



MSc in Artificial Intelligence

Machine Learning

Genre classification and popularity prediction based on album cover

Bochalis Christodoulos

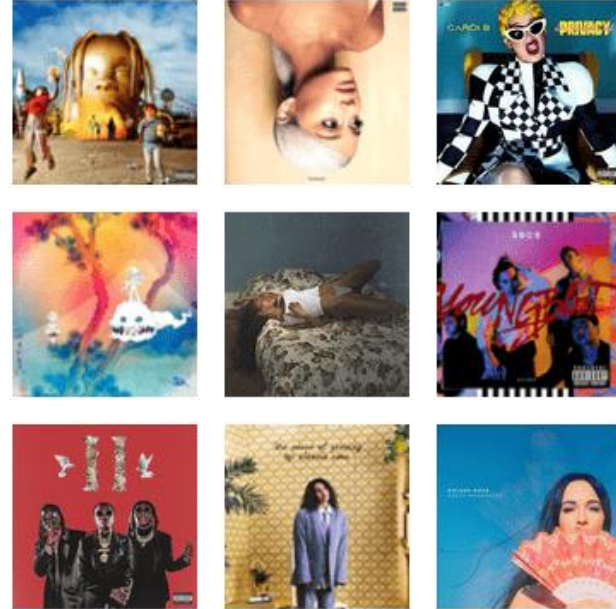
February 16, 2021

Project Idea

Judge an album by it's cover?

Using Album Cover Art:

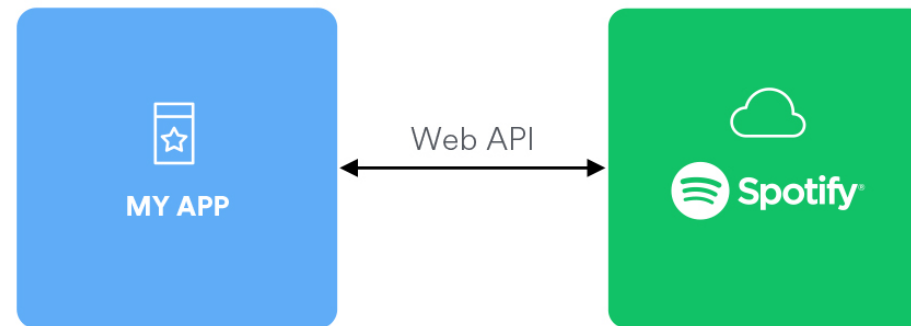
- Classify album genre
- Predict Song Popularity



Get Data

Use Spotify API [link](#) to build the dataset

- The Client Credentials Flow is used for authentication
- Various endpoints requests
 - Search queries
 - Get Album Art (various sizes available)
 - Get Album Genre, Popularity etc.



Get Data

The playlist [*Most Popular Songs of All Time*](#) from Spotify was used.

3.300 songs and **204 hours** of music

Key points

- Removed duplicate albums
- Genres maintained at artist level
- The final dataset is stored in a json file

