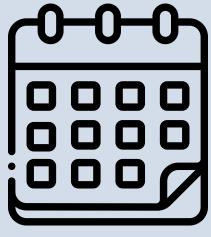
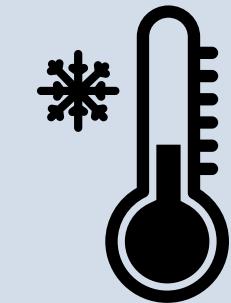
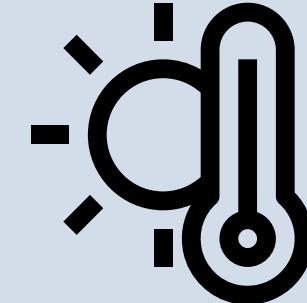
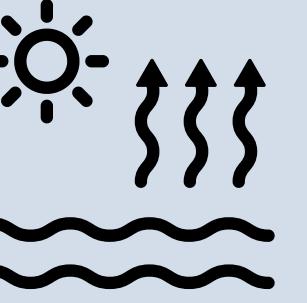
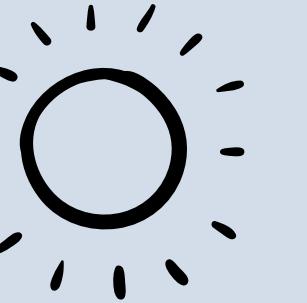
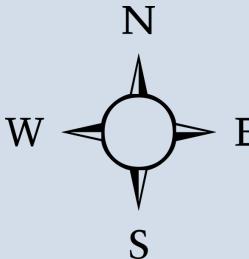
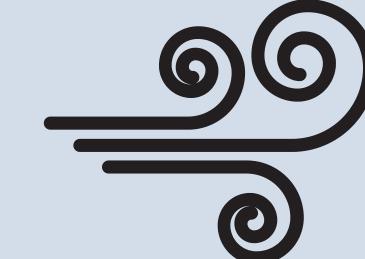
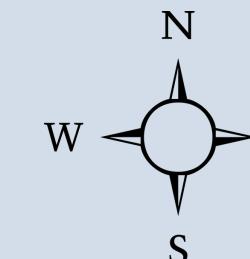
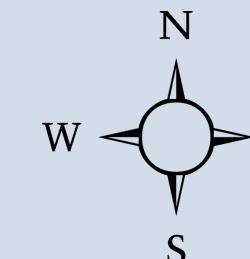
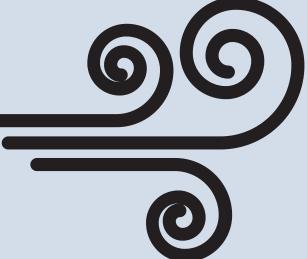
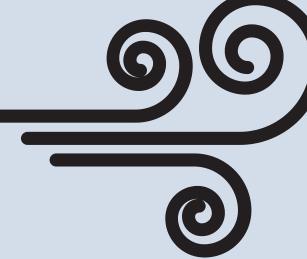
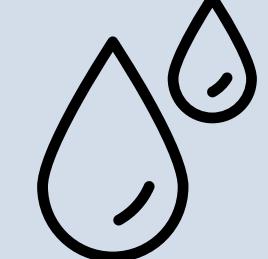
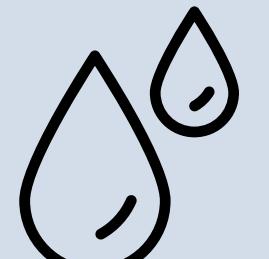
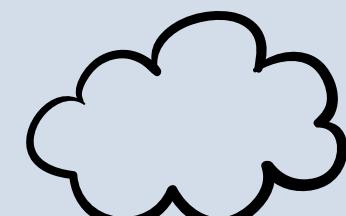
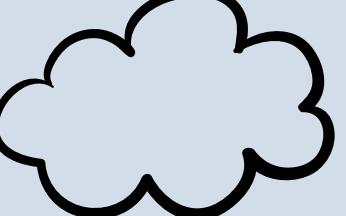
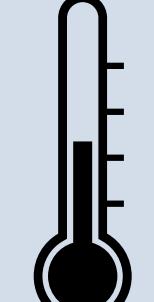
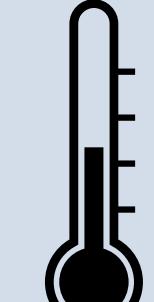


# Rain in Australia

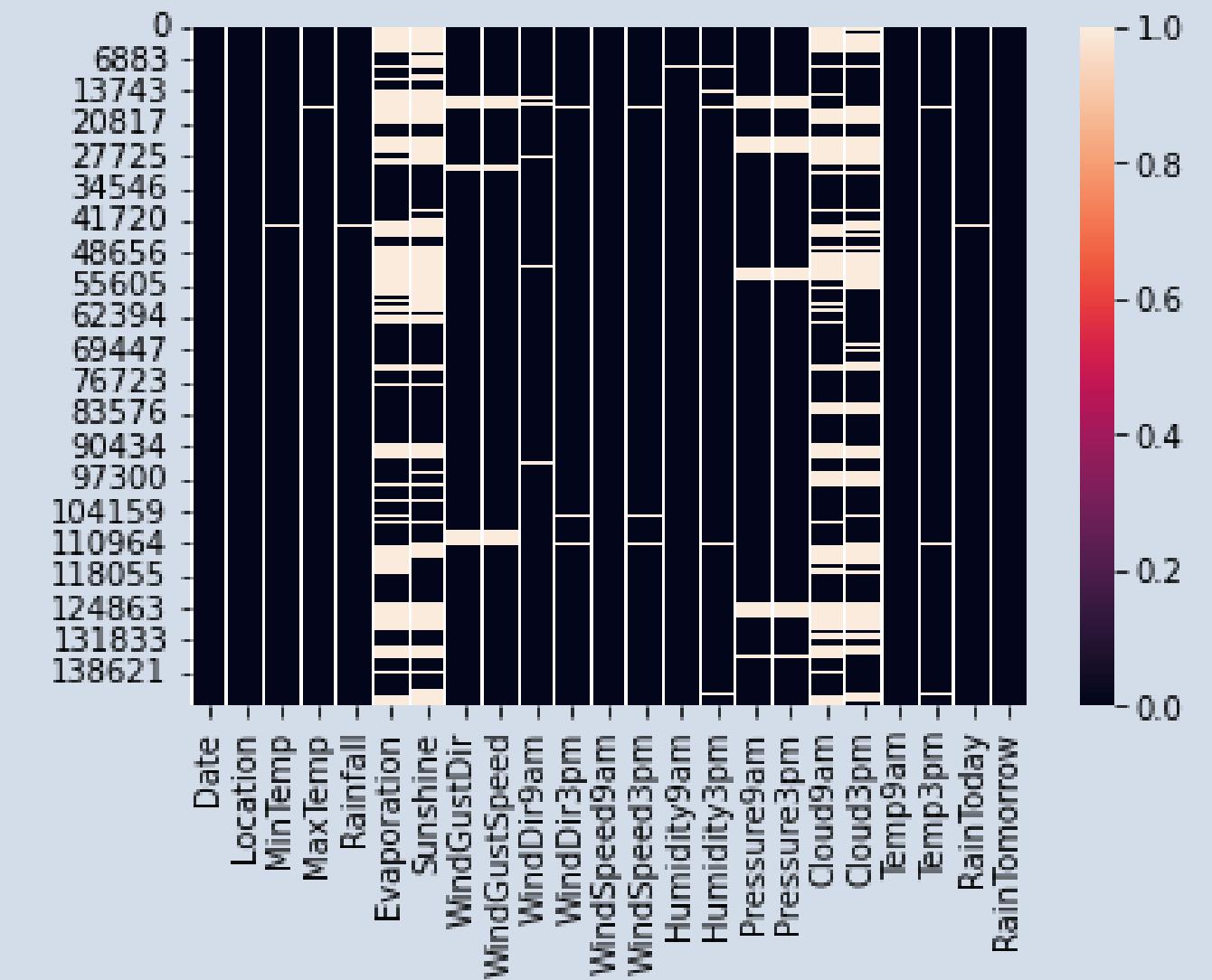
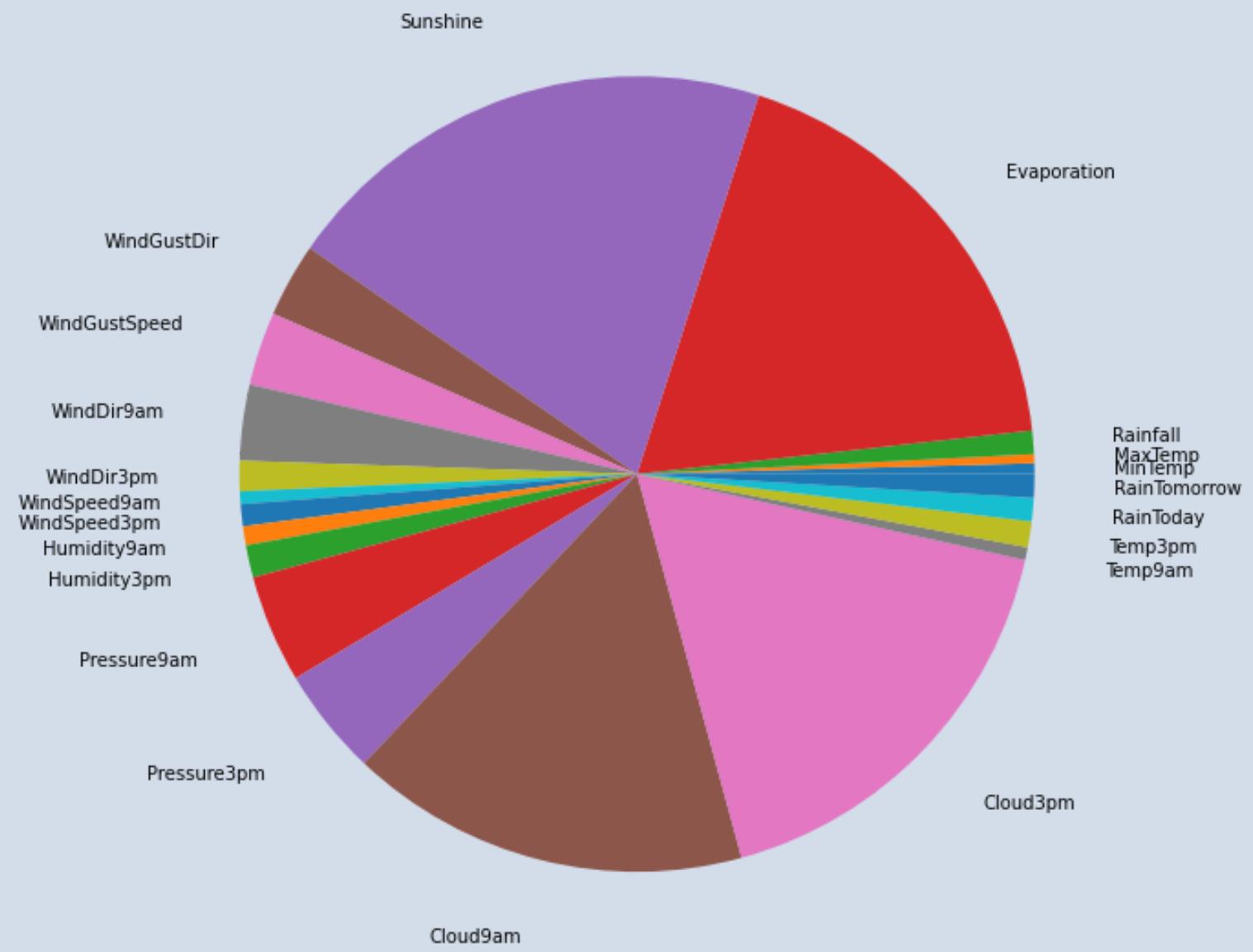


