JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MUSIC GENRE CLASSIFICATION

Nidhi Verma[1], Niharika Deepti[2], Mr. Mahendra Kumar[3]

nidhiverma21june@gmail.com, tiwarinikku98@gmail.com, mahe.gupta.it@gmail.com

Student, Student, Guide

Computer Science & Engineering

Shri Ramswaroop Memorial College of Engineering & Management, Lucknow, India

Abstract: Today's music is very important part of people's daily lives. There are various genre of music and because the genres are different, the taste of music is also different. This increases the need for proper cataloging and greater reach. Radio stations and music TV channels have archives of millions of music tapes. Gigabytes of music files are also spread across the web. Automating the search and organization of music files by their genre is a daunting task. In this paper, we describe an automatic classification system model for musical genres. Our research objective is to find the best machine learning algorithm that predicts genre using K-Nearest Neighbour machine (KNN) and Support Vector Machine (SVM). In this we present the GTZAN music dataset. It has ten different genres of music.

Index term- Music, genre, classification, features, Mel Frequency Cepstral Coefficients (MFCC), K- Nearest Neighbour (KNN), Support Vector Machine (SVM), Librossa, GTZAN Dataset.

I. INTRODUCTION

The musical genre is a label created and used by humans to classify and describe the vast musical world. Music genres are created through complex interactions between public, marketing, historical and cultural factors, so there are no strict definitions or boundaries.

Genre classification can be helpful in explaining some interesting real-world problems like creating song references, finding related songs, finding companies that will like that particular song. Music classification is still considered as one of the research areas due to the challenges of selecting and extracting optimal acoustic (audio) features. The classification of musical genres has been a difficult task in the field of music information retrieval (MIR). Music genres are inherently subjective and therefore difficult to describe systematically and coherently. However, even on the current music genre, it is clear that members of a particular genre share it. Certain characteristics that are usually associated with musical instruments, rhythmic structures and pitch content. Automatic extraction of music information is becoming increasingly important for structuring and organizing the growing number of music files available digitally on the web. It is very likely that in the near future all music recorded in human history will be available on the internet. Genre classification has, so far, been done manually by adding it to the metadata store of the audio files. This paper is intended for content-based classification with an emphasis on the information in the audio. Using a traditional machine learning approach, we found the right features for the audio signal, trained the classifier with the features data, and made prediction of song genre using nearest neighbour machine (KNN) and support vector machine (SVM).

II. RELATED WORK

In this music genre classification project [1], we have developed a classifier on audio files to predict its genre. It explains how to extract important features from audio files.

In this paper [2], KNN had difficulty but SVM is a more effective classifier which gave 77% accuracy.

We feature extraction and selection, and finally classification process [3]. KNN is a supervised learning classifier that is easy to implement. From results we found that the k-Nearest Neighbours classifier gave more accurate outcomes compared to support vector machine classifiers.

In this literature review [4] we have discussed how the neural network works to classify music genres in general. The application uses a K-Nearest Neighbours algorithm to perform the classification. This is performed by using the libROSA library package for music and audio analysis.

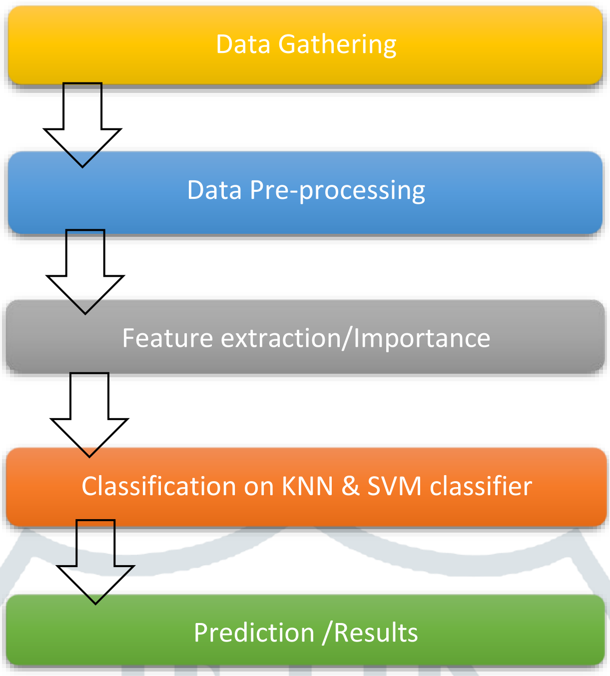
In this letter [5], we describe our research into music stretching resistance, in which an applicable classification method using audio features and genre information based on the metric learning technique has been proposed.

In this paper [6], the automatic classification of audio signals into a hierarchy of musical genres is explored. More specifically, three feature sets for representing timbral texture, rhythmic content and pitch content are proposed. [7] used multiple layers of SVMs to achieve over 90% accuracy on a dataset containing only four genres.

