Music recommendation system

112001054 - Pratham Nagpure
112001028 - Nideesh N
112001033 - Priyanshu Gupta

Problem statement

Music recommendations system -

Given a .wav audio file we will recommend songs that are similar to the audio. We consider two songs to be similar if they are of the same genre. We are classifying into 10 genres - 'hip hop', 'disco', 'blues', 'pop', 'reggae', 'jazz', 'country', 'metal', 'rock', 'classical'

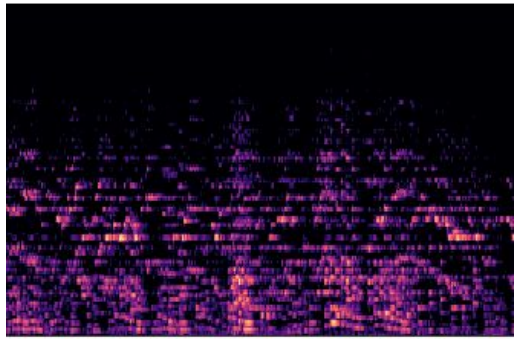
Dataset

We have used the GTZAN dataset from Kaggle. It has two parts ,

- genres original - A collection of 10 genres with 100 audio files each, all having a length of 30 seconds .
- images original - A visual representation for each audio file. The audio files were converted to Mel Spectrograms to make this possible.

Audio File

Spectrogram



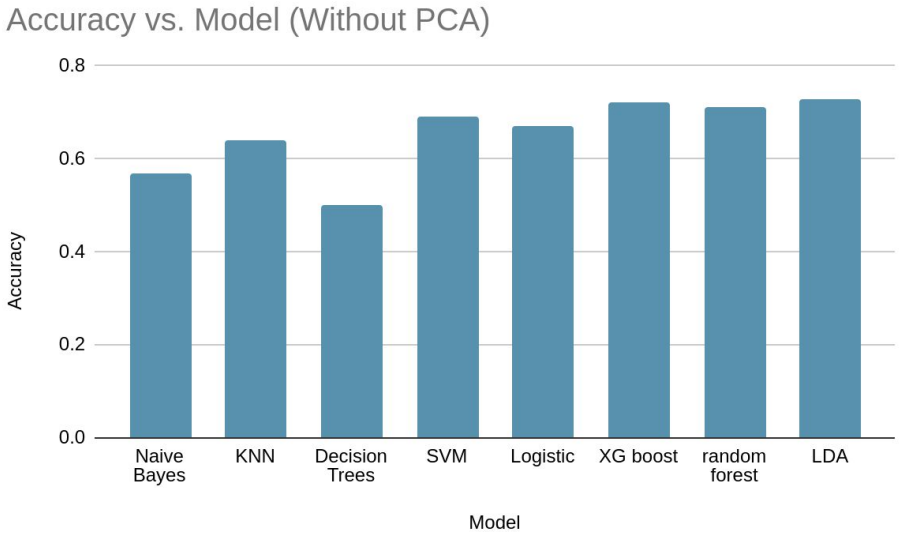
Features used

- Mel-frequency cepstral coefficients (MFCC)
- Chroma_stft
- Spectral centroid
- Spectral rolloff
- Tempo
- Zero Crossing Rate (ZCR)
- Onset strength
- Pitches and Magnitudes

We used python librosa library to extract these features. Librosa is a python package for music and audio analysis. We added mean and variance of these features. Total number of features used is 75.

