Music Genre Classification

Shivani Mishra
MT20062
IIIT Delhi
shivani20062@iiitd.ac.in

Aditi Sharma
MT20100
IIIT Delhi
aditi20100@iiitd.ac.in

Abstract

This project compares machine learning algorithms in their ability to automatically classify songs into their musical genres depending on the features extracted from the music data. First, a review of existing techniques and approaches is carried out, in terms of algorithms. The songs are collected manually from Free Music Archive for genres- Pop, Rock, Hiphop, Rnb and Blues. Fifty songs of each genre are collected to form training samples. Another twenty five songs of these categories are collected to form testing data. Twenty Nine features are extracted from each song using libROSA library. Each song is divided into chunks of 5, 10, 20 seconds to measure the impact of length of a sample on the amount of information represented by the features. Six classifiers are created - a logistic regression classifier, a support-vector machine, a k-nearest neighbor (k-NN) classifier, a decision tree, an ensemble of the above classifiers and a neural network composed of several dense layers. Training and testing performance is evaluated using 5-fold cross validation. Results indicate that support-vector machine and ensemble classifier performs far better than neural network. Also, SVM and ensemble performs best on samples of length 20 seconds indicating the spread of information across the length of the song rather than presence of discrete information in small chunks.

1 Project Introduction

”Music is a moral law. It gives soul to the universe, wings to the mind, flight to the imagination, and charm and gaiety to life and to everything” - Plato

The words of Plato rightly describe the importance of music in the world. As the field of music is evolving, huge number of songs are being published every day. The bookkeeping of such huge amount of data requires intense manual effort. The aim of this project is to automate the process of classifying this data so as to ease its organizing process.

This study explores the application of machine learning algorithms to identify and classify the genre of a given audio file.

Each song is divided into chunks of 5 seconds, 10 seconds and 20 seconds. Twenty Nine features are extracted from each chunk separately. These features are then fed to machine learning models namely a logistic regression model, a support-vector machine, a k-nearest neighbor (k-NN) model, a decision tree, an ensemble of the above classifiers and a neural network, which are trained to classify the genres: Rock, Pop, RnB, Blues and Hip-Hop.

The models are evaluated on the manually created dataset. The performance is evaluated using 5-split cross validation. We compare the proposed models on different length of data samples and also study the relative importance of different features and the relation of length of the sample to the relevant information it carries, which may help in better performance.

The rest of this paper is organized as follows. Section 2 describes the importance of the project. Section 3 describes the existing methods in the literature for the task of music genre classification. Section 4 is an overview of the the dataset used in this project and how it was created. The proposed models and the implementation details are discussed in Section 5. The results are reported in Section 6, followed by the conclusions from this study in Section 7 and a brief idea about the future work for this project in Section 8.

2 Importance

Use of music classification helps us understand music creations in greater context and makes it easier to identify patterns, recommend new artists to one another, and find creations that are the most satisfying to our individual tastes.

Music genres can greatly enhance our personal listening enjoyment and allow us to recognize and honor the creative decisions of the hardworking artists who make the music, including and especially those who branch out and experiment with new styles.

3 Literature Survey

Automatic genre classification of music can be attempted in a variety of ways. A typical approach involves processing a dataset of audio files, extracting features from them, and then using a dataset of these extracted features to train a machine learning classifier.

((Silla Jr et al., 2007)) collected a dataset of 3000 Latin music samples and segmented each sample into 3 pieces, each of length 30 seconds. He observed that the best information about each song lied in the middle piece and not the starting and ending pieces.