*Ivana Nestorović*

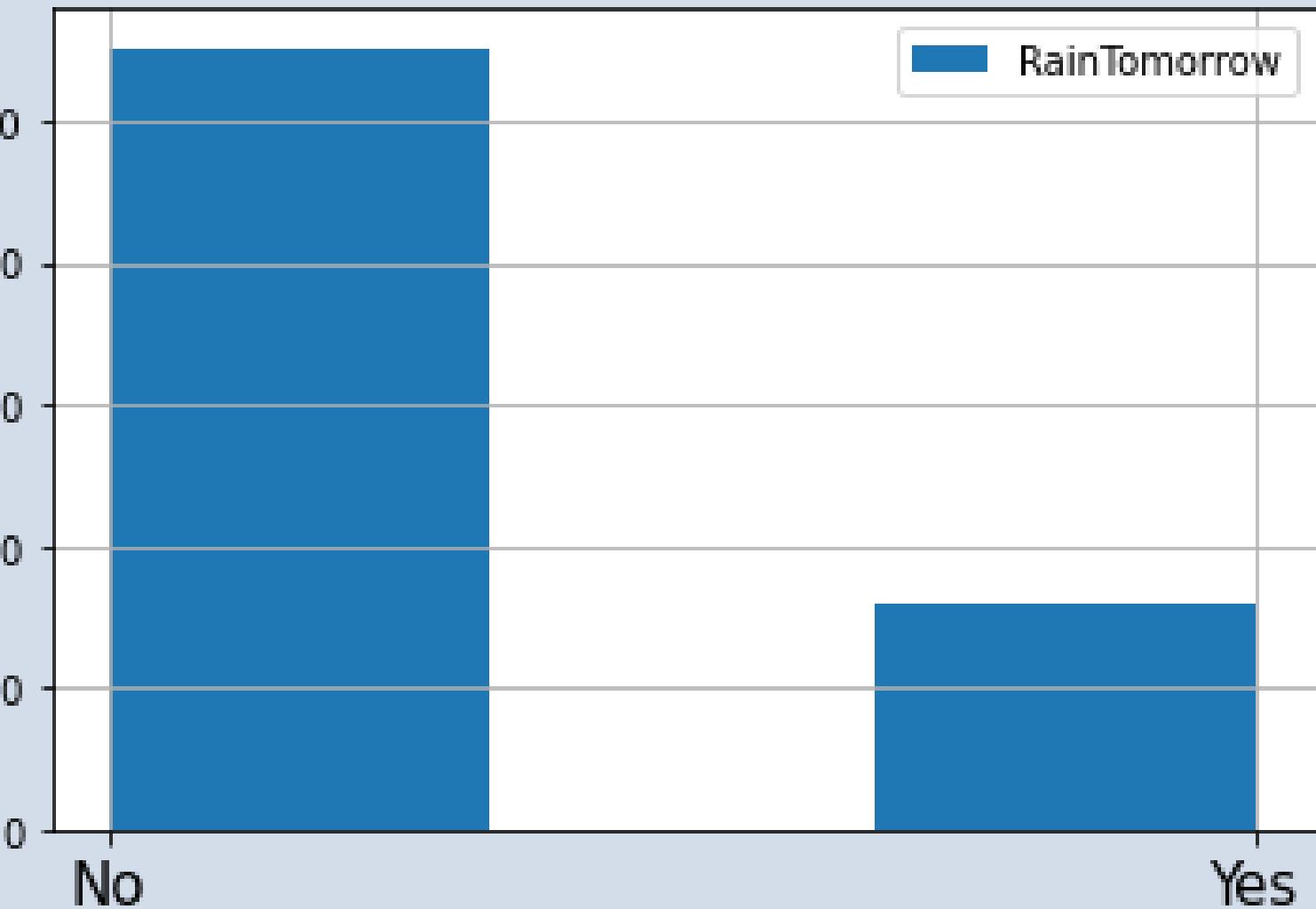
# Atributi

Atributi		Atributi		Atributi		Atributi	
Date	Location	Minimal Temeprature	Maximal Temeprature	Rainfall	Evaporation	Sunshine	Sunshine
							
Wind dir.	Wind speed	Wind dir.	Wind dir.	Wind speed	Wind speed	Humidity	Humidity
Max	Max	9am	3pm	9am	3pm	9am	3pm
							
Pressure	Pressure	Cloud	Cloud	Temperature	Temperature	Rain	Rain
9am	3pm	9pm	3pm	9pm	9pm	Today	Tomorrow
							

