DIMITRA MANOURA

SPOTIFY GENRE CLASSIFICATION

DATA PREPROCESSING

- Our Dataset contains features like Instrumentalness, Acousticness, Danceability, Key, Mode etc. Total: 232,725 tracks
- Firstly we keep only the genres we want to classify, in our case: Ska, Hip-Hop and Classical. Total: 27,425 tracks
- We make sure that there are no nulls and no duplicated samples
- Now we can drop the features we don't need, like track_id, artist song name

DATA PREPROCESSING

- Then, we convert strings to integers to all the features that have strings, like key, mode and finally our label, genre.

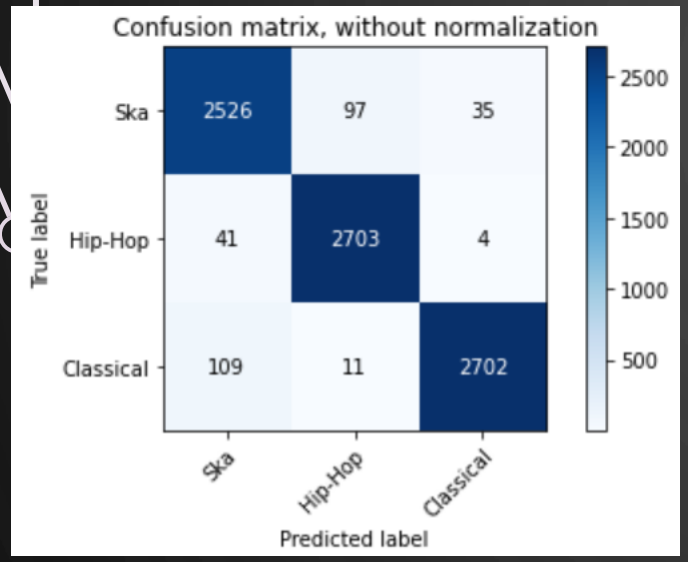
- We make sure that out Dataset is balanced by checking the quantity of each genre.
- Now it's time to make a correlation heatmap. We decide to exclude the feature 'key'.

CLASSIFICATION

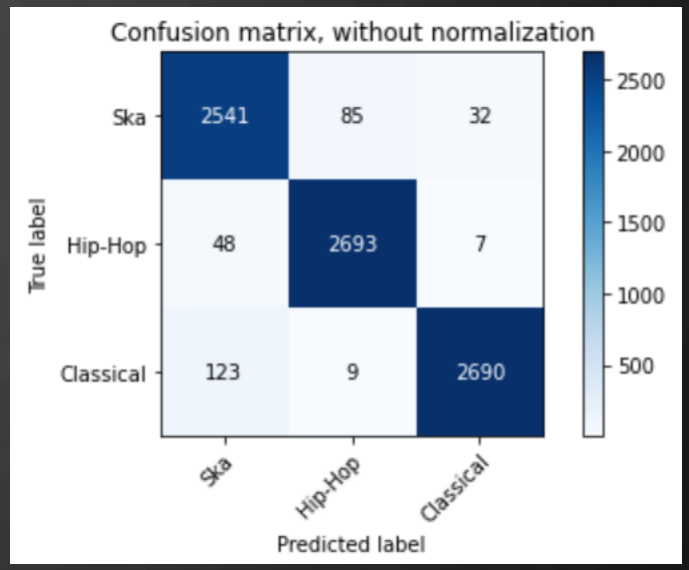
- First we split our Dataset in train set and test set
- Then, we try 4 classifiers that we have tuned
- We calculate scores and plot a confusion matrix for each classifier
- We perform cross-validation to evaluate our models' performance

		Random Forest	SVM	KNN	GaussianNB
Precision	macro avg	0.96	0.96	0.96	0.94
	weighted avg	0.96	0.96	0.96	0.94
Recall	macro avg	0.96	0.96	0.96	0.93
	weighted avg	0.96	0.96	0.96	0.94
F1-score	macro avg	0.96	0.96	0.96	0.93
	weighted avg	0.96	0.96	0.96	0.94
Accuracy		0.96	0.96	0.96	0.94

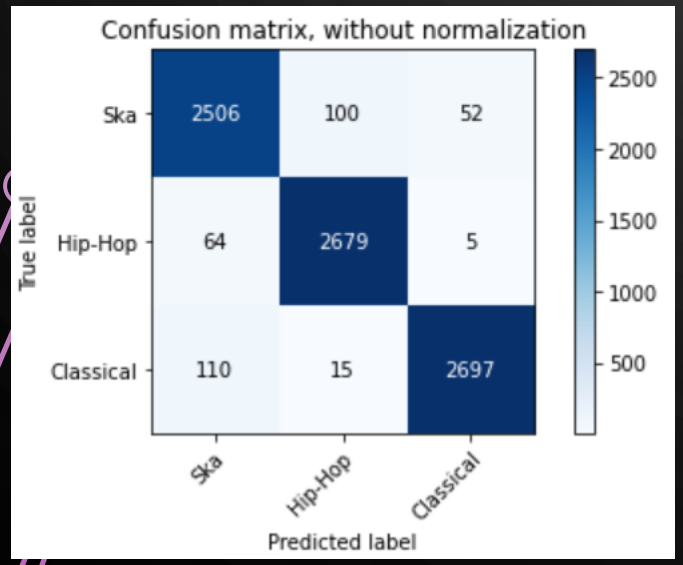
RANDOM FOREST



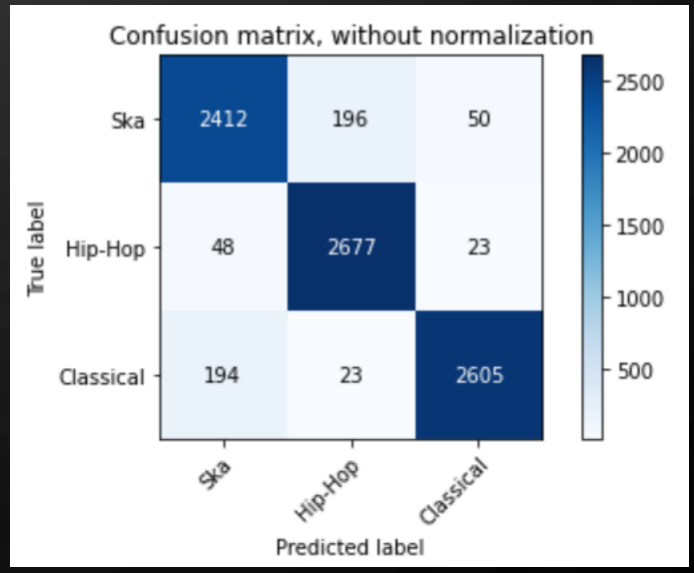
SVM



KNN



GNB



CONCLUSIONS

- Overall the performance of our classifiers' was satisfying.
- Interestingly enough, if we were to pinpoint to a mistake of our classifiers, it would be that they all tend to misclassify Classical as Ska and Ska as Hip-Hop.
- Future work: Classify more genres

THANK YOU