2.370 unique album covers

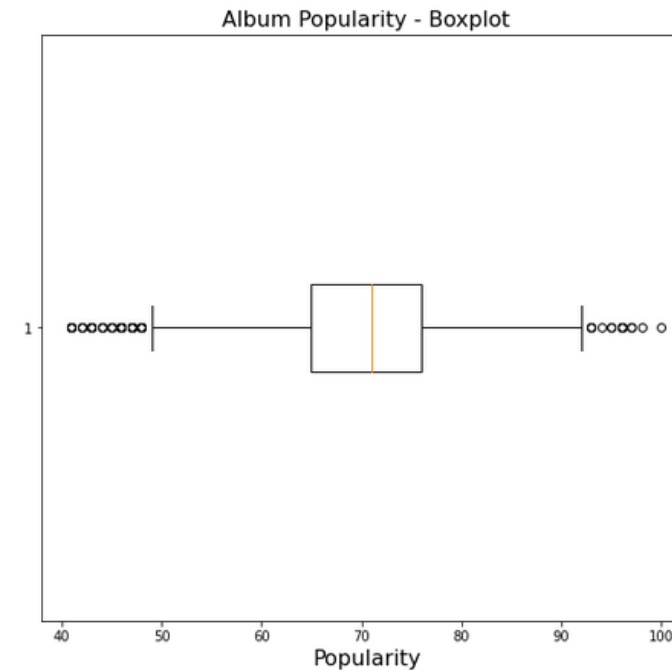
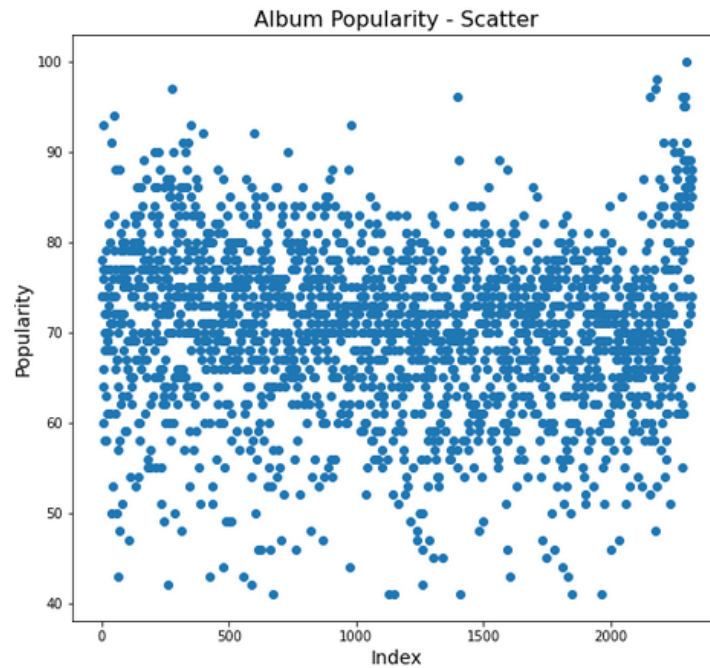
Get Data

A sample from the dataset:

```
▼ 0      {8}
  artist_id : 5me0Irg2ANcsgc93uaYrpb
  artist_name : The Notorious B.I.G.
  album_id : 7dRdaGSxgcBdJnrOviQRuB
  album_name : Life After Death (2014 Remastered Edition)
  album_popularity : 78
▼ genres [5]
  0 : east coast hip hop
  1 : gangster rap
  2 : hardcore hip hop
  3 : hip hop
  4 : rap
  album_url : https://i.scdn.co/image/ab67616d0000b273fde79b88e2a659c394c5ae30
  filepath : .\\images\\7dRdaGSxgcBdJnrOviQRuB.jpg
```

