Istraživanje skupa podataka “Anuran Calls”

Stefanija Marković 306/2018

Osnovno o skupu “Anuran Calls”

- Kreiran segmentacijom 60 audio zapisa zvukova žaba.
- Žabe su iz četiri različite porodice, osam rodova i deset vrsti.
- Nakon segmentacije dobijeno je 7195 slogova.
- Za svaki slog izračunato je 22 kepstralnih koeficijenata mel skale (MFCC).

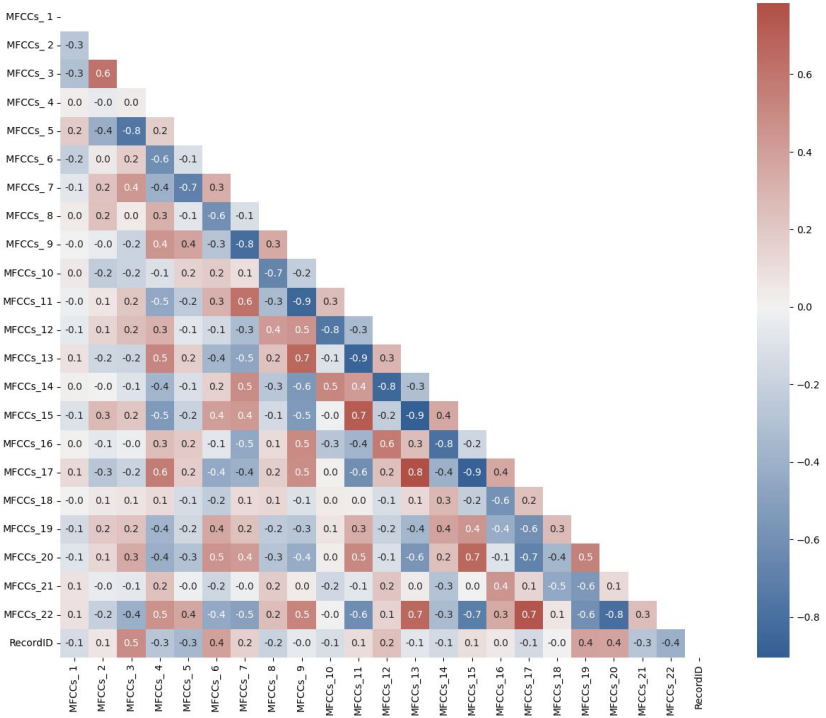
Eksploracijska analiza podataka

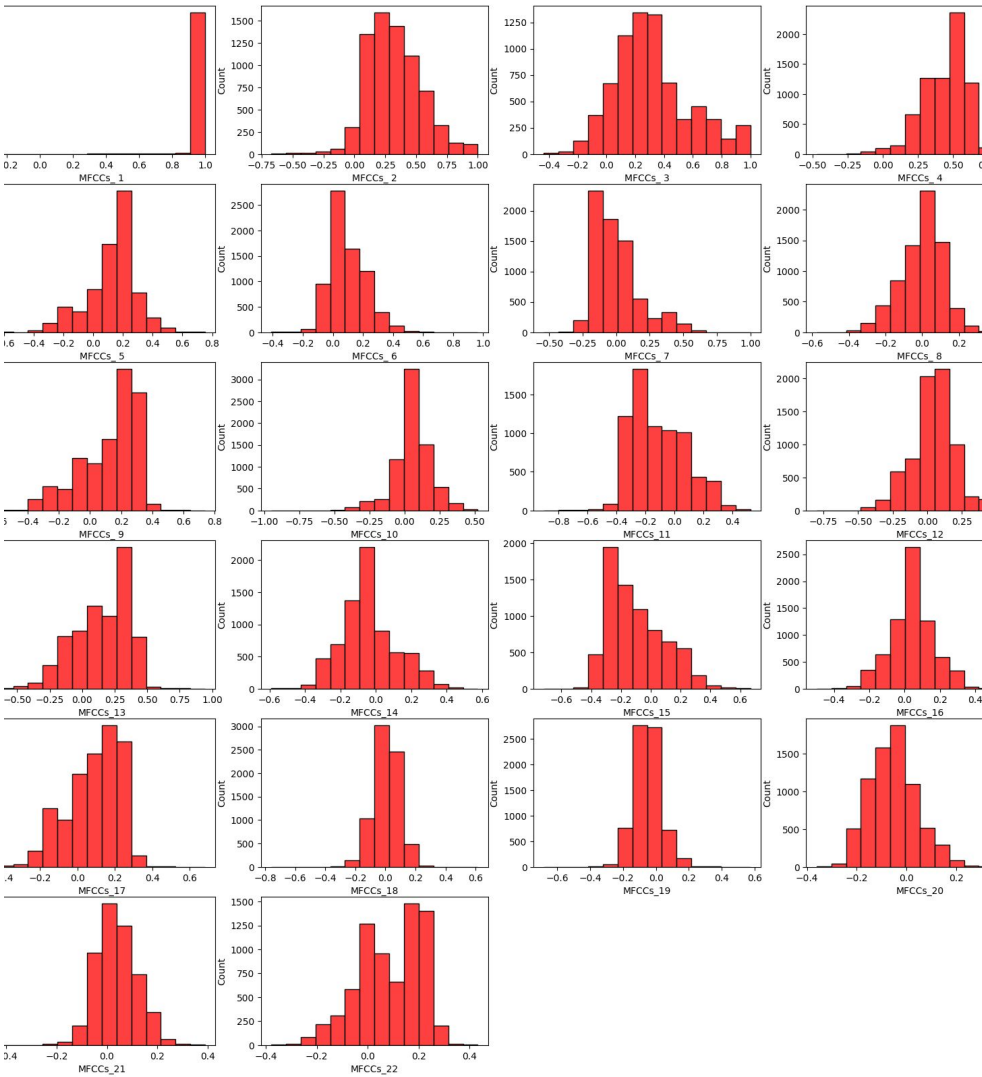
	MFCCs_ 1	MFCCs_ 2	MFCCs_ 3	MFCCs_ 4	MFCCs_ 5	MFCCs_ 6	MFCCs_ 7	MFCCs_ 8	MFCCs_ 9	MFCCs_10	...	MFCCs_17	MFCCs_18	MFCCs_19	MFCCs_20	MFCCs_21	MFCCs_22	Family	Genus	Species	RecordID
0	1.0	0.152936	-0.105586	0.200722	0.317201	0.260764	0.100945	-0.150063	-0.171128	0.124676	...	-0.108351	-0.077623	-0.009568	0.057684	0.118680	0.014038	Leptodactylidae	Adenomera	AdenomeraAndre	1
1	1.0	0.171534	-0.098975	0.268425	0.338672	0.268353	0.060835	-0.222475	-0.207693	0.170883	...	-0.090974	-0.056510	-0.035303	0.020140	0.082263	0.029056	Leptodactylidae	Adenomera	AdenomeraAndre	1
2	1.0	0.152317	-0.082973	0.287128	0.276014	0.189867	0.008714	-0.242234	-0.219153	0.232538	...	-0.050691	-0.023590	-0.066722	-0.025083	0.099108	0.077162	Leptodactylidae	Adenomera	AdenomeraAndre	1
3	1.0	0.224392	0.118985	0.329432	0.372088	0.361005	0.015501	-0.194347	-0.098181	0.270375	...	-0.136009	-0.177037	-0.130498	-0.054766	-0.018691	0.023954	Leptodactylidae	Adenomera	AdenomeraAndre	1
4	1.0	0.087817	-0.068345	0.306967	0.330923	0.249144	0.006884	-0.265423	-0.172700	0.266434	...	-0.048885	-0.053074	-0.088550	-0.031346	0.108610	0.079244	Leptodactylidae	Adenomera	AdenomeraAndre	1

5 rows × 26 columns

	MFCCs_ 1	MFCCs_ 2	MFCCs_ 3	MFCCs_ 4	MFCCs_ 5	MFCCs_ 6	MFCCs_ 7	MFCCs_ 8	MFCCs_ 9	MFCCs_10	...	MFCCs_14	MFCCs_15	MFCCs_16	MFCCs_17	MFCCs_18	MFCCs_19	MFCCs_20	MFCCs_21	MFCCs_22
count	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	...	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000	7195.000000
mean	0.989885	0.323584	0.311224	0.445997	0.127046	0.097939	-0.001397	-0.000370	0.128213	0.055998	...	-0.039244	-0.101748	0.042062	0.088680	0.007755	-0.049474	-0.053244	0.037313	0.087567
std	0.069016	0.218653	0.263527	0.160328	0.162722	0.120412	0.171404	0.116302	0.179008	0.127099	...	0.152515	0.187618	0.119915	0.138055	0.084733	0.082546	0.094181	0.079470	0.123442
min	-0.251179	-0.673025	-0.436028	-0.472676	-0.636012	-0.410417	-0.538982	-0.576506	-0.587313	-0.952266	...	-0.590380	-0.717156	-0.498675	-0.421480	-0.759322	-0.680745	-0.361649	-0.430812	-0.379304
25%	1.000000	0.165945	0.138445	0.336737	0.051717	0.012581	-0.125737	-0.063109	0.004648	-0.001132	...	-0.132980	-0.255929	-0.019549	-0.001764	-0.042122	-0.106079	-0.120971	-0.017620	0.000533
50%	1.000000	0.302184	0.274626	0.481463	0.161361	0.072079	-0.052630	0.013265	0.189317	0.063478	...	-0.050715	-0.143259	0.041081	0.112769	0.011820	-0.052626	-0.055180	0.031274	0.105373
75%	1.000000	0.466566	0.430695	0.559861	0.222592	0.175957	0.085580	0.075108	0.265395	0.117725	...	0.039157	0.017348	0.107046	0.201932	0.061889	0.006321	0.001342	0.089619	0.194819
max	1.000000	1.000000	1.000000	1.000000	0.752246	0.964240	1.000000	0.551762	0.738033	0.522768	...	0.575749	0.668924	0.670700	0.681157	0.614064	0.574209	0.467831	0.389797	0.432207

