e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:07/July-2021 Impact Factor- 5.354 www.irjmets.com

MUSIC GENRE CLASSIFICATION USING MACHINE LEARNING

Prajwal R*1, Shubham Sharma*2, Prasanna Naik*3, Mrs. Sugna Mk*4

*1,2,3,4Dept. Of Information Science And Engineering, Sir MVIT College, Bangalore, Karnataka, India.

ABSTRACT

Music Genre Classification Model is a model to classify songs or an audio music based on variety of features of it into the corresponding genre. As the lifestyle of the people in the world is more depending on the music, the technology and the internet becoming cheaper to the end users, there is a much need of developing an efficient and more accurate model for this classification. So this project aims at developing a model that can classify and organise the audio music to its belonged genre out of many genre. Machine Learning techniques are employed to do this task. The Classification is performed based on acoustic features of the audio music and the spectrogram of the audio. This model is designed and developed to produce a better accuracy for Music Genre Classification.

I. INTRODUCTION

Music has been the major part of human life from many generations and many centuries. Music has got very complex diversity based on many factors like culture, history, public, marketing etc. There is no proper definition for the classification of these music. But for the attempt to do so, to make music more defined and organised to the current generation and for the upcoming too, these music are made to be classified between many tags called as GENRES.

Machine Learning is a tool recognised and used widely in the world of automation. Our project on Music Genre Classification use the same and its techniques. There are many real-time applications that provide music stock in various ways to its users who listen and enjoy. And in the real-world there are billions of songs with different target audience like community, age group, language, and acoustic features of the song. So to classify these billion records a model has to be built that automatically do this job of replacing manual classification of the songs into its corresponding genres. The classification has to be made not on the user-end, but by the software.

Sound is defined based on parameters like frequency, decibel, bandwidth etc. Different combination of these parameters results in spectrogram of the sound. This spectrogram is a better tool to analyse the characteristics of the audio files. The genre of the music is defined by the characteristics it possess. The characteristics are likely to be instrumentation, Rhythmic structure, harmonic content etc.

II. PROBLEM STATEMENT

The Genre of the song is one of the main category to classify millions of songs. There are handful of genres that categorise the songs. To the current generation and to the upcoming, a futuristic model that brings out better classification of the songs in music industry has to be built with the most efficient techniques and algorithms we have in today's world.

To produce a model based on machine Learning that solves the challenge of automatic classification of songs into it's corresponding genre.

To bring a better accuracy rate in the produced model/project as compared to the early attempts and pre-existing models.

With the fast growing technologies of Machine Learning, a useful model to the Entertainment industry has to be built.

III. OBJECTIVES

- To develop a model that classifies the input audio sample or song into the music genre which it belong.
- To reduce or replace the involvement of manual tasks in the music genre classification.
- To produce a model with better accuracy rate so that it can take part in real-time applications.
- To develop a model that is better than many early research.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:07/July-2021 Impact Factor- 5.354 www.irjmets.com

IV. RELATED WORK

There are few pre-existing work on the above described model. In 2008, Hareesh Bahuleyan worked on Music Genre Classification using Machine Learning techniques. His work is to provide tags to the music automatically to the songs in the library. Both Neural network and traditional Machine Learning Algorithm were made use to reach the goal. These two approaches used different set of features. After comparing both these approaches , model with Convolutional Neural Network (CNN) approach gave the highest accuracy.

In 2002, Tzanetakis G. et al. worked on classifying audio signals into hierarchy of genres. They believed that characteristics of song depends on the instrumentation, the rhythmic structures, the harmonic content of the song. They proposed three main feature sets: timbral texture, the rhythmic content and the pitch content. Using the proposed feature sets, their model can classify almost 61% of the songs into the corresponding genre out of ten music genre correctly.

Again in 2002, Lu L. et al. The work on Content analysis for audio classification and segmentation. They divided their work to two steps. The first step is to discriminate and separate the speech and non-speech audio. Algorithms are developed using K-nearest-neighbour (KNN) and linear spectral pairs-vector quantisation (LSP-VQ). The second step is to classify the audio into music and other form of sounds.

In 2010, Tom LH Li et al. Automatic musical pattern feature extraction using convolutional neural network, where they tried to understand better features the helps in developing more accurate Music Genre Classification model. Their work was to extract musical features of the audio using Convolutional Neural Network. They proved that CNN is better tool to read informative features from the varying musical pattern. Dataset GTZAN was considered.

V. DATASET

GTZAN is a public dataset collected in 2000-2001 form different sources. There are 1000 audio files in the dataset. It's a collection of 10 genres and each genre consists of 100 audio files. Each audio file is of 30 seconds length, enough for the machine to read its features. The audio files are in .wav format. The size of whole data set is approximately 1.2GB. GTZAN dataset is found to be mostly used in the study of Music Genre Classification giving good results.