This paper[8], named Comparison of music genre classification using Nearest Centroid Classifier and K-Nearest Neighbors is published by Elizabeth Nurmiyati Tamatiita; Aditya Wikan Mahatama. They research audio files to classify into 12 different genres. The comparison is done between two algorithms Nearest Centroid Classifier and K Nearest Neighbours algorithms. And the KNN algorithm gives the better result with 56.3 % accuracy.

The paper "Neural Network Music Genre Classification" was published by Nikki Pelchat and Craig M. Gelowitz[9]. They used a machine learning language named CNN. The dataset they used had 1880 songs and they split the dataset into 70% training data and 30% testing data. The CNN algorithm gives the accuracy of 67%.

III. METHODOLOGY



Flowchart: Step by step of the method

According to Flowchart in more detail the process of the methods used as follows:

- Data Gathering-

Get a GTZAN dataset using the website www.Kaggle.com. GTZAN dataset contains a total of 1000 tracks, covering 10 genres and each with 18 features. Because the number of features and the number of categories will greatly affect the complexity and computation time required.

- Pre-Processing-

In this stage, the train data and test data is distributed. Here, 75% of the training data and 25% of the testing data.

- Feature Extraction-

The following 18 metadata features are extracted, excluding the gender features, then 17 features are converted to numeric values.

- Classification-

At this stage, the training is done first based on the data and selection of the defined features, then the testing is done using each method.

- Measurement-

The results of the testing for each method are then measured for accuracy.

IV. DATASET

Music Analysis, Retrieval and Synthesis for Audio Signals (Marsays) is an open-source World Wide Web (www) for audio management with a specific add on using audio data. For our research, we used the GTZAN dataset containing a collection of thousands of audio files. Each file lasts thirty seconds. The ten genres included in this dataset contain one hundred tracks each. Each track has a 16-bit 22050Hz Mono audio file in .au format. We've selected ten genres; Blues, Classical, Rock, Jazz, Reggae, Metal, Country, Pop, Disco and Hip-hop. Our total data is 1000 songs.

GENRE	NUMBER OF GENRES
Blues	100
Classical	100
Country	100
Disco	100
Hip-Hop	100
Jazz	100
Metal	100
Pop	100
Reggae	100
Rock	100
TOTAL	**1000**

Table: Distribution of the dataset

V. FEATURE EXTRACTION

It is the process of computing a compact digital representation that can be used to characterize an audio clip. Feature vector extraction is done using the LibROSA package in python. LibROSA is a python package for music and audio analysis that provides the basic building blocks needed to create a music information retrieval (MIR) system. The characteristics we choose must be able to distinguish and describe different forms of the audio signal taking into account amplitude, time and spectrum and include Mel Frequency Spectral coefficient (MFCC), Spectral Roll-off, Root Mean Square Error (RMSE), Short-time Fourier transform (STFT), Spectral Bandwidth, Spectral Centroid, and Zero Crossing Rate.

- Mel Frequency Spectral coefficient**:**

Mel Frequency cepstral coefficients (MFCCs) are perceptual directional features also based on STFT. After taking the magnitude of the amplitude spectrum, the FFT boxes are grouped and smoothed according to the perceptually controlled Mel Frequency ratio.

Finally, for decoration of the obtained features vectors, a discrete cosine transform is performed. Although typically 13 coefficients are used for speech representation, we found that the first 5 provide the best genre classification performance.

- Spectral Roll-off:

It defines the magnitude of the frequency where the higher level of the frequency drops to below. The spectral rolloff frequency is described as spectrogram where spectrum energy of roll-off is 85%.

- Root Mean Square Error:

This feature environment detects the difference between the values we predicted with the actual value on training our model.

- <u>Spectral Centroid</u>**:**

It is defined as the center of gravity of the magnitude spectrum of the Short-time Fourier transform. It captures in a long time the descriptive form of the central mass of the music. These are musical characteristics that basically measure the weighted frequencies that exist in musical sound.

- <u>Zero Crossing Rate:</u>

The zero crossing in the time domain provide a signal change in a given time period. A change of sign is defined as the transition of a signal between positive and negative values. More detailed explanation can be found in [3].

