Analiza skupa podataka CS:GO Round Winner Classification

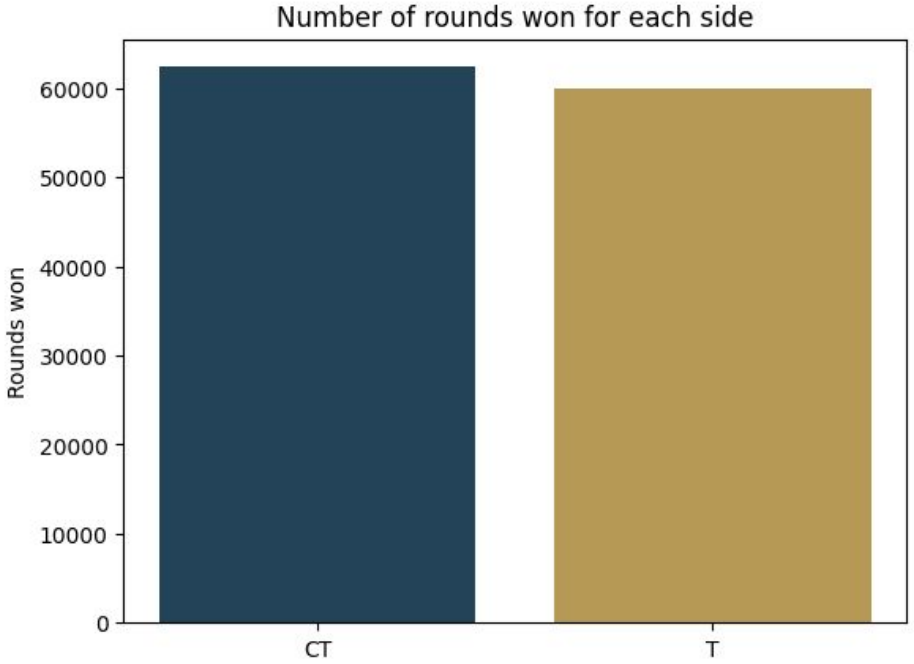
Bogdan Stojadinović
mi20073@alas.matf.bg.ac.rs

Uvod

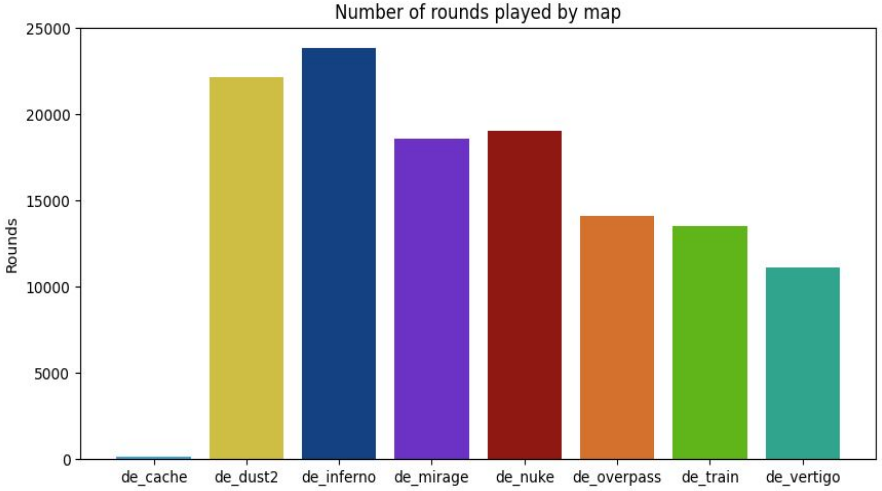
Eksplorativna analiza podataka

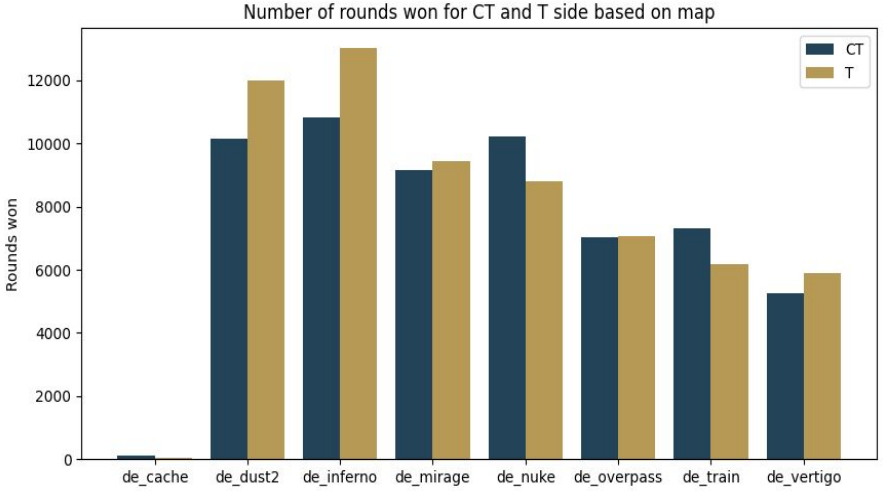
time_left	Vreme preostalo do kraja trenutne runde
ct_score	Broj rundi koje je pobedila CT strana
t_score	Broj rundi koje je pobedila T strana
map	Mapa na kojoj se igra runda
bomb_planted	Indikator da li je bomba postavljena
ct_health	Ukupan broj životnih poena na CT strani
t_health	Ukupan broj životnih poena na T strani
ct_armor	Ukupan broj oklopnih poena na CT strani
t_armor	Ukupan broj životnih poena na T strani
ct_money	Ukupna količina novca na CT strani
t_money	Ukupna količina novca na T strani
ct_helmets	Ukupan broj kaciga na CT strani
t_helmets	Ukupan broj kaciga na T strani
ct_defuse_kits	Ukupan broj kompleta za deaktiviranje bombe na CT strani
ct_players_alive	Ukupan broj živih igrača na CT strani
t_players_alive	Ukupan broj živih igrača na T strani
ct_weapon_X	Ukupan broj za svako od oružja X na CT strani
t_weapon_X	Ukupan broj za svako od oružja X na T strani
ct_grenade_X	Ukupan broj za svaku od granata X na CT strani
t_grenade_X	Ukupan broj za svaku od granata X na T strani
round_winner	Pobednik runde