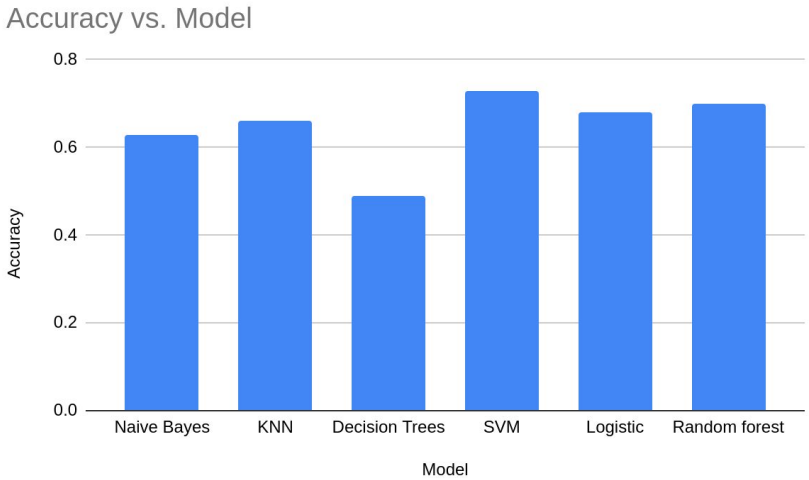
Models used

Naive Bayes
K Nearest Neighbour (KNN)
Support Vector Machine (SVM)
Logistic Regression
Linear Discriminant Analysis (LDA)
Random Forest
Decision Trees
XGBoost



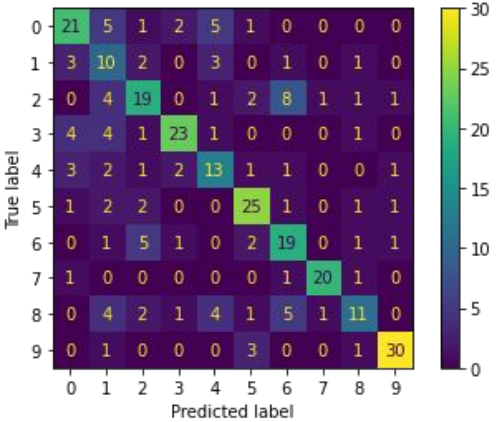
PCA

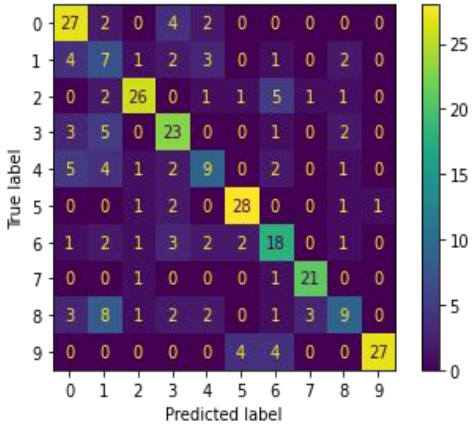
We did PCA for feature extraction. For each model we found the optimal number of components and then found the accuracies.