((Ali and Siddiqui, 2017)) extracted the MFCC for each song as features and compared the performance of SVM and kNN on the features. He also compared the classifiers after applying dimensionality reduction on the features. He found SVM to be the better performing classifier with an accuracy of 77 %

((Bahuleyan, 2018)) applied ensemble of Neural networks and different machine learning models on the Audio Set dataset and compares the performance to choose the most valuable features out of the initially taken features in the time and frequency domain. He finds the ensemble to be better performing and reports an accuracy of 89 %

((Lansdown, 2019)) collected data using online music archive and performed undersampling and oversampling to handle the imbalance of classes in the data, thus creating two datasets. He extracted fifty-four features manually from each sample, using audio analysis libraries LibROSA and Aubio and scikit-learn functions and created five classifiers - a neural network, a support-vector machine, a random forest, KNN and logistic regression and training and testing was performed on each, with each of the two unique datasets created.He found SVM to be the best performing.

4 Dataset and Data Analysis

The choice of dataset plays an important role in any machine learning project.For this project, we manually downloaded three hundred songs from genres-Pop, Hip-Hop, RnB, Blues and Rock, fifty songs of each genre. Each song is of at least 1 minute. We use these to form train dataset. We also downloaded twenty five songs for making test dataset. For each of these, we then split each song into 5, 10 and 20 second chunks using pydub. The motivation behind this is to gather as much information possible from each song. We want to analyse the effect of segmenting the song into smaller chunks, and thereby enchancing the local regions of the song, on the classification performance. Also, the use of smaller song excerpts makes the dataset much easier to acquire than if more number of full tracks had to be downloaded.

4.1 Data Preprocessing

Several stages of preprocessing were required in order to ready the dataset for classification. First we have downloaded the data on the basis of genre like rock, hiphop, pop, blues and RnB. Then we have divided these data into chunks of 5, 10 and 20 seconds of music segments.

Length	5 sec	10 sec	20 sec
POP	3000	1500	900
RnB	3000	1500	900
BLUES	3000	1500	900
HIPHOP	3000	1500	900
ROCK	3000	1500	900

Number of samples of each length

4.2 Feature Extraction

In order to represent the tracks numerically, twenty nine audio features were extracted from each track after careful analysis of importance of audio features and the selecting the best features.Features can be broadly classified as time domain and

frequency domain features. The feature extraction was done using libROSA a Python library.

1) Time Domain Features

These are features which were extracted from the raw audio signal.

1) Zero Crossing Rate (ZCR) : A zero crossing point refers to one where the signal changes sign from positive to negative. The average of the ZCR across all frames are chosen as representative features.

2) Root Mean Square Energy (RMSE) : The root mean square value can be computed as:

$$\sqrt{\frac{1}{N}\sum_{n=1}^{n} x(n)^2 = 1}$$

RMSE is calculated frame by frame and then we take the average across all frames.

3) Tempo : In general terms, tempo refers to the how fast or slow a piece of music is; it is expressed in terms of Beats Per Minute (BPM).

2) Frequency Domain Features

The audio signal can be transformed into the frequency domain by using the Fourier Transform. We then extract the following features.

1) Mel-Frequency Cepstral Coefficients (MFCC) : It is representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

2) Chroma Features : This is a vector which corresponds to the total energy of the signal in each of the classes.

3) Spectral Centroid : For each frame, this corresponds to the frequency around which most of the energy is centered.

4) Spectral Roll-off : This feature corresponds to the value of frequency below which 85% (this threshold can be defined by the user) of the total energy in the spectrum lies

There are many more features available. For each of the spectral features described above, the mean of the values taken across frames is considered as the representative final feature that is fed to the model.

5 Proposed Model

This section provides a brief overview of the machine learning classifiers adopted in this project till date.

1) Logistic Regression (LR) : This linear classifier is generally used for binary classification tasks.

2) Support Vector Machines (SVM) : SVMs transform the original input data into a high dimensional space using a kernel trick. The transformed data can be linearly separated using a hyperplane. The optimal hyperplane maximizes the margin. For this multi-class classification task, the SVM is implemented as a one-vs-one method.

3) K-Nearest Neighbour (k-NN) : It is very famous for its simplicity of execution. The k-NN is by design non-linear and it can detect direct or indirect spread information. It also slants with a huge amount of data. The essential computation in our k-NN is to measure the distance between two tunes. We implement 7-KNN model after careful analysis of the effect of the neighbors to choose the optimal k.

