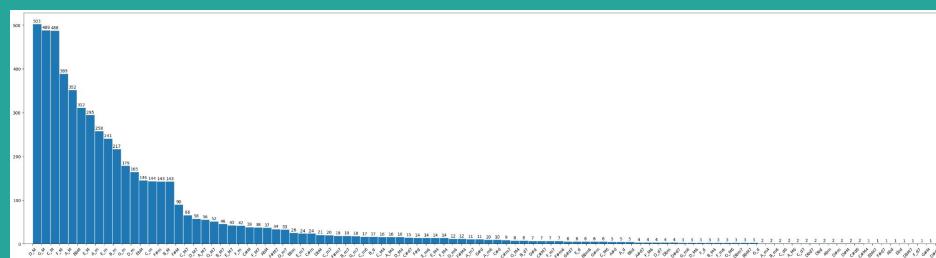
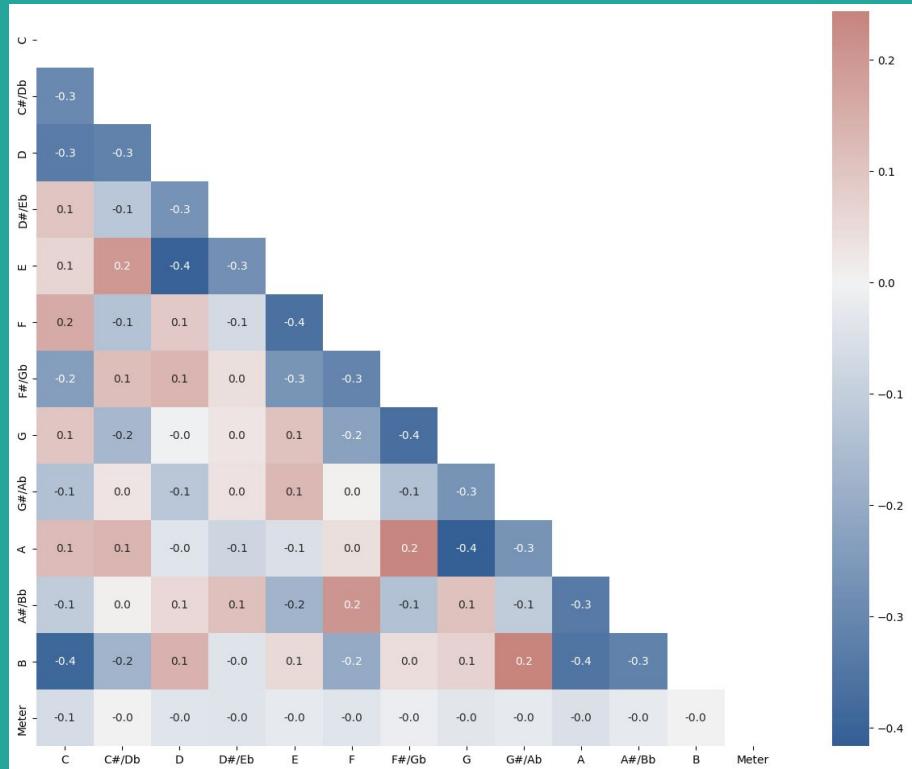


