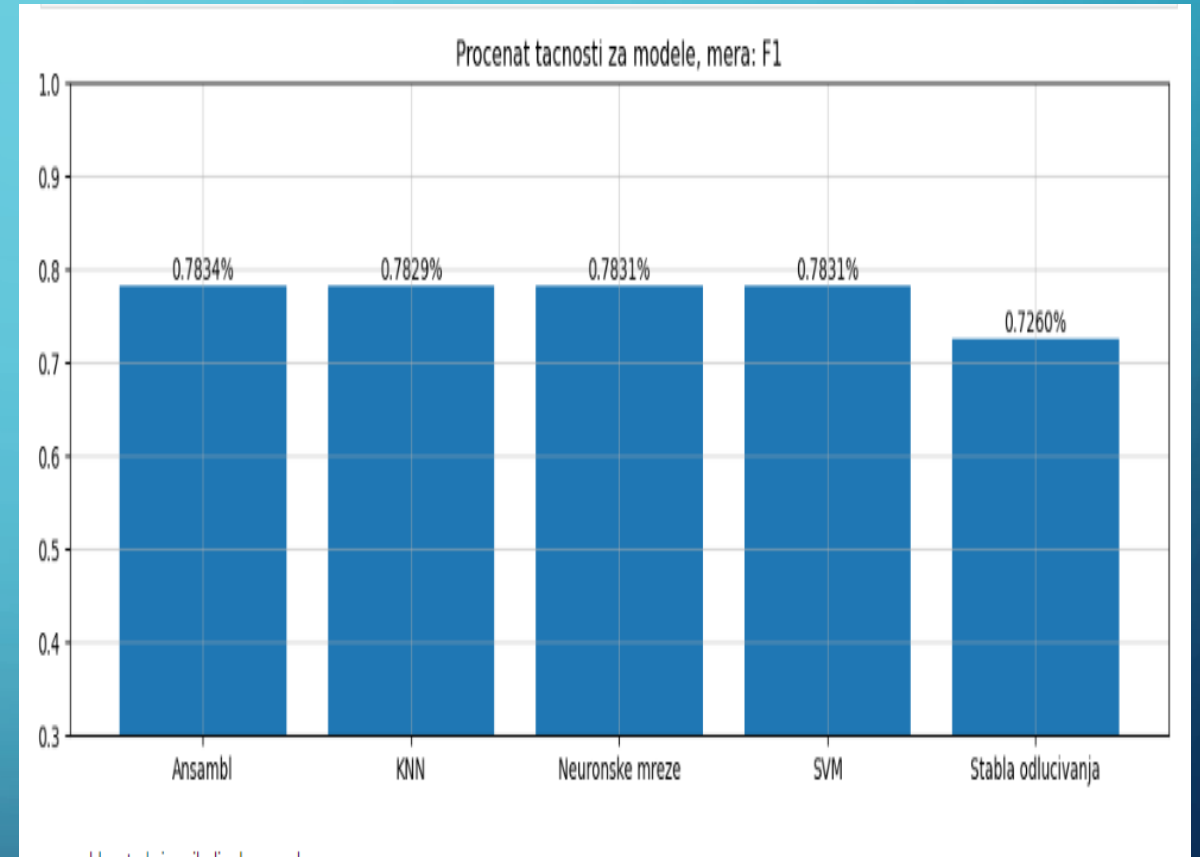


Rezultati

REZULTATI



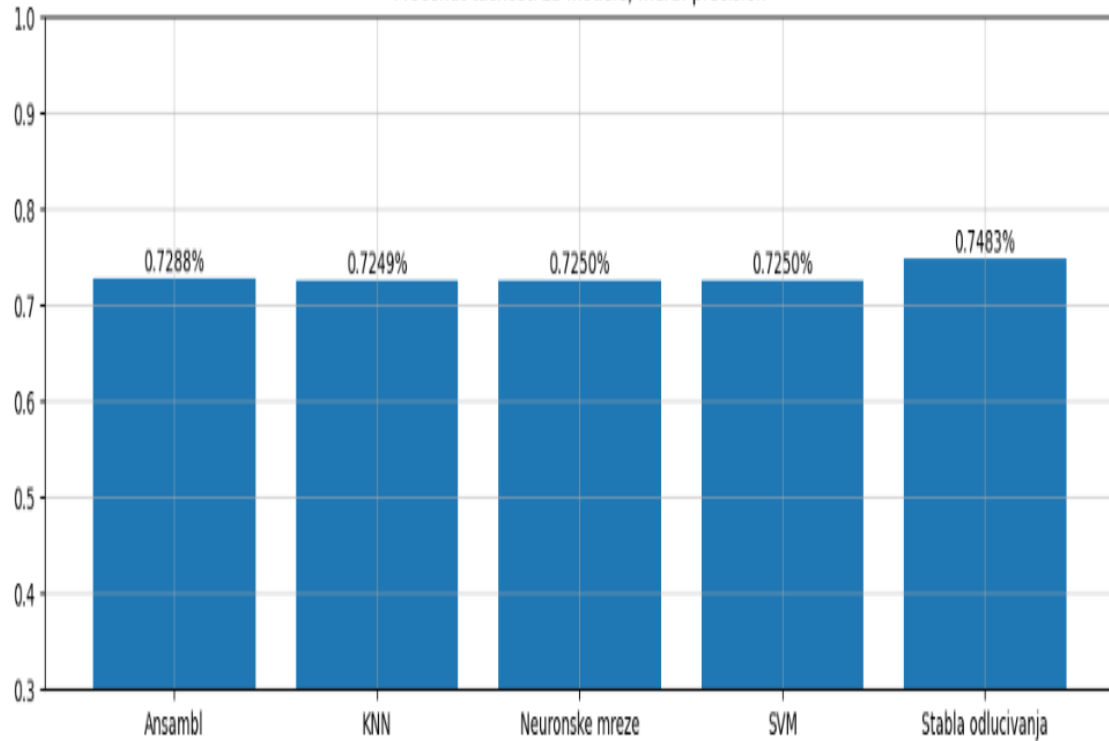
Accuracy mera modela



F1-score modela

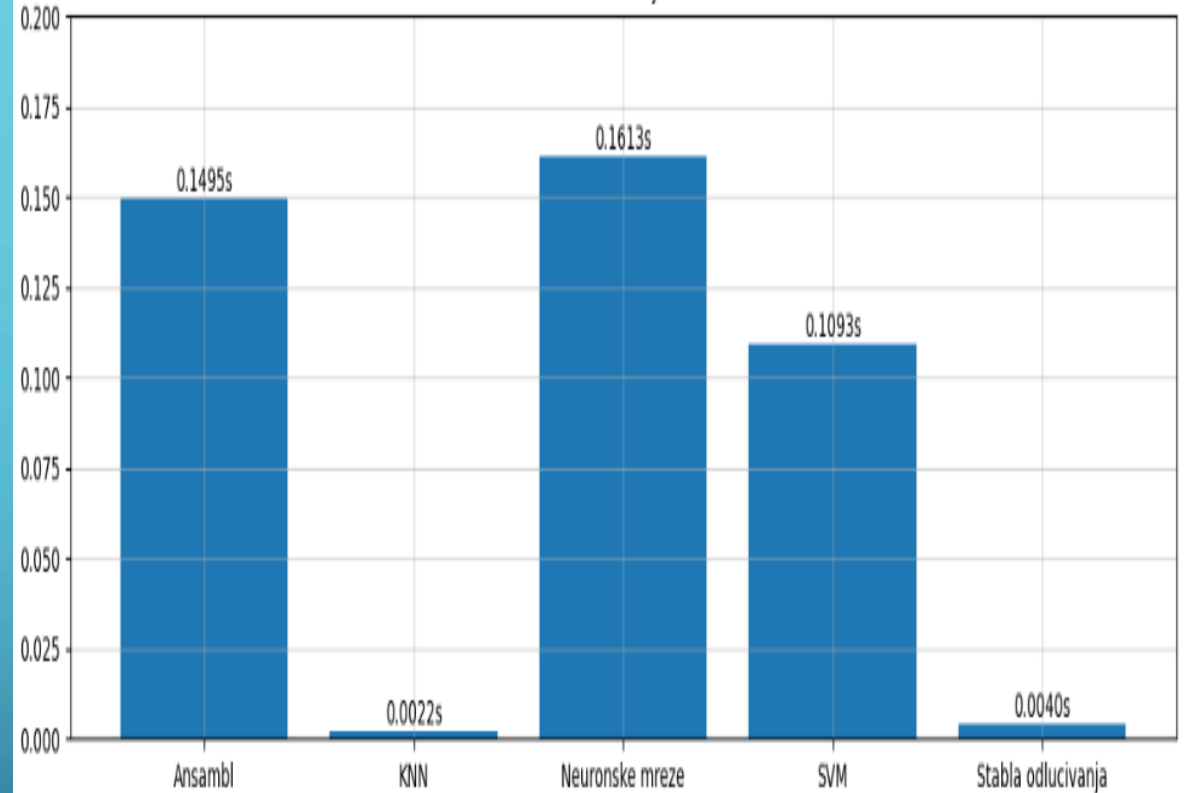
REZULTATI

Procentat tacnosti za modele, mera: precision