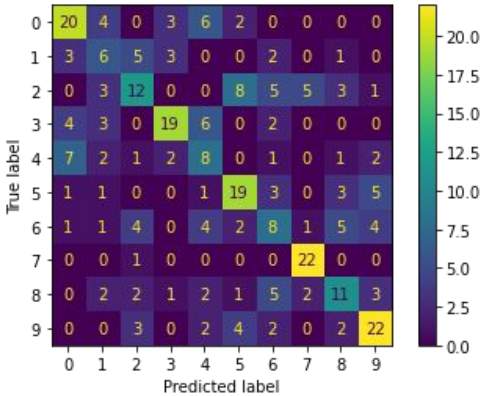


SVM has highest accuracy of 73%

Observations

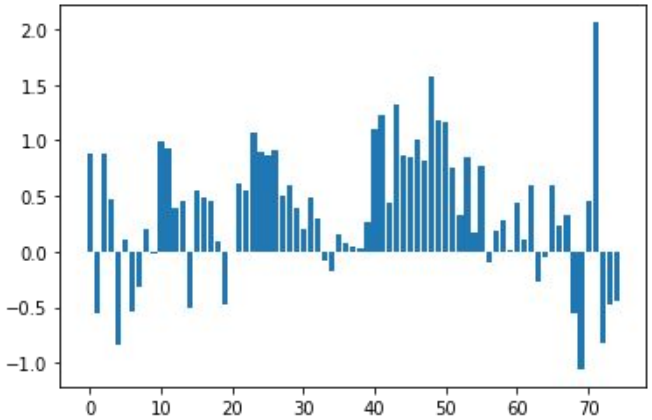


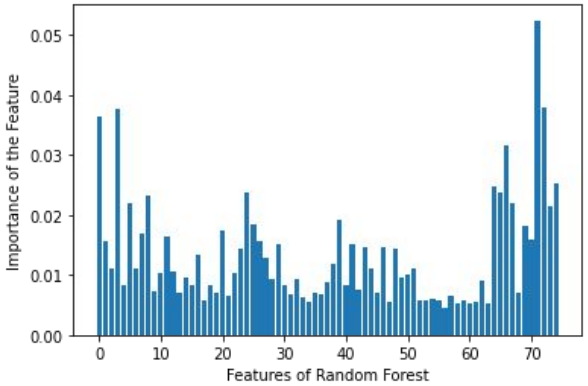


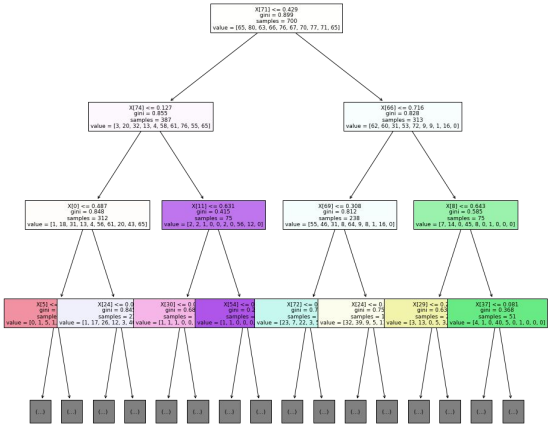


We can see that for the 9th genre which is classical the true predictions are very high (30/35) in the first figure. We observed that only classical genre is having high accuracy of predictions in every model. We also observed that rock genre has low accuracy of predictions

Feature Importance







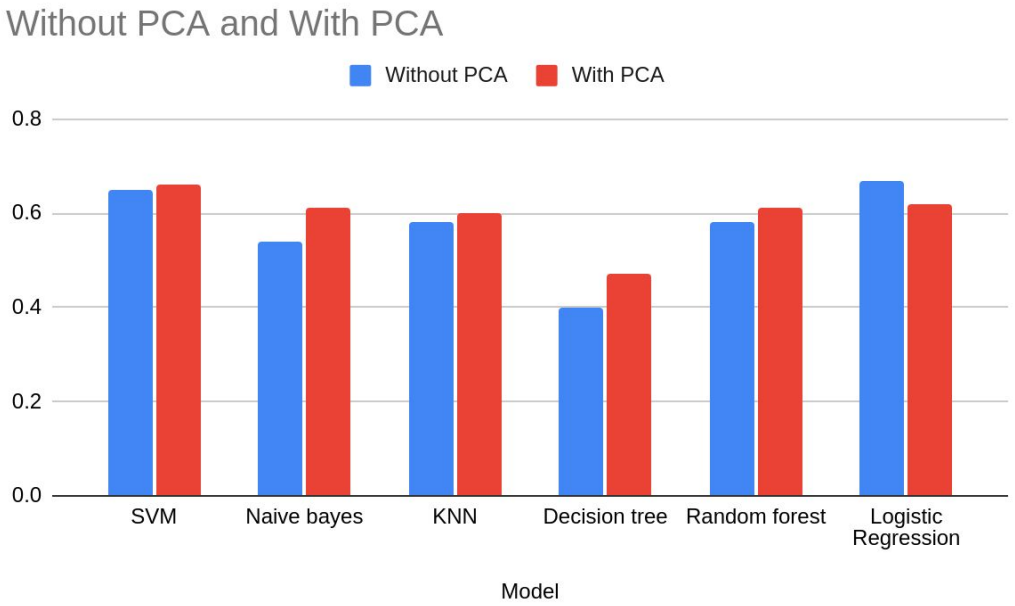
Feature number 71 has the highest importance which corresponds to mean of onset strength.

We took the top 15 features and trained our models based on that, but the accuracy decreased.

Pre-trained model for feature extraction

- We used VGG16 model for feature extraction.
- The features were extracted from the spectrogram images which are also available in the dataset.
- Initially the number of features extracted was 25,088.
- To reduce this number we removed columns which had more than 80% zero entries which reduced the features to 2570.
- Various models were trained with these features with and without Principal Component Analysis (PCA).

Accuracies by using Pre-trained model



Recommendation System

User gives a .wav audio file and we first classify it as one of the ten genres. To classify it we are using SVM model with PCA as it has the highest accuracy.

The following steps were performed:

- Extract the features from .wav file
- Scale the features using the min-max scaler fitted with train data
- Transform the features using PCA which was fitted for the SVM Model
- Predict the genre using SVM Model
- Recommend random 5 songs of the predicted genre