# Istraživanje skupa podataka o horalima J.S. Baha

---

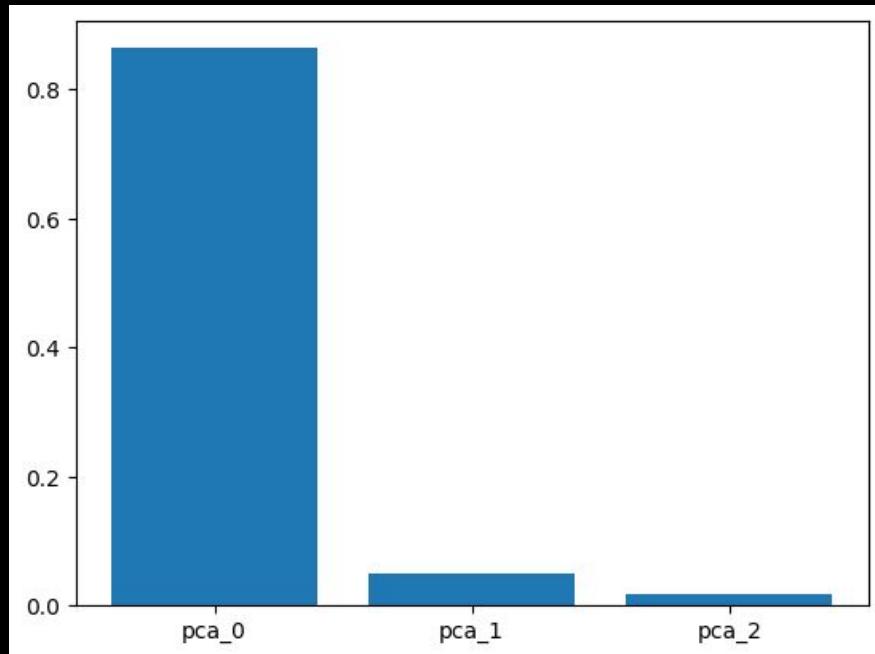
Mihailo Simić

# Eksplorativna analiza skupa podataka

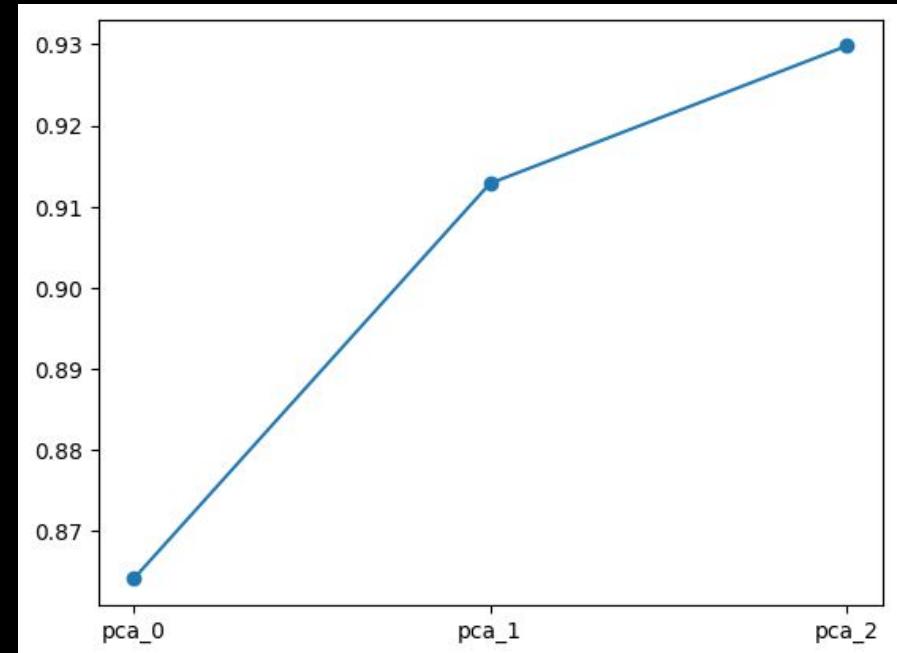


# Pretprocesiranje

Explained variance ratio

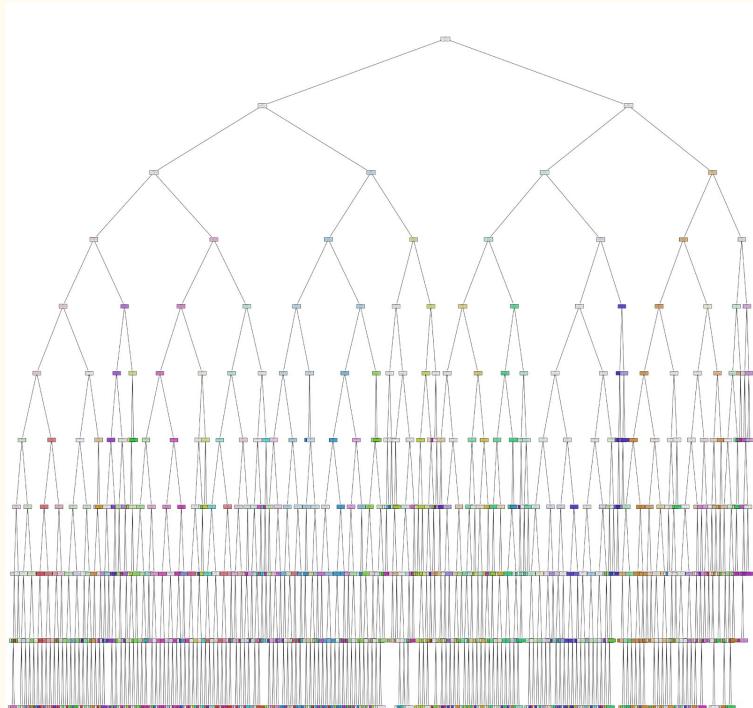


Cumulative variance ratio

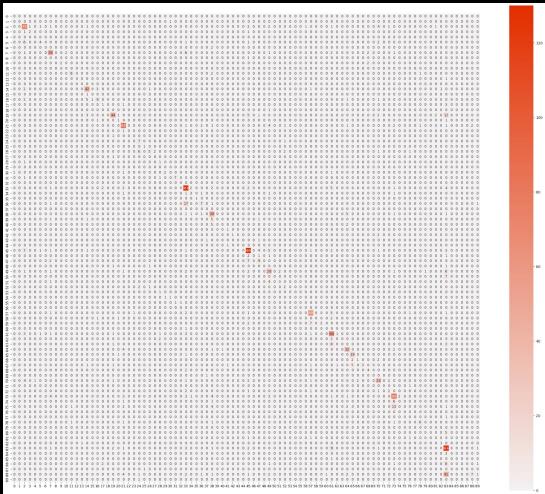


# Klasifikacija

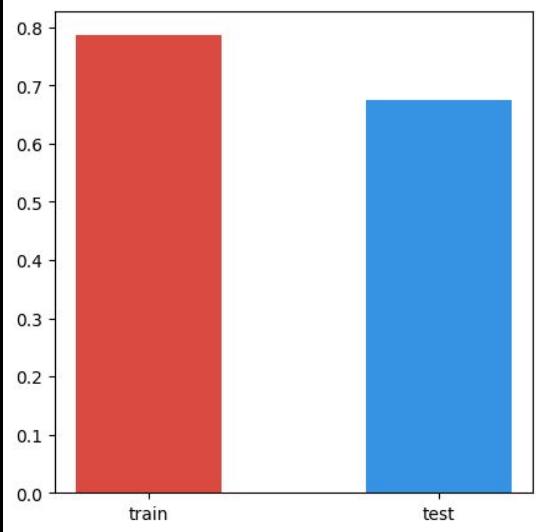
## Stablo odlučivanja (Decision Tree)