Eksplorativna analiza podataka

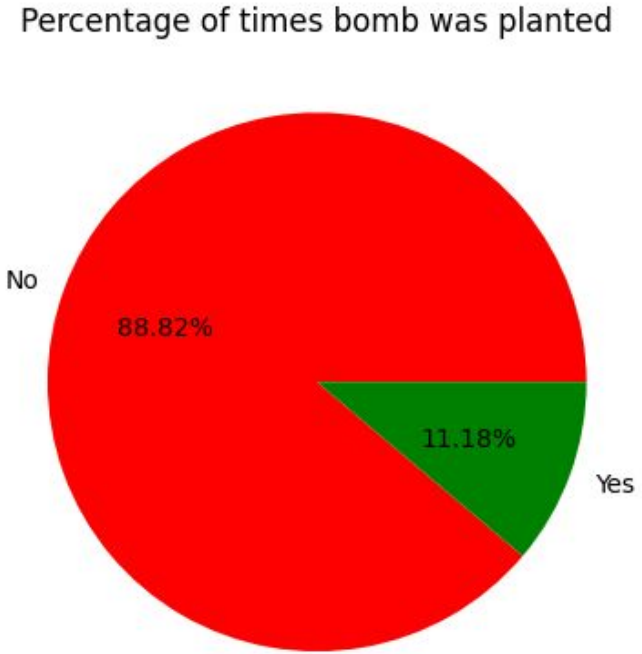


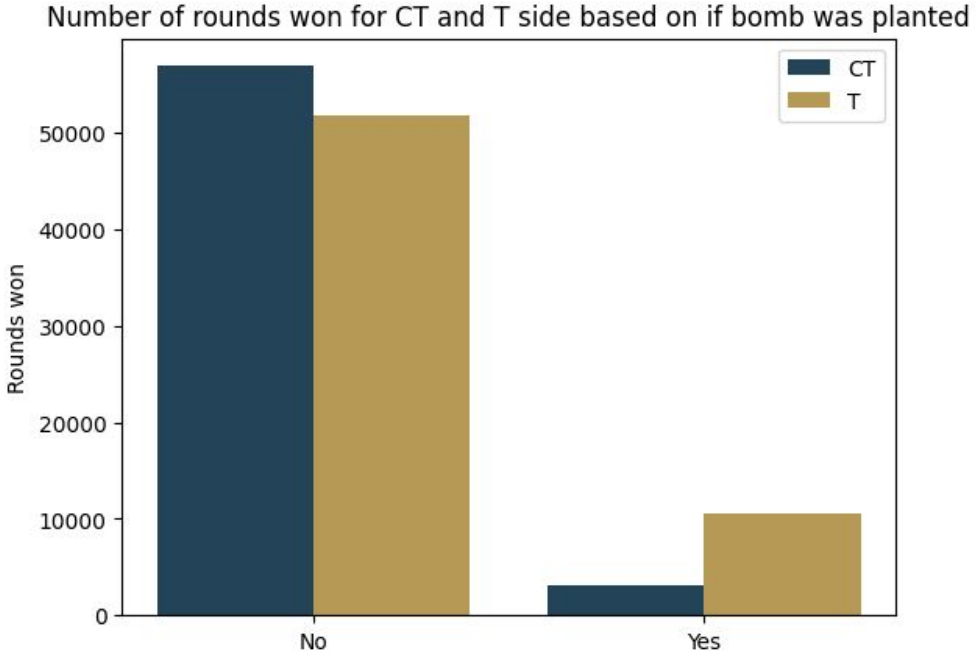
Eksplorativna analiza podataka



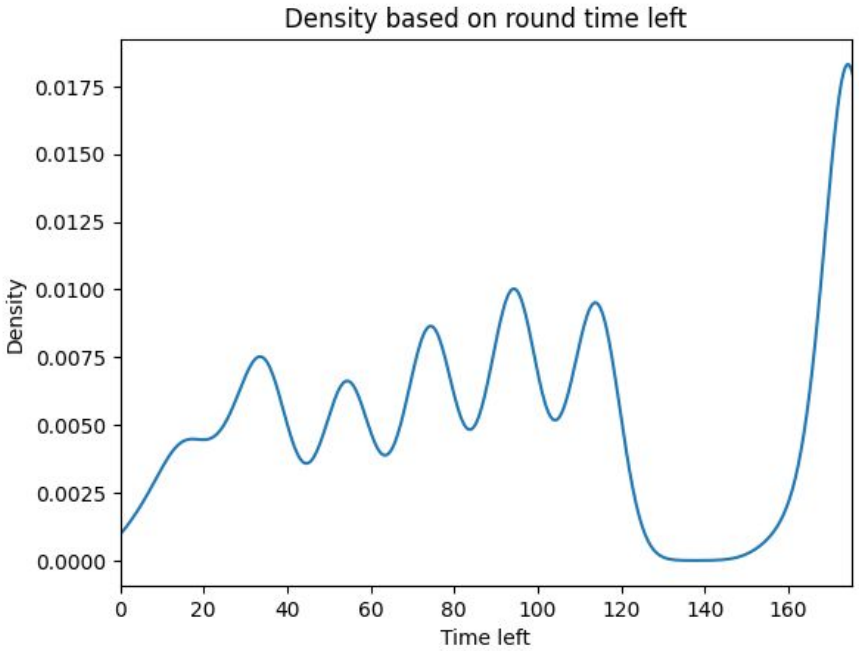


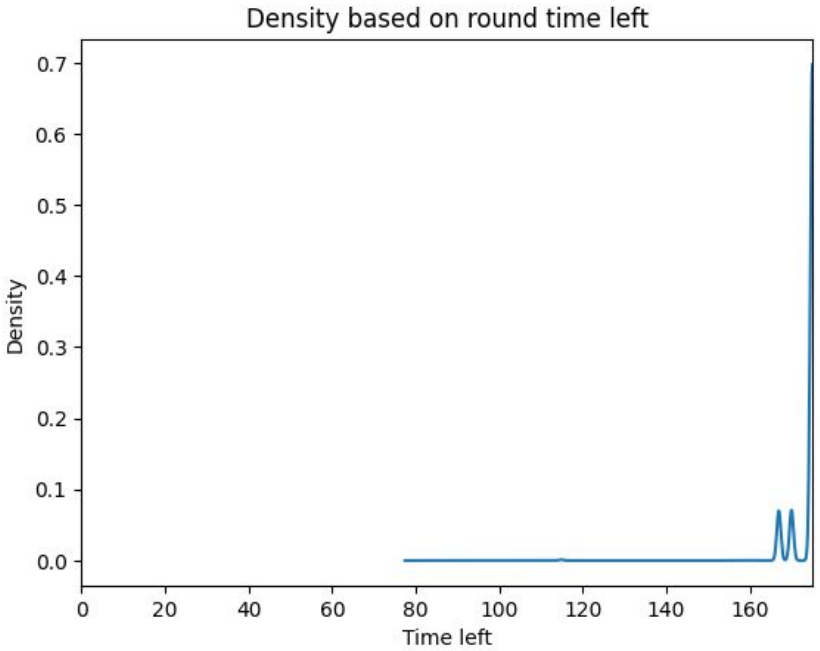
Eksplorativna analiza podataka



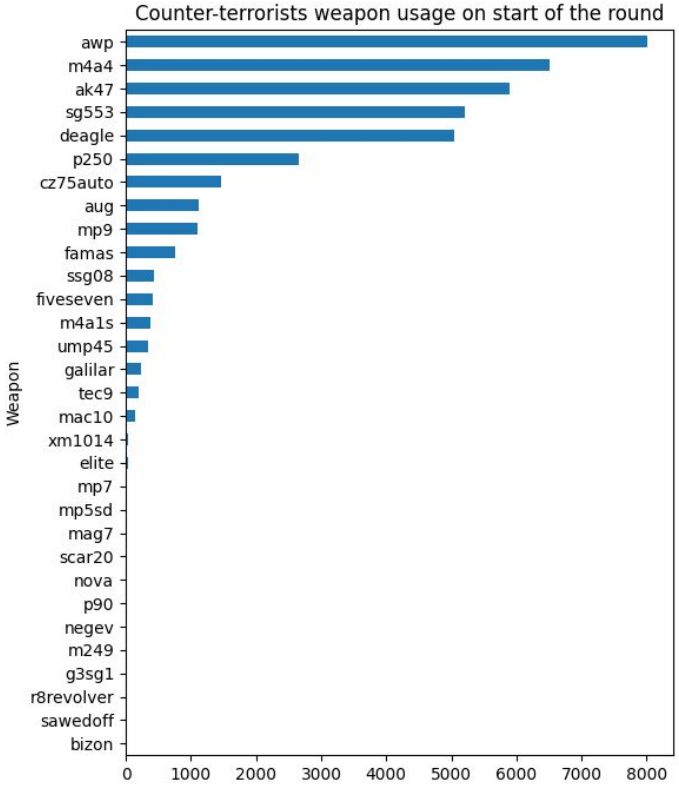


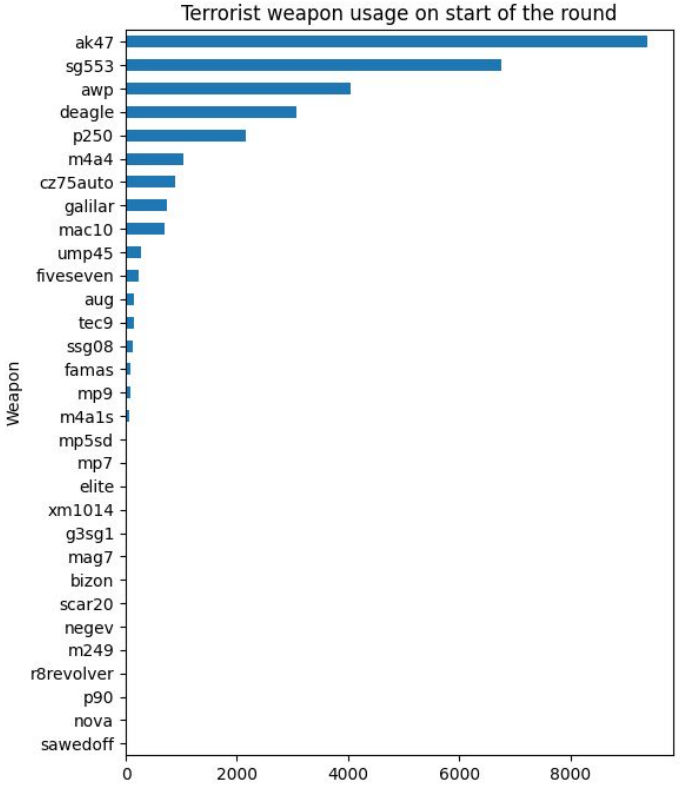
Eksplorativna analiza podataka





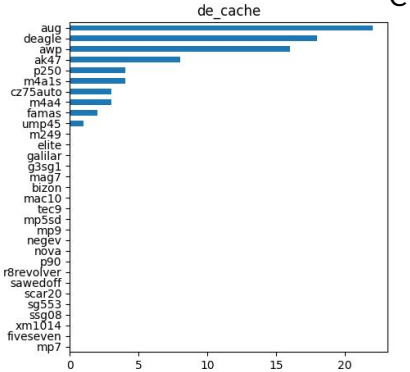
Eksplorativna analiza podataka

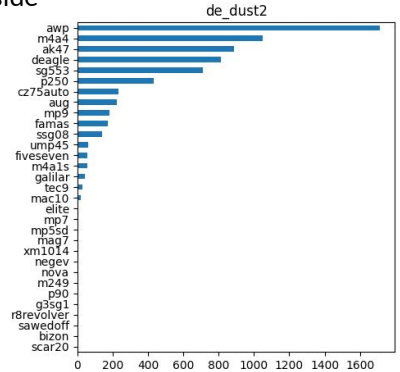


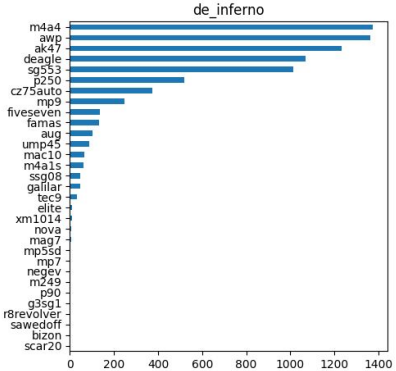


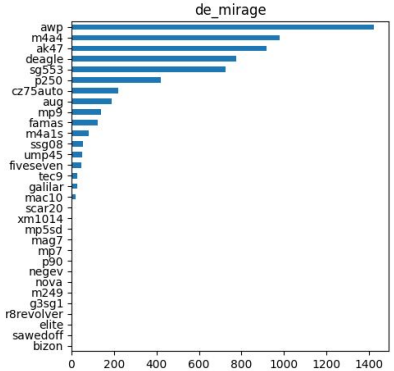
Eksplorativna analiza podataka

CT side

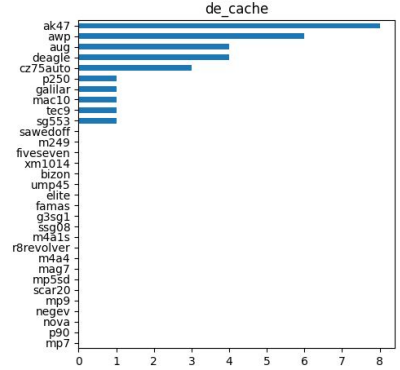


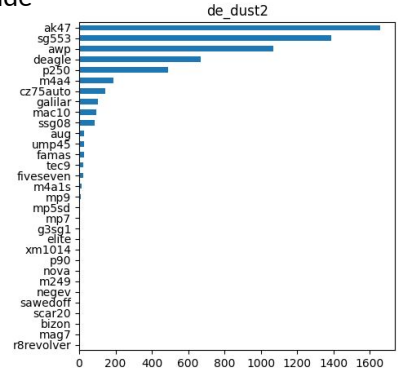


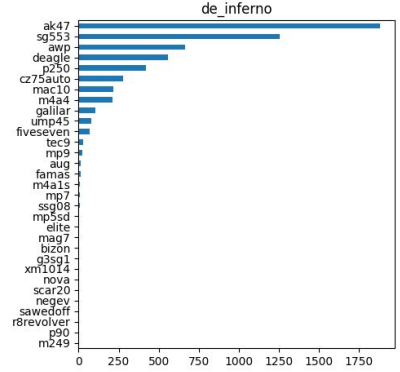


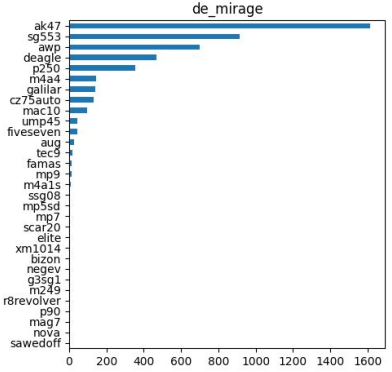


T side









Pretprocesiranje podataka

- Nedostajuće vrednosti
- Elementi van granica, autlajeri
- Podela na ulazne i ciljne atribute
- Provera balansiranosti klasa
- Enkodiranje kategoričkih atributa
- Normalizacija
- Podela na trening i test skupove (klasifikacija)
- Čuvanje podataka za dalju upotrebu

Klasifikacija – Decision Trees

DecisionTreeClassifier()

Train data:

Confusion matrix:
