MUSIC GENRE CLASSIFICATION

VINUTHNA NEKKANTI (G01333348)
SRIVAMSI PRIYANKA CHAGANTI (G01339921)
KRISHNA KOUSHIK MADDUKURI (G01322729)

1. Abstract

The purpose of this project is to create a classifier using several machine learning approaches and determine which sort of classifier accurately predicts music genre. To differentiate one tone from another, we are going to analyze the features extracted from the GTZAN dataset and build different types of ensemble models to see how better we can differentiate one genre from another. We plan to experiment with various machine learning models like K-Nearest Neighbors, Logistic Regression, Multiclass support vector machines, Random Forest, boosting classifiers like adaboost and xgboost. Finally, compare each other to determine which model is a best fit for classifying such kinds of Audio data.

2. Introduction

This application is a machine learning classifier which classifies music into different kinds of genres by predicting a piece of audio clip into multiple target classes. Many music industries like Apple music, Spotify, Amazon Music use similar applications to recommend music to users based on their interest and several other parameters. With increasing amounts of data and music releases, classifying each song is very tedious and often done manually, however we can use this large amount of music data and make predictions using various machine learning techniques. In this project we implemented a classifier using several machine learning approaches to determine which sort of classifier accurately predicts music genre. We normalized the audio data for each song to remove volume difference in the audio clips which are collected from different sources and does not affect their genres. We used two compact forms of data representations namely Fourier-transform coefficients and Mel-frequency cepstral coefficients (MFCC) for audio data representation. Using a Librosa library (A Library designed for sound) the MFCC was performed for each segment of audio resulting from Fourier transformation. Then we encoded the data and split the data into test and train sets, using these sets we implemented different Machine learning models. We started with simpler and familiar Machine Learning Models like Logistic Regression and then moved to complex Machine Learning Models like KNN, Random Forest, SVM. We used evaluation metrics like confusion matrix, precision, Recall and F1 Score on test train split sets to determine the accuracy of each and every model. Based on these evaluation metrics we found that SVM and XG boost gave better results of all the models we experimented.

3. Method

For this project we used the GTZAN dataset. The GTZAN genre collection dataset was collected in 2000-2001. It consists of 1000 audio files each having 30 seconds duration. There are 10 classes (10 music genres) each containing 100 audio tracks. It contains audio files of the following 10 genres:

Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock. The Dataset can be downloaded from http://marsyas.info/downloads/datasets.html.

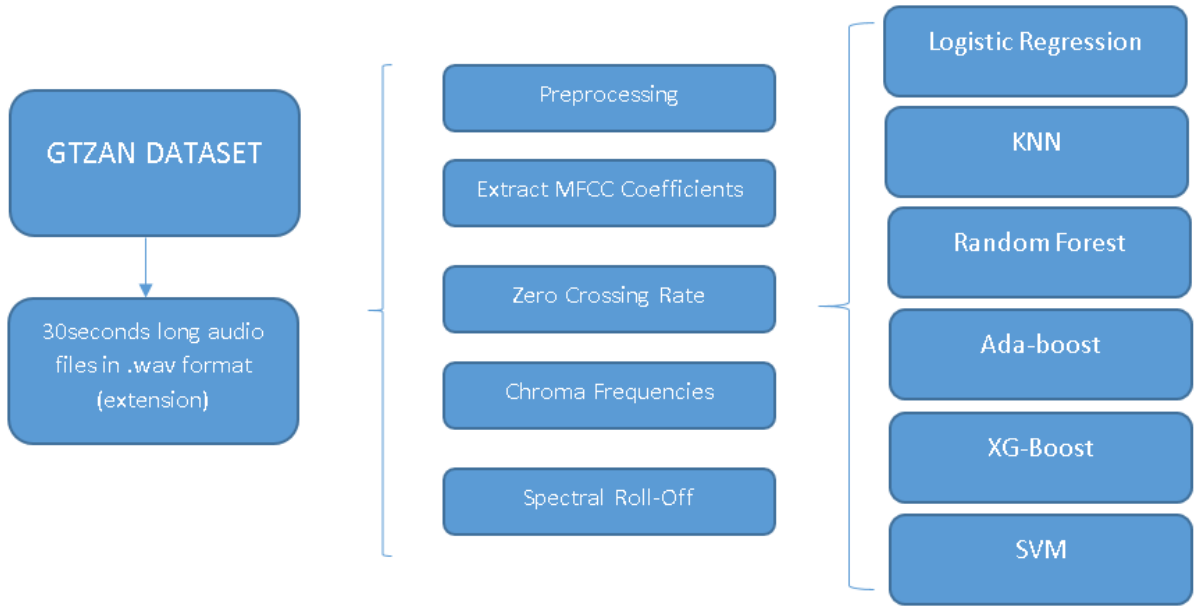
The GTZAN dataset consists of 1000 audio files. It also contains a CSV file which has values from 58 acoustic features of each song. There are also images of Mel Spectrograms for every 30-second audio file. Using all the data, we applied above mentioned Machine Learning techniques to classify the songs into ten genres. Using the librosa library we performed several transformations on the data including audio feature extraction and MFCC. After the features extracted from MFCC we validate our dataset by splitting the data into train and test.We tried to remove null values or missing values from the extracted features from audio files, it turns out there are no NAN values in the extracted features. Later we sorted the dataset in ascending order and calculated the 1st and 3rd quartiles(Q1, Q3) using which we computed IQR = Q1 - Q3 finally found the upper and lower bounds. Finally we checked every data and determined in which bound they fell in and specified the outliers. Since there is no categorical data in the extracted features no duplicate value creation is needed.