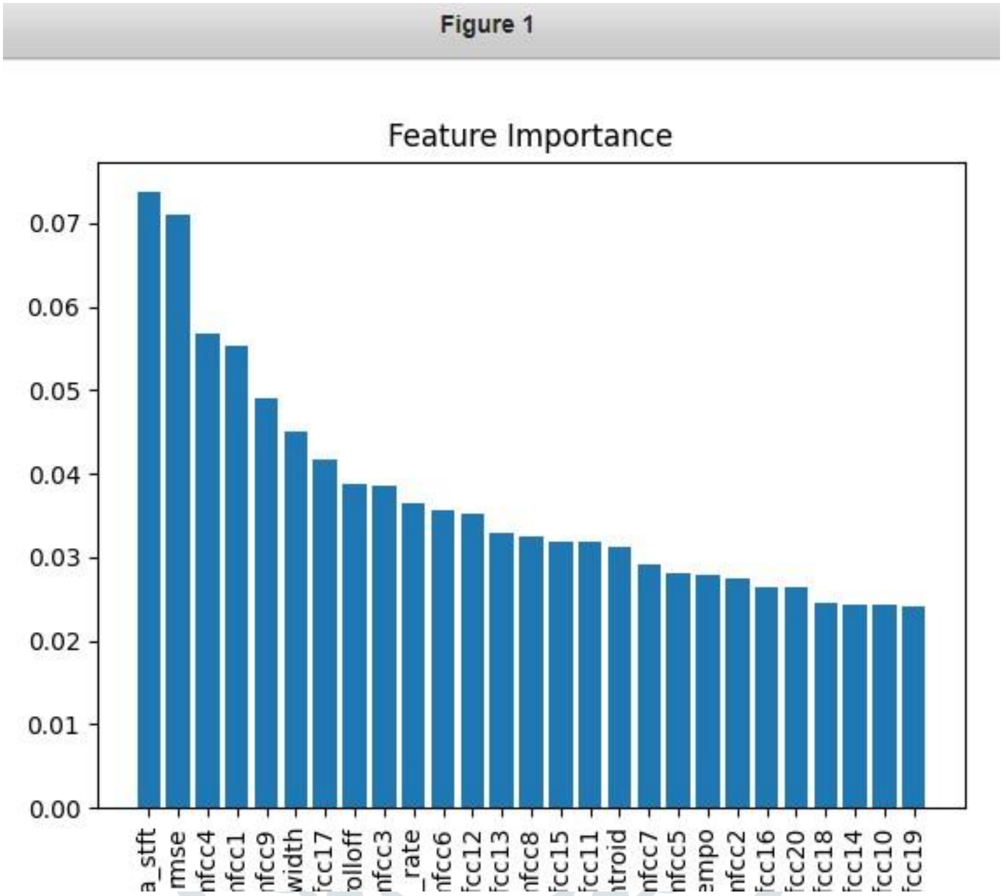


Figure1: Feature Importance using Random Forest

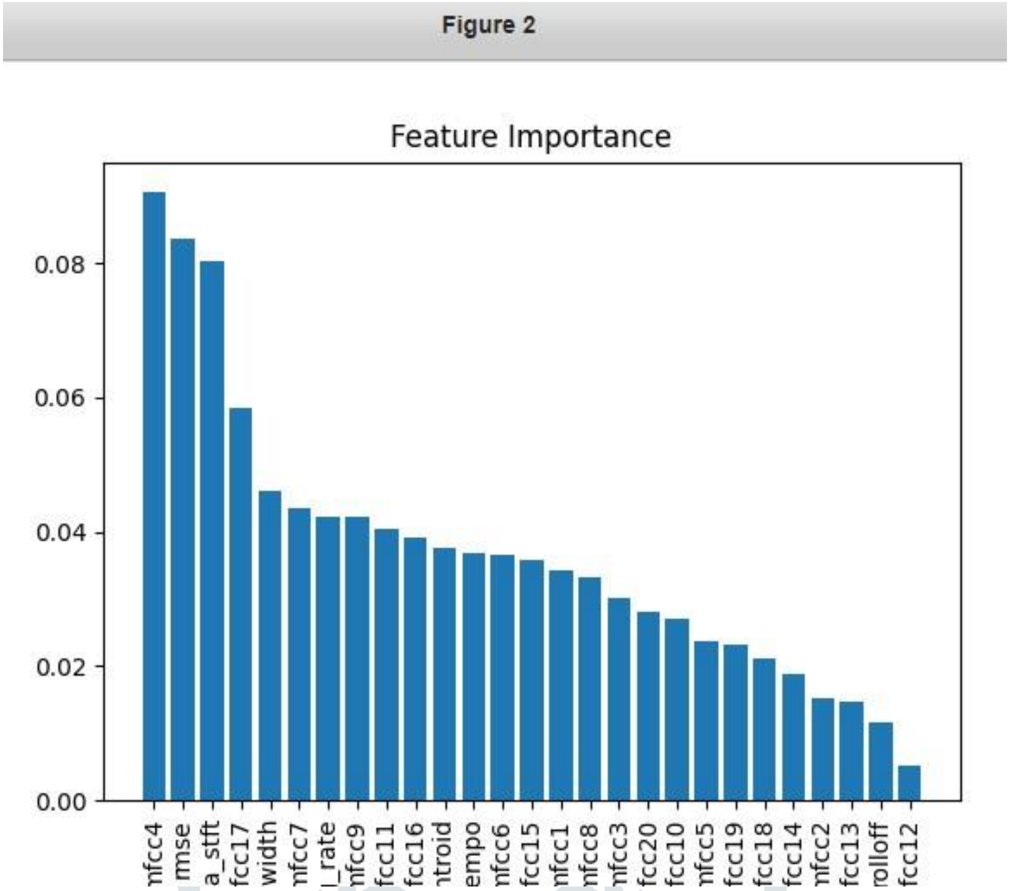


Figure 2: Feature Importance using Decision tree

VI. EVOLUTION

A. Classification:

In this section, to classify music according to their genre, several machine learning algorithms used like K – Nearest Neighbor, Random Forest, Support Vector machine, Decision Tree. These model is trained using the obtain feature vector data set.

- K-Nearest Neighbor Algorithm (KNN)-

The first machine learning techniques we executed was the K Nearest Neighbor Algorithm, is the simplest machine learning algorithms based on Supervised Learning technique. When the KNN algorithm is used to classify music genres, it looks for songs that are similar and assumes they belongs in the same group because they are close in proximity. This strategy has produced the best results among the several other techniques that prevail in this notion.

The KNN algorithm is one of the most basic machine learning approaches. It interprets data in such a way that when fresh data is supplied, the machine identifies it and categorize it based on feature similarity.

- Random Forest (RF)-

Random Forest (RF) is also used for the same feature set to study the success of ensemble techniques in classifying musical genres. RF can be used as a combination of multiple decision trees with a packed sampling strategy [].

- Support Vector Machine (SVM)-

The second algorithm, Support Vector Machine [6], is a supervised learning algorithm and a directed organization method that determine the extreme boundary splitting two groups of information. If the information is not directly distinguishable in the feature space during this time, it can be put into an upper dimensional space using the Mercer kernel approach.

- Decision Tree (DT)-

Decision Tree are learning algorithms that provide a model-based, supervised approach. It tries to identify the most distinctive feature in the data set which is the root node of the tree. The entropy calculation is performed when the most distinct feature is found. There are also different metrics in the literature that provide distinct characteristics [5].

B. Prediction / Result