[[47918 114]
[181 49697]]
Accuracy score: 0.9969870289040956
Precision score: 0.9977113488988376
Recall score: 0.9963711455952524
F1 score: 0.9970407968782915

Test data:

Confusion matrix:
[[9812 2152]
[2118 10396]]
Accuracy score: 0.8255576435983332
Precision score: 0.8284985655084476
Recall score: 0.8307495604922487
F1 score: 0.829622536110446

DecisionTreeClassifier()

GridSearchCV()

Confusion matrix:
[[47687 345]
[221 49657]]
Accuracy score: 0.9942191808804004
Precision score: 0.9931002759889604
Recall score: 0.9955691888207225
F1 score: 0.9943331998398078

Confusion matrix:
[[9776 2188]
[2077 10437]]
Accuracy score: 0.8257619086526677
Precision score: 0.8266930693069307
Recall score: 0.8340258910020777
F1 score: 0.8303432913003699

Klasifikacija – Random Forest

`RandomForestClassifier()`

Train data:	Test data:
Confusion matrix:	Confusion matrix:
[[47880 152]	[[10598 1366]
[143 49735]]	[1626 10888]]
Accuracy score: 0.9969870289040956	Accuracy score: 0.8777677914862325
Precision score: 0.9969531140377252	Precision score: 0.8885261955279908
Recall score: 0.9971330045310558	Recall score: 0.8700655266101965
F1 score: 0.99704305117025	F1 score: 0.8791989664082687