Precision mera modela

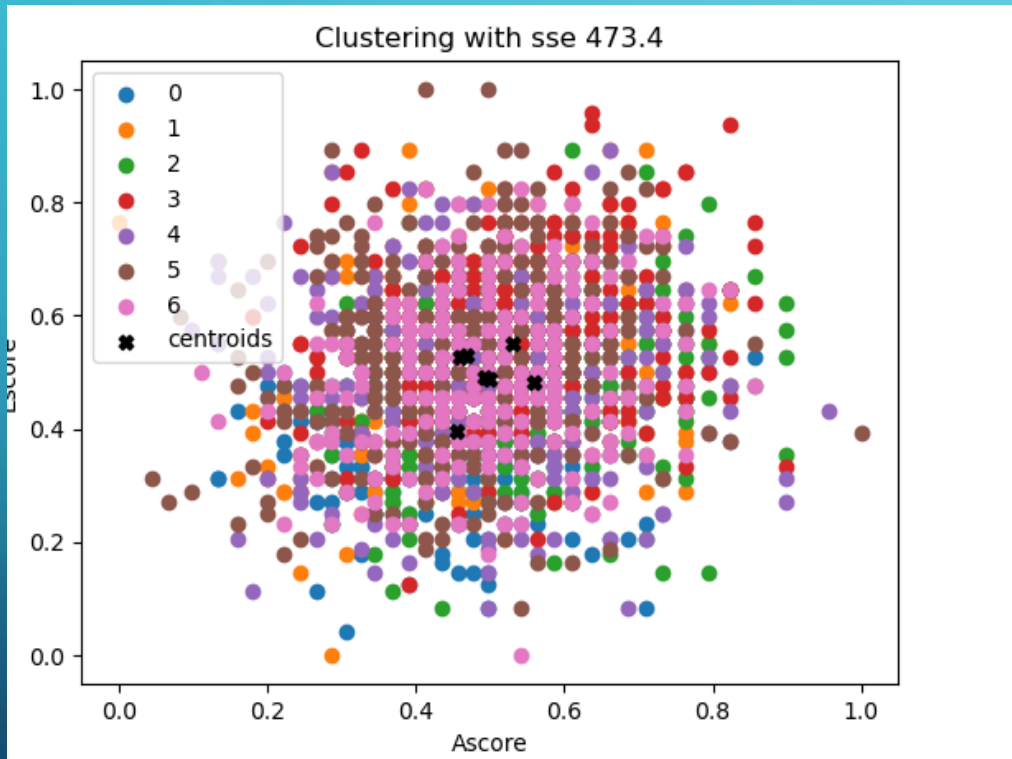
Vreme izvrshavanja modela



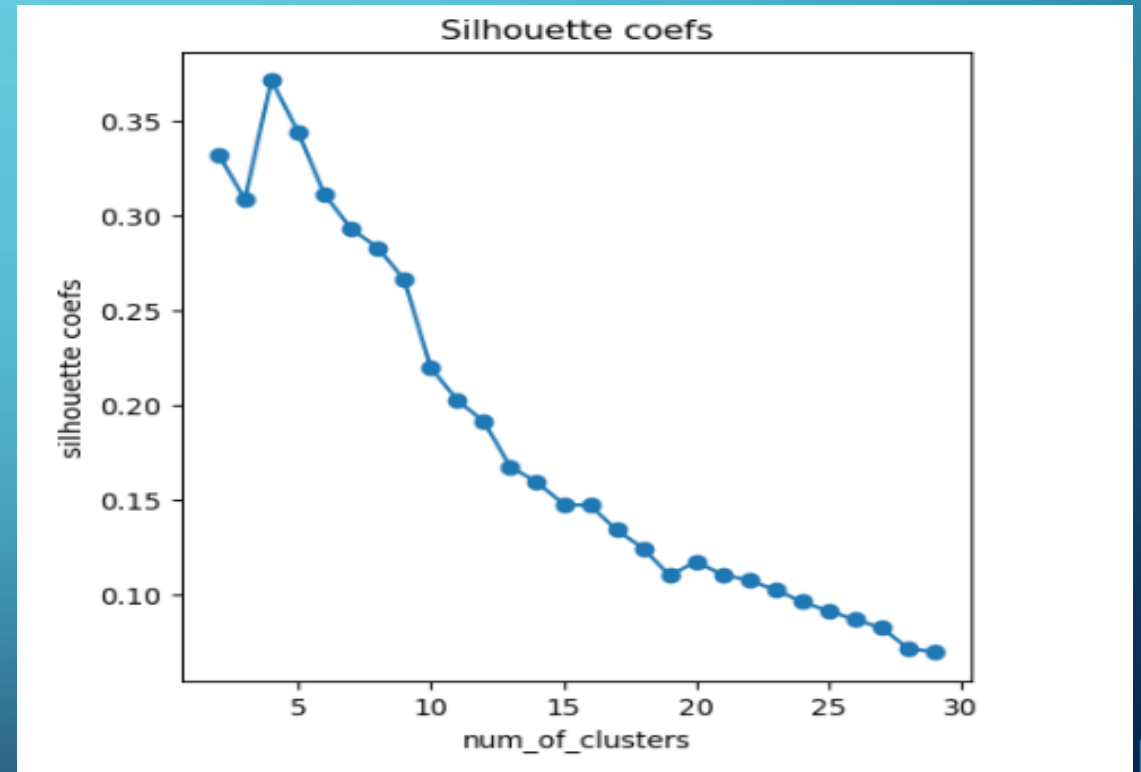
Brzina treniranja

KLASTEROVANJE

- Nema mnogo podataka, iako radi sporo ne treba uzorkovanje

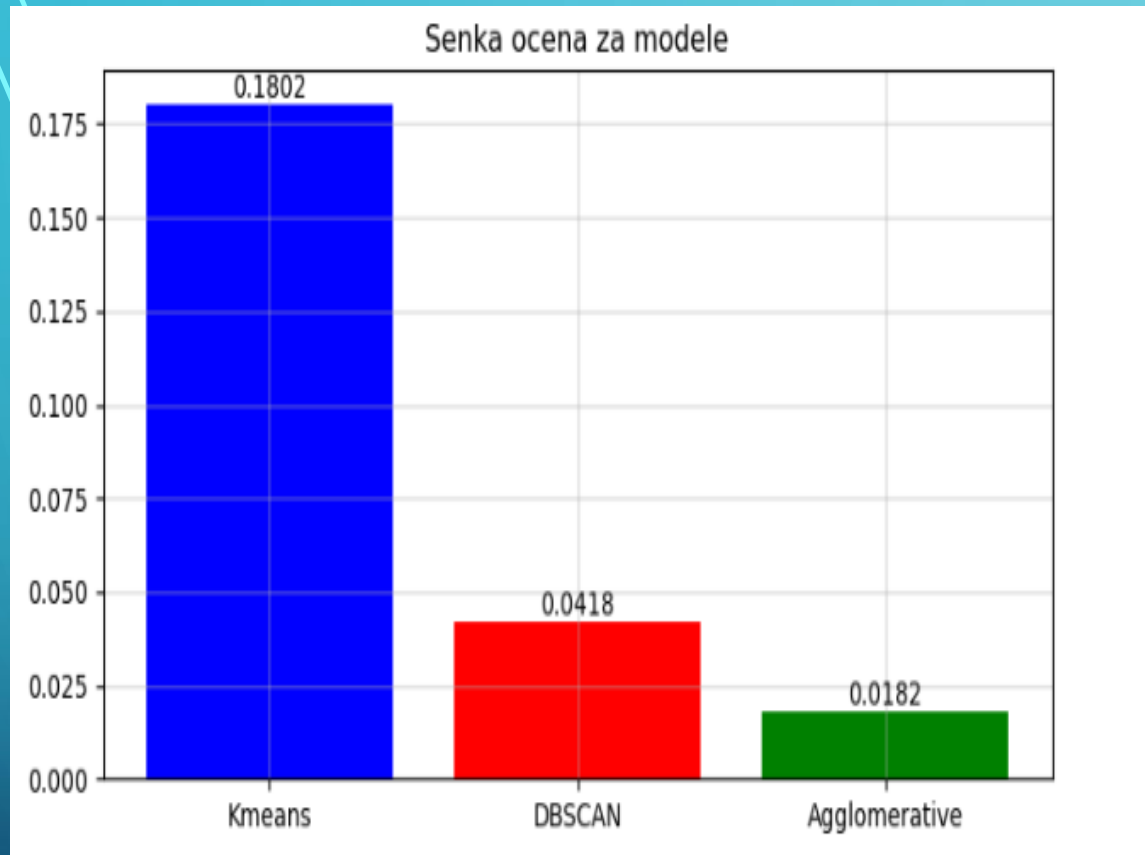