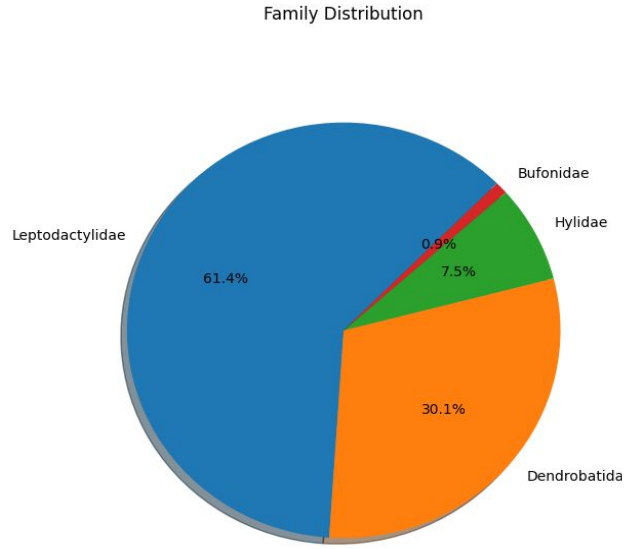
Eksplorativna analiza podataka

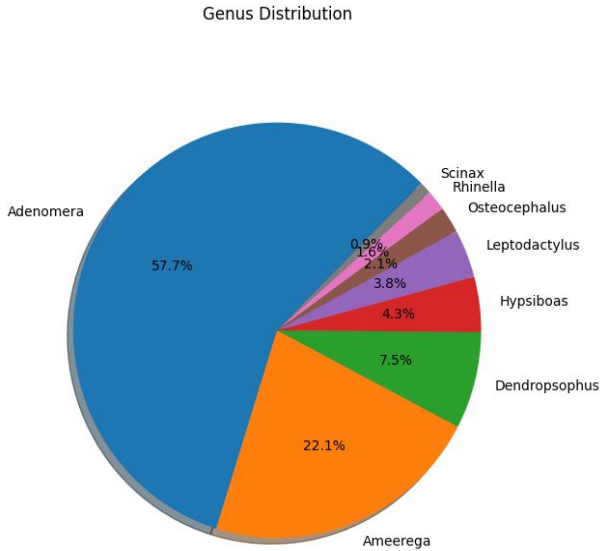


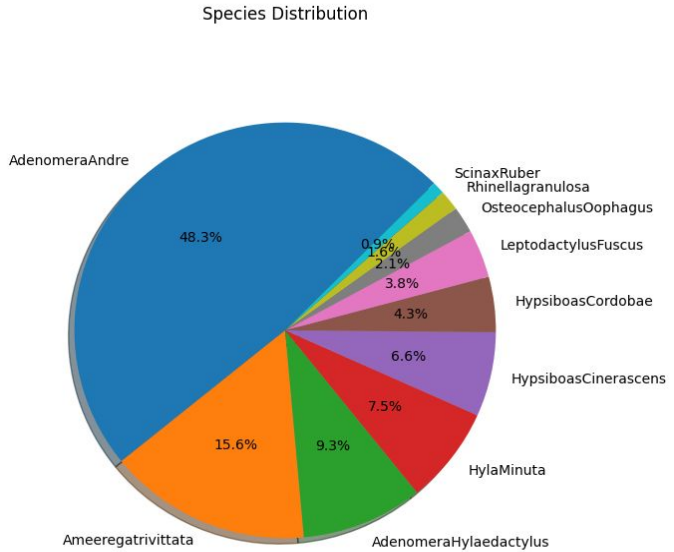


Eksplorativna analiza skupa podataka

Raspoređenost jedinki

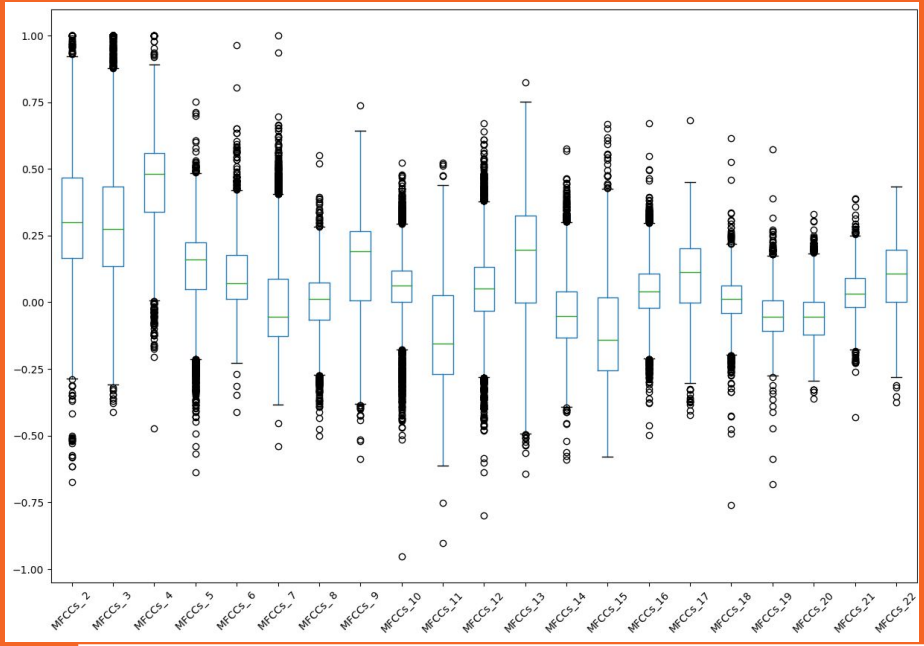


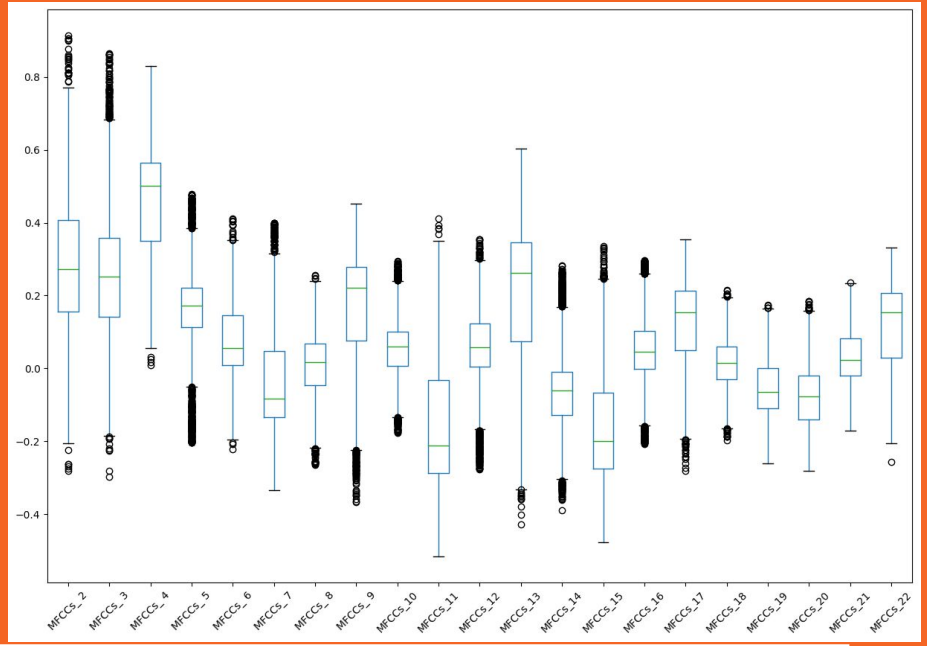




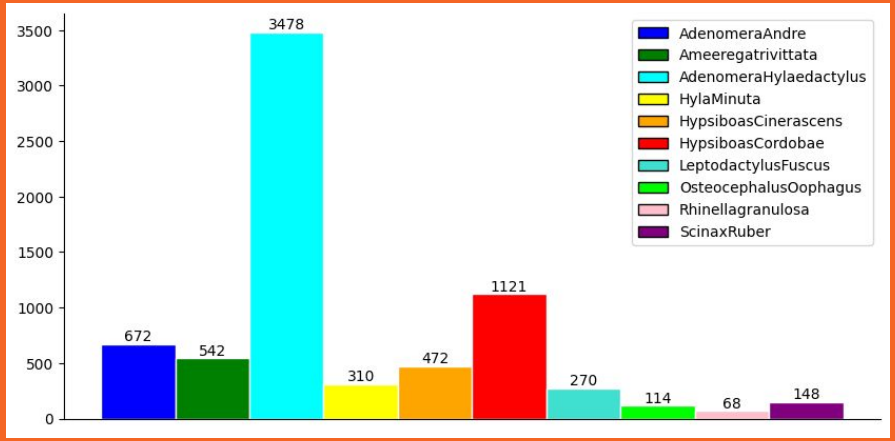
Klasifikacija

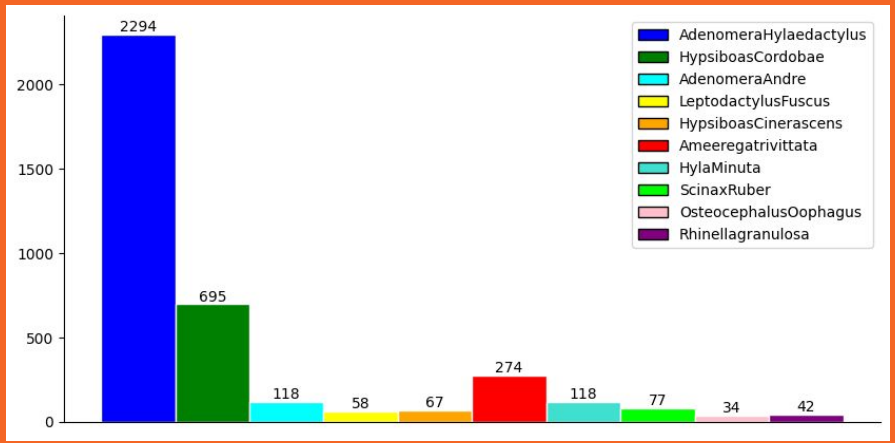
Preptrocesiranje - rad sa elementima van granica