# Nedostajuće vrednosti



# *Balansiranost ciljnog atributa*



## *Tipovi atributa*

```
7 kategorickih :  
['Date', 'Location', 'WindGustDir', 'WindDir9am', 'WindDir3pm', 'RainToday', 'RainTomorrow']  
  
16 numerickih :  
['MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine', 'WindGustSpeed', 'WindSpeed9am', 'WindSpeed3pm', 'Hu  
midity9am', 'Humidity3pm', 'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am', 'Temp3pm']
```

# Stablo odlučivanja

## Bez podešavanja hiperparametara

```
Classification report for model DecisionTreeClassifier on training data
-----
precision    recall   f1-score  support
0           1.00    1.00     1.00    84993
1           1.00    1.00     1.00    24544

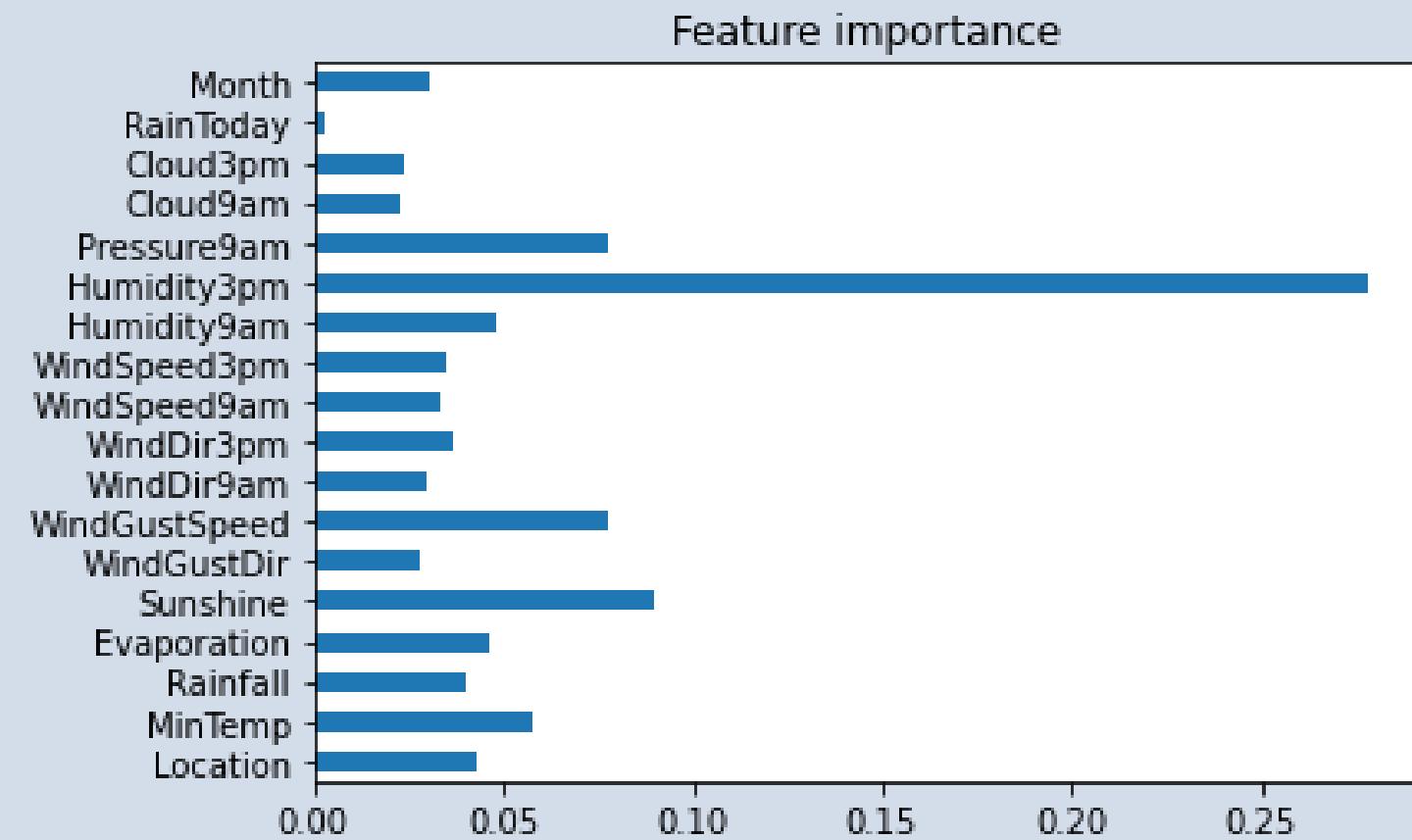
accuracy                           1.00    109537
macro avg      1.00    1.00     1.00    109537
weighted avg    1.00    1.00     1.00    109537

Confusion matrix for model DecisionTreeClassifier on training data
-----
```

```
Classification report for model DecisionTreeClassifier on test data
-----
precision    recall   f1-score  support
0           0.86    0.86     0.86    22064
1           0.52    0.51     0.51    6375

accuracy                           0.78    28439
macro avg      0.69    0.69     0.69    28439
weighted avg    0.78    0.78     0.78    28439

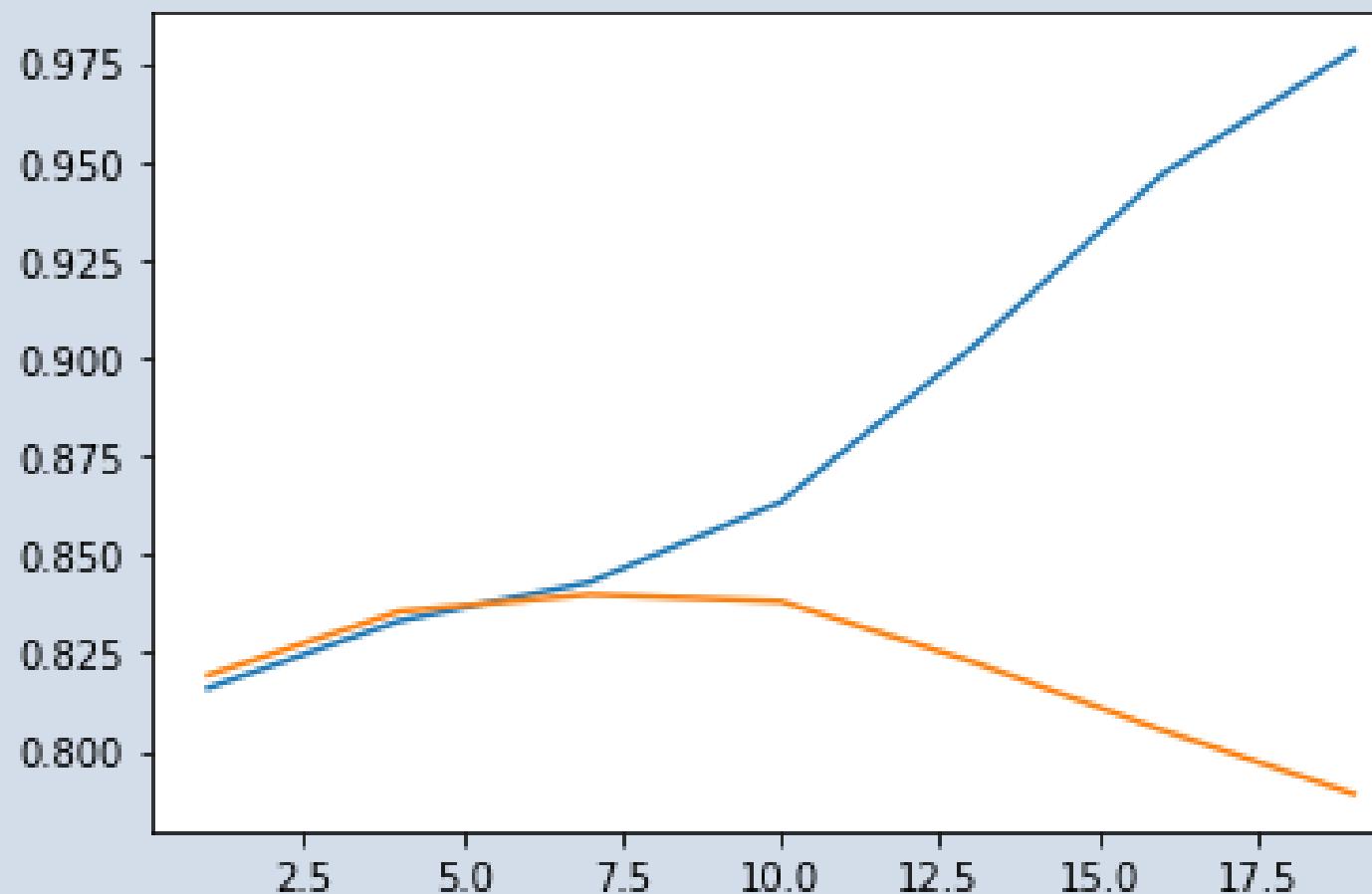
Confusion matrix for model DecisionTreeClassifier on test data
-----
```