Crtanje centroida za Kmeans

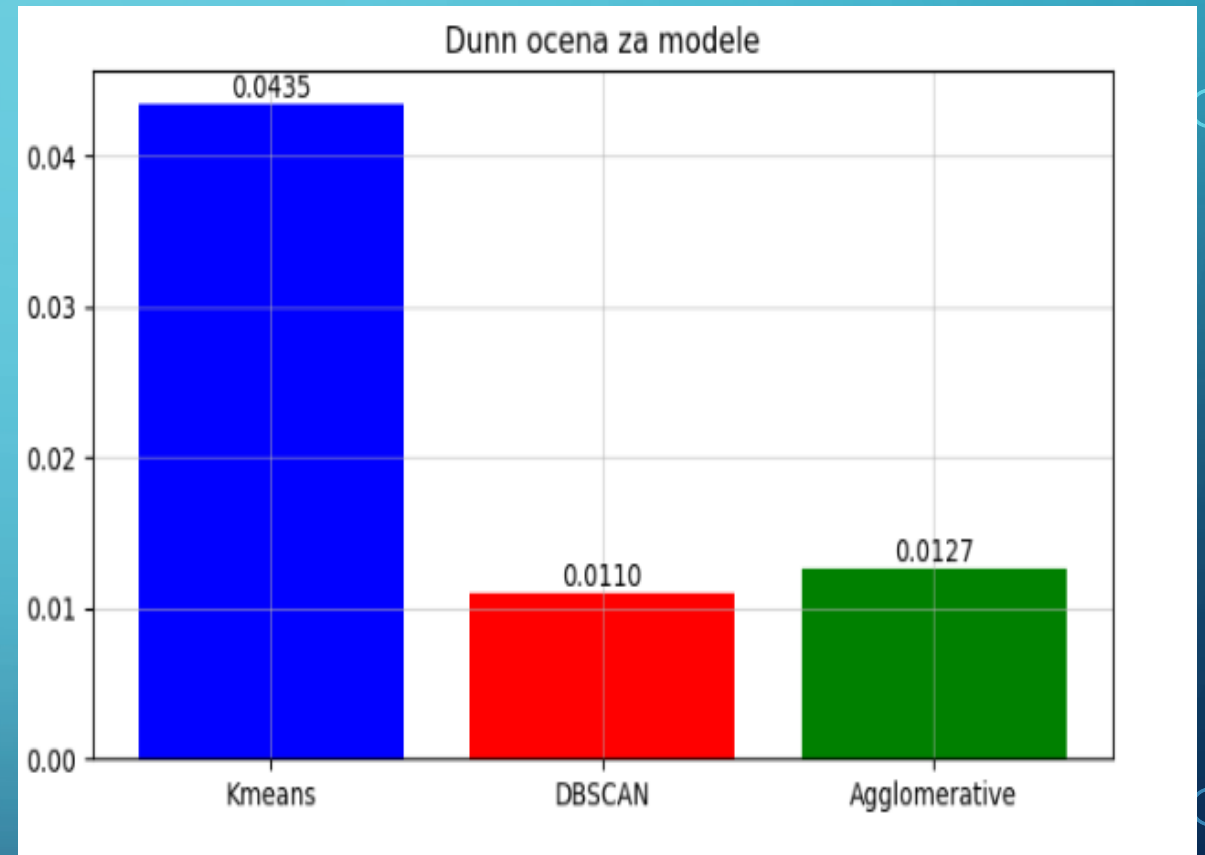


Broj klastera hijerarhijskog klasterovanja

REZULTATI

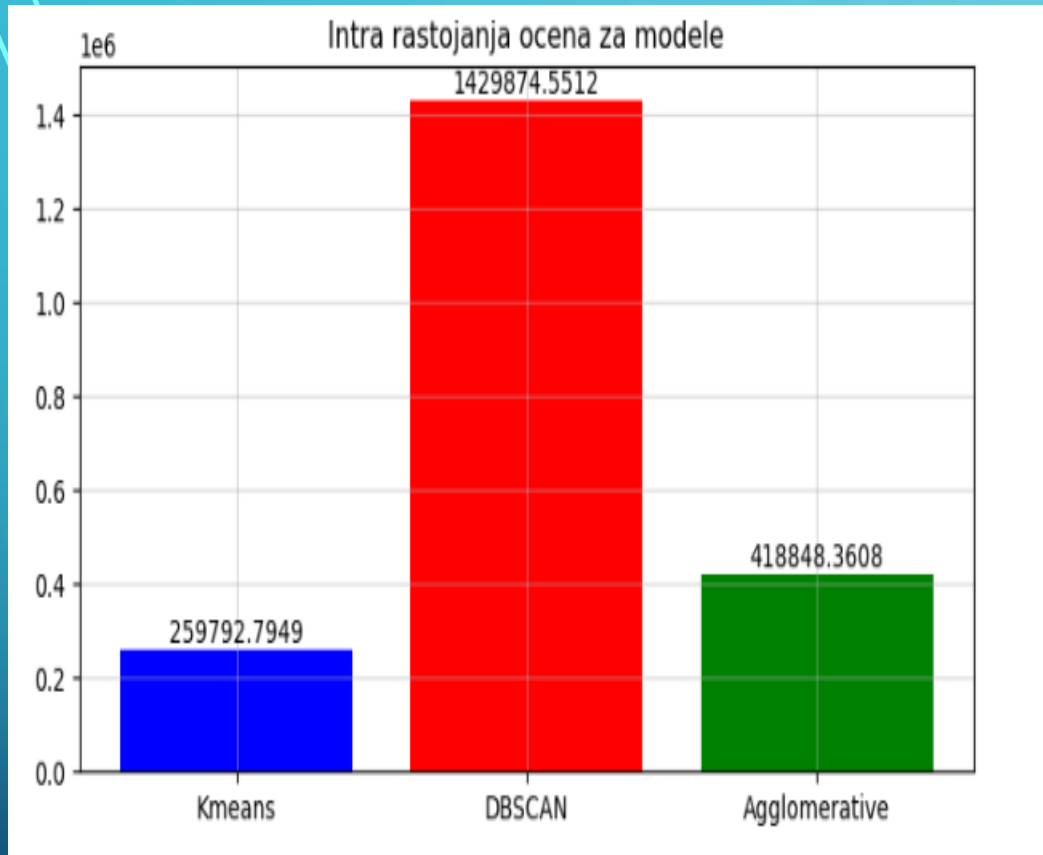


Senka ocene modela

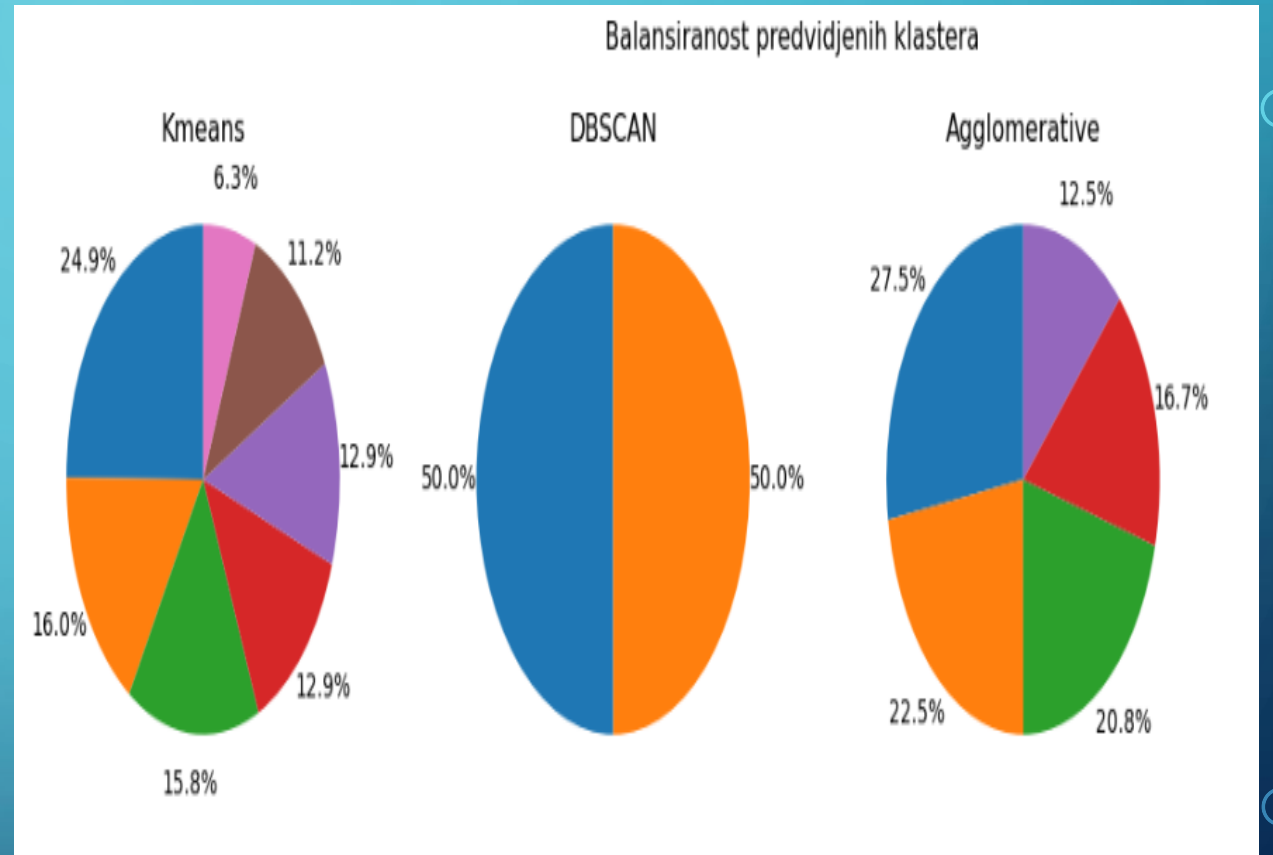


