Music Genre Classification

Deep[1], Rajat Soni[2], and H Sriram[3]

[1]B20CS087
[2]B20CS050
[3]B20EE020

May 2, 2022

Abstract

The music industry has undergone major changes from its conventional existence and also in the form of music created in last few years. The ever-growing customer base has also increased the market for different music styles. Therefore, it is essential to classify the music according to the genres to satisfy the needs of the people categorically. The manual ranking of music is a repetitive, lengthy task and the duty lies with the listener. This paper reports our experience with building different classification models.The **GTZAN dataset** is used which contains collection of 10 genres with 100 audio files each, all having a length of 30 seconds. We use various classification algorithms and compare their results in this report. A website is implemented that takes audio files and predict the genre of the audio file

1 Introduction

Audio processing is at the heart of recording,storing and transmitting audio content.One such application of audio processing is *Music Genre Classification*. The aim of this project is to classify the audio files to certain categories of genre to which they belong to. The objective of automating the music classification is to make the selection of songs quick and less cumbersome. This will help us find valuable data such as trends,genre and artists easily.

2 Dataset

The GTZAN dataset contains 100 audio files of 30 seconds each of all 10 genres. To increase the proportion of data the audio files are further broken into 10 parts each of 3 seconds.The csv file which contains mean and variance of various features extracted from the audio files are used. The dataset contains 9990 rows and 60 columns which includes the following:

- Mean and variance of features like MFCC,zero crossing rate,chroma stft,rolloff etc
- 1 label column which contains names of the 10 genre.

3 Methodology

3.1 Overview

The Methodology includes a two part process where in the first part the following models were run in the base model

- Decision Tree Classifier
- Gradient Boosting Classifier

- Random Forest Classifier
- Extra Tree Classifier
- Bagging Classifier
- LGBM Classifier

In the second part the model that produced the best result for the evaluation metrics is hyper-parameterised.This model is used to make prediction on the audio file uploaded on the web.

3.2 Exploring the data and Preprocessing

On counting the null values in the dataset it was found that there are no *NULL* values present in the dataset.
The column data types of the features were in float which suits our requirement. Hence there is no need for data preprocessing and the data is well built

3.3 Visualization of the dataset

Violin plot:The violin plot helps us understand the distribution of numerical data.(*Figure 1*)

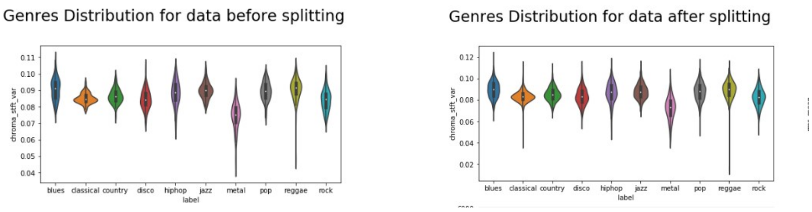

Figure 1: Violin Plot

MFCC's plot:The MFCC plot for different genre are plotted.The MFCC plot tries to capture the capture the time power spectrum of the audio signals.Further the plots of coorealation,chroma shift,zero crossing rate,roll off and spectral features were plotted to help us understand the features which helped us in feature extraction.

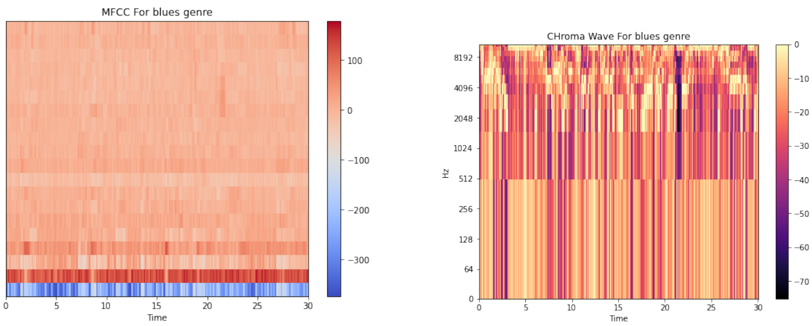

Figure 2: MFCC and Chroma plot

3.4 Evaluation of the Models

The models implemented were evaluated using techniques like -accuracy,precision , recall ,f1 score, Confusion matrix , ROC plots and cross validation scores. *Table 1* contains the score of the mean of the cross validation scores on different models. *Figure 4* shows the classwise accuracy plot for the differernt models.

Model	*Accuracy score*	*F1 score*	*Recall*	*Precision*
Decision Tree	0.636937	0.636973	0.637702	0.636937
Random Forest	0.857758	0.856972	0.858427	0.857758
Gradient Boost	0.818919	0.818879	0.819912	0.818919
Extra Trees	0.879279	0.878565	0.819912	0.879279
Bagging Classifier	0.773373	0.77184	0.775097	0.773373
LGBM Classifier	0.898198	0.898267	0.898993	0.898198