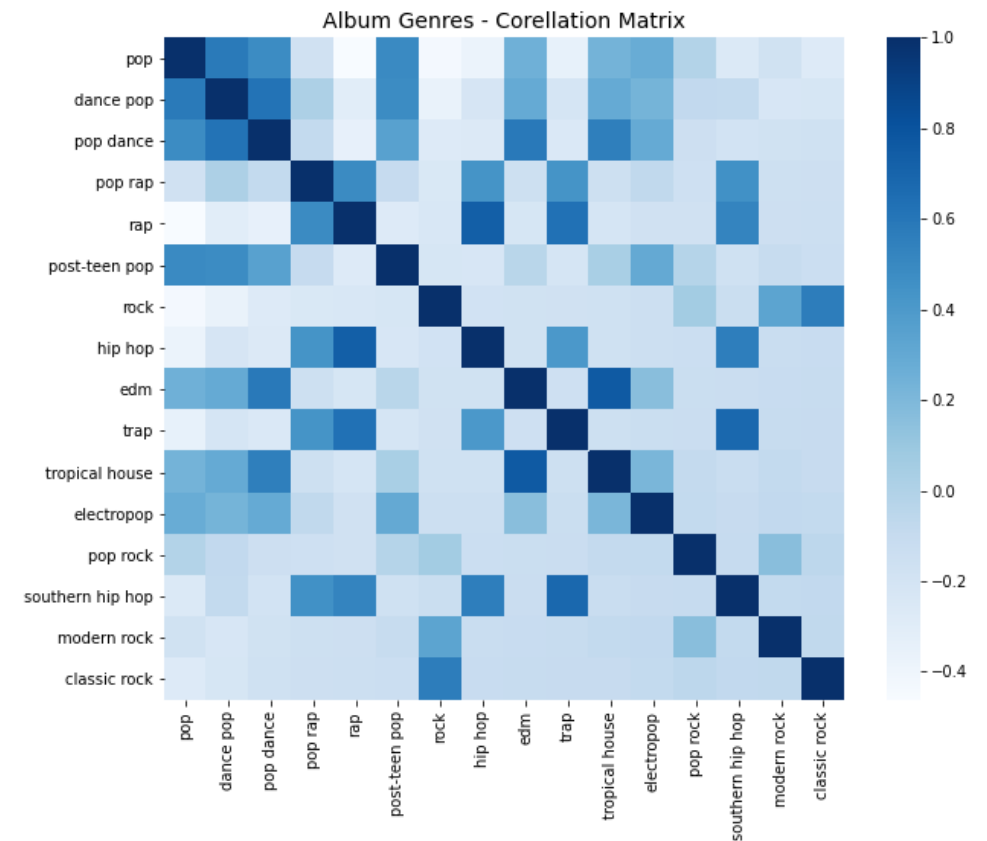
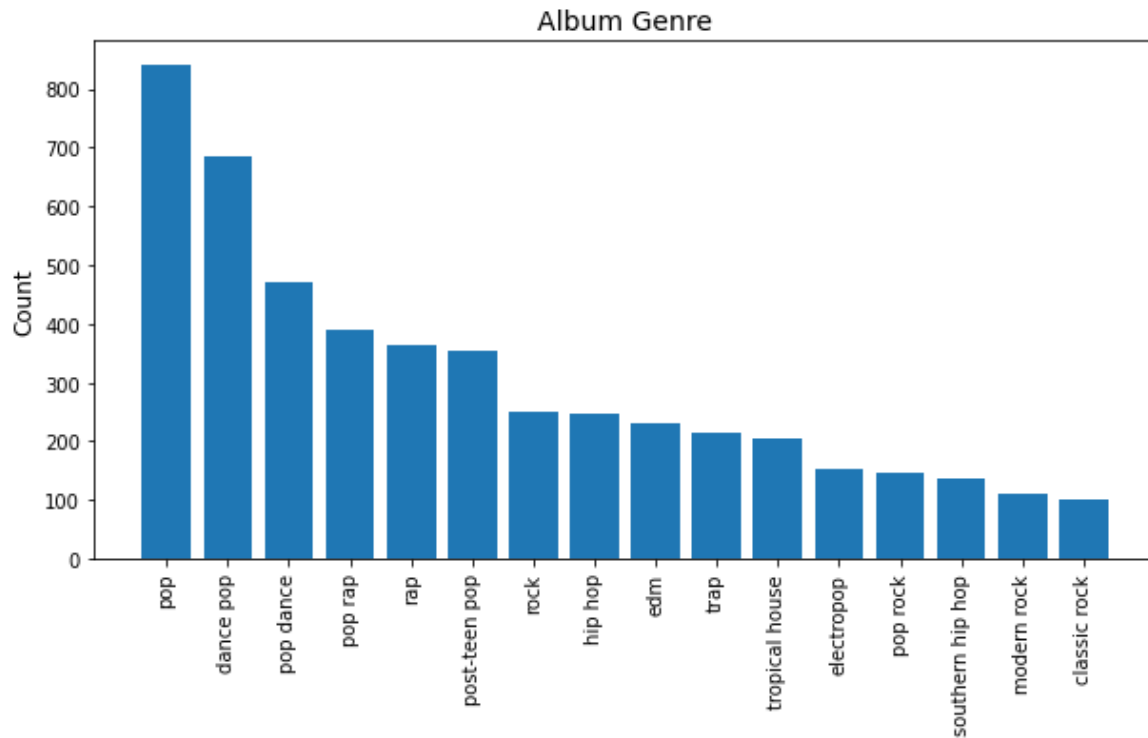
Prepare Data - Popularity

- Spotify API returns a popularity value between 0-100.
- Filter out albums with popularity less than 40.