4) Decision Tree : The decision trees are the robust, non linear classifiers that use entropy and information gain to classify the features.

5) Ensemble Model : All the above listed models are combined in the sense that predictions from each of the above are collected and the performance is evaluated using majority voting.

6) Neural Network : Neural Networks have the ability to learn by themselves and then produce the output that is not limited to the input provided to them.

We have used the following structure for the Neural Network.

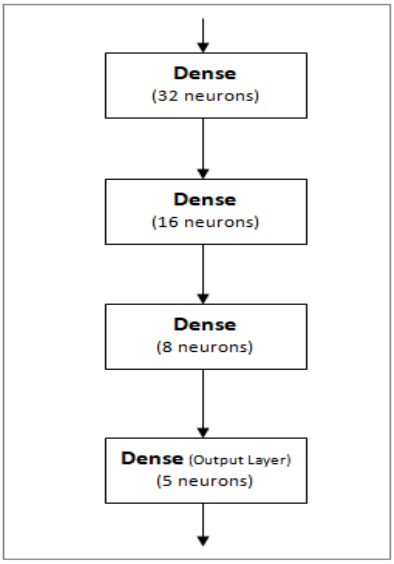
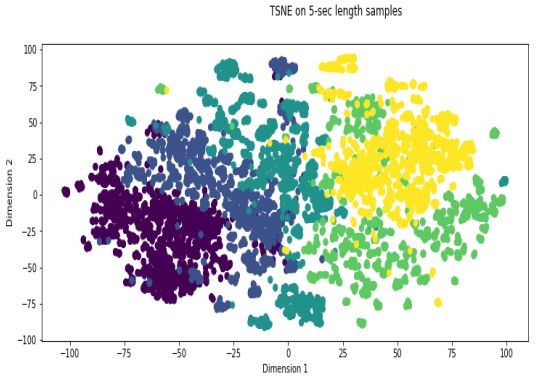
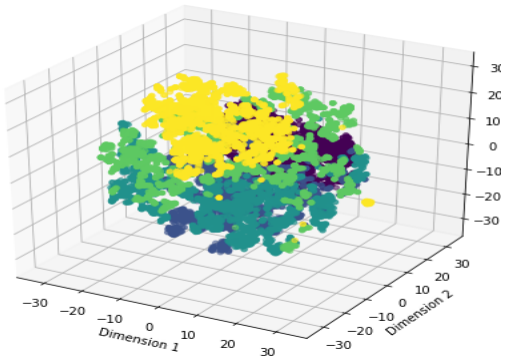


Fig: Structure of the Neural Network

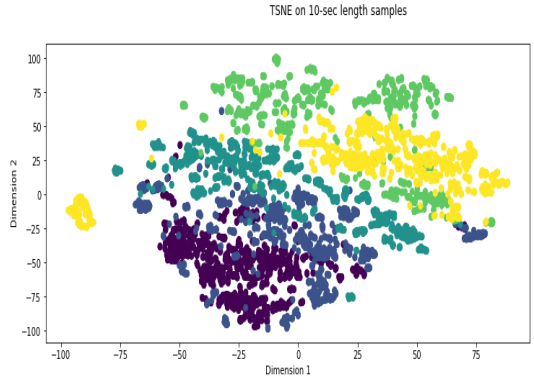
6 Results

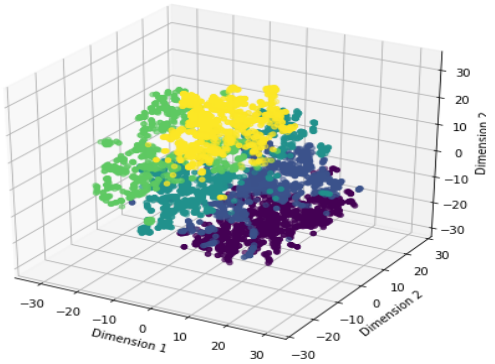
This section displays the results in the form of 5-fold cross validation accuracy achieved using each classifier.
Firstly, we plot one waveform of a song to visualize it. Then we analyse the TSNE plot in 2D and 3D to visualize the data perfectly.
Following is the TSNE plot for 5 second length data.