Matrica konfuzije

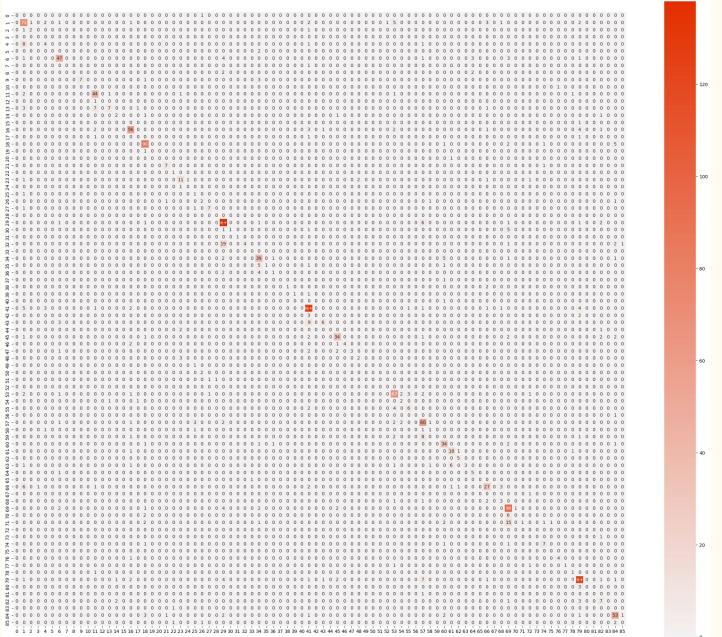


Poređenje tačnosti (accuracy) na trening i test skupu



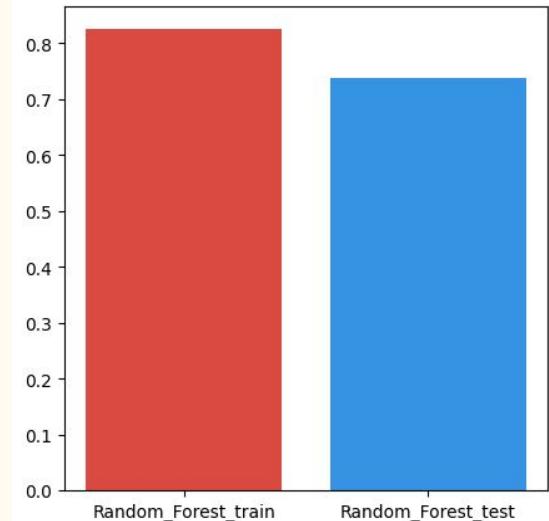
# Klasifikacija

## Slučajna šuma (Random forest)

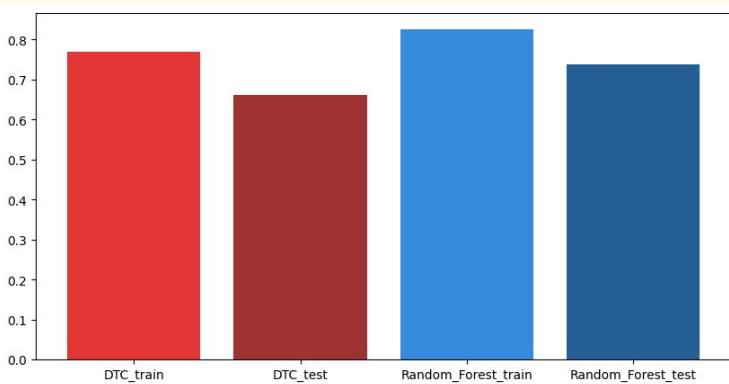


Matrica konfuzije

Poređenje tačnosti  
na trening i test  
skupu.

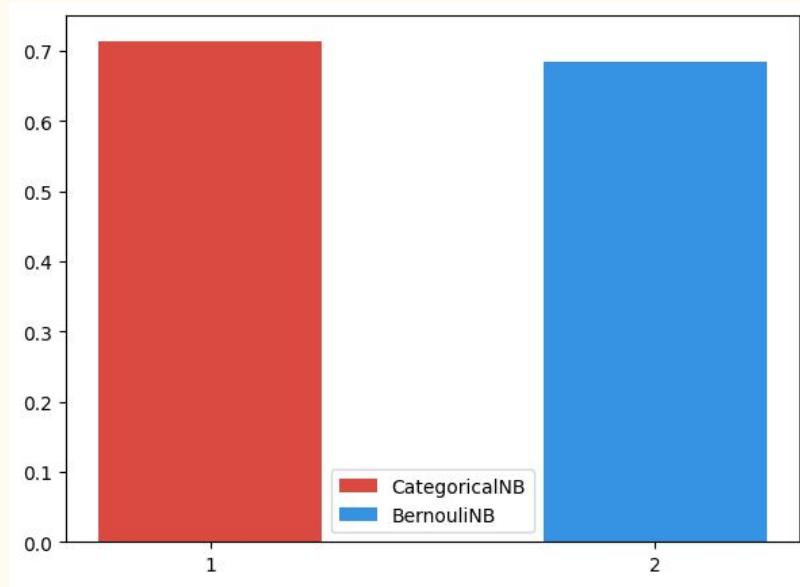


Poređenje tačnosti  
stabla odlučivanja  
i slučajne šume.



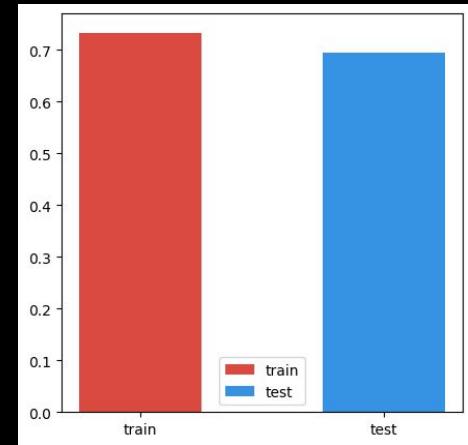
# Klasifikacija

## Naivni Bajes (Naive Bayes)

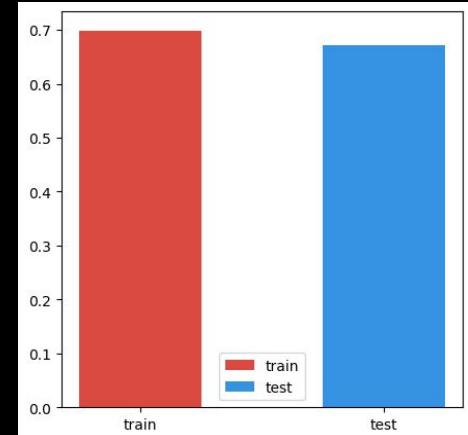


Poređenje kategoričkog i Bernulijevog naivnog Bajesa

Kategorički  
naivni Bajes

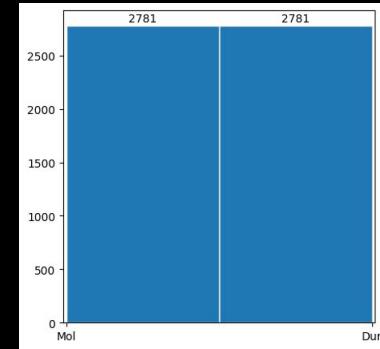
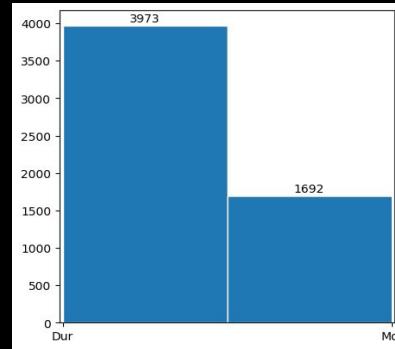
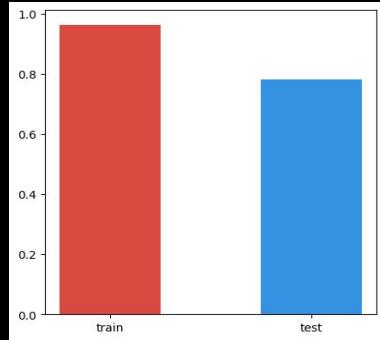
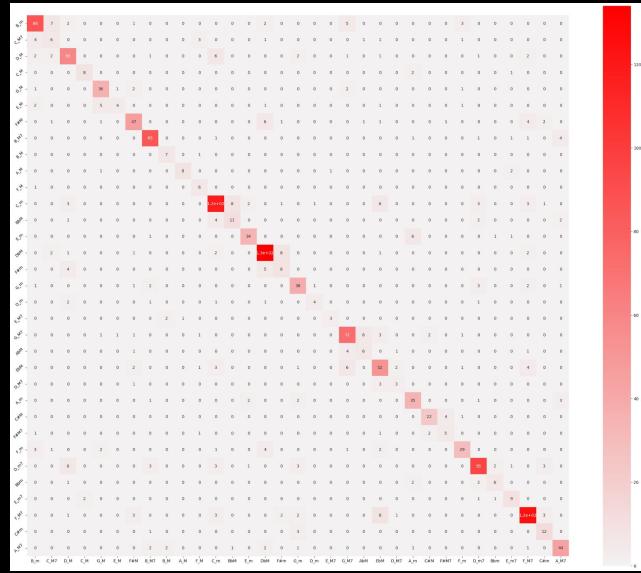