`RandomForestClassifier()`

`GridSearchCV()`

Confusion matrix:	Confusion matrix:
[[47877 155]	[[10641 1323]
[140 49738]]	[1558 10956]]
Accuracy score: 0.9969870289040956	Accuracy score: 0.8823024756924586
Precision score: 0.9968933517727938	Precision score: 0.8922550696310775
Recall score: 0.9971931512891455	Recall score: 0.8754994406264983
F1 score: 0.9970432289943971	F1 score: 0.8837978461662567

Klasifikacija – KNN

KNeighborsClassifier()

Train data:

Confusion matrix:
[[42912 5120]
[5445 44433]]
Accuracy score: 0.8920947809212542
Precision score: 0.8966762859968115
Recall score: 0.8908336340671238
F1 score: 0.8937454113908138

Test data:

Confusion matrix:
[[9816 2148]
[2288 10226]]
Accuracy score: 0.8187760437944277
Precision score: 0.8264102149668661
Recall score: 0.8171647754514944
F1 score: 0.8217614914818387

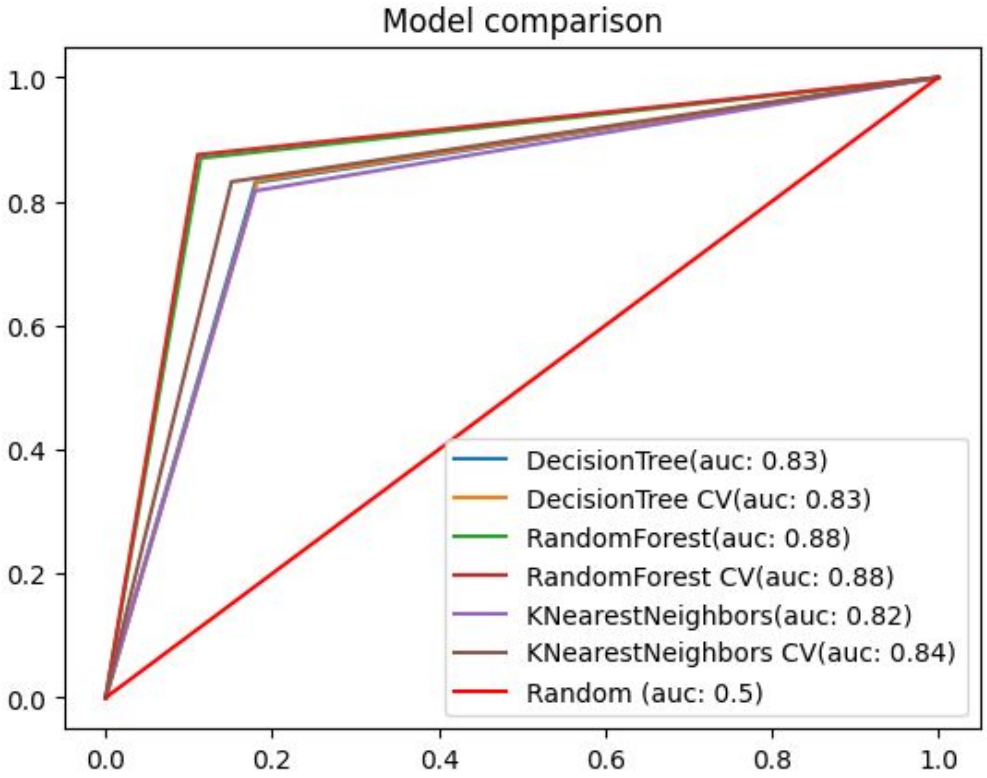
KNeighborsClassifier()

GridSearchCV()

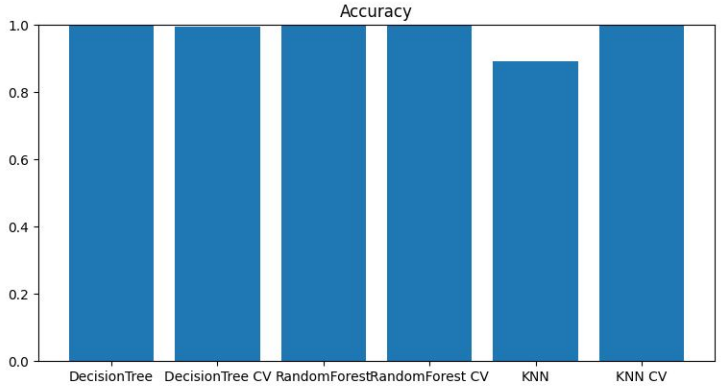
Confusion matrix:
[[47932 100]
[197 49681]]
Accuracy score: 0.9969666019814115
Precision score: 0.9979912014624054
Recall score: 0.9960503628854405
F1 score: 0.9970198376463741

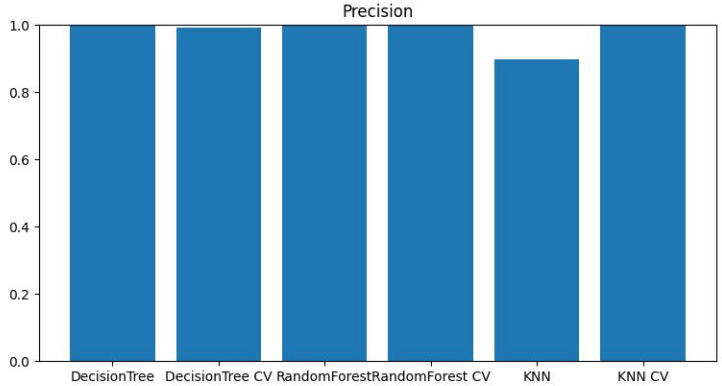
Confusion matrix:
[[10156 1808]
[2104 10410]]
Accuracy score: 0.8401830214886837
Precision score: 0.8520216074643968
Recall score: 0.8318683074956049
F1 score: 0.8418243571081999