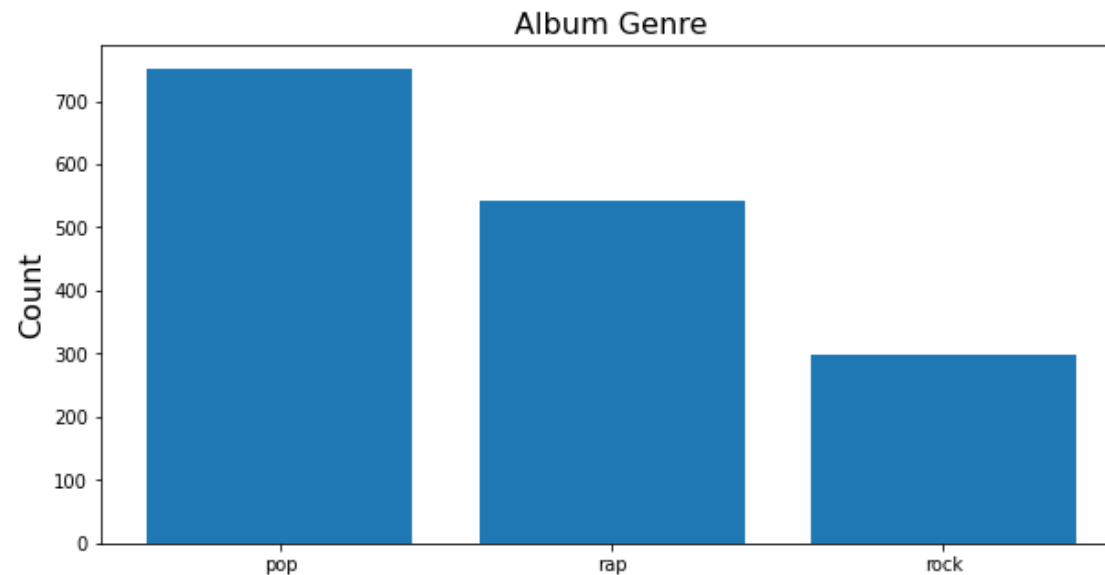
Prepare Data - Genres

- Spotify API returns **multiple genres per artist**.
- Keep the most common genres



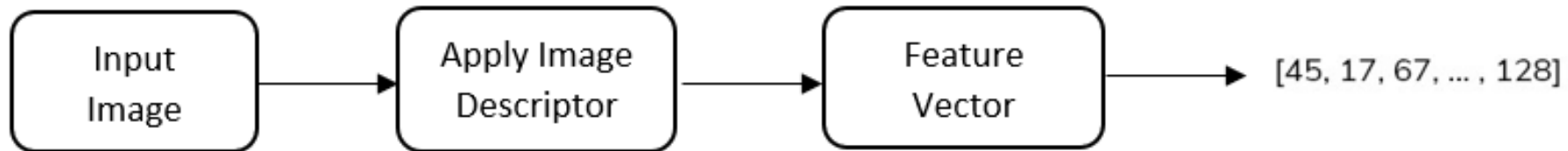
Explore Data - Genres

- The albums are annotated as **pop**, **rap** and **rock**.
- The albums that don't belong to one of these genres are removed.
- The final dataset contains **1.592 samples**