Bernulijev  
naivni Bajes

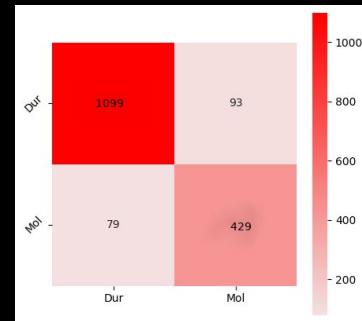
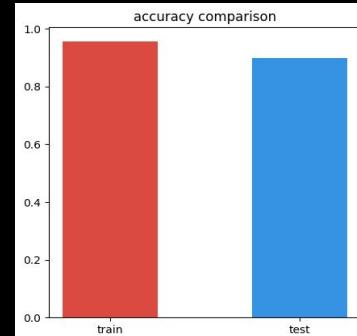


# Klasifikacija

Poboljšanje modela slučajne šume



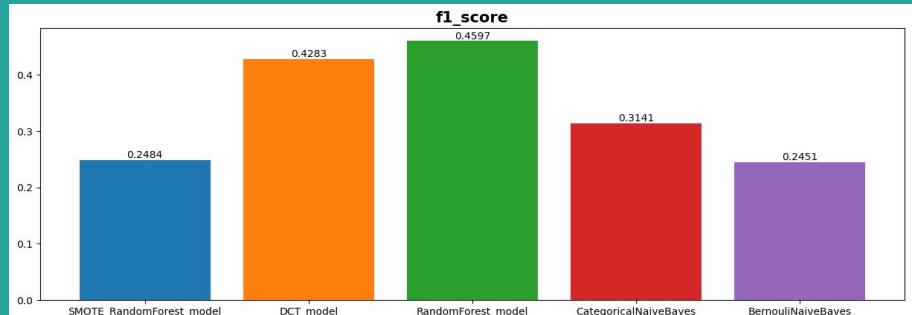
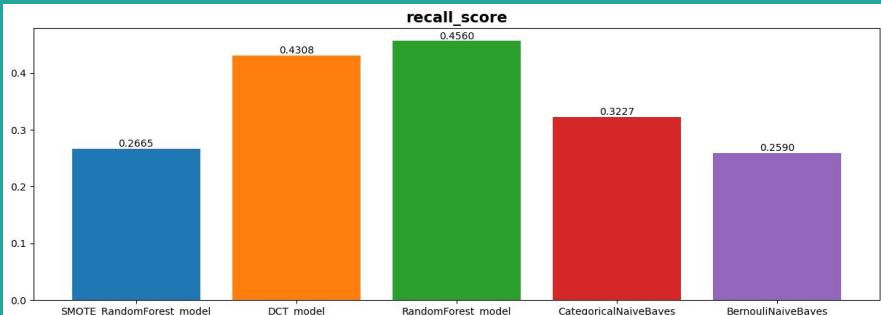
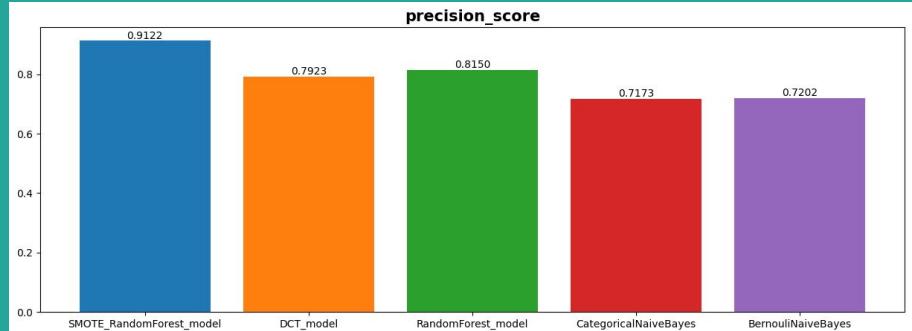
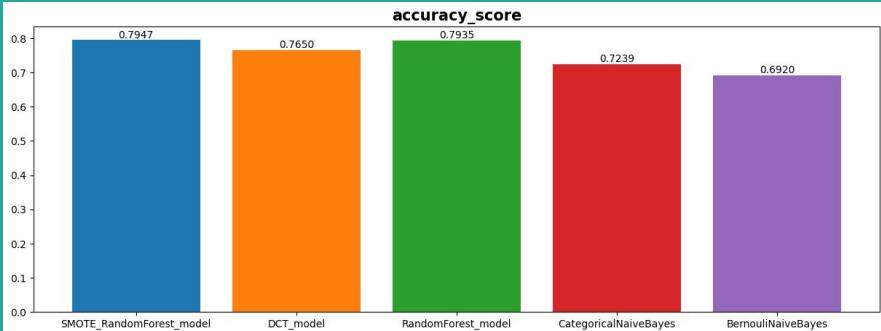
Rezultati nakon grupisanja klasa u dve klase



Matrica konfuzije i tačnost na trening i test skupu nakon odsecanja klasa i balansiranja pomoću SMOTE algoritma