For the GTZAN dataset, the K- Nearest Algorithm model achieves training accuracy of about 62% and Support Vector Machine model achieves training accuracy of about 73%. Python is the language used for model development. Some packages like Keras, Numpy, Pandas was used to build the model. Tenserflow package is used for Deep learning.

VII. CONCLUSION & FUTURE SCOPE

This study provides insights into an application that classifies musical genres using machine learning techniques. The application uses a Support Vector Machine model to perform the classification. Obtain the Mel spectrum of each track in the GTZAN dataset. This is done using python's libROSA package. The software is implemented to classify a huge database of songs into their respective genres. An extension of this work would be to look at larger datasets and also track in different formats like mp3, au, etc. In addition, over time, the representative style of each genre will continue to change.

Therefore, the goal for the future will be to update the evolution of genres styles and extend our software to work on theses updated styles. This work could also be extended to act as a music recommendation system based on the person's mood.

REFERENCES

1. Congyue Chen, Xin Steven, "Combined Transfer and Active Learning for High Accuracy Music Genre Classification Method", IEEE 2nd ICBAIE 2021.
2. Lakshman Kumar Puppala, Siva Sankar Reddy Muvva, "A Novel Music Genre Classification Using Convolutional Neural Network", IEEE 6th ICCES 2021.
3. Jitesh Kumar Bhatia, Rishabh Dev Singh, Sanket Kumar, "Music Genre Classification", IEEE Oct 2021
4. Matthew Crème, Charles Bublin, Raphael Lenain, "Music Genre Classification", Stanford University, 15, dec 2016.
5. G. Tzanetakis, P. Cook, "Musical genre classification of audio signal", IEEE Transactions on Speech and Audio Processing, Vol. 19, Issue 5, July 2002.
6. M.D.Nevetha, A.Nithyasree, A.Parveenbanu, Mrs. Jetlin CP, "Music Genre Classification using Machine Learning", IJEDR, Vol.9, Issue 2, 2021.
7. Muhammad Asim Ali, Zain Ahmad Siddiqui, "Automatic Music Genres Classification using Machine Learning", International Journal of Advanced Computer Science and Application, Volume 8, 22017.
8. Jitesh Kumar Bhatia, Rishabh Dev Singh, Sanket Kumar, Music Genre Classification, Oct. 2021.
9. Changsheng Xu, N. C. Maddage, Xi Shao, Fang Cao, Qi Tian, "Musical genre classification using support vector machines", IEEE International Conference on Acoustics, speech and Signal Processing, Volume 5, April 2003.