Figure 3: Table 1

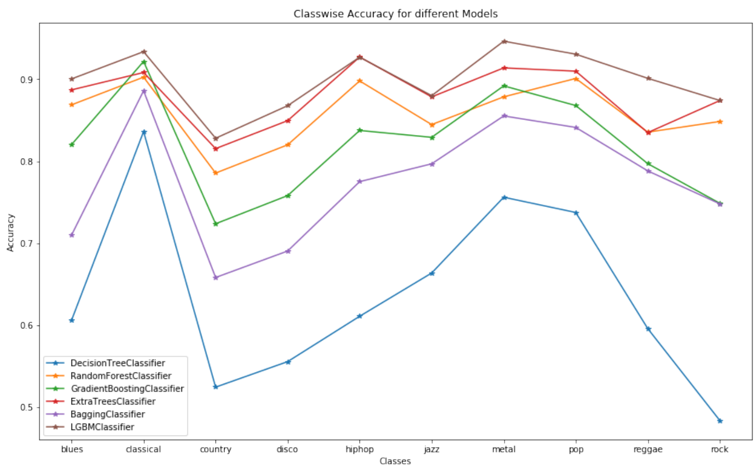

Figure 4

CV fold	1	2	3	4
LGBM Classifier	0.92473979	0.9307446	0.92190629	0.91509812

Figure 5

The best scoring model *LIGHTGBM* was hyper-parameterised further.*Figure 5* gives the result of the hyper-parameterised model on 4 fold-cross validation.

4 Web Deployment

A web server was created using streamlit which took an audio file as input from the user and predicted the genre using the tuned LightGBM model.Screenshots of the working model are attached in *Figure*

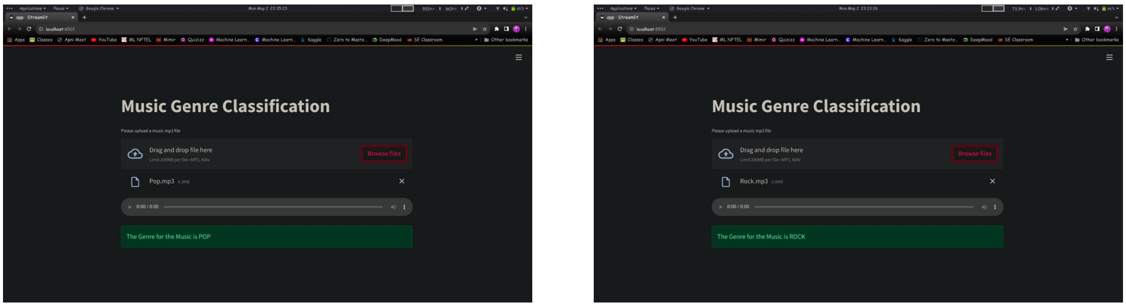

Figure 6: Screen shot of the Working Model on Web Server

5 Results and Analysis

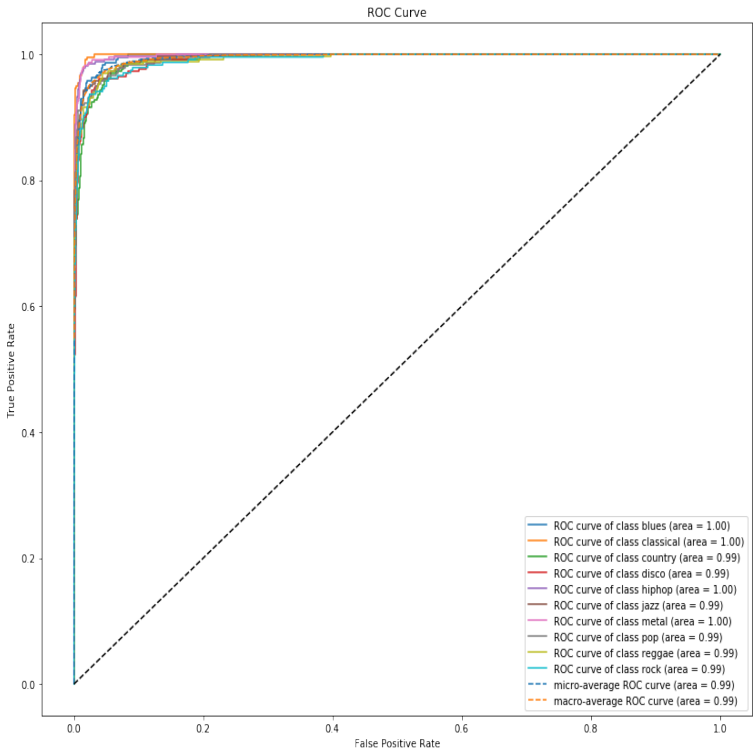

Figure 7

For very less false positive rate we have achieved more than 0.8 accuracy for almost all classes of music given in the dataset.Even the area under the curve for all the classes is greater than 0.99. As the false positive rate increases forward the true positive rate tends to 1. The saturation level is achieved when the false positive rate is in between 0.4 to 0.6. The consequences of all these observations leads us to the conclusion that the model is not overfitted and can be considered as a good model.

Contributions

The project selection,learning and planning was done as a team. The individual contributions are as follows:

- Deep Patel *B20CS087* : Decision Tree,web server deployment,report,readme
- Rajat Soni *B20CS050* : LightGBM hyper-parametertuning,Bagging Classifier,Extra TreesClassifier,Gradient Boost,Report
- H Sriram *B20EE020* : Data Pre-processing and exploratory analysis,Random Forest,feature extraction,Report

References

[1] Pattern Classification *2nd Edition*,Richard O Duda,Peter E Hart,David G Stork.

[2] Dataset link:*https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification*,

[3] LightGBM—lightgbm.readthedocs.io

[4] Evaluation Metrics for Classification Problem—AnalyticsVidya

Made in Latex