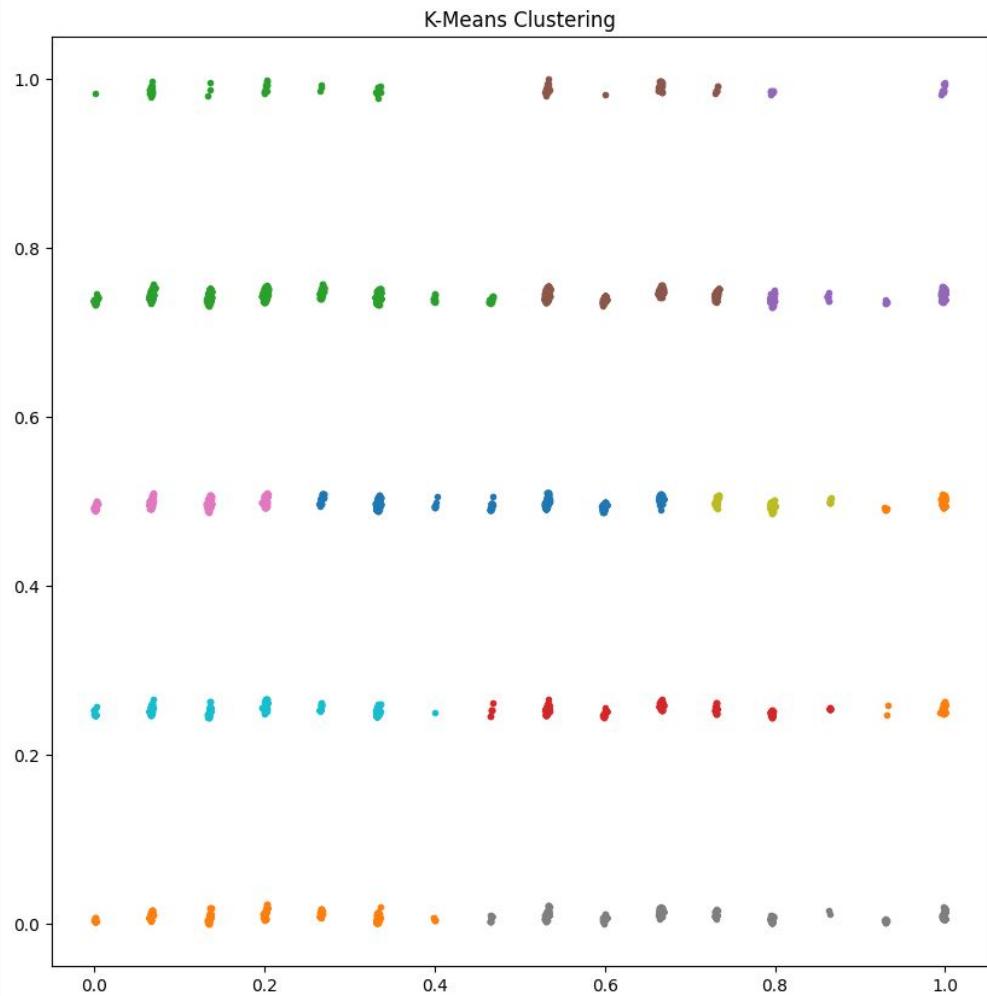
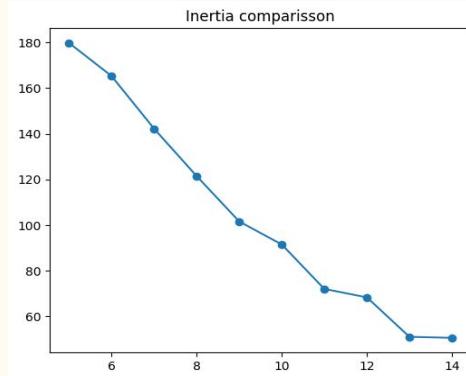
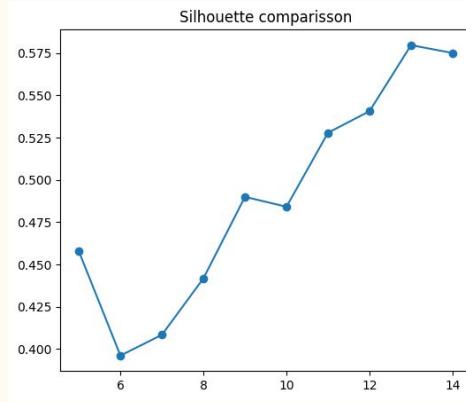
# Klasifikacija