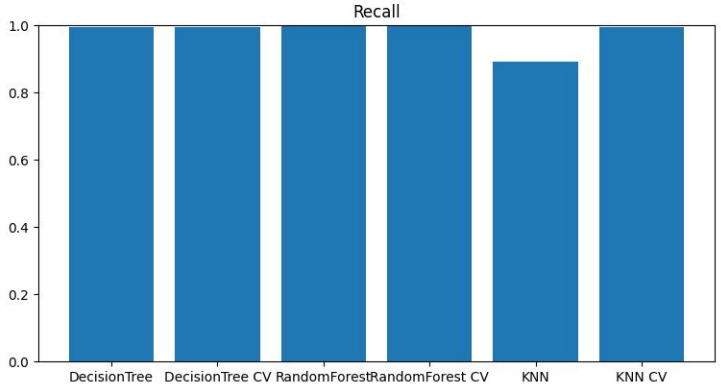
Poređenje modela – ROC Curve

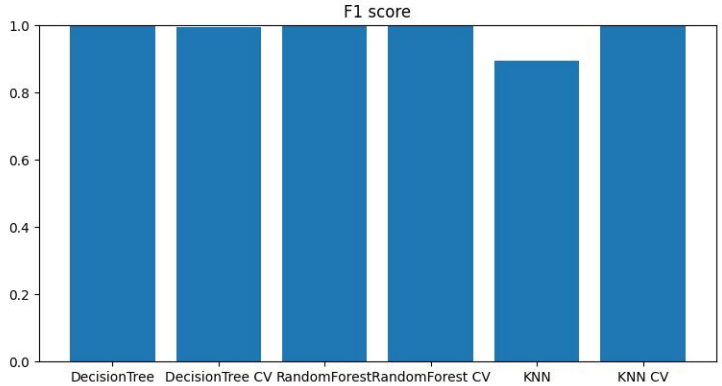


Poređenje modela – Train

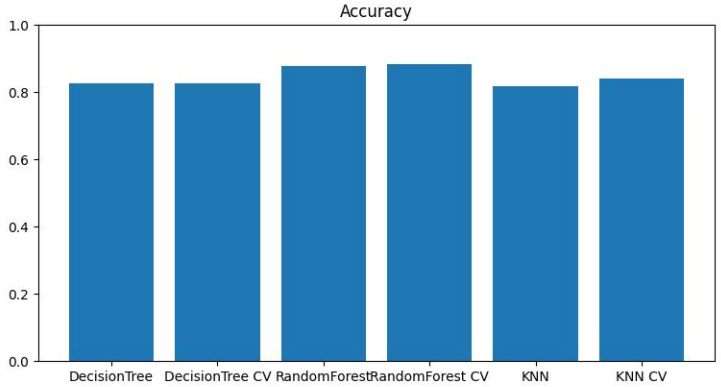


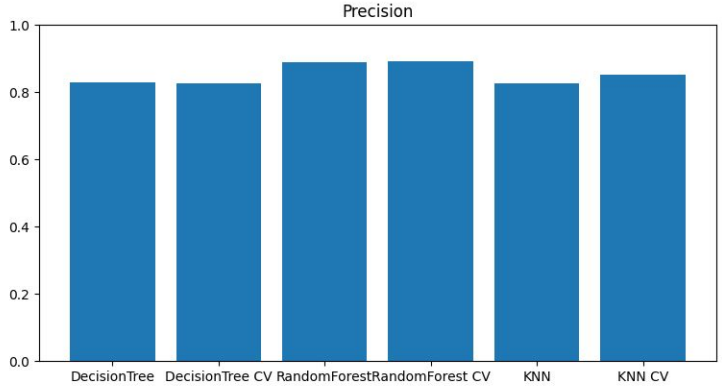


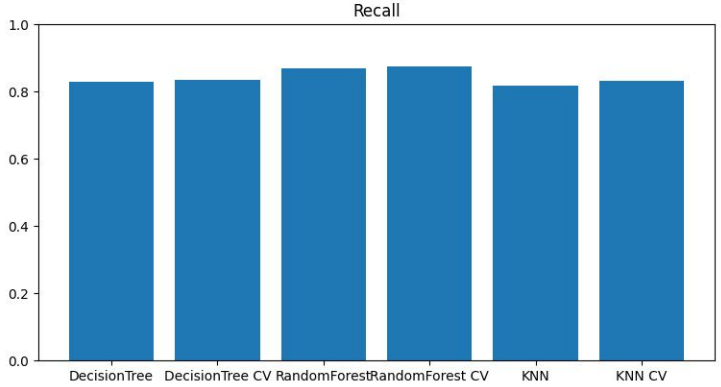


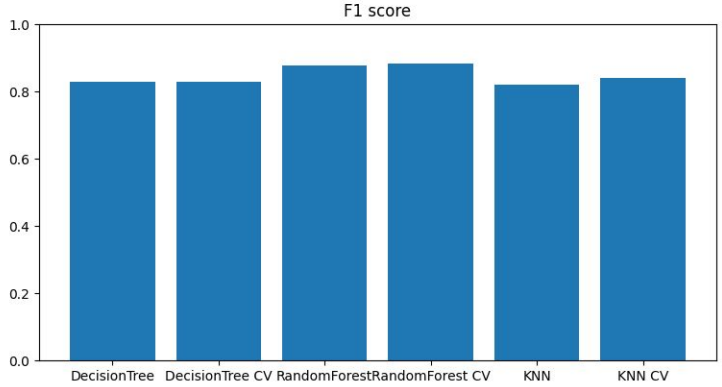


Poređenje modela – Test

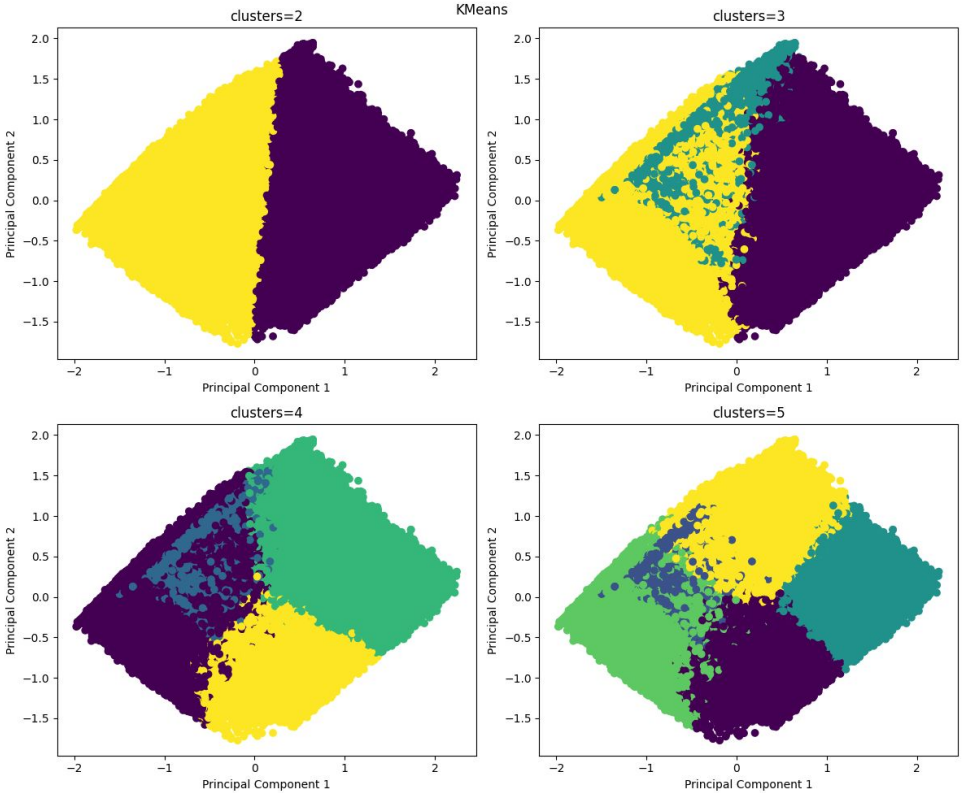




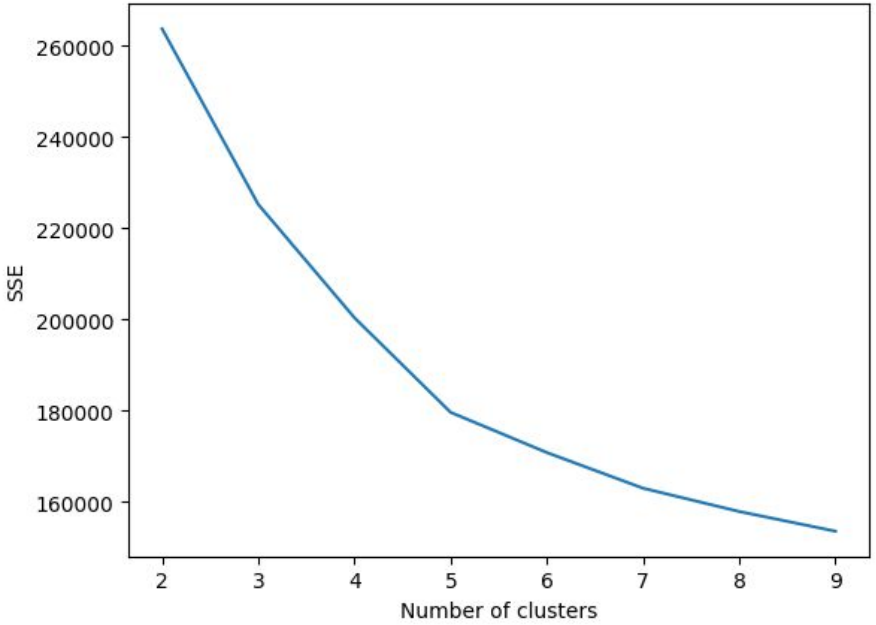


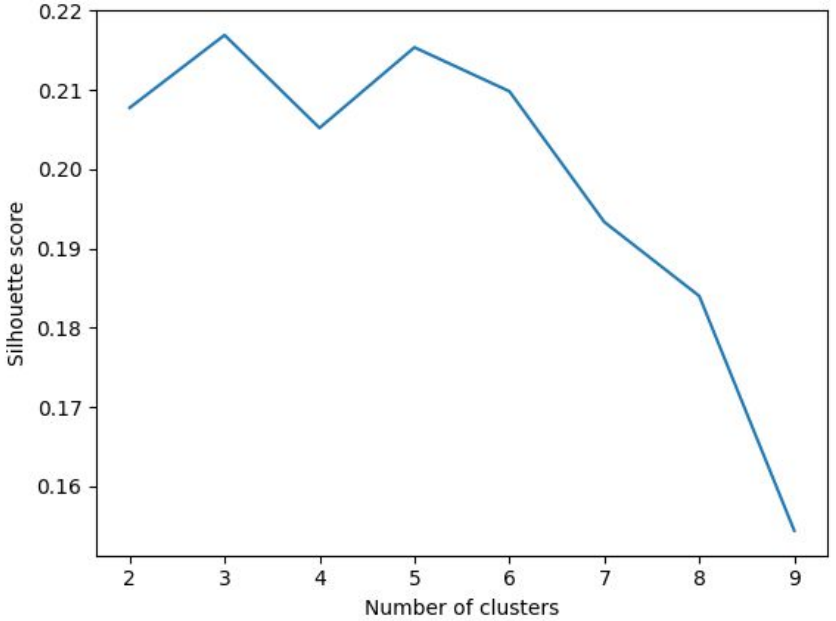


Klasterovanje – K-means

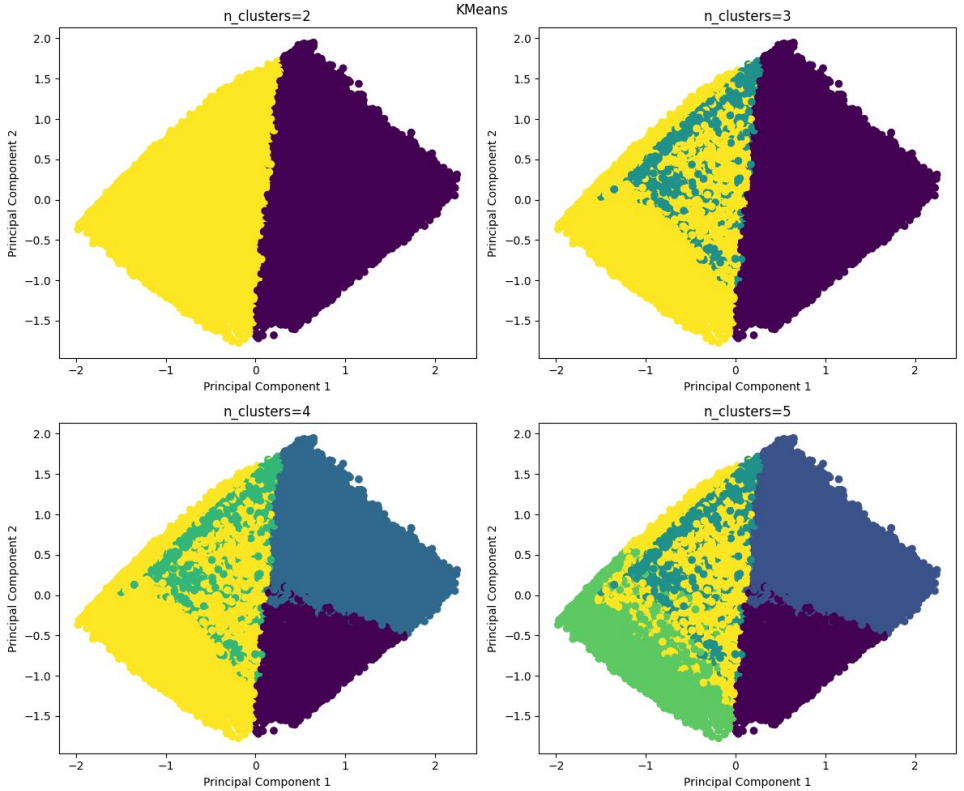


Klasterovanje – K-means

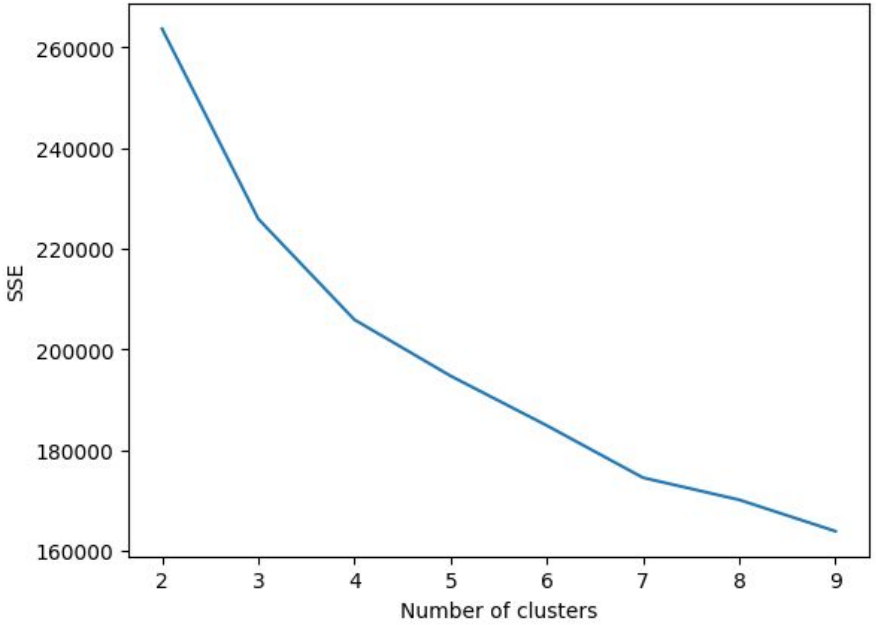


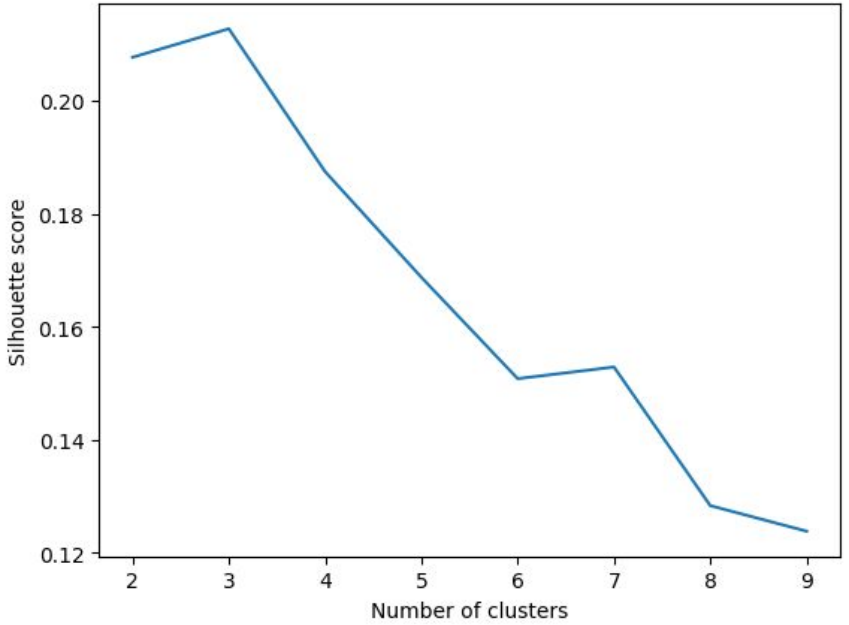


Klasterovanje – Bisecting K-means

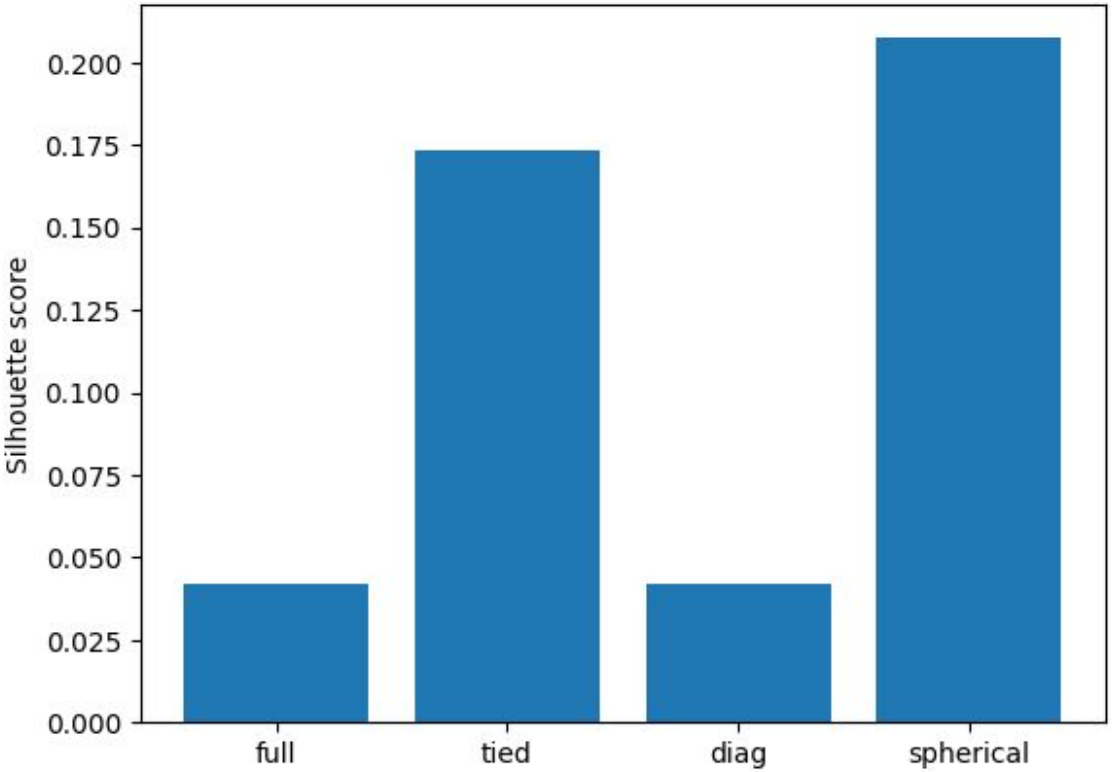


Klasterovanje – Bisecting K-means