Balansiranje



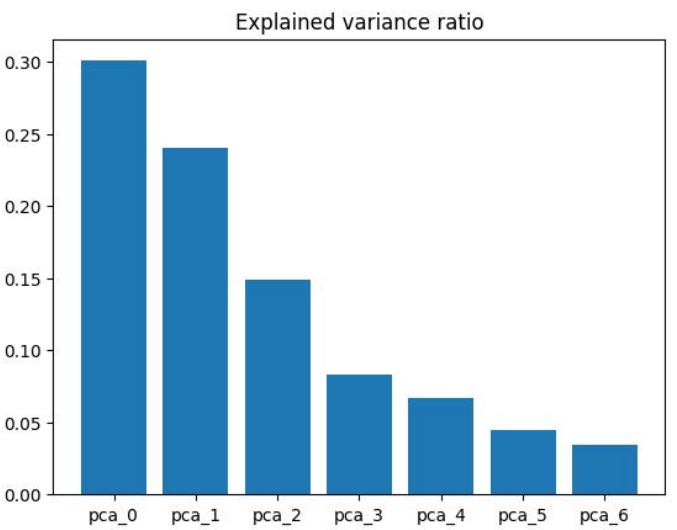


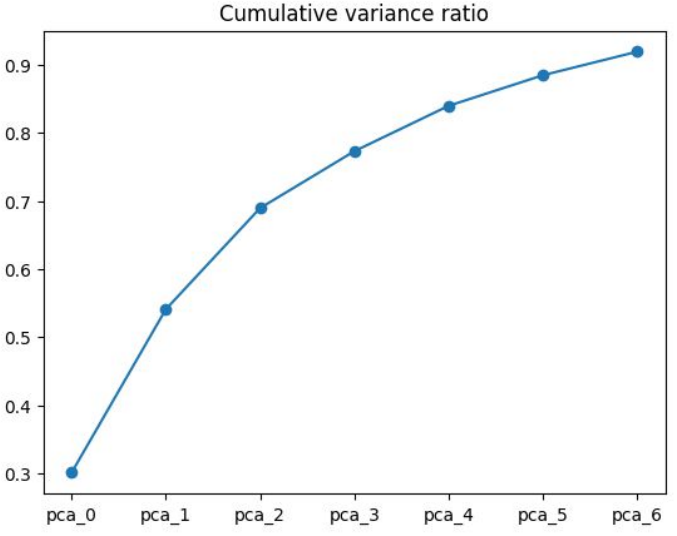
Redukcija dimenzionalnosti

Eliminisani su atributi ***Species***, ***Family***, ***Genus***, **RecordID** i **MFCC_1**.

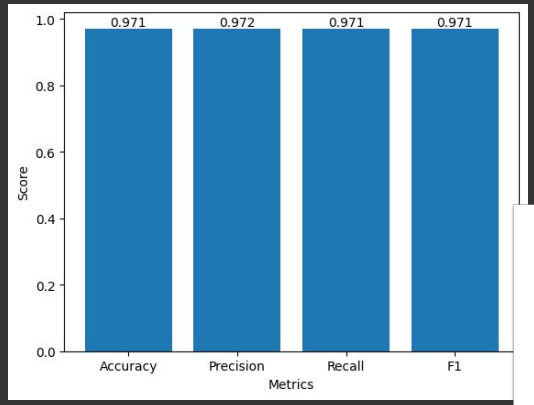
Primenjena je tehnika **PCA** (Principal component analysis).

Na prvom grafiku je predstavljeno koliko varijanse svaki atribut opisuje, dok je na drugom predstavljena njihova kumulativna suma.



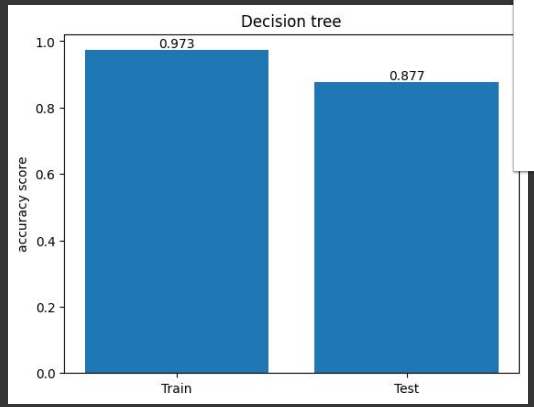


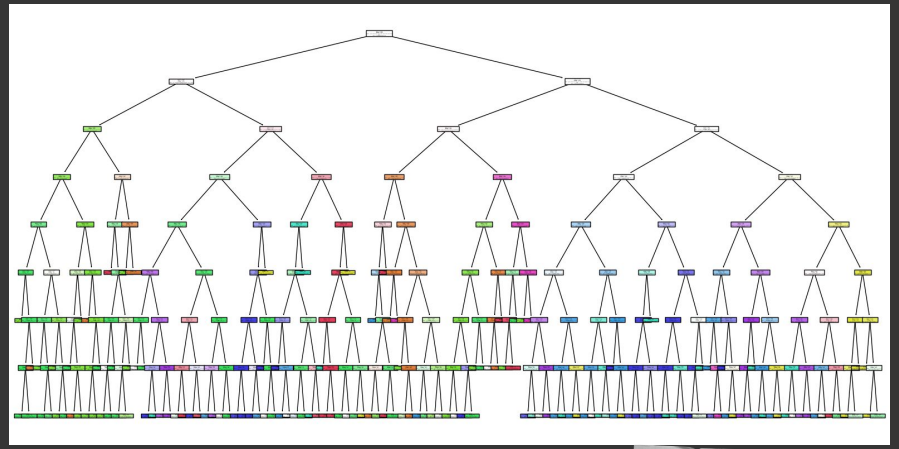
Stablo odlučivanja

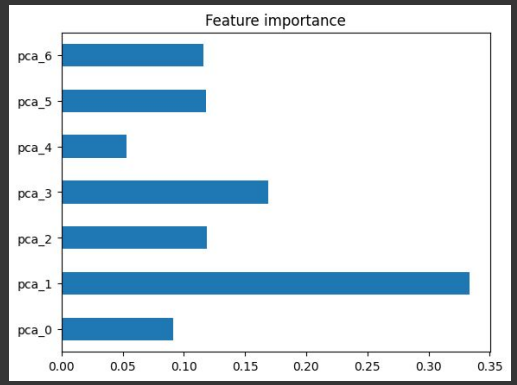


Gornja slika pokazuje rezultate različitih **mera**.

Donja slika pokazuje **tačnost** na trening i test skupu.



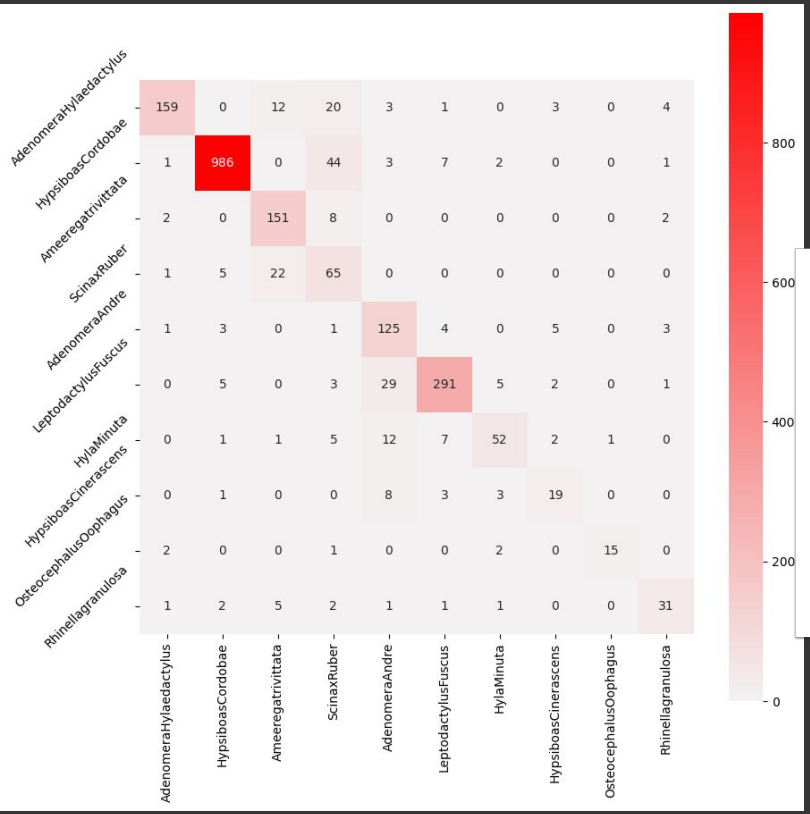




Gore je prikazano kreirano **stablo odlučivanja**.