## Poređenje modela



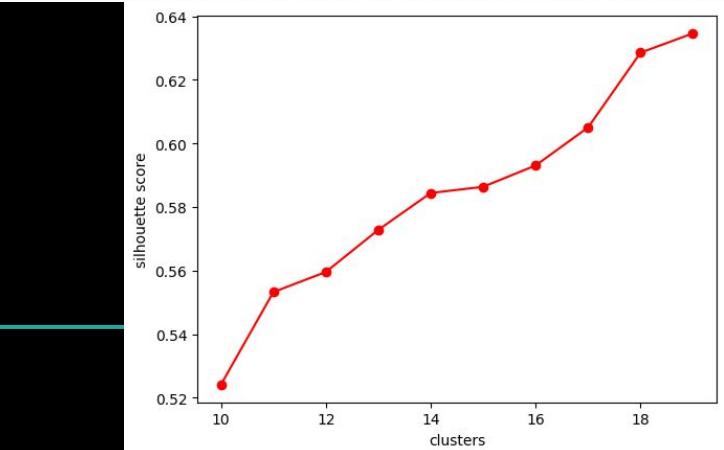
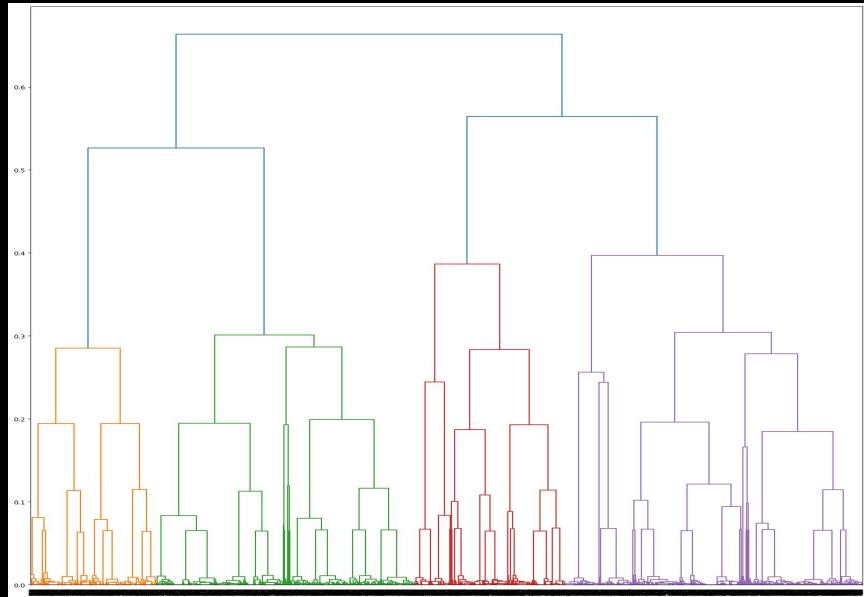
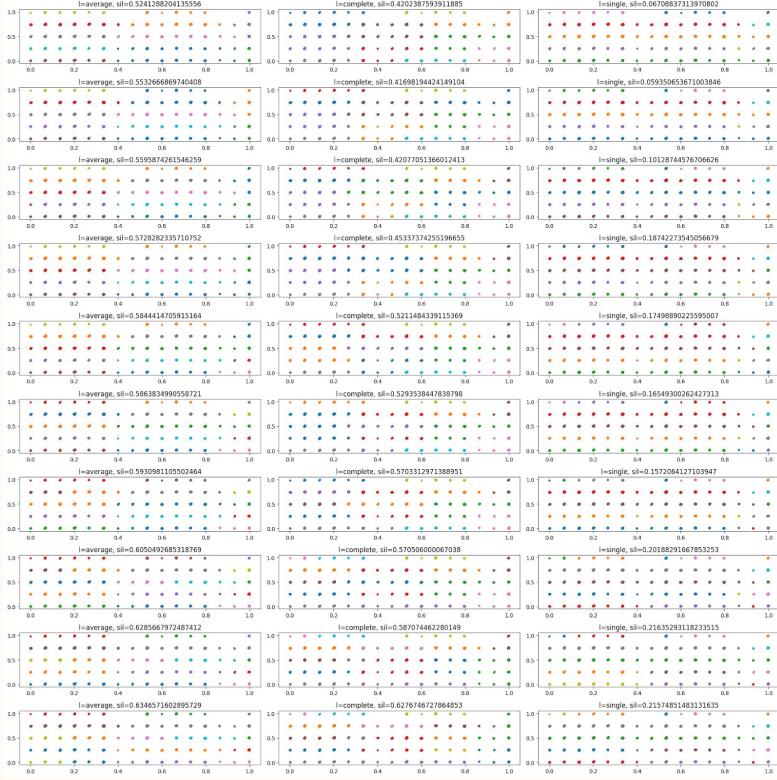
# KlastEROVANJE

## $K$ sredina ( $K$ -means)



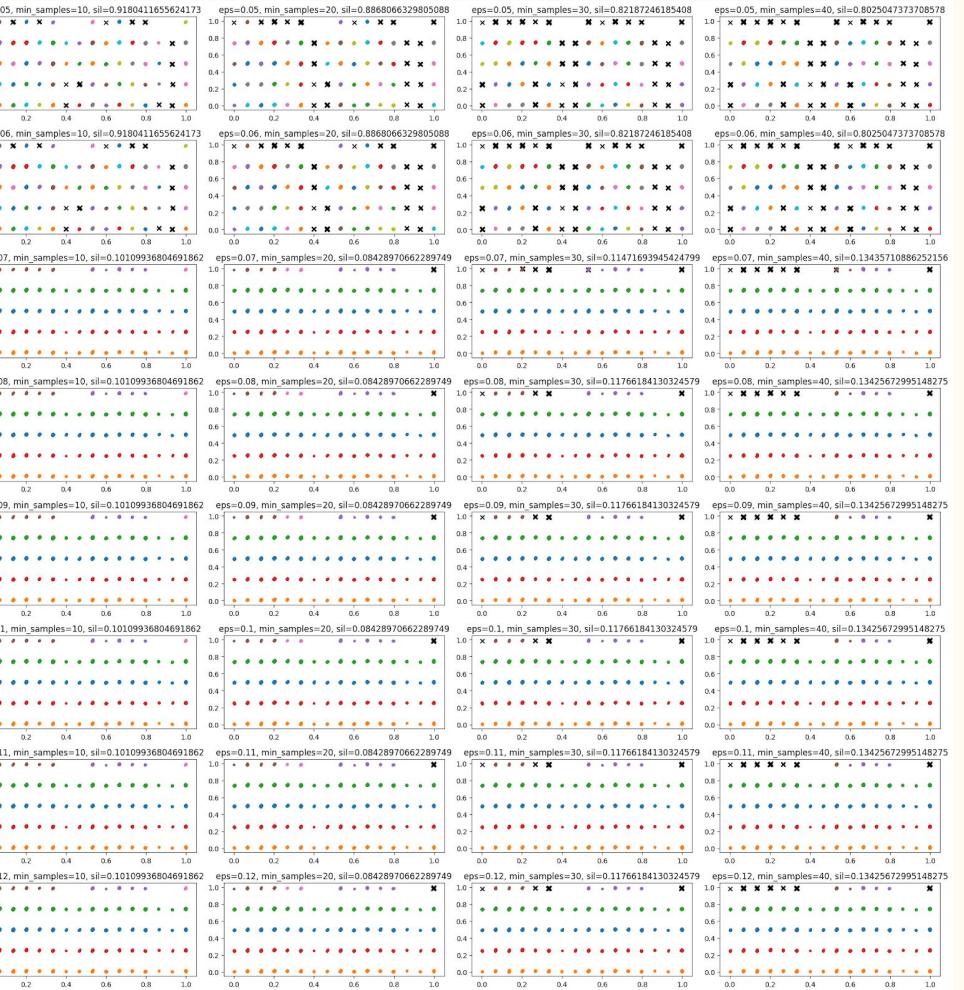
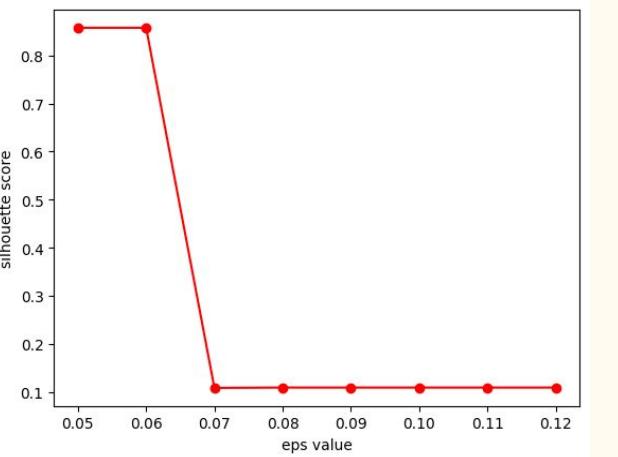
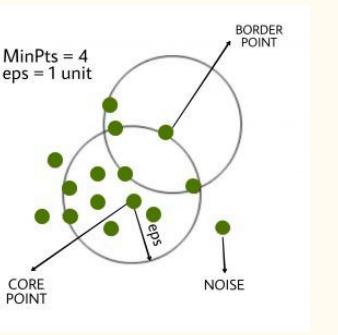
# Klasterovanje