MFCC transform involves the following steps:

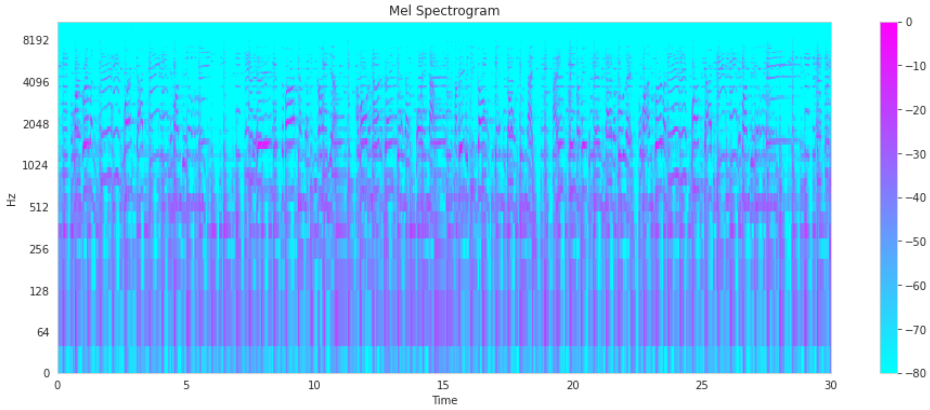
1. Take the Fourier transform of each segment of audio.
2. Map the powers of the spectrum obtained from Fourier transform onto the Mel-scale, using triangular overlapping windows.
3. Take the log of the powers at each of the Mel-frequencies.
4. Take the discrete cosine-transform of the list of Mel-log powers and the resulting spectrum amplitudes are the MFCCs.

Mel Scale

$$-- \; Mel(f) = 1125\log(1 + f/700) \; --$$



The Mel spectrogram with decibel-log of the audio clips can be visualized below



Since our features extracted from all the audio clips are huge we needed to reduce the size, for this we used principal component analysis dimensionality reduction technique. Firstly in order to standardize the data we used min-max scalar to normalize the data now using z-score calculated corresponding mean and standard deviation for each column and found the covariance matrix for the scaled data from this matrix we calculated the eigenvalues and extracted eigen vectors which are nothing but the reduced form of huge data in vectors with single dimensions. Finally we split the data into test and train using test train split from sklearn library and feed this data into different machine learning models.