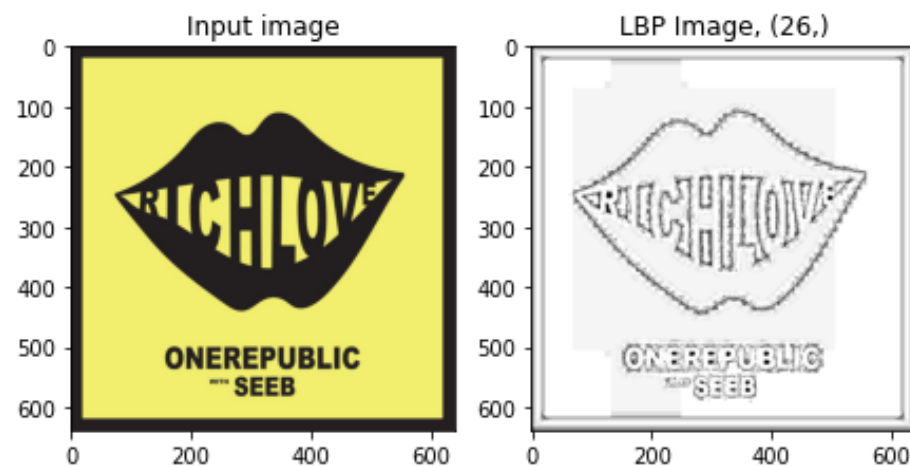
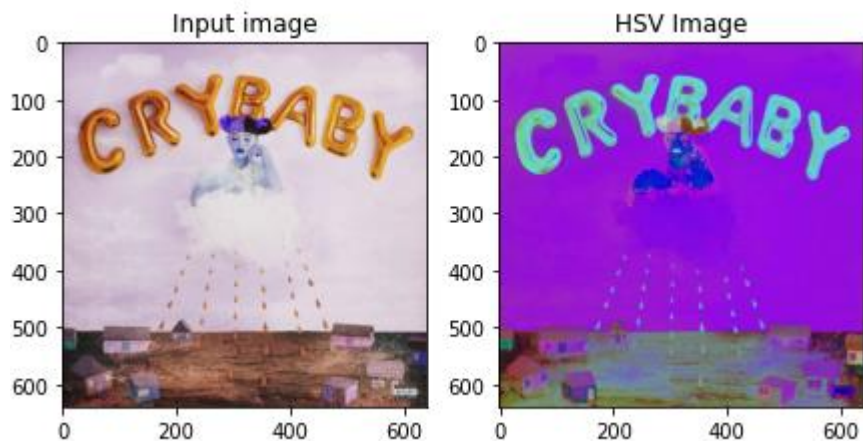
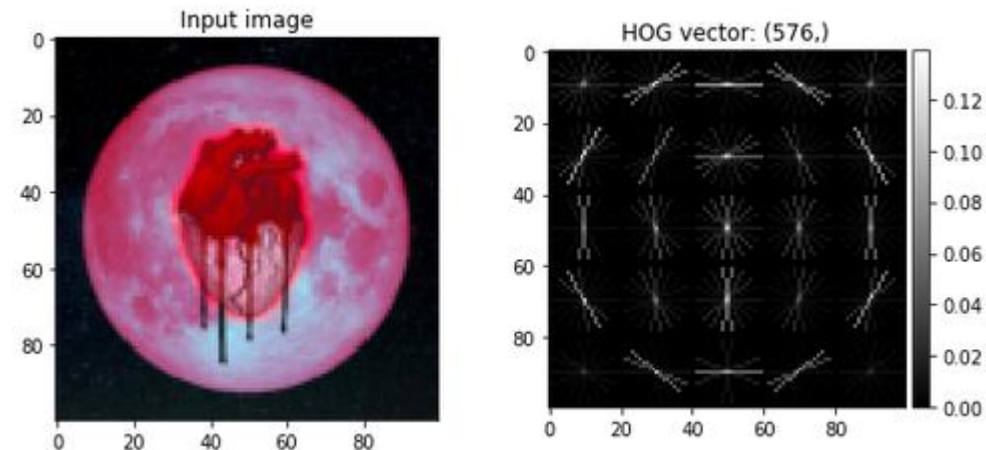
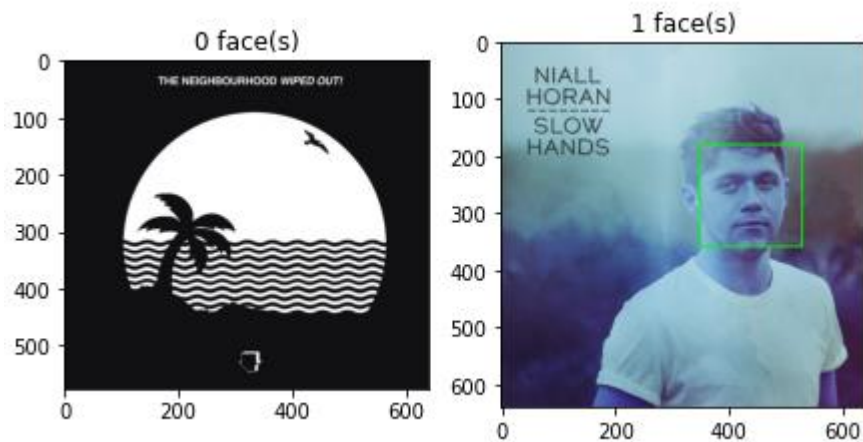