## Hijerarhijsko (Agglomerative)



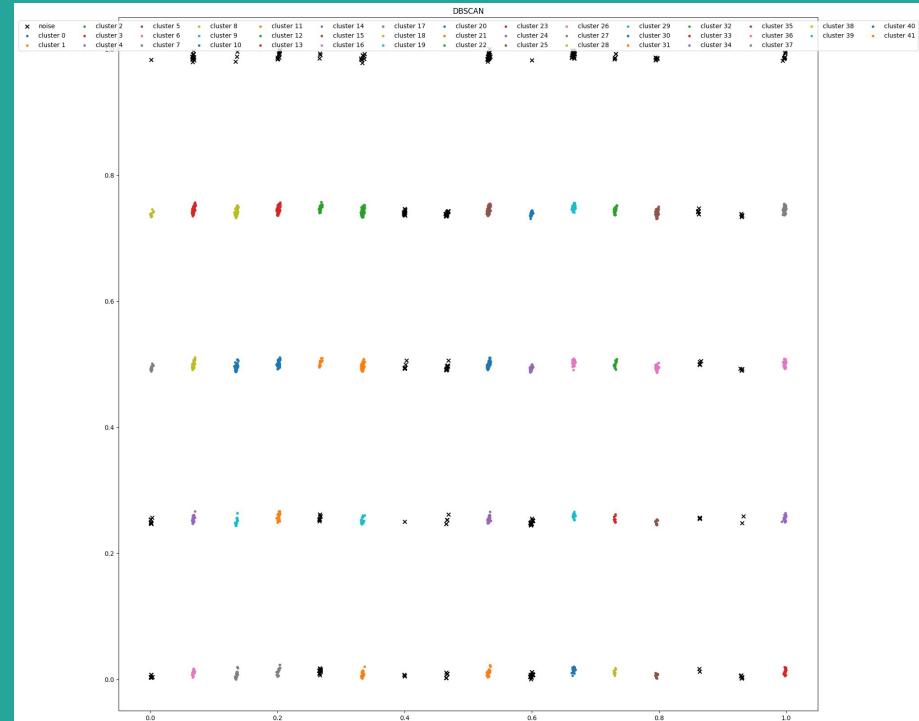
# KlastEROVANJE

## DBSCAN

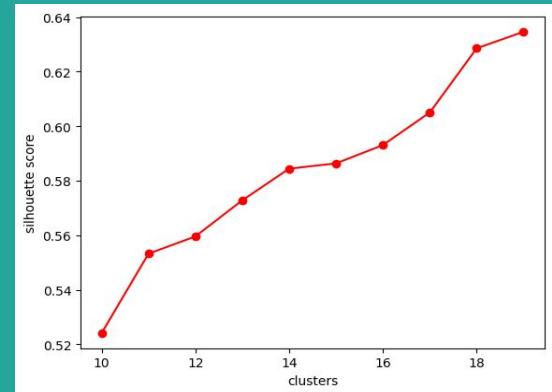


# Klasterovanje

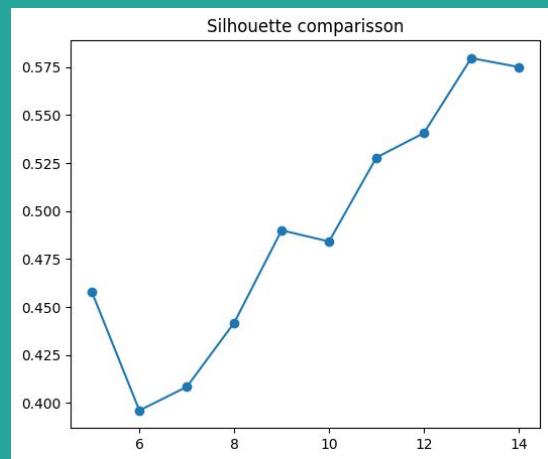
## Poređenje modela



Hijerarhijsko  
klasterovanje



$K$  sredina



DBSCAN klasterovanje

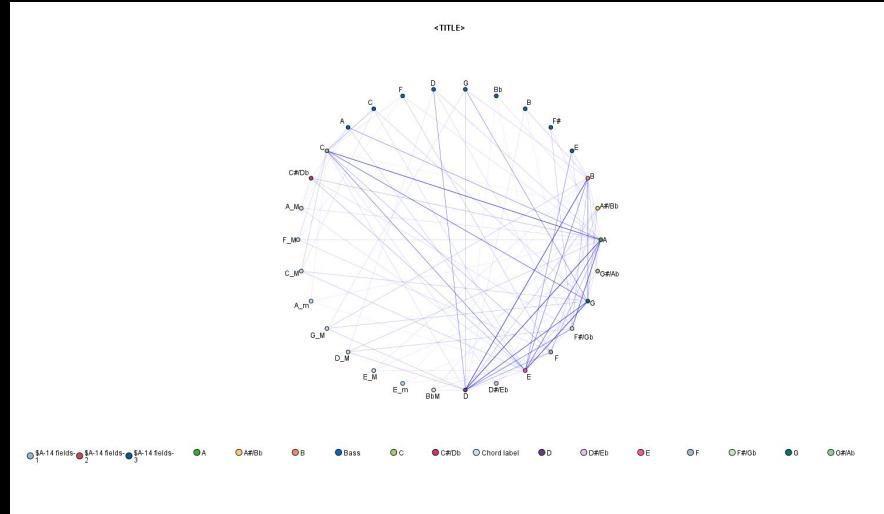
# Pravila pridruživanja

## Apriori

Consequent	Antecedent	Support %	Confidence %	Lift
Chord label = A_m	Bass = A			
	C	2.365	85.075	18.68
	E			
	A			
Chord label = A_m	Bass = A			
	C	2.383	84.444	18.542
	E			
Chord label = BbM	Bass = Bb			
	F	3.054	94.22	17.108
	D			
Chord label = BbM	Bass = Bb			
	A#/Bb	3.054	94.22	17.108
	F			
	D			
Chord label = E_M	G#/Ab			
	Bass = E	3.16	86.592	16.629
	B			
Chord label = E_M	G#/Ab			
	Bass = E	3.16	86.592	16.629
	B			
	E			