VI. METHODOLOGY

Various kinds of time domain and frequency domains features are extracted from our dataset. We will choose the best features out of all extracted features. All the features which are finalised will be appended to a CSV file which will be used by our machine learning classifiers to classify different genres.

The GTZAN dataset contains the audio file in .wav format. To train our machine learning models there is a need to extract features from that data. For features extraction and required initial preprocessing a python library is used for audio analysis call Librosa.

After which, various techniques are used to clean and pre-process our data to make it suitable for ML classifiers. Finally, we optimise our model's performance by selecting the beat hyper-parameters. This is extremely important and a deciding factor in what kind of predictions our model gives.

METHODS

K-Nearest Neighbour (K-NN)

K-NN is a supervised Machine Learning Algorithm. K-NN reads the similarity between new case and the pre-existing case and categorise the new case into the most similar category in the available cases.

Support Vector Machine (SVM)

SVM is another supervised machine learning algorithm that classifies linearly based on margin maximisation principle. SVM creates a hyperplane that distinguish between two different classes. The algorithm improves the complexity of the classifier by performing structural risk minimisation to achieve good generalisation performance.

Logistic Regression

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:07/July-2021 Impact Factor- 5.354 www.irjmets.com

Logistic Regression is a supervised Machine Learning Classifier Algorithm that predicts the target variable probability. This algorithm make use of already identified variables unlike SVM and works on Statistical approach.

Random Forests

Random forest is a Machine Learning Algorithm that is used in classification and to solve regression problems. The algorithm works based on decision trees. Random forest creates uncorrelated forest of trees whose prediction is more accurate than a that of a single tree.

VII. FEATURE EXTRACTION AND DATA PROCESSING

For data processing and features extraction, we used one of the very popular python library for audio analysis, Librosa. The Librosa library has all the necessary tools and techniques for the data extraction from any audio sample. The tasks of extracting the data can be achieved using many inbuilt functions in Librosa library. It extracts all the necessary data for training and classification.

We have to extract relevant characteristics out of many, which are necessary to solve our problem. The process of extraction of features from the audio samples is called feature extraction. Some of the features are studied below.

Zero-Crossing Rate: It's the measure of rate of sign changes along the signal. It is the rate at which the graph moves from negative to positive or back. The feature is widely used in speech recognition and now to classify the audio samples.

Spectral Centroid: It calculates the weighted mean of the frequencies in the sound and locates the 'center of mass' of the sound. For example consider a song of genre metal which has more frequencies towards the end compared to a song belonging to the blues genre whose frequencies are same throughout the length. So the spectral centroid for the blues genre will be somewhere in the middle and for the metal, it will be in the end.

Spectral Rolloff: It measures the shape of a signal. It indicates the frequency below which lies the specified percentage of spectral energy.

Mel Frequency Capstrum Coefficients(MFCC's): Are the set of features that describe the overall shape of the spectral envelope. There around 10-20 features in the set. Used in modelling the characteristics of human voice.

20 MFCCs are calculated over 97 frames and feature scaling is also conducted on our dataset.

Chroma Feature: 12 distinct semitones (or chroma) of the musical octave by projecting the entire spectrum onto 12 bins. This is one of the most used and powerful to represent the audio or music feature.

We can extract various different features mentioned from the dataset based on time domain and frequency. Out of these extracted features the best features are chosen and these finalized features are appended to a CSV file which is used by our algorithms to conduct the Music Genre Classification.

Feature Selection

Features selection is extremely important that we have so many features that can be considered for music classification according to different genres. And that's where selecting the best and relevant features becomes so important because irrelevant features can lead to an unnecessary increase in complexity of the model and it may become extremely hard to make sense out of the model. With too many features our model will take a longer time to train and also will require high computational power. Training on unnecessary and inappropriate features can lead to bad predictions.

The feature selection is extremely important as it helps to remove the irrelevant features from our data and also to reduce noise. If there are a lot of irrelevant variables it may lead to overfitting. Feature selection also helps in reducing the computing power as well. So in a nutshell by selecting the only relevant features we can save time, money, and resources. And by selecting only those features that contribute directly to our target variable we can achieve better accuracy for our model.

There are three major feature selection techniques namely:-

- Filter methods
- Wrapper methods

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:07/July-2021 Impact Factor- 5.354 www.irjmets.com

- Embedded methods

In our model we have used random forest and XGBoost for feature selection. In the below figure you can see the features are all features are arranged in order of their importance from highest to lowest.