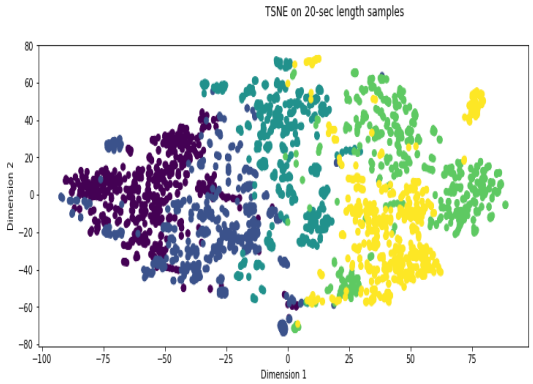


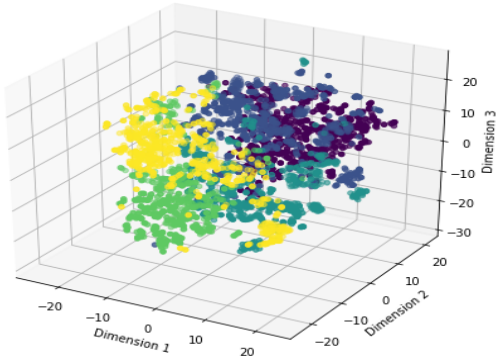
Following is the TSNE plot for 10 second length data.





Following is the TSNE plot for 20 second length data.





Then we recorded the accuracy of classification by various models. The table given below lists out the accuracy for the different models used using 5-fold cross validation. For neural network, we use the separately formed test dataset for measurement of its performance.

MODEL	5 SECONDS	10 SECONDS	20 SECONDS
	CV-A	CV-A	CV-A
LOGISTIC REGRESSION	89%	88.7%	90%
SVM-Linear kernel	91.7%	96%	96.4%
DECISION TREE	84%	84%	83.9%
7-KNN	88.5%	86%	87%
ENSEMBLE	92.2%	93%	95.2%
NEURAL NETWORK	89.6%	82.7%	85.2%

CV-A: 5-Fold CROSS VALIDATION ACCURACY;

Fig: Accuracies of various models for samples of different length

7 Conclusion

Accuracy of classification by different genres and different machine learning algorithms is varied. In this work, the task of music genre classification is studied using the manually created dataset. Here, we use the six algorithms namely logistic regression, a support-vector machine, a decision tree, k-nearest neighbor (k-NN), an ensemble of the above classifiers and at last neural network. We have used different genres such as rock, hiphop, pop, blues and RnB.
According to our estimations as given above, SVM and Ensemble classifier have best performances which are comparable to each other and both perform far better than neural network and other classifiers.

8 Future Work

1. Collection of more data which may improve performance of used models.
2. Expanding the dataset to include more genres and sub genres which would increase the capacity to classify greater varieties of songs.
3. Analysing the effect of presence and absence of vocals on the classification.
4. Making an end-to-end pipeline for better user experience

9 Link to code and data

The following link hosts the code files and dataset created for the project.

https://drive.google.com/drive/folders/1aYFsHKRmMuI9Y0SwOdtx0SDb_pwCIMQ8?usp=sharing

References

M. A. Ali and Z. A. Siddiqui. Automatic music genres classification using machine learning. *Int. J. Adv. Comput. Sci. Appl*, 8(8):337–344, 2017.

H. Bahuleyan. Music genre classification using machine learning techniques. 04 2018.

B. Lansdown. *Machine Learning for Music Genre Classification*. PhD thesis, 09 2019.

C. N. Silla Jr, C. A. Kaestner, and A. L. Koerich. Automatic music genreclassification using ensemble of classifiers. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 1687–1692. IEEE, 2007.