Consequent	Antecedent	Support %	Confidence %	Lift
D	Bass = D	12.162	99.129	2.374
A	Bass = A	12.145	99.419	2.371
G	Bass = G	12.021	99.559	2.633
E	Bass = E	11.968	99.705	2.658
C	Bass = C	10.591	98.5	3.117
F	Bass = F	8.155	97.835	4.316
F#/Gb	Bass = F#	8.049	99.342	3.986
B	Bass = B	7.926	98.441	3.114
G	Bass = G	7.82	99.323	2.627
D	Bass = D	7.538	99.063	2.373
	A			
C	Bass = C	7.22	98.533	3.118
	G			

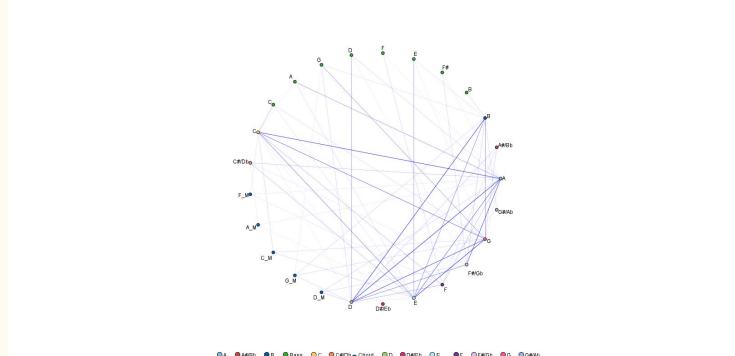


# Pravila pridruživanja

## FP-Growth

Rule Statistics				
Measurements	Minimum	Maximum	Mean	Standard Deviation
Condition Support (%)	5.08	41.92	16.77	11.20
Confidence (%)	14.02	100.00	56.55	28.46
Rule Support (%)	5.01	12.07	7.03	1.94
Lift	2.00	11.12	4.05	1.96
Deployability (%)	0.00	36.05	9.74	10.53
Information for Most Frequent Items				
Item name	Records (%)	Conditions (%)	Predictions (%)	
A	41.92	24.47	3.19	
D	41.75	22.34	3.19	
G	37.81	19.15	4.26	
E	37.51	18.09	3.19	
B	31.62	13.83	4.26	
C	31.60	15.96	5.32	
F#Gb	24.92	8.51	3.19	
F	22.67	10.64	4.26	
A#Bb	18.02	3.19	2.13	
C#Db	16.84	3.19	1.06	
Bass = D	12.16	3.19	3.19	
Bass = A	12.14	3.19	3.19	
Bass = G	12.02	8.51	5.32	
Bass = E	11.97	3.19	3.19	
G#Ab	11.63	1.06	1.06	
Bass = C	10.59	3.19	3.19	
Chord_label = D_M	8.88	0.00	7.45	
Chord_label = G_M	8.63	0.00	9.57	
Chord_label = C_M	8.61	0.00	7.45	
Bass = F	8.16	4.26	3.19	
Bass = F#	8.05	2.13	2.13	
Bass = B	7.93	1.06	1.06	
Chord_label = F_M	6.87	0.00	7.45	
Chord_label = A_M	6.21	0.00	6.38	
Chord_label = BbM	5.51	0.00	1.06	
Bass = Bb	5.10	1.06	1.06	

Most Interesting Rules by Confidence								
Rank	Rule ID	Condition	Prediction	Sorted By Confidence(%)	Other Evaluation Statistics			
					Condition Support (%)	Rule Support (%)	Lift	Deployability (%)
1	40	A Bass = F#	F#Gb	100.00	5.49	5.49	4.01	0.00
2	59	G Bass = E	E	100.00	5.68	5.68	2.67	0.00
3	60	B Bass = E	E	100.00	6.65	6.65	2.67	0.00
4	76	F#Gb Bass = D	D	100.00	5.68	5.68	2.40	0.00
5	61	Bass = E	E	99.71	11.97	11.93	2.66	0.04
6	77	C Bass = A	A	99.68	5.44	5.42	2.38	0.02
7	17	Bass = Bb	A#Bb	99.65	5.10	5.08	5.53	0.02
8	65	Bass = G	G	99.56	12.02	11.97	2.63	0.05
9	68	B Bass = G	G	99.47	6.71	6.67	2.63	0.04
10	81	Bass = A	A	99.42	12.14	12.07	2.37	0.07
11	41	Bass = F#	F#Gb	99.34	8.05	8.00	3.99	0.05
12	69	D Bass = G	G	99.32	7.82	7.77	2.63	0.05
Most Interesting Rules by Lift								
Rank	Rule ID	Condition	Prediction	Sorted By Lift	Other Evaluation Statistics			
					Condition Support (%)	Confidence (%)	Rule Support (%)	Deployability (%)
1	1	A C F	Chord label = F_M	11.12	7.10	76.37	5.42	1.68
2	2	G E C	Chord label = C_M	9.14	8.37	78.69	6.58	1.78
3	3	A C#Db	Chord label = A_M	9.06	9.53	56.30	5.37	4.17
4	4	A D F#Gb	Chord label = D_M	8.67	9.29	77.00	7.15	2.14
5	5	D G B	Chord label = G_M	8.65	8.70	74.65	6.50	2.21
6	6	C F	Chord label = F_M	8.45	10.22	58.03	5.93	4.29
7	7	E C#Db	Chord label = A_M	8.13	9.92	50.53	5.01	4.91
8	8	A F	Chord label = F_M	8.11	10.36	55.71	5.77	4.59
9	9	D F#Gb	Chord label = D_M	6.51	13.26	57.79	7.66	5.60
10	10	E C	Chord label = C_M	6.49	13.26	55.93	7.41	5.84
11	11	A F	Bass = F	6.31	10.36	51.45	5.33	5.03
12	12	G	Chord label = G_M	6.16	13.54	53.19	7.20	6.34



Hvala na pažnji!

---