Dunn ocena modela

REZULTATI



Intra-rastojanje modela



Balansiranost klastera

PRAVILA PRIDRUŽIVANJA

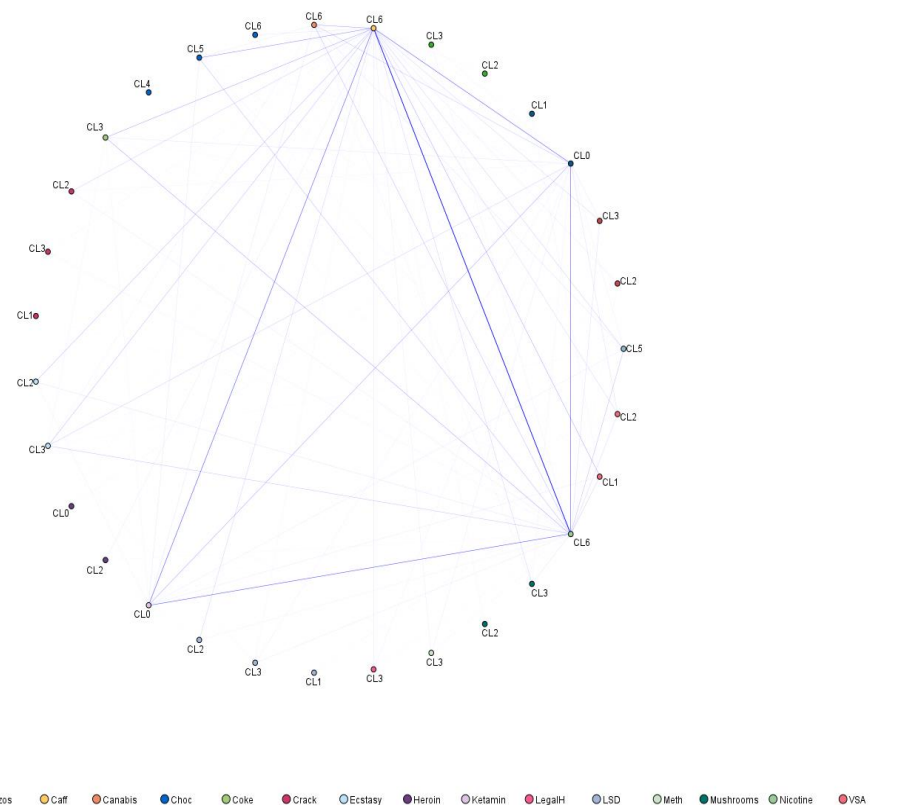
Consequent	Antecedent	Support %	Confidence %	Lift
Mushrooms = CL1	LSD = CL1 VSA = CL1	18.072	73.333	5.072