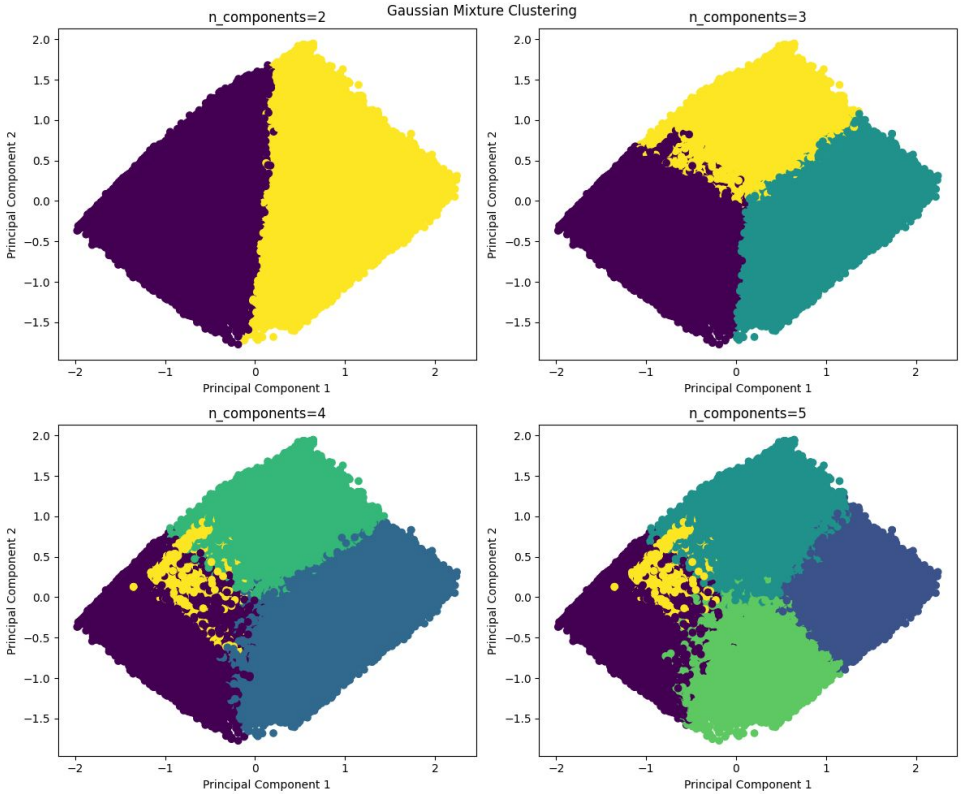




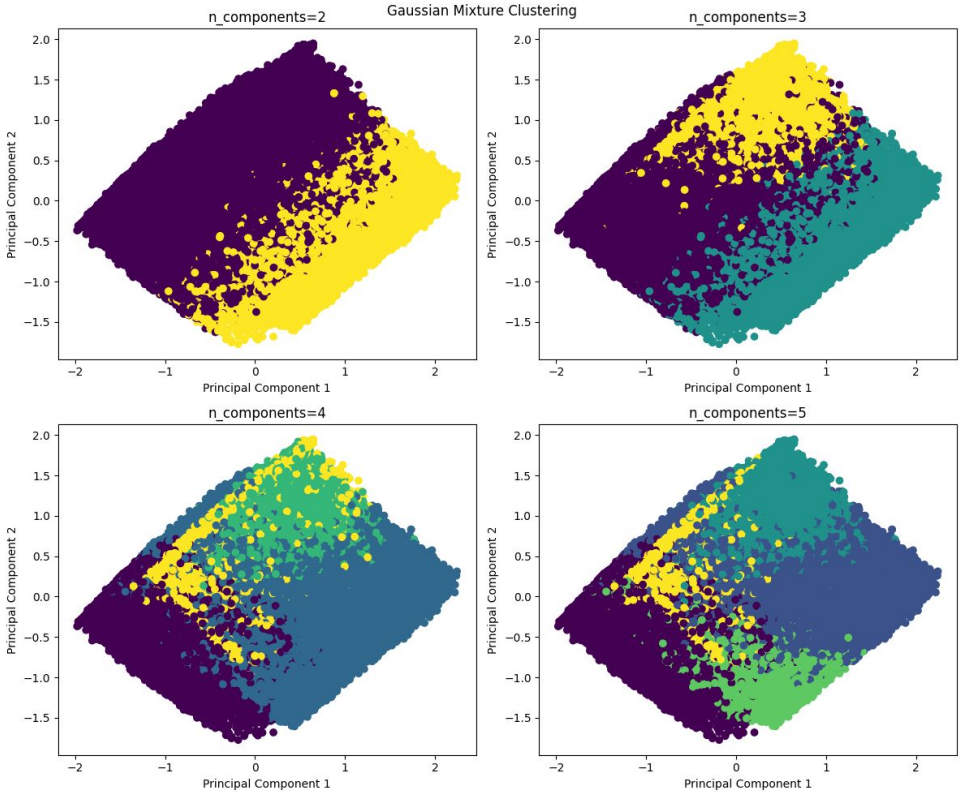
Klasterovanje – Gaussian Mixture



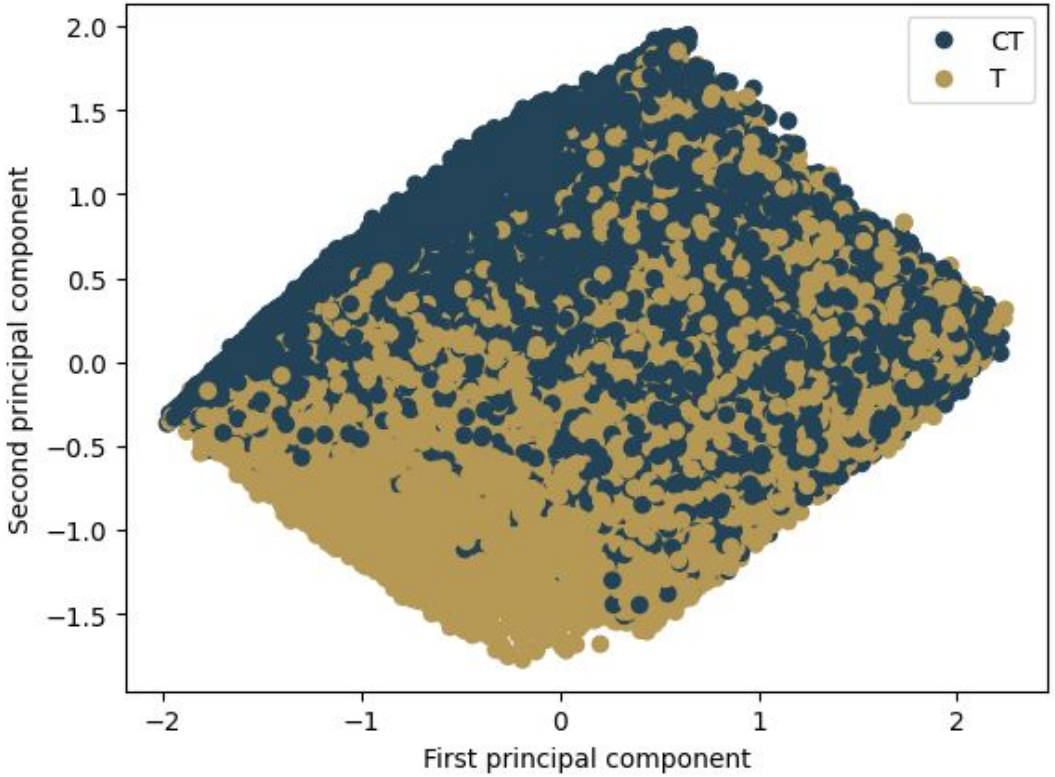
Klasterovanje – Gaussian Mixture (spherical)



Klasterovanje – Gaussian Mixture (tied)

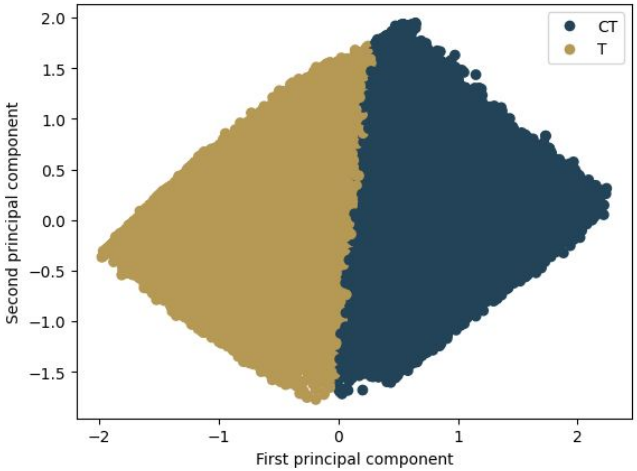


Poređenje modela

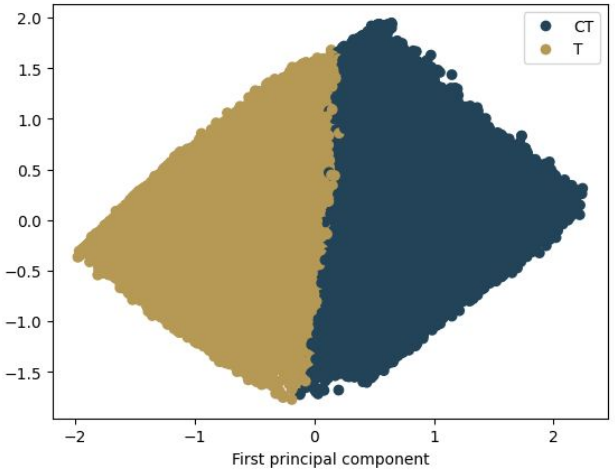


Poređenje modela

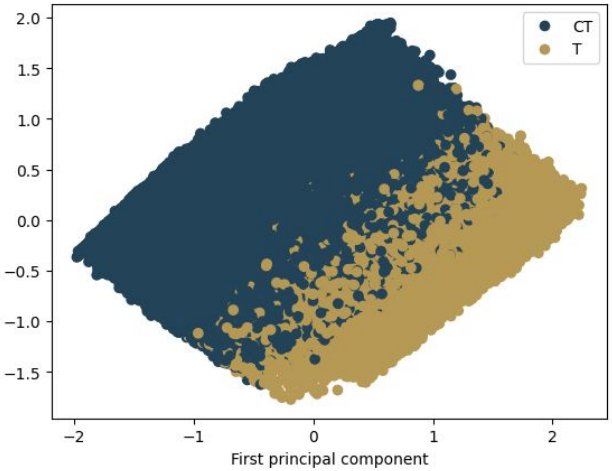
K-means



GMM spherical



GMM tied



Pravila pridruživanja – Apriori

Sort by: Lift 4127 of 4127

Consequent	Antecedent	Support %	Confidence %	Lift
ct_grenade_incen...	ct_grenade_hegre... ct_weapon_awp ct_weapon_m4a4 ct_grenade_smok... ct_defuse_kits	22,526	89,824	2,061