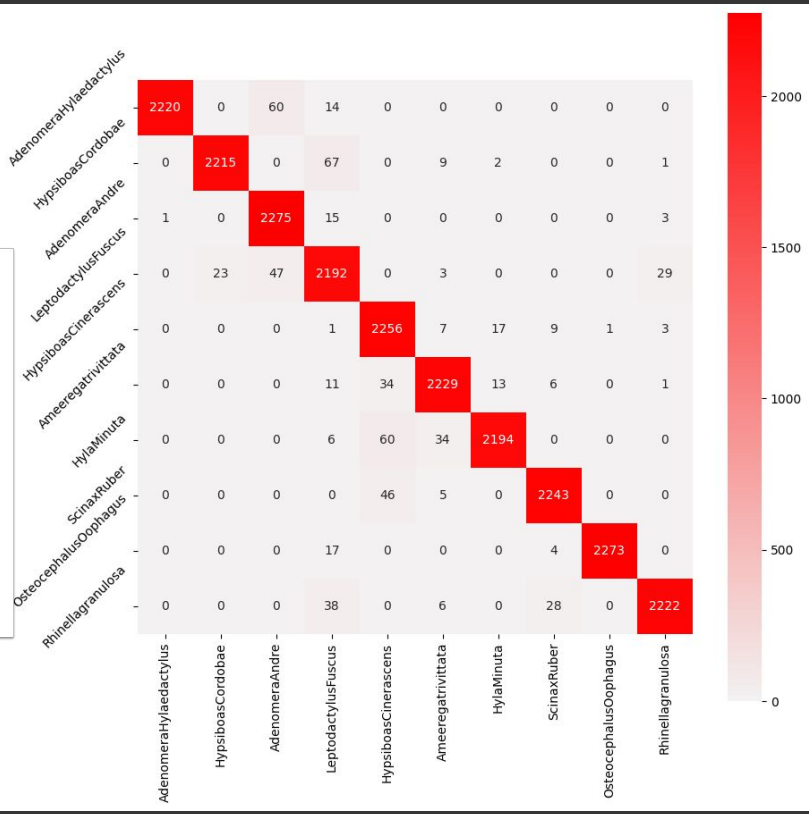
Levo vidimo **važnost atributa** pri odlučivanju.

Stablo odlučivanja

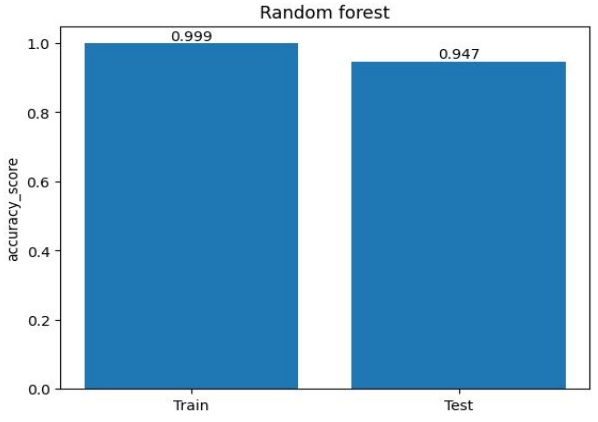


Matrica konfuzije

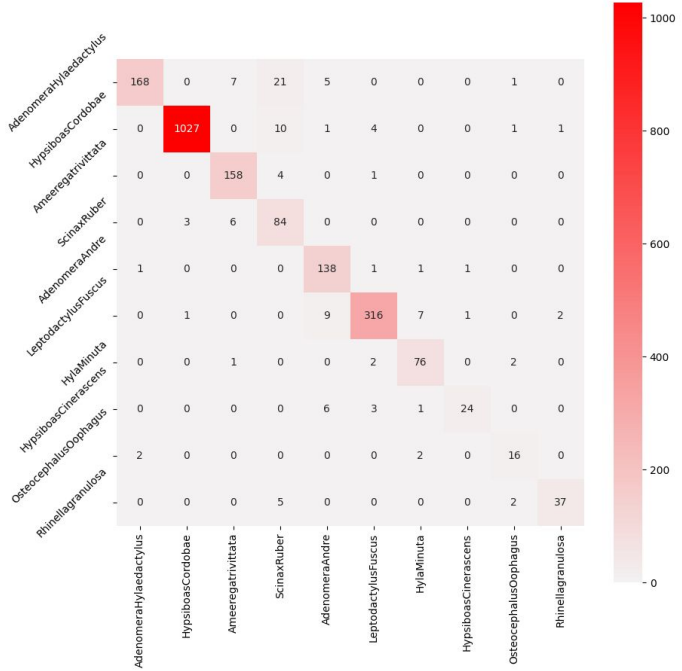
Leva se odnosi na **test** skup, a desna na **trening** skup.

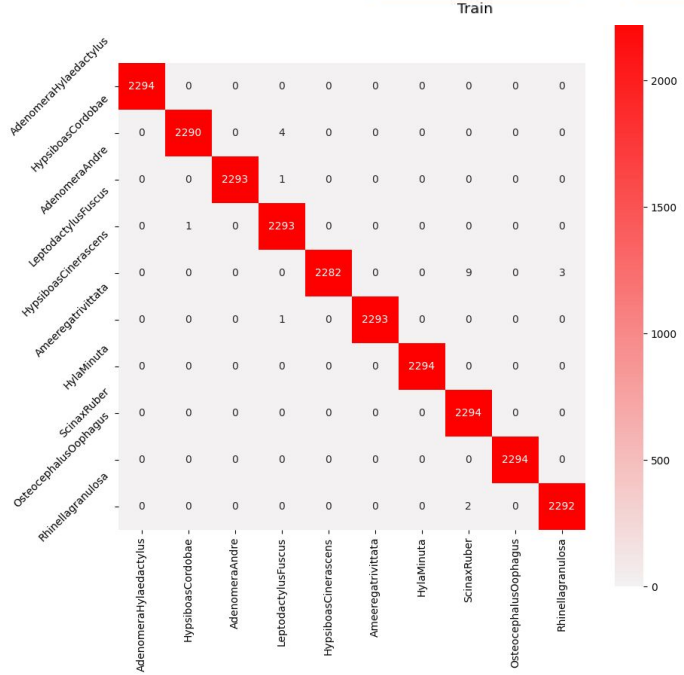


Slučajna šuma



Tačnost slučajne šume na trening i test skupu.

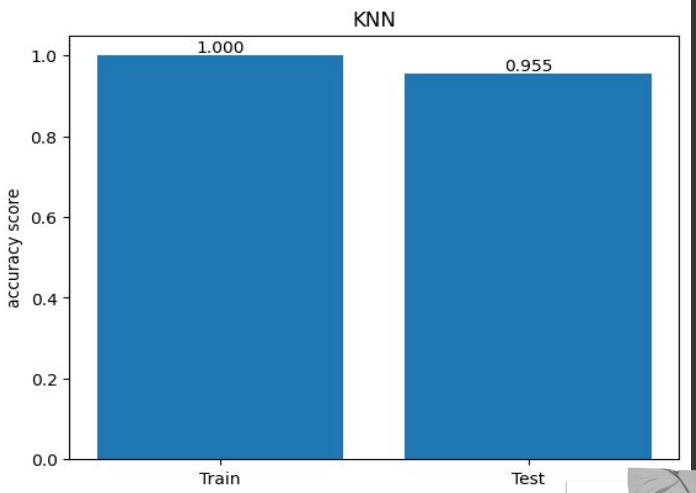


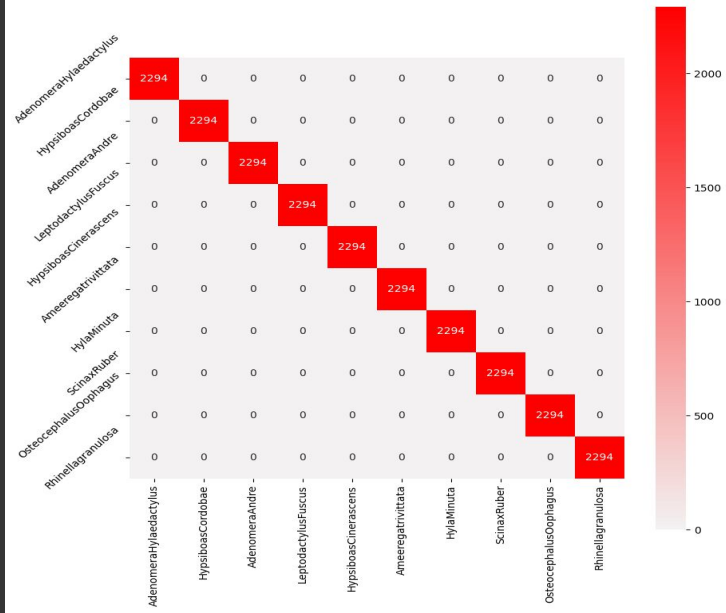


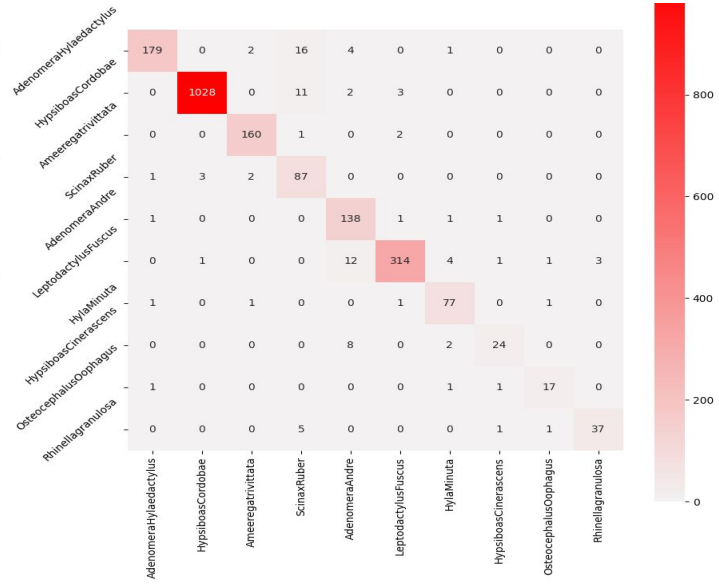
Matrica konfuzije

Leva se odnosi na **test** skup, a desna na **trening** skup.