Meth = CL1	Crack = CL1 VSA = CL1	18.072	80.0	4.743
Crack = CL1	Meth = CL1 VSA = CL1	15.663	92.308	4.507
Meth = CL1	Crack = CL1	20.482	70.588	4.185
Crack = CL1	Meth = CL1	16.867	85.714	4.185
LSD = CL1	Meth = CL1	16.867	71.429	3.294
Amyl = CL1	Crack = CL1 VSA = CL1	18.072	73.333	3.204
Amyl = CL1	Crack = CL1	20.482	70.588	3.084
Coke = CL2	Canabis = CL2	15.663	76.923	2.902
LSD = CL3	Ketamin = CL3	18.072	73.333	2.536
Ecstasy = CL2	Coke = CL2 Caff = CL6	16.867	78.571	2.329
Coke = CL3	Amphet = CL3 Amyl = CL0 Nicotine = CL6	15.663	84.615	2.266
VSA = CL1	Amyl = CL1 Caff = CL6	18.072	93.333	2.213
VSA = CL1	Meth = CL1	16.867	92.857	2.202
Coke = CL3	Ecstasy = CL3 Amyl = CL0 Nicotine = CL6	18.072	80.0	2.142
VSA = CL1	Amyl = CL1	22.892	89.474	2.122