ct_grenade_incen...	ct_grenade_hegre... ct_weapon_awp ct_weapon_m4a4 ct_grenade_smok... ct_helmets	23,068	89,202	2,046
ct_grenade_incen...	ct_grenade_hegre... ct_weapon_awp ct_weapon_m4a4 ct_grenade_smok... ct_grenade_flash...	23,888	88,985	2,041
ct_grenade_incen...	ct_grenade_hegre... ct_weapon_awp ct_grenade_smok... ct_defuse_kits ct_grenade_flash...	25,393	88,984	2,041
ct_grenade_incen...	ct_grenade_hegre... ct_weapon_awp ct_grenade_smok... ct_defuse_kits ct_helmets	25,003	88,842	2,038
ct_grenade_incen...	ct_weapon_p250 ct_grenade_hegre... ct_grenade_smok... ct_helmets	10,01	88,737	2,036

Sort by: Lift 1480 of 1480

Consequent	Antecedent	Support %	Confidence %	Lift
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_weapon_ak47 t_grenade_smoke... t_grenade_flashb...	10,72	96,388	1,861
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_grenade_smoke... t_grenade_flashb... t_helmets	12,202	95,99	1,853
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_grenade_smoke... t_grenade_flashb... t_armor	12,203	95,983	1,853
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_grenade_smoke... t_grenade_flashb...	12,22	95,862	1,851
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_weapon_ak47 t_grenade_smoke... t_armor	10,93	95,725	1,848
t_grenade_moloto...	t_grenade_hegre... t_weapon_awp t_weapon_ak47 t_grenade_smoke... t_helmets	10,928	95,724	1,848

Zaključak i pitanja

- Nimalo naivan zadatak
- Ljudski faktor
- Nepredvidivost rundi
- Više sreće drugi put – Counter Strike 2?

Hvala na pažnji