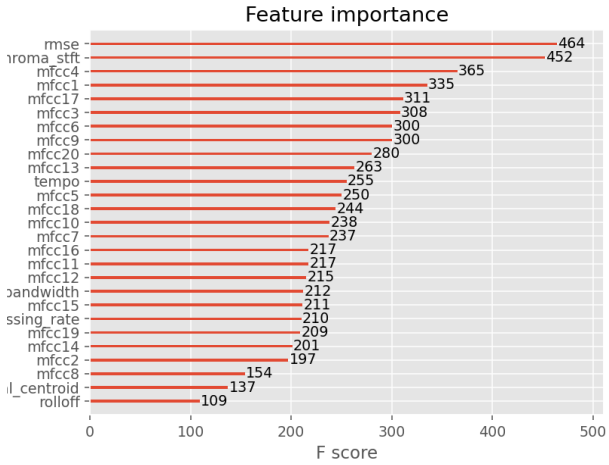

Figure 1: Feature importance

VIII. IMPLEMENTATION

After doing all the processing and feature selection we tried the different models on our extracted data from the CSV files, the first one being k nearest neighbour.

In K-NN, Initially we chose the hype parameter n_neighbours(k) as one. After series of experiments we found that our model works best when the value is 5 after plotting accurate accuracy vs k value.

To find out the best parameters on which the model gives best accuracy we used GridSearchCV, which is a technique for hyper parameter optimisation. GridSearchCV is a library function that belongs to sklearn's model_selection package. Using this library function we can select the best parameter of all the hyper parameters that fit our model.

IX. RESULTS AND DISCUSSION

Accuracy is the metric we used to judge our model. Accuracy is the percentage of predicted output that match the real output, where in here our output is the genre of the song. Confusion matrices are our visual to measure the performance of the best model as the genres are uniformly distributed through our dataset.

We have used various techniques for selecting the best hyper parameters depending upon the best technique for that particular algorithm.

This is the confusion matrix for logistic regression for all 10 genres in our dataset.

Table 1. Accuracy of different models

	Model	Without hyper-parameters tuning .	After hyper-parameters tuning
1	K-Nearest Neighbour(K-NN)	60.4%	66.4%
2	Support Vector Machine (SVM)	73.2%	76.4%
3	Logistic	62.4%	67.2%

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:07/July-2021 Impact Factor- 5.354 www.irjmets.com

	Regression		
4	Random Forests	66.8%	69.6%

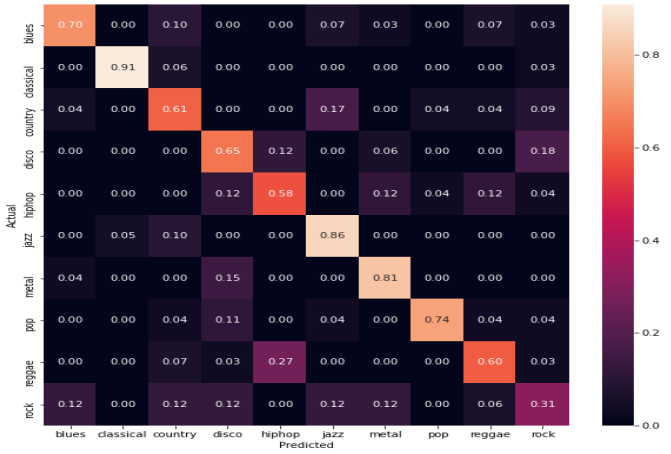

Figure 2: Confusion matrix

X. CONCLUSION

In this paper, we primarily used four machine learning models knn, SVM, Logistic Regression, and random forest. Initially, when we didn't employ any techniques our models were not giving good performance because we were choosing hyperparameters randomly.

But when we employ different techniques to find the best combination of hyperparameters on which our models would perform optimally our accuracy improves a lot. So, by using the techniques like GridSearchCV, RandomizedSearchCV, and finding the optimal K value for k-nearest neighbors, etc we can achieve the maximum accuracy for our models.

And also we found that by selecting relevant features by using various methods and by training the model with the most relevant features the performance model can be greatly increased.

In the future, we will try to use other deep learning approaches on our project, and then we will try to see how parameter optimization helps in improving our deep learning model's predictions.

XI. REFERENCES

[1] A Porter, D Bogdanov, R Kaye, R Tsukanov, and X Serra. Acousticbrainz: a community platform for gathering music information obtained from audio. In ISMIR, 2015.

[2] Michael I. Mandel and Daniel P.W. Ellis, Song-level Features and Support Vector Machines for Music Classification, Queen Mary, University of London, 2005.

[3] Jean-Julien Aucouturier and Francois Pachet. Improving timbre similarity : How high's the sky. Journal of Negative Results in Speech and Audio Sciences, 1(1), 2004

[4] Alan V. Oppenheim. A speech analysis-synthesis system based on homomorphic filtering. Journal of the Acostical Society of America, 45:458–465, Februar Kristopher West and Stephen Cox. Features and classifiers for the automatic classification of musical audio signals. In International Symposium on Music Information Retrieval, 2004.

[5] Dan Ellis, Adam Berenzweig, and Brian Whitman. The "uspop2002" pop music data set, 2005. http://labrosa.ee.columbia.edu/projects/musicsim/ uspop2002.html.

[6] Beth Logan and Ariel Salomon. A music similarity function based on signal analysis. In ICME 2001, Tokyo, Japan, 2001.