Pravila nastala samo klasama

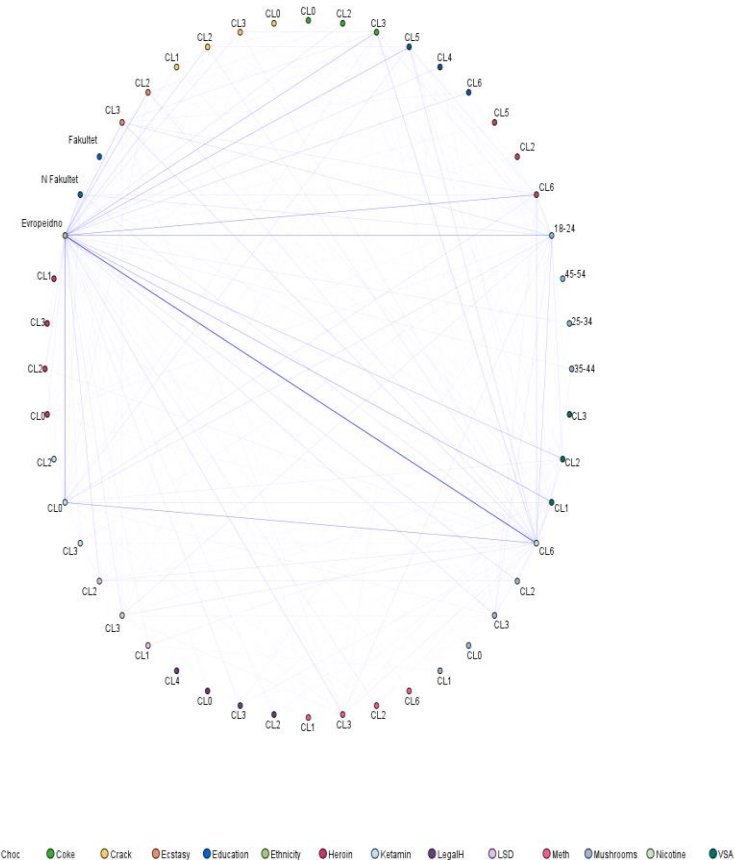
Consequent	Antecedent	Support %	Confidence %	Lift
VSA = CL2	Education = N Fakultet Age = 18-24 Country = USA Ethnicity = Evropeidno	15.663	76.923	2.28
VSA = CL1	Age = 35-44	20.482	94.118	2.232
VSA = CL1	Age = 35-44 Ethnicity = Evropeidno	18.072	93.333	2.213
VSA = CL2	Education = N Fakultet Age = 18-24 Country = USA	18.072	73.333	2.174
VSA = CL2	Education = N Fakultet Age = 18-24 Ethnicity = Evropeidno	20.482	70.588	2.092
VSA = CL1	Country = UK	21.687	77.778	1.844
VSA = CL1	Country = UK Ethnicity = Evropeidno	21.687	77.778	1.844
Canabis = CL6	Age = 18-24 Country = USA Ethnicity = Evropeidno	30.12	72.0	1.66
Amyl = CL0	Age = 18-24 Country = USA	36.145	80.0	1.581
Amyl = CL0	Age = 18-24	45.783	76.316	1.508
Amyl = CL0	Age = 18-24 Country = USA Ethnicity = Evropeidno	30.12	76.0	1.502
Ketamin = CL0	Gender Country = USA	15.663	76.923	1.485
Amyl = CL0	Education = N Fakultet Age = 18-24	22.892	73.684	1.456
Amyl = CL0	Education = N Fakultet Age = 18-24 Country = USA	18.072	73.333	1.449

Pravila nastala atributima

REZULTATI



Grafik korišćenja samo klasa



Pravila nastala atributima

The background is a blue gradient with a faint, abstract circuit board pattern. The pattern consists of white lines and circles, resembling a network or data flow, located in the corners and along the edges of the image.

PITANJA