K najbližih suseda







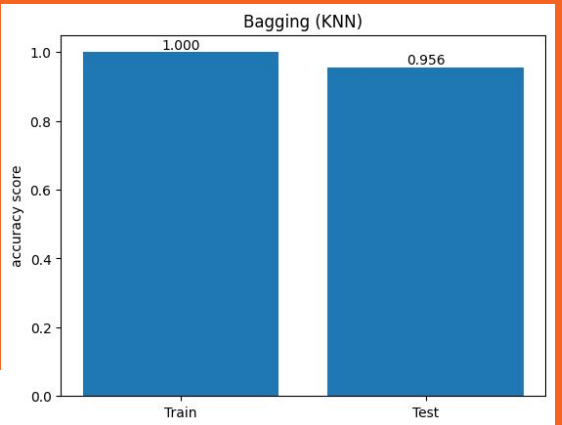
Gornja slika predstavlja **tačnost** na trening i test skupovima.

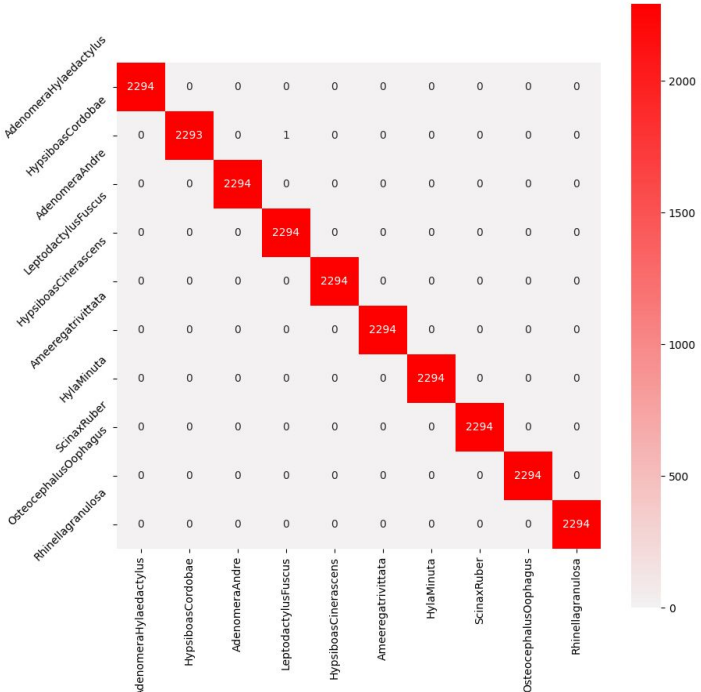
Matrica konfuzije

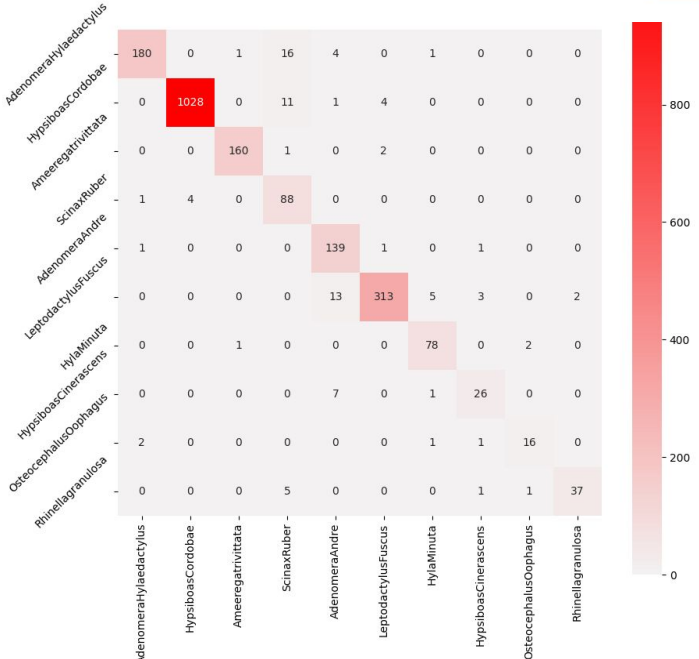
Leva se odnosi na **trening** skup, a desna na **test** skup.

Pakovanje

Tačnost ansambla na trening i test skupu.



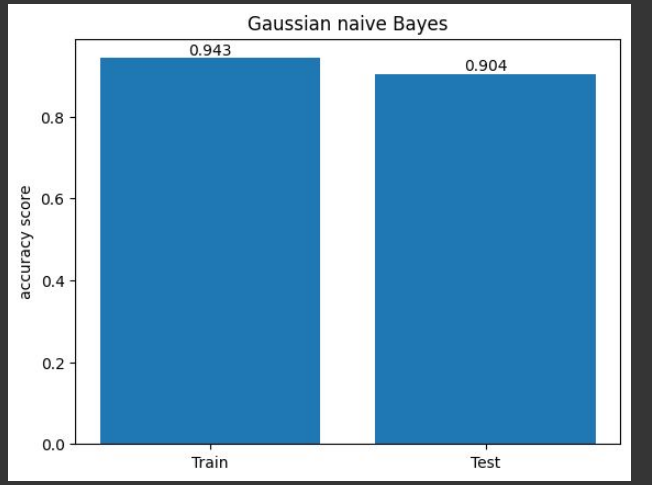




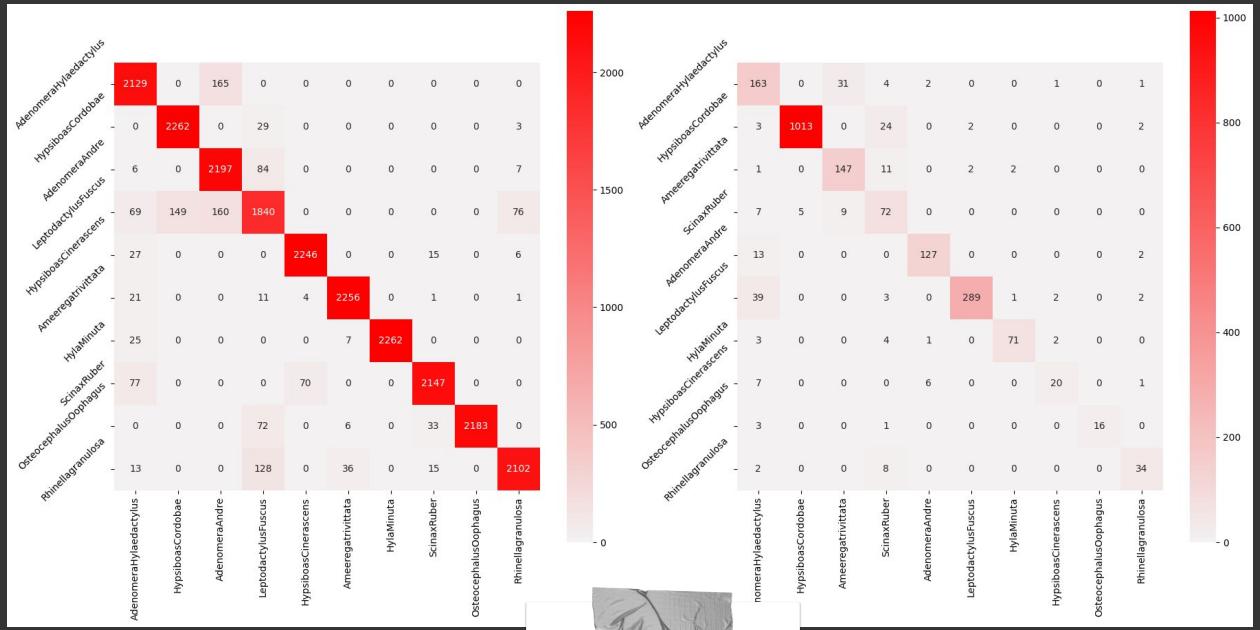
Matrica konfuzije

Leva se odnosi na **trening** skup, a desna na **test** skup.

Naivni Bajes



Gornja slika predstavlja **tačnost** na trening i test skupovima.