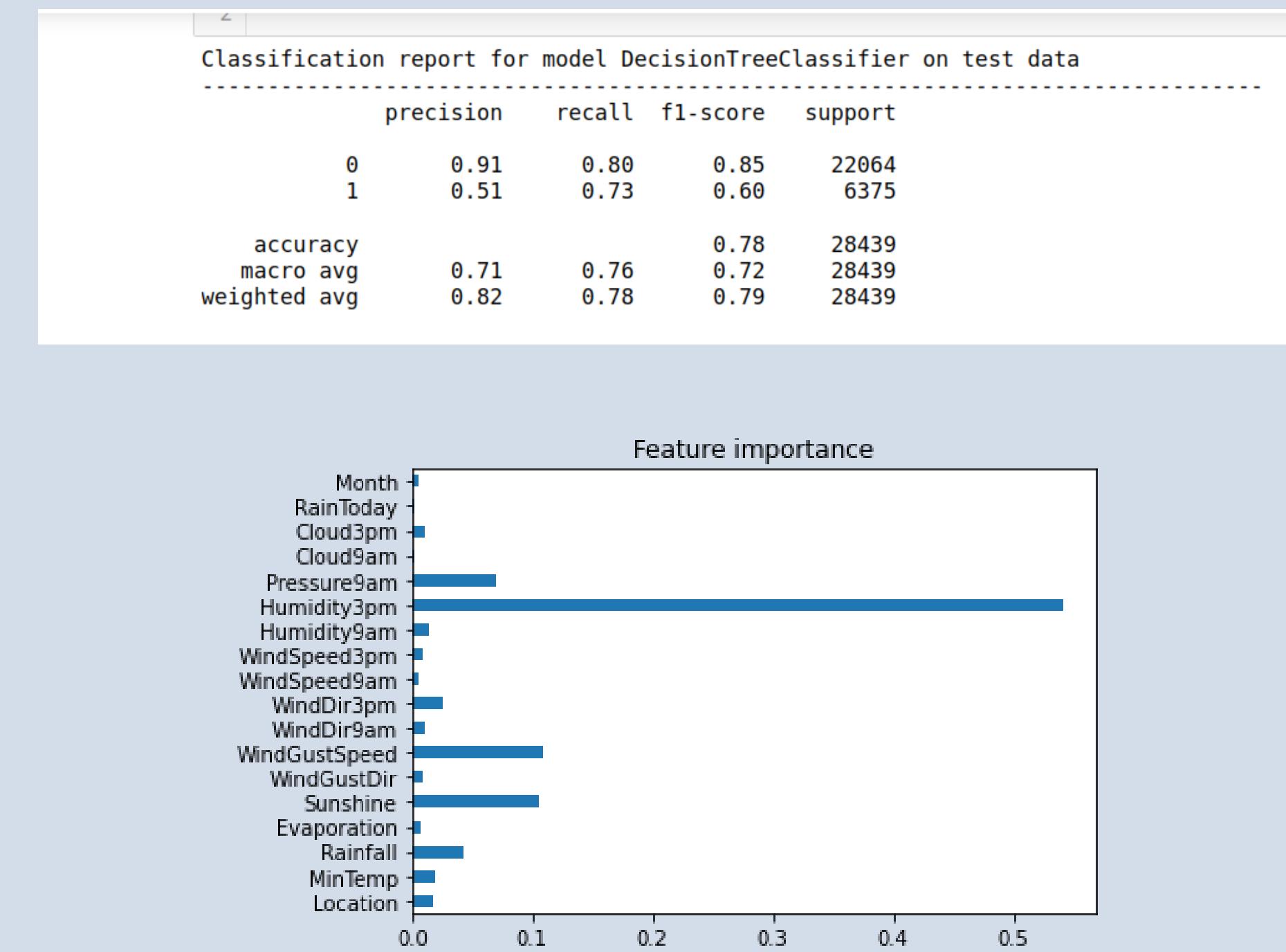
Dubina stabla 36  
Broj listova 15729

# *Stablo odlučivanja*

## GridSearchCV



```
: {'criterion': 'gini',
 'max_depth': 7,
 'min_samples_leaf': 5,
 'min_samples_split': 2,
 'random_state': 42}
```



# *K najблиžih suseda*

Classification report for model KNeighborsClassifier on test data

	precision	recall	f1-score	support
0	0.91	0.78	0.84	27580
1	0.49	0.75	0.59	7969
accuracy			0.77	35549
macro avg	0.70	0.76	0.72	35549
weighted avg	0.82	0.77	0.78	35549

Confusion matrix for model KNeighborsClassifier on test data

	No	Yes
No	21387	6193
Yes	2026	5943

Classification report for model KNeighborsClassifier on test data

	precision	recall	f1-score	support
0	0.86	0.94	0.90	27580
1	0.69	0.49	0.57	7969
accuracy			0.84	35549
macro avg	0.77	0.71	0.73	35549
weighted avg	0.82	0.84	0.82	35549