Extract Features

- The following features are extracted from each image:
- **has_face** : that indicates if a **face** is present in the album cover
- **hog_descriptor** : that quantifies the **structure** of the image
- **hsv_hist** : that quantifies the **color** of the image
- **lbp_hist** : that quantifies the **texture** of the image

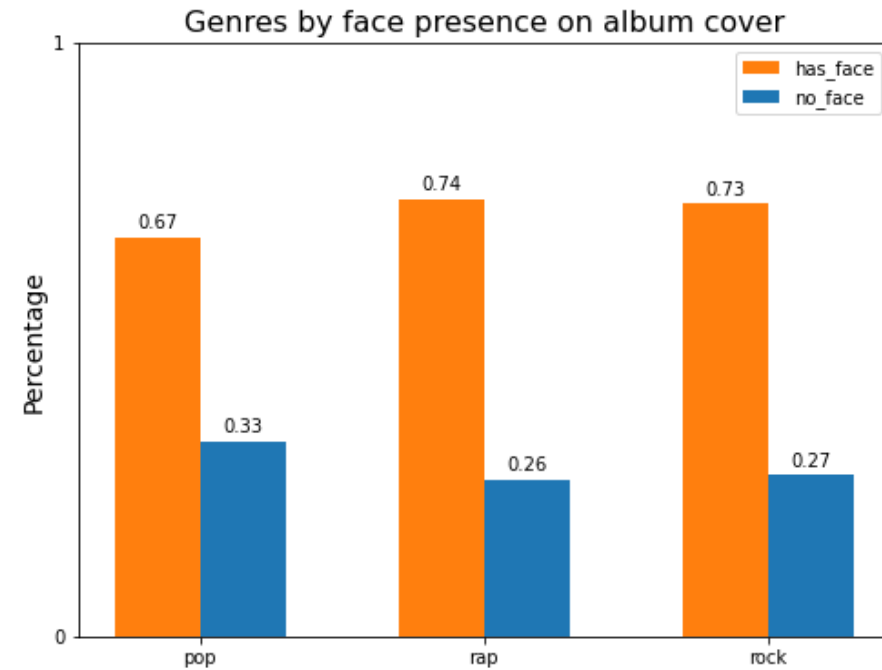
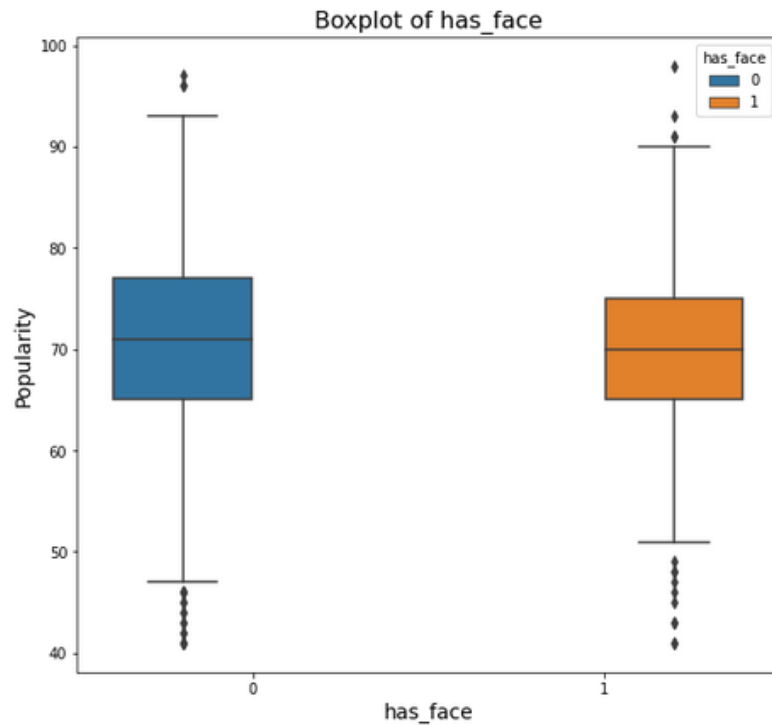