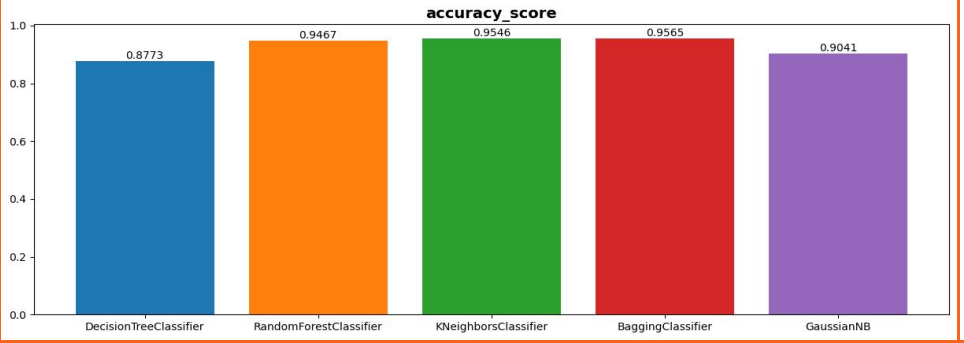


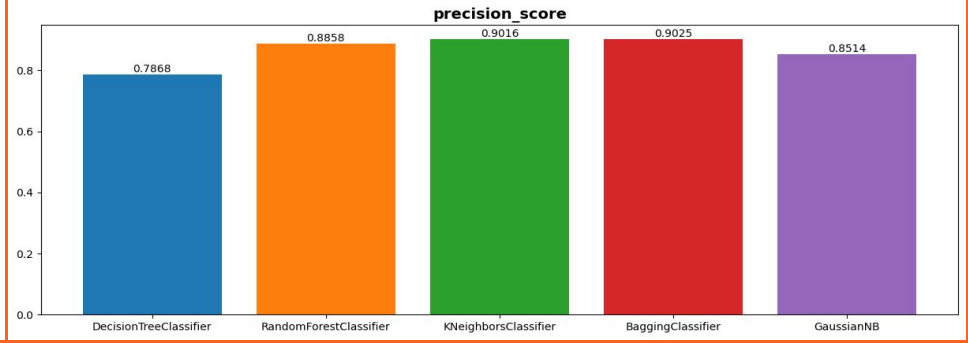
Matrica konfuzije

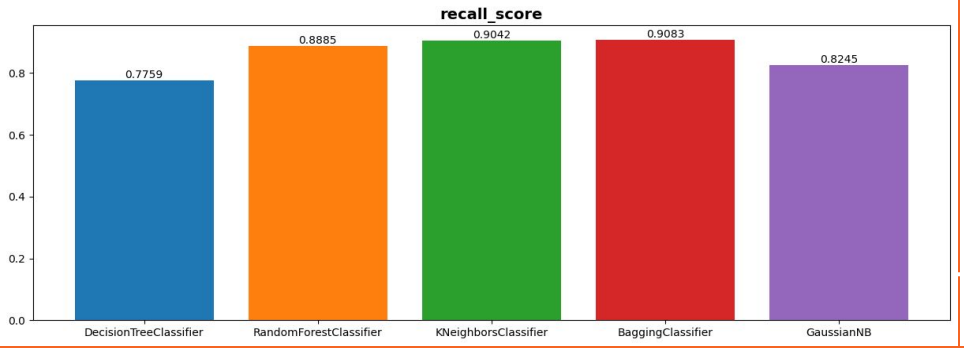
Leva se odnosi na **trening** skup, a desna na **test** skup.