Confusion matrix for model KNeighborsClassifier on test data

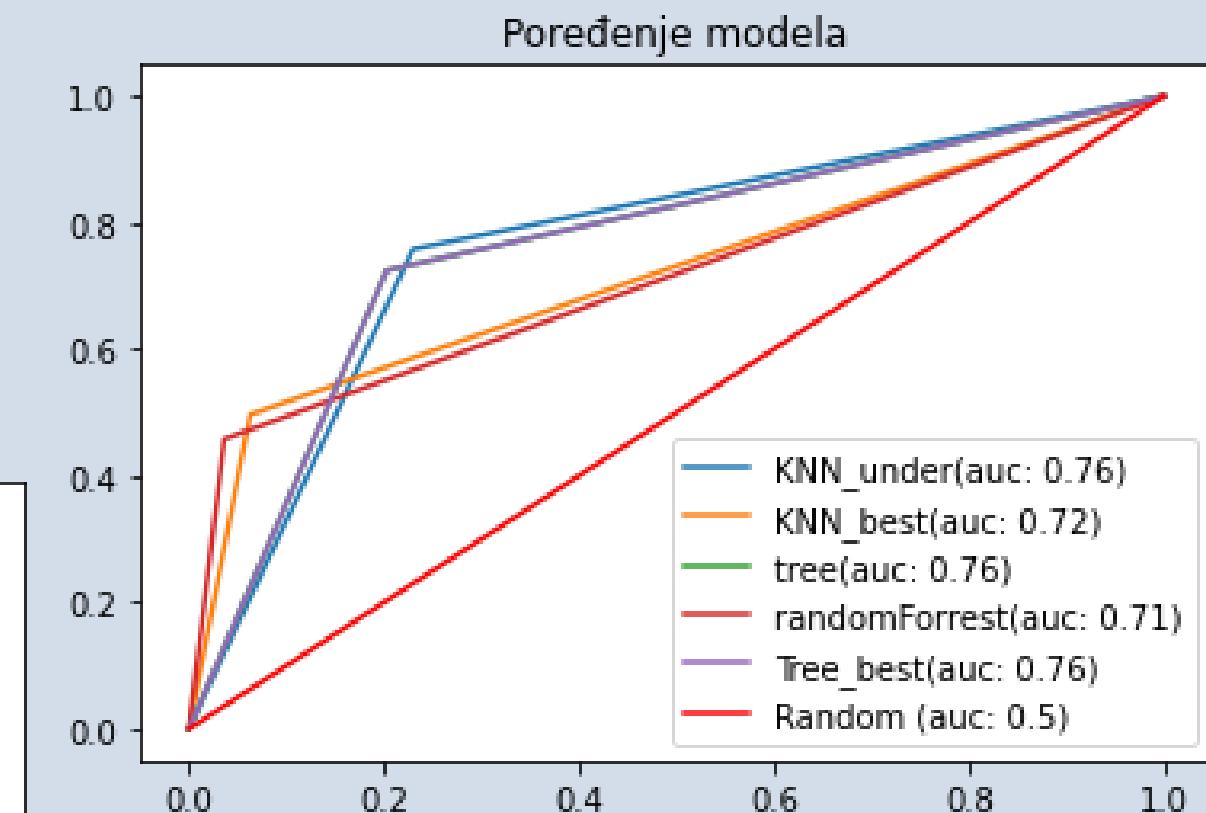
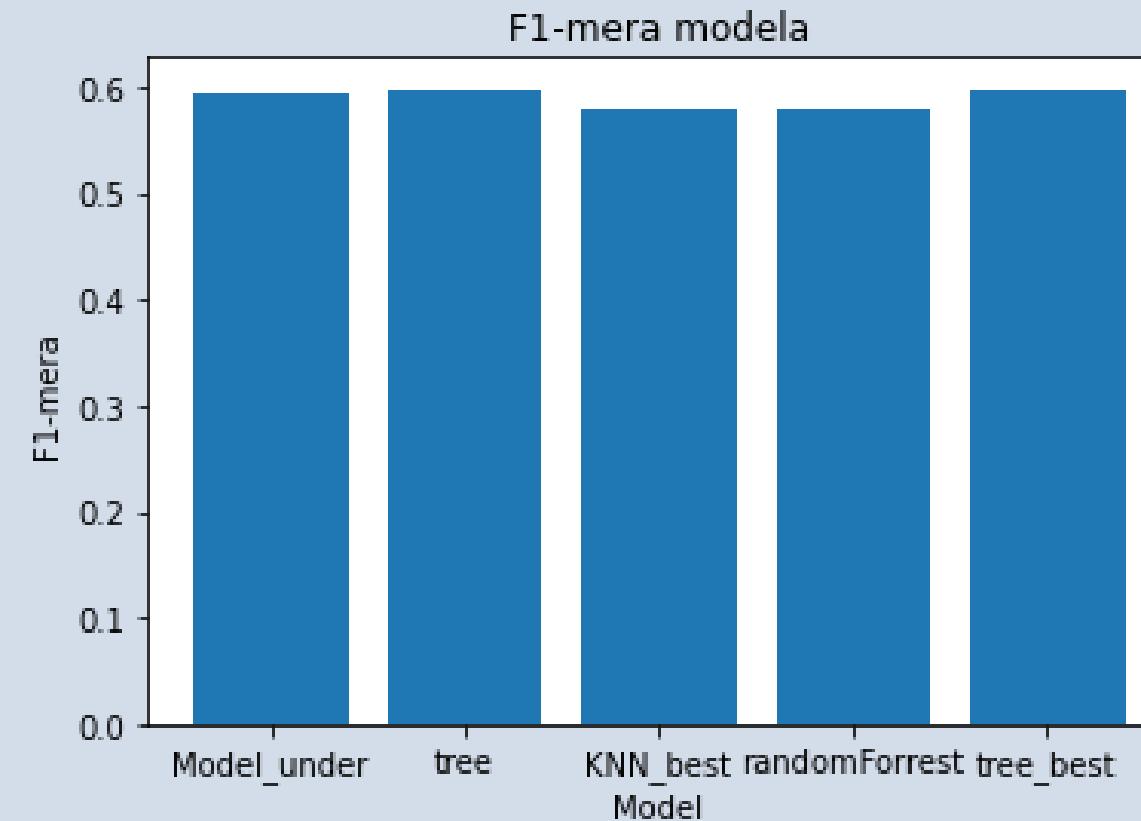
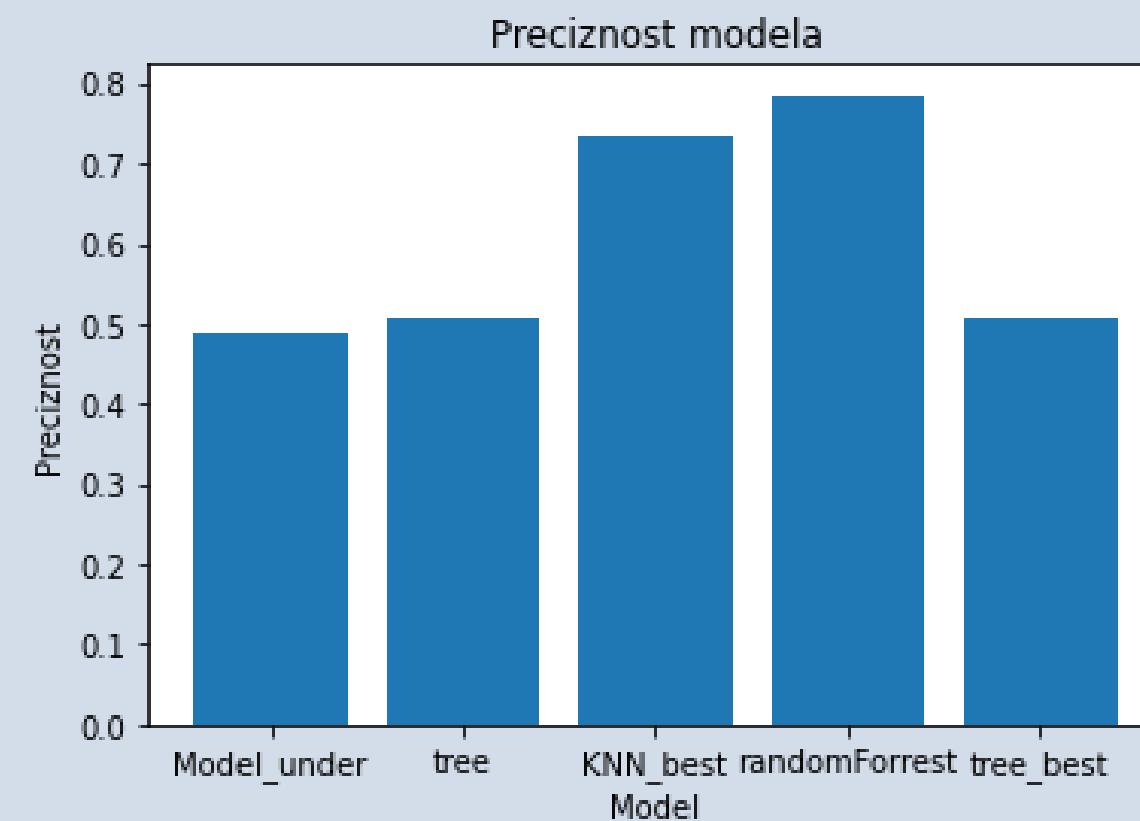
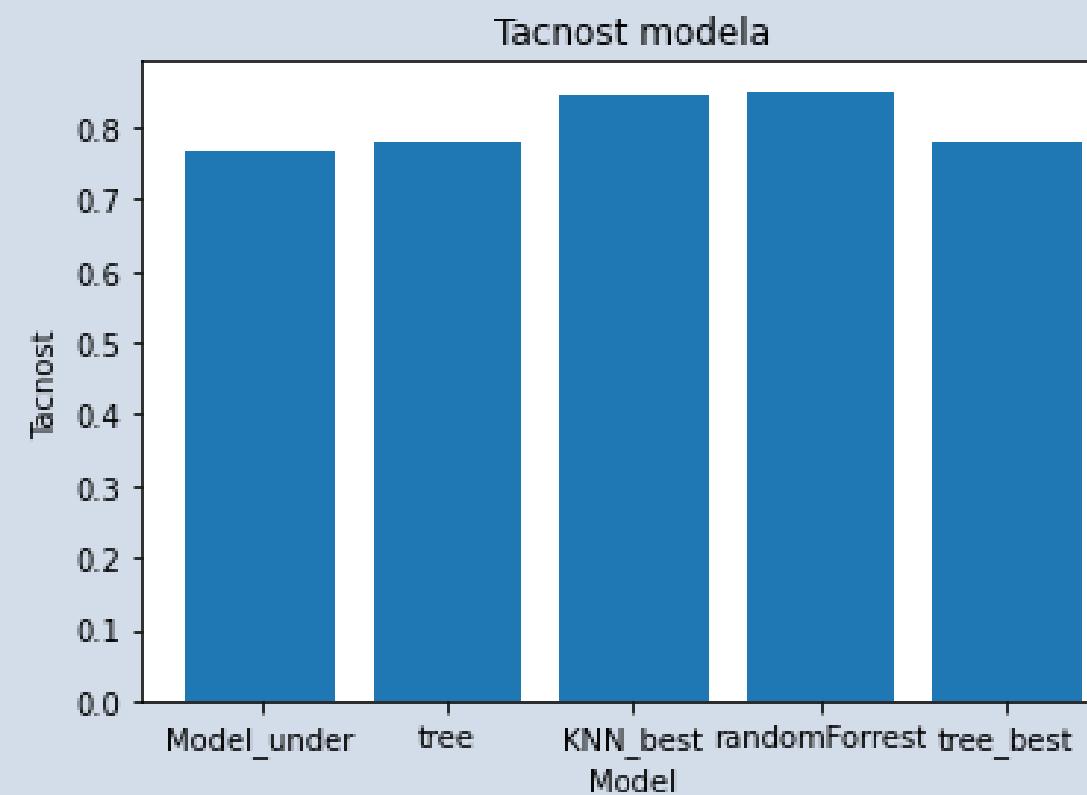
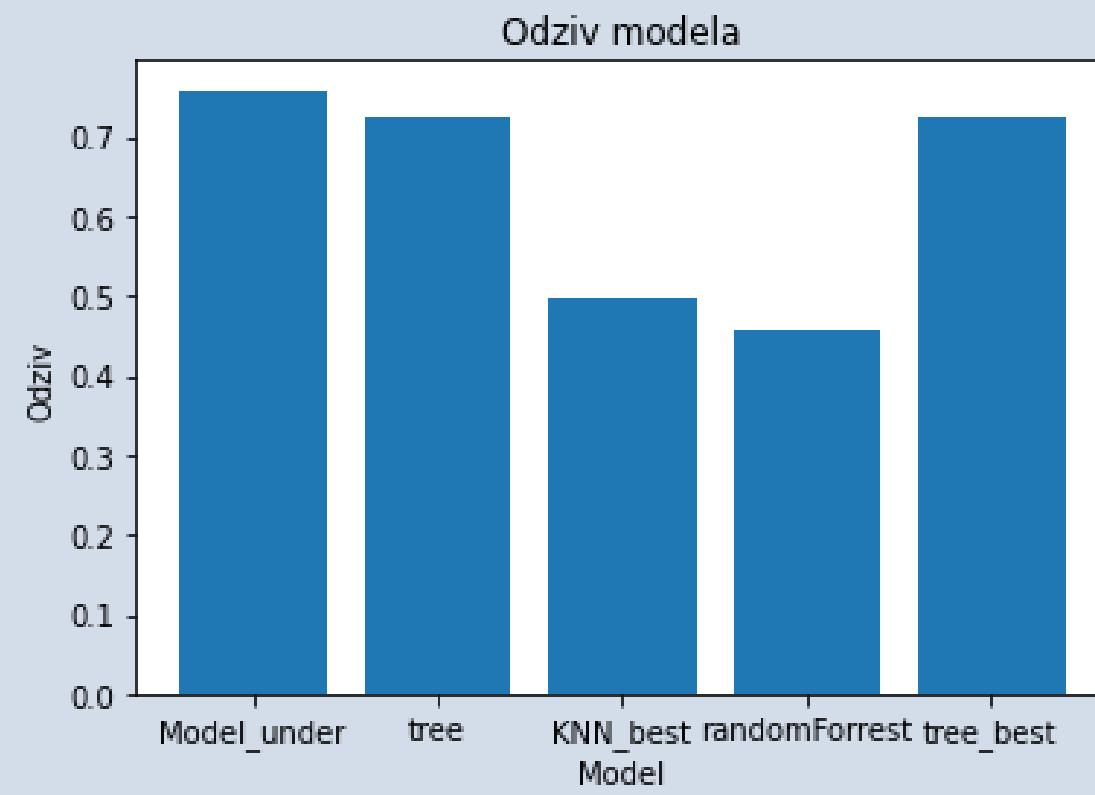
	No	Yes
No	25796	1784
Yes	4068	3901