Extract Features



Extract Features

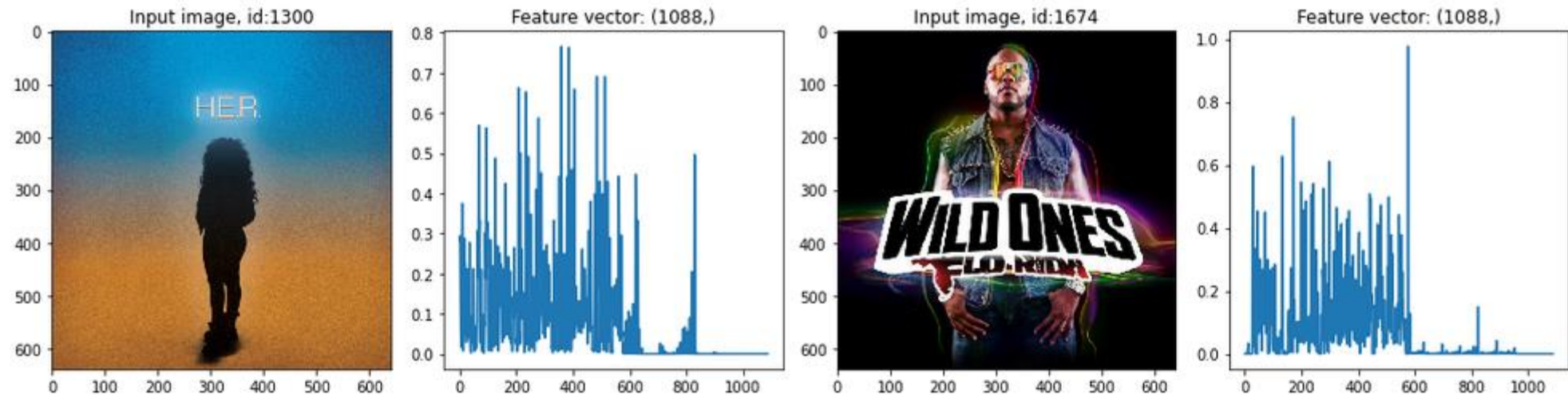
- Face presence on the album cover, has no correlation over popularity or genre



Extract Features

- After several tests, the following combination gave the best results

hog_descriptor + hsv_hist



Train Model

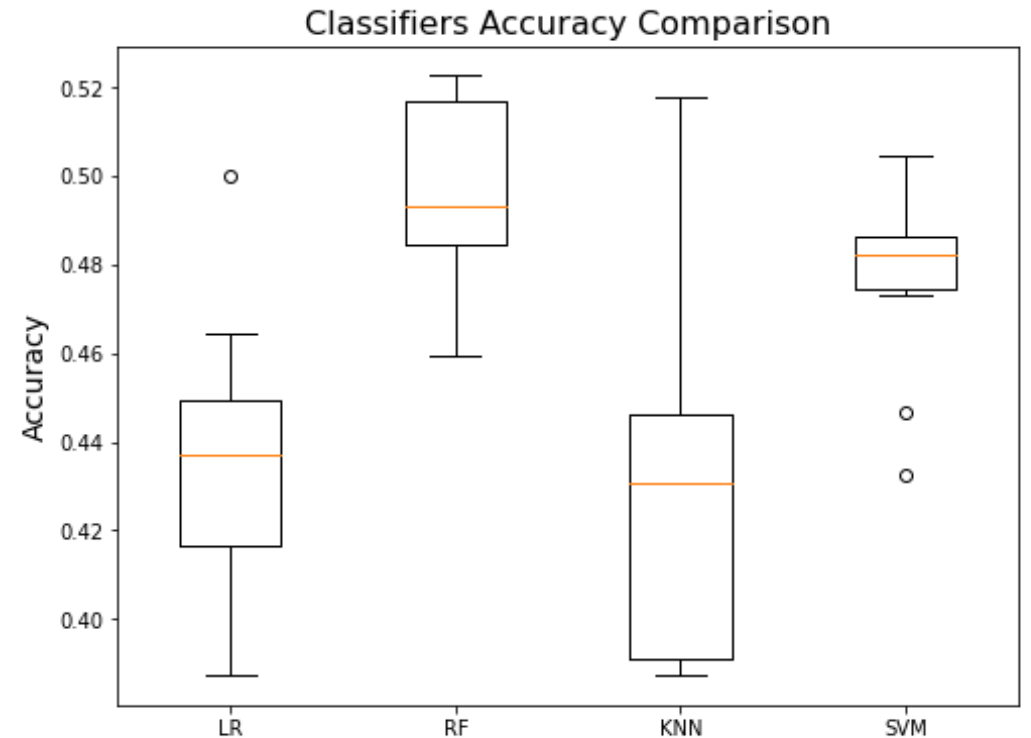
- Test Dataset size 20%
- Problem with MinMax scaler from sklearn
- The following models tested using Stratified K-Fold cross validation:
 - Logistic Regression
 - Random Forest Classifier
 - Kneighbors Classifier
 - SVM
- More samples from pop class and low number of samples overall