Classifiers used:

- K-Nearest Neighbors
- Logistic Regression
- Multiclass support vector machines
- Random Forest
- Adaboost
- Xgboost

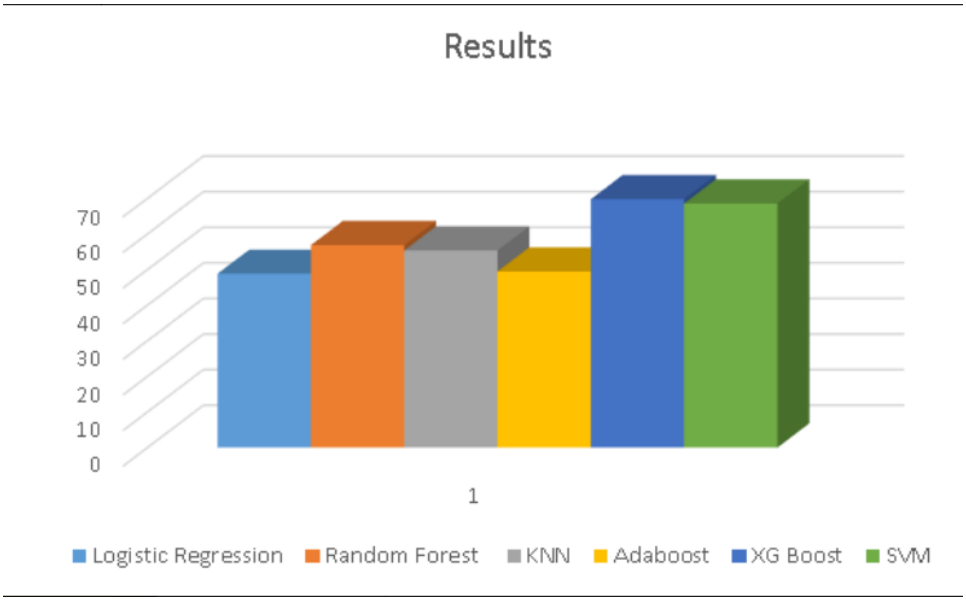
It was also observed that the accuracy for each classifier reduced as the number of genres increased

4.RESULTS

Following are the results observed from various classifiers for ten genres:

Model	Accuracy
Logistic Regression	48.83
Random Forest	56.85
KNN	55.34
Adaboost	49.46
XG Boost	69.76
SVM	68.56

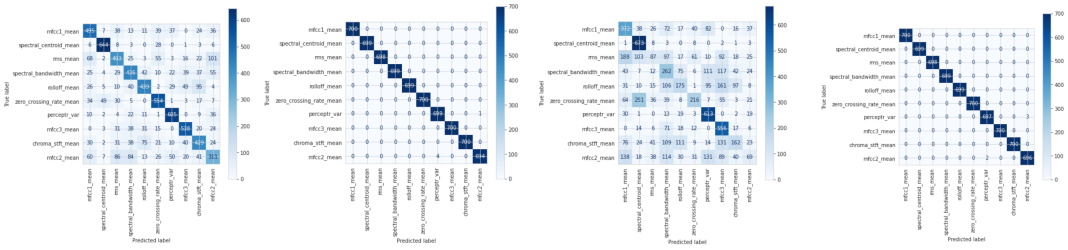
Accuracies obtained by different classifiers are represented as follows:



The key metric used to evaluate our models is Accuracy and also we have developed confusion matrices to represent the performance of the best model among 6 of them.

5.CONCLUSION

In Conclusion, using the models we have built in this project, we can easily classify music genres of various audio clips. Of all the models that we have built, SVM and XG-Boost provided us with better results when compared to the rest. These look the longest time to train but with increase in accuracy the computation cost also increased. Due to time constraints, we could only implement the above 6 Machine Learning algorithms. If given a chance to work from scratch again, we would start by implementing Deep Learning techniques like Convolution Neural Networks and Gradient Descent Algorithms. We can also experiment with other data models in addition to Mel-Spectrogram and also extend to various Big Data techniques for the feature extraction. This would also help us classify audio data of larger length. We could improve our algorithm so that they can recommend songs to user not only based on genre but also various other factors like mood, occasions etc.



6.Project Video Link

The project can be found at: **https://youtu.be/k1U-wA27Y9c**