**KNN bez balansiranja**

**KNN sa  
RandomUnderSampler**

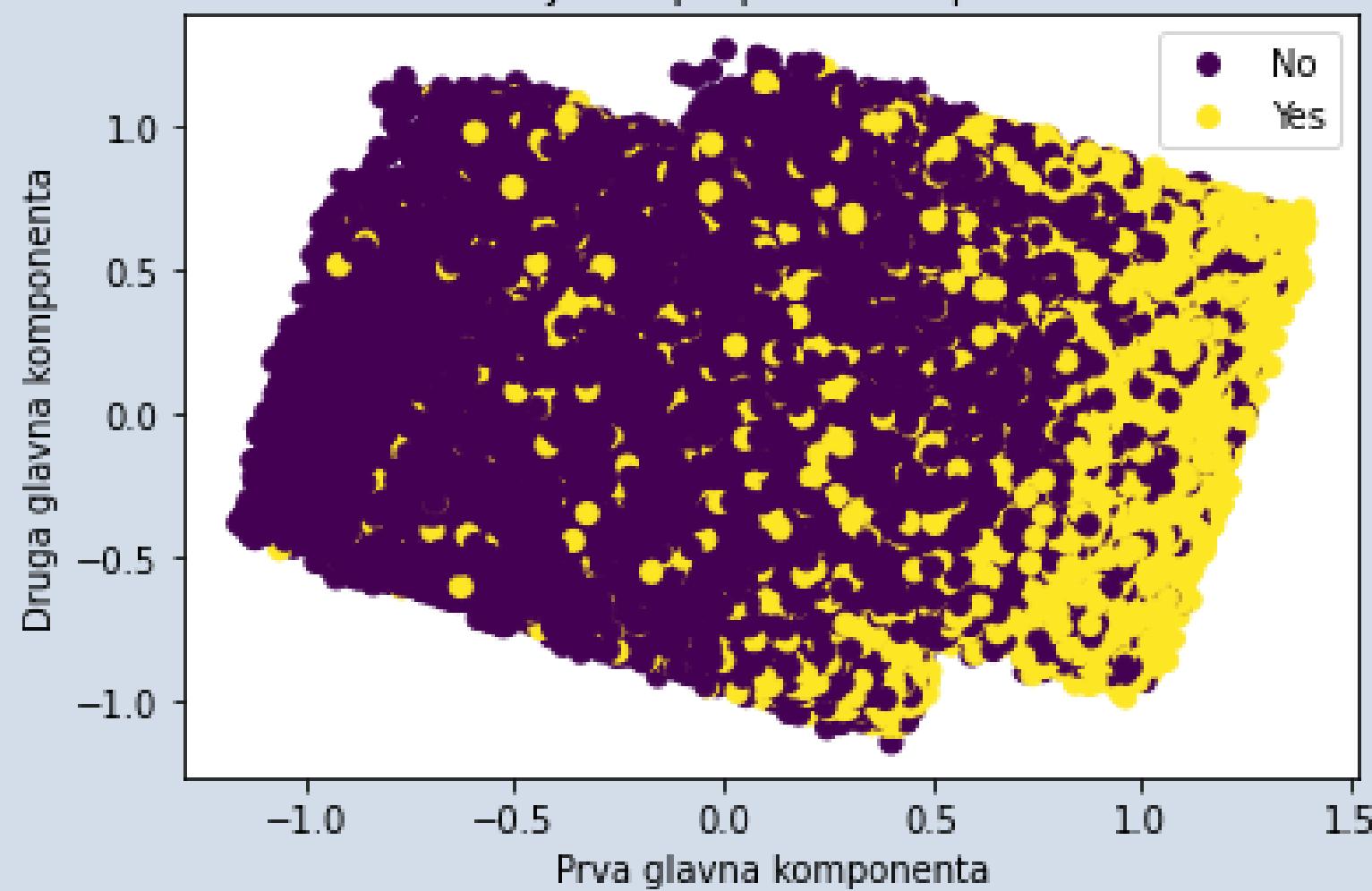
```
{'metric': 'manhattan', 'n_neighbors': 5, 'weights': 'distance'}
```