Tune Model

Grid Search on various parameters for the best performing models:

- Random Forest Classifier
 - the number of trees
 - number of random features to sample at each split point
- SVM
 - kernel
 - penalty C



Validate Model

Random Forest Classifier

Train Data (overfit)

	precision	recall	f1-score	support
pop	1.00	1.00	1.00	601
rap	1.00	1.00	1.00	433
rock	1.00	1.00	1.00	239
accuracy			1.00	1273
macro avg	1.00	1.00	1.00	1273
weighted avg	1.00	1.00	1.00	1273

Test Data

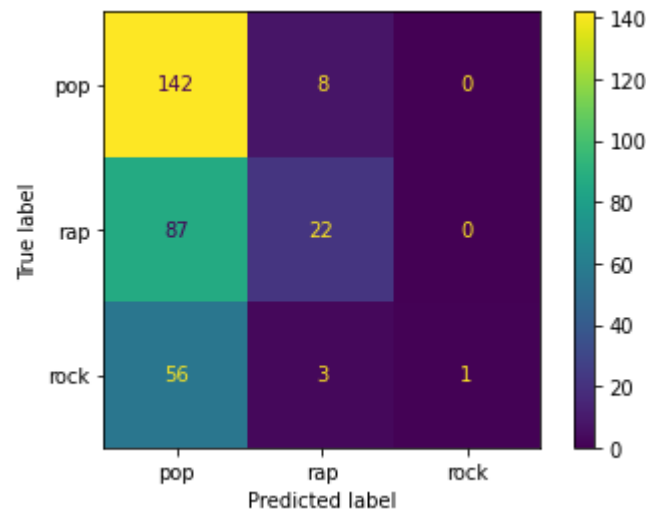
	precision	recall	f1-score	support
pop	0.50	0.95	0.65	150
rap	0.67	0.20	0.31	109
rock	1.00	0.02	0.03	60
accuracy			0.52	319
macro avg	0.72	0.39	0.33	319
weighted avg	0.65	0.52	0.42	319

Validate Model

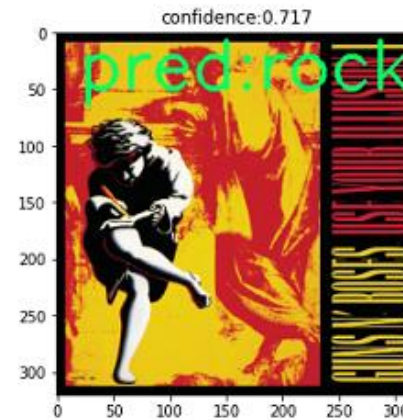
Random Forest Classifier gives the best overall accuracy but **overfits** the train dataset.

Our model has **Low Bias** but **High Variance** and does not generalize well.

The most notable result is the F1-Score for pop genre: 65%



Sample Execution on Test Data



Next Steps

To address the overfitting issue, the following steps can be performed:

- Collect more data for each class. Try 500-1000 images per class.
- Data augmentation using various transformations on the dataset images.
- Use also Local Image features with the Bag of Visual Words(BOVW) technique. For example SIFT, SURF, KAZE.
- Consider using a dimensionality reduction technique.

Thank you

Questions?

```
git clone https://github.com/bojito/ml-album-popularity.git
```