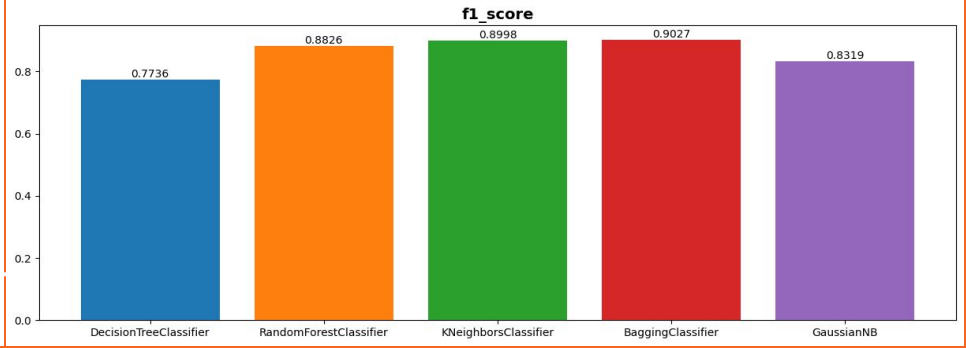
Klasifikacija

Poređenje modela



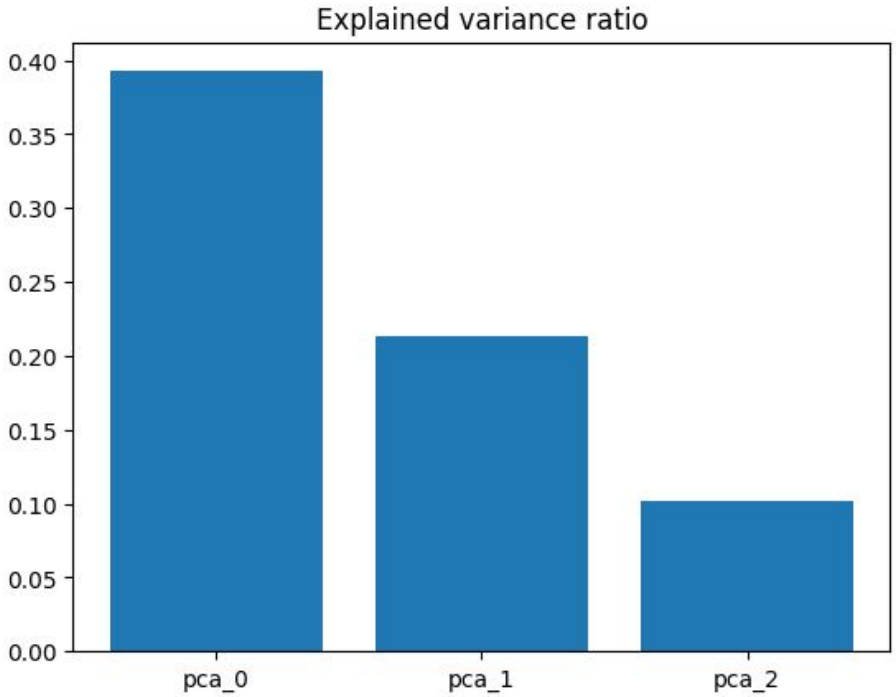


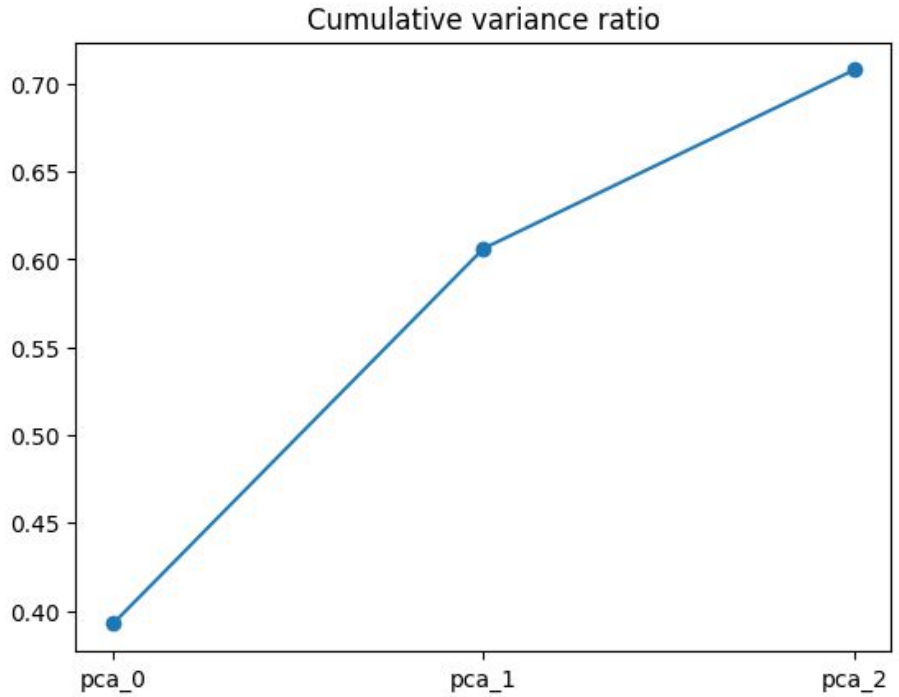




Klasterovanje

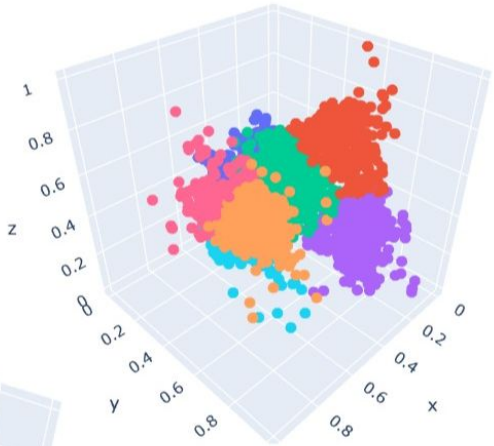
Pretprocesiranje

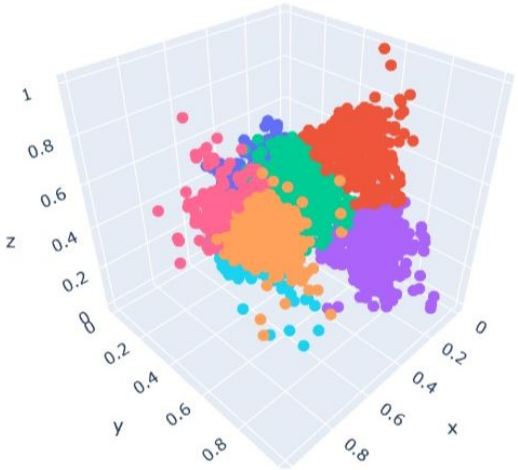


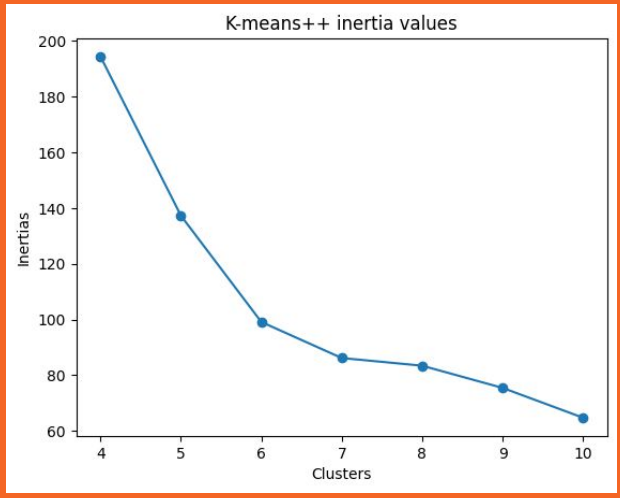


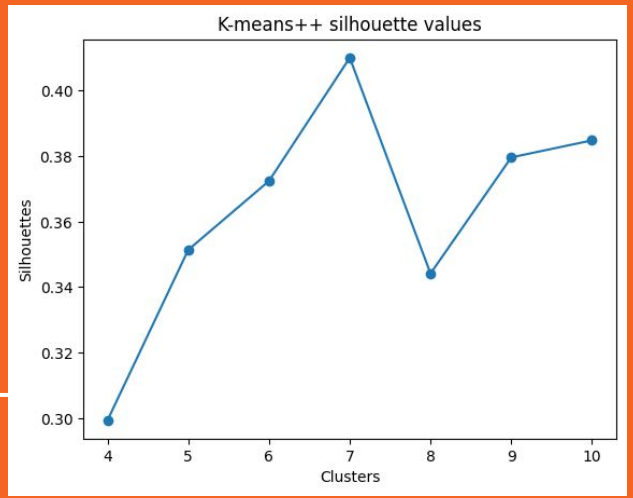
K sredina

Prikaz sedam klastera koristeći *k-means*++ i *random* načine za odabir centroida.

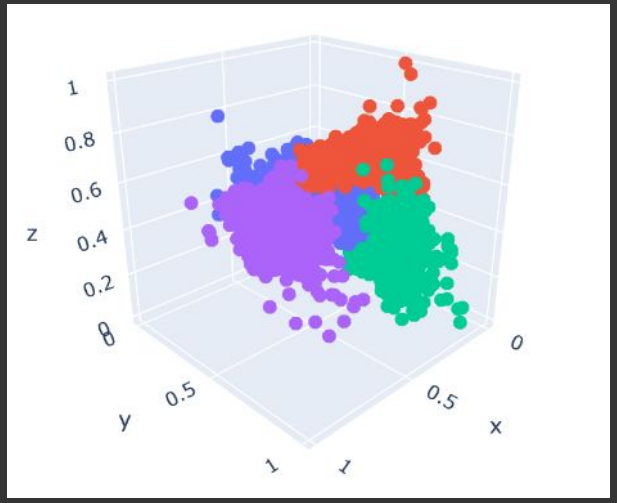


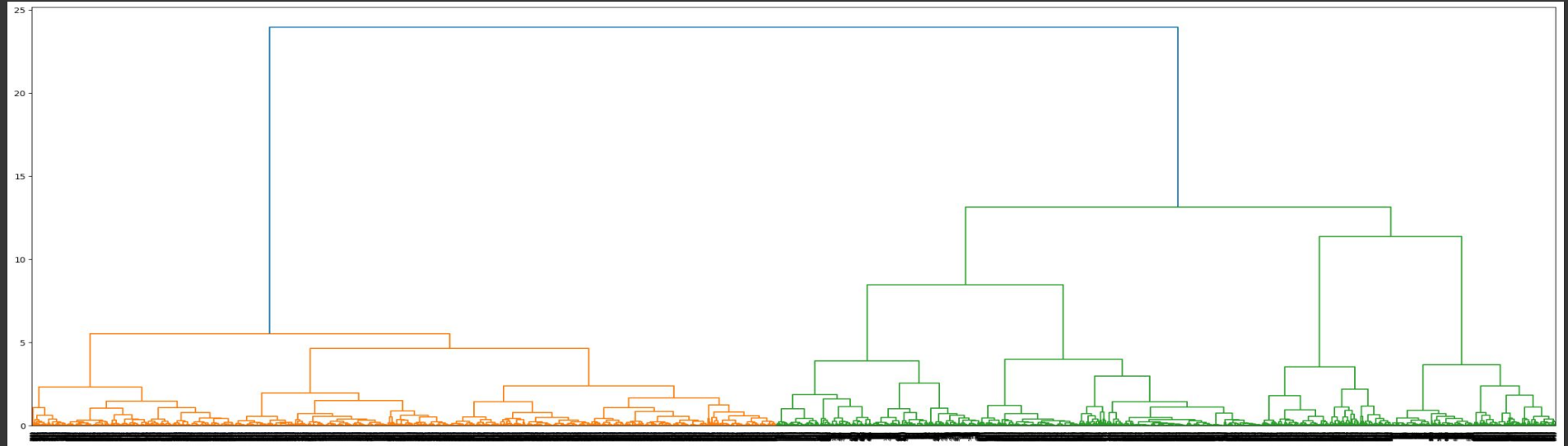




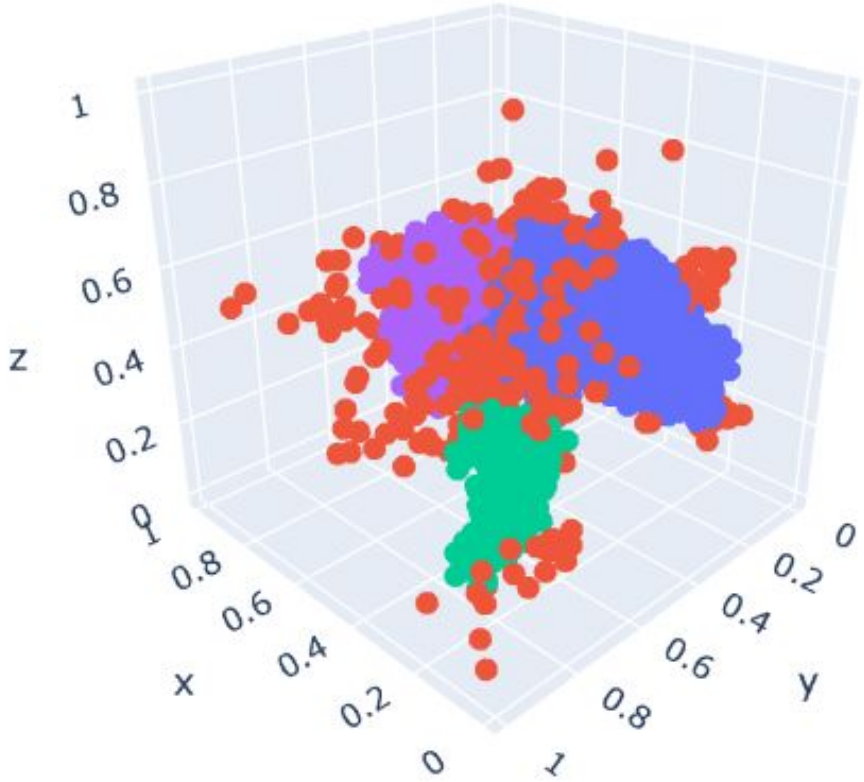


Hijerarhijsko klasterovanje





DBSCAN

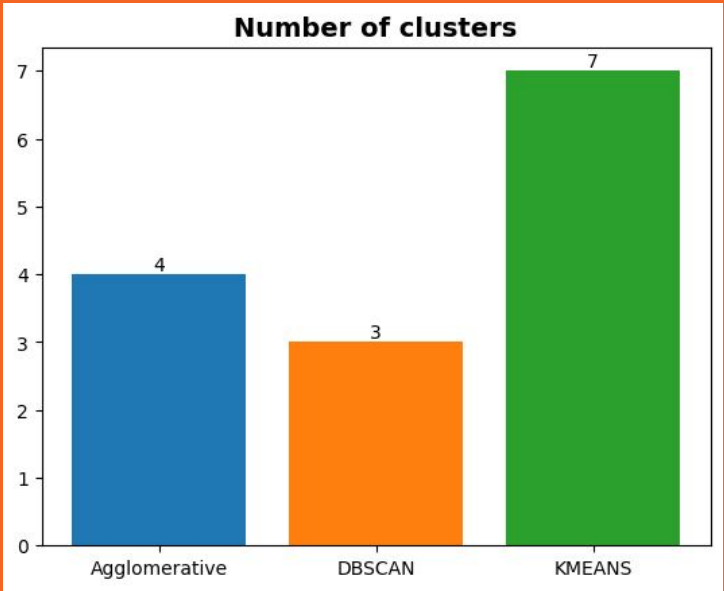


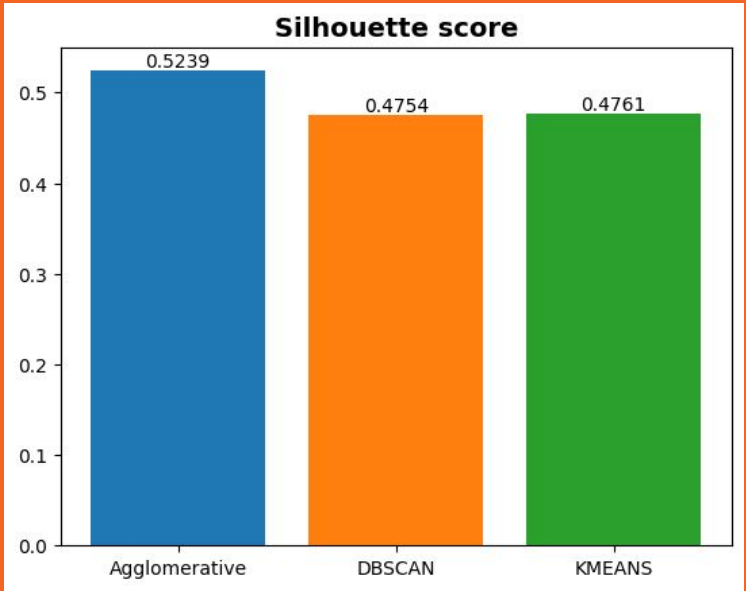
Klasteri

Algoritam definiše tri klastera, obojeni zelenom, plavom i ljubičastom bojom, i elemente van granica obojene crvenom bojom.