# Poredjenje modela klasifikacije

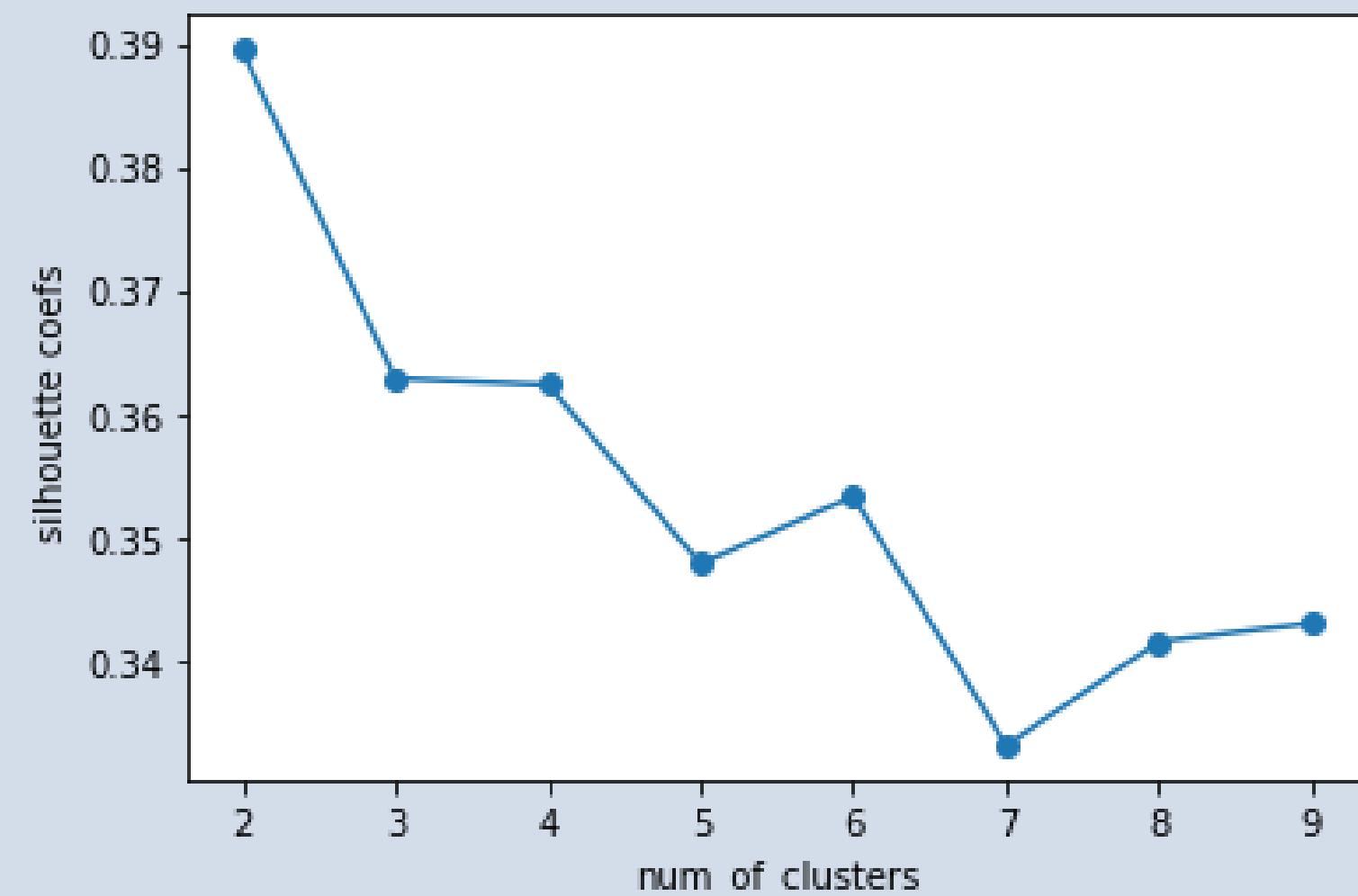


# Kmeans

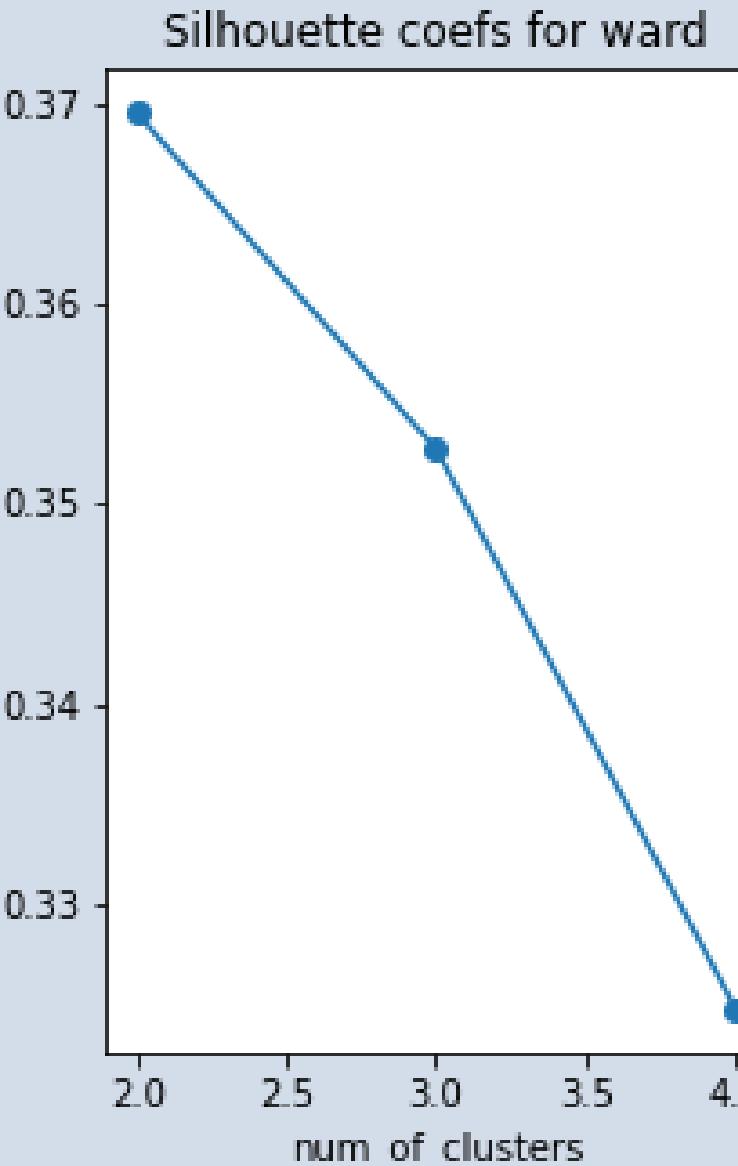
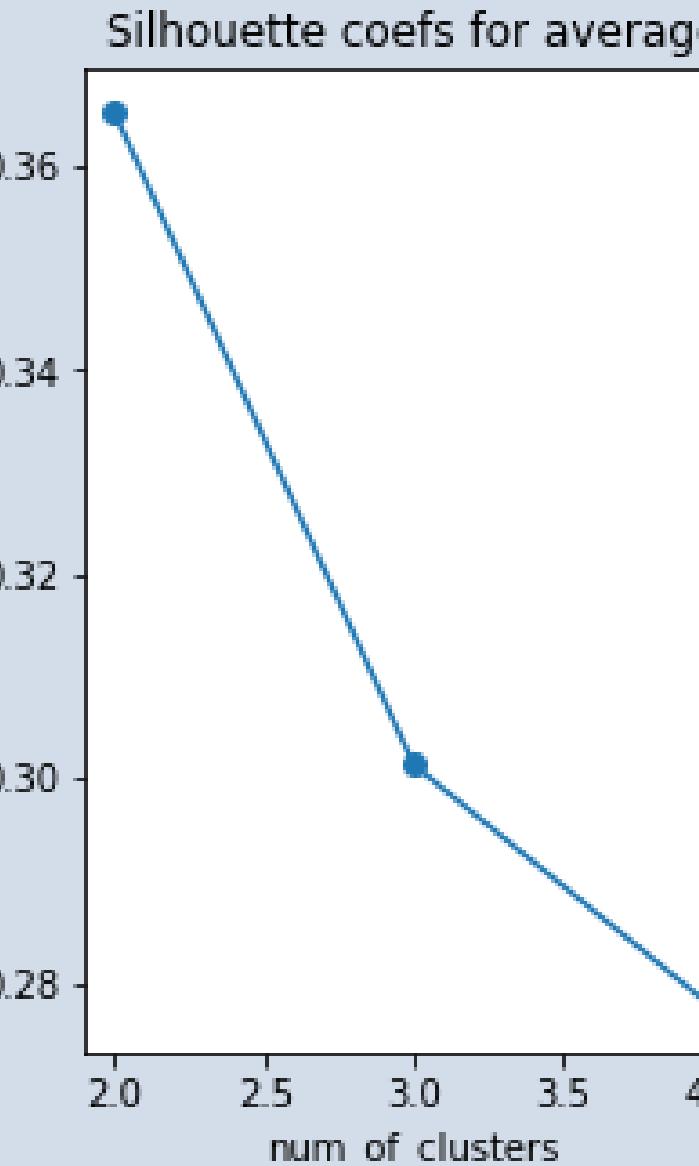
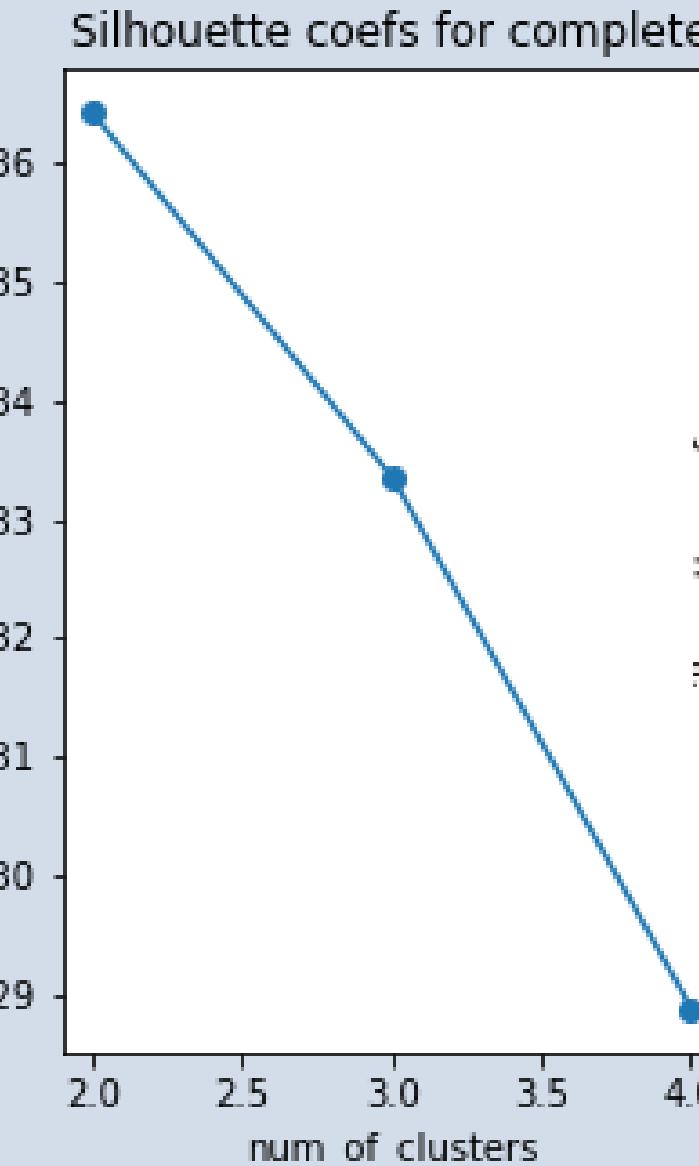
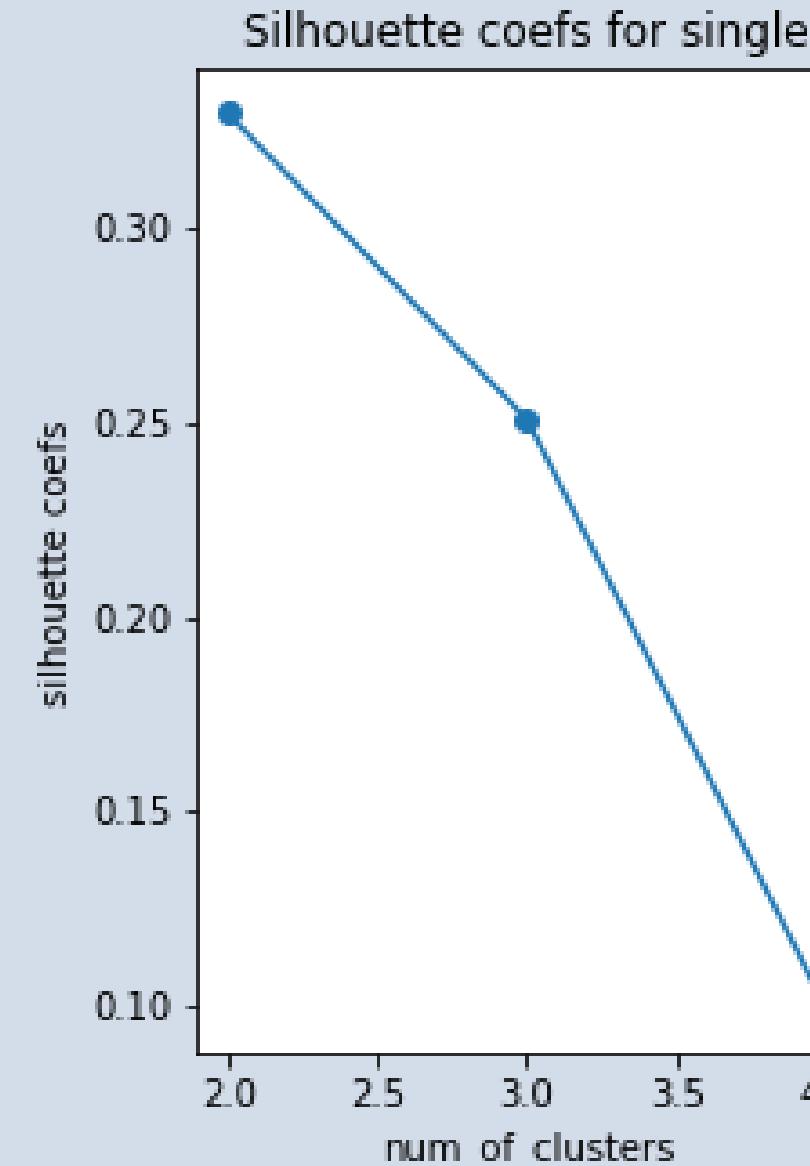
Vizuelizacija skupa podataka pomoću PCA



Silhouette coefs



# Hijerarhijsko klasterovanje



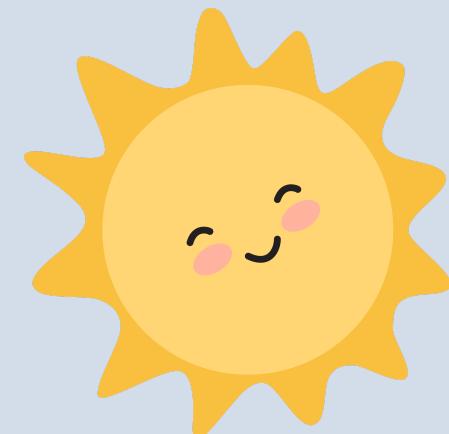
**Na osnovu ovog istraživanja, dolazimo do zaključka da sa datim atributima nije lako predvideti da li će sledećeg dana padati kiša.**

**Svi modeli sa velikom tačnošću predvidjaju da narednog dana neće biti padavina, ali problem nastaje pri predvidjanju kišnog dana.**

**Iz matrica konfuzije možemo doneti zaključak da naši modeli često pogrešno klasifikuju stvarno negativne instance kao pozitivne, što znači da iako smo predvideli da će biti padavina, postoji mogućnost da ipak neće.**

**Za preciznije predvidjanje potrebni su nam još neki atributi kao što su na primer smer i brzina kretanja oblaka.**

**Ukratko, ako se vodite našom prognozom, i za naredni dan nije predvidjena kiša, velika je šansa da vas neće iznenaditi. Ali ukoliko su najavljenе padavine, imamo lepe vesti za vas, postoji mogućnost da vam kišobran neće biti potreban!**





*Hvala na pažnji!*