Klasterovanje

Poređenje modela

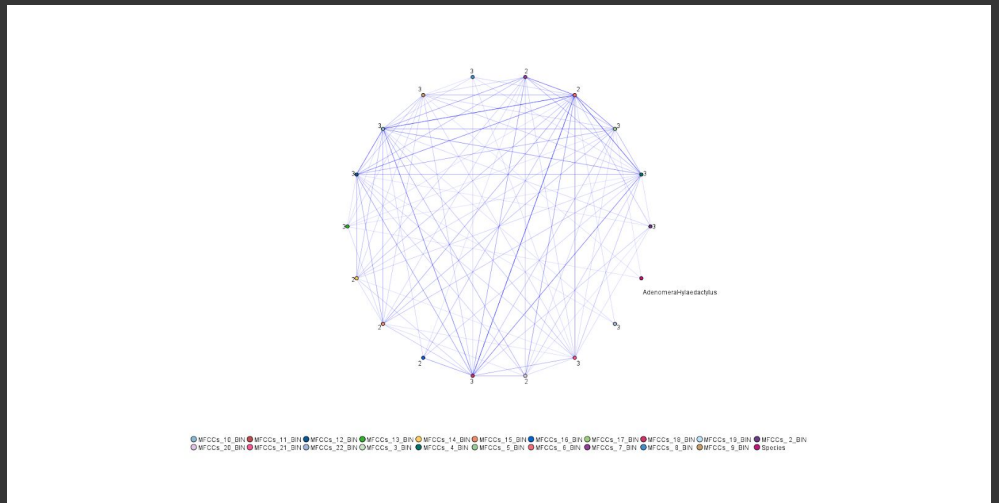




Pravila pridruživanja

Apriori

Consequent	Antecedent	Support %	Confidence %	Lift
Species = AdenomeraHylaedactylus	MFCCs_ 9_BIN = 3 MFCCs_10_BIN = 3	58.277	80.014	1.655
Species = AdenomeraHylaedactylus	MFCCs_ 9_BIN = 3 MFCCs_12_BIN = 3	56.511	82.612	1.709
Species = AdenomeraHylaedactylus	MFCCs_15_BIN = 2 MFCCs_ 5_BIN = 3 MFCCs_12_BIN = 3	55.942	80.547	1.666
Species = AdenomeraHylaedactylus	MFCCs_15_BIN = 2 MFCCs_ 5_BIN = 3 MFCCs_12_BIN = 3 MFCCs_10_BIN = 3	54.454	80.807	1.672
Species = AdenomeraHylaedactylus	MFCCs_ 9_BIN = 3 MFCCs_10_BIN = 3 MFCCs_ 6_BIN = 2	54.149	82.161	1.7

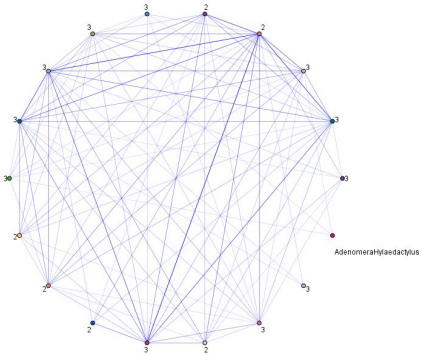


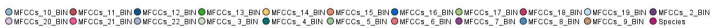
Consequent	Antecedent	Support %	Confidence %	Lift
Species = AdenomeraHylaedactylus	MFCCs_20_BIN = 1 MFCCs_11_BIN = 2 MFCCs_17_BIN = 3 MFCCs_ 3_BIN = 2	12.203	100.0	2.069
Species = AdenomeraHylaedactylus	MFCCs_20_BIN = 1 MFCCs_11_BIN = 2 MFCCs_17_BIN = 3 MFCCs_19_BIN = 2	12.064	100.0	2.069
Species = AdenomeraHylaedactylus	MFCCs_20_BIN = 1 MFCCs_11_BIN = 2 MFCCs_17_BIN = 3 MFCCs_13_BIN = 3	13.468	100.0	2.069
Species = AdenomeraHylaedactylus	MFCCs_20_BIN = 1 MFCCs_11_BIN = 2 MFCCs_17_BIN = 3 MFCCs_14_BIN = 2	11.383	100.0	2.069
Species = AdenomeraHylaedactylus	MFCCs_20_BIN = 1 MFCCs_11_BIN = 2 MFCCs_17_BIN = 3 MFCCs_16_BIN = 2	11.744	100.0	2.069

Consequent	Antecedent	Support %	Confidence %	Lift
Species = HypsiboasCordobae	MFCCs_ 3_BIN = 3 MFCCs_ 8_BIN = 2 MFCCs_11_BIN = 3 MFCCs_14_BIN = 2 MFCCs_12_BIN = 3	11.049	99.371	6.378
Species = HypsiboasCordobae	MFCCs_ 3_BIN = 3 MFCCs_ 8_BIN = 2 MFCCs_13_BIN = 2 MFCCs_14_BIN = 2 MFCCs_12_BIN = 3	10.716	99.351	6.377
Species = HypsiboasCordobae	MFCCs_ 3_BIN = 3 MFCCs_ 8_BIN = 2 MFCCs_13_BIN = 2 MFCCs_17_BIN = 2 MFCCs_14_BIN = 2	10.021	99.307	6.374
Species = HypsiboasCordobae	MFCCs_ 3_BIN = 3 MFCCs_ 8_BIN = 2 MFCCs_11_BIN = 3 MFCCs_17_BIN = 2 MFCCs_14_BIN = 2	10.354	99.195	6.367
Species = HypsiboasCordobae	MFCCs_ 3_BIN = 3 MFCCs_ 8_BIN = 2 MFCCs_13_BIN = 2 MFCCs_14_BIN = 2 MFCCs_20_BIN = 2	10.688	99.09	6.36

Pravila pridruživanja

FP-Growth





Most Interesting Rules by Rule Support

				Other Evaluation Statistics				
Rank	Rule ID	Condition	Prediction	Sorted By Rule Support(%)	Condition Support (%)	Confidence (%)	Lift	Deployability (%)
1	984	MFCCs_16_BIN = 2 MFCCs_12_BIN = 2	MFCCs_14_BIN = 3	15.82	18.28	86.54	2.83	2.46
2	990	MFCCs_4_BIN = 3 MFCCs_12_BIN = 2	MFCCs_14_BIN = 3	14.64	17.08	85.68	2.80	2.45
3	970	MFCCs_18_BIN = 3 MFCCs_12_BIN = 2	MFCCs_14_BIN = 3	14.50	16.50	87.87	2.88	2.00
4	952	MFCCs_17_BIN = 2 MFCCs_12_BIN = 2	MFCCs_14_BIN = 3	14.32	15.98	89.57	2.93	1.67
5	954	MFCCs_18_BIN = 3 MFCCs_16_BIN = 2 MFCCs_12_BIN = 2	MFCCs_14_BIN = 3	13.50	15.08	89.49	2.93	1.58

Most Interesting Rules by Confidence

				Other Evaluation Statistics				
Rank	Rule ID	Condition	Prediction	Sorted By Confidence(%)	Condition Support (%)	Rule Support (%)	Lift	Deployability (%)
1	1	MFCCs_12_BIN = 3 MFCCs_14_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	99.37	11.05	10.98	6.38	0.07
2	2	MFCCs_12_BIN = 3 MFCCs_14_BIN = 2 MFCCs_13_BIN = 2 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	99.35	10.72	10.65	6.38	0.07
3	3	MFCCs_14_BIN = 2 MFCCs_17_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	99.19	10.35	10.27	6.37	0.08
4	4	MFCCs_20_BIN = 2 MFCCs_14_BIN = 2 MFCCs_13_BIN = 2 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	99.09	10.69	10.59	6.36	0.10
5	5	MFCCs_20_BIN = 2 MFCCs_14_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	98.99	11.04	10.92	6.35	0.11

Most Interesting Rules by Lift

				Other Evaluation Statistics				
Rank	Rule ID	Condition	Prediction	Sorted By Lift	Condition Support (%)	Confidence (%)	Rule Support (%)	Deployability (%)
1	1	MFCCs_12_BIN = 3 MFCCs_14_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	6.38	11.05	99.37	10.98	0.07
2	2	MFCCs_12_BIN = 3 MFCCs_14_BIN = 2 MFCCs_13_BIN = 2 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	6.38	10.72	99.35	10.65	0.07
3	3	MFCCs_14_BIN = 2 MFCCs_17_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	6.37	10.35	99.19	10.27	0.08
4	4	MFCCs_20_BIN = 2 MFCCs_14_BIN = 2 MFCCs_13_BIN = 2 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	6.36	10.69	99.09	10.59	0.10
5	5	MFCCs_20_BIN = 2 MFCCs_14_BIN = 2 MFCCs_11_BIN = 3 MFCCs_8_BIN = 2 MFCCs_3_BIN = 3	Species = HypsiboasCordobae	6.35	11.04	98.99	10.92	0.11