**Music Genre Classification Techniques**

**Samarth Gupta**

**Lovely Professional University, Phagwara, India**

**samarth.11911223@lpu.in**

**Abstract:** One of the most challenging tasks in Music Information Retrieval (MIR) is categorizing music files according to there genre. Which can be very useful for music recommendation. In this study, we compare performance of various techniques and will find out which works best. Which includes KNN, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, SVM and Artificial Neural Networks. These all the techniques use hand-crafted features both from time domain and frequency domain. We will also use CNN, which will use Spectrograms to classify music.

**Keywords:** Music Genre classification Techniques.

1. **Introduction**

Music Industry is rising every day. The companies like Spotify and iTunes have contributed a lot to it. One of the biggest problems they face is Analyse and Categories the music. Classification based on Genre of Music is one of the best categorizations. To do this Automatically, we must analyse the music and then classify into different categories. It has various use cases. One of them is music recommendation.

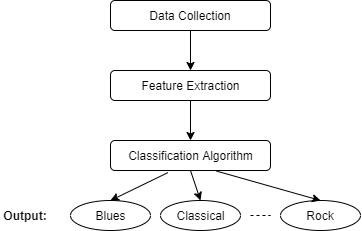
This study explores the application of Machine Learning (ML) algorithms to identify and classify the genre of a given audio file. Also, will put some flash on important features of Audio file.

1. **Dataset**

In this work, we make use of GTZAN Dataset, which is a very popular Dataset for Genre classification. This dataset contains 10 Genre classes (blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock). Each class have 100 Music files in it.

1. **Methodology**

The whole process includes the following architecture.



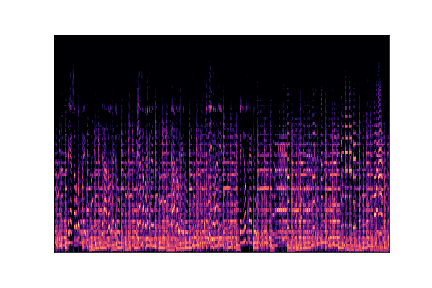
**3.1 Data Collection**

This includes collecting dataset as described in section 2.

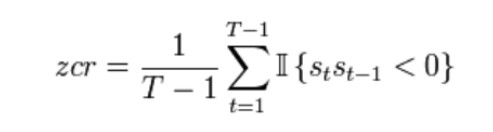
* 1. **Feature Extraction**

To classify music, we need features of that music file. So, the main features a music file are -

* **Spectrogram**: A visual representation of the signal strength or loudness or pitch of a signal over time at different frequencies present in a waveform. By this we can see energy levels varying over time.



* **Mel Spectrogram**: The Mel Scale, is the result of some non-linear transformation of the frequency scale. The Mel Spectrogram is a normal Spectrogram, but with a Mel Scale on y axis. This is what we are using for CNN.
* **Zero Crossing Rate**: The rate at which the signal changes from positive to negative or number of zero crossing withing a segment of a signal. It is used to measure smoothness of signal.



* **Spectral Centroid**: It Indicates where the “centre of mass” for a sound is located and is calculated as the weighted mean of the frequencies present in the sound.

Text

Description automatically generated

* **Spectral RollOff**: It is a measure of the shape of signal. It represents the frequency at which high frequencies decline to 0.

A picture containing arrow

Description automatically generated

* **Mel-Frequency Cepstral Coefficients**: MFCCs of a signal are small set of features which concisely describe the overall shape of a spectral envelope. It models the characteristics of the human voice.

Background pattern

Description automatically generated

* **Chroma Frequencies**: Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones of the musical octave.
* **Harmonics and Perceptrual:** Harmonics are characteristics that human ears can't distinguish It represents sound colour. Perceptrual understanding shock wave represents the sound rhythm and emotion
* **Tempo:** It represents how fast or slow a music piece is. It is expressed in Beats Per Minute (BMP).
  1. **Classification Algorithms**

We are using various classifiers to find which works best.

**1 KNN:** A supervised machine learning algorithm, that takes labelled input data to process unlabelled data in future. When using in Music genre classification, it looks at familiar music and assumes they belongs to same category because they seems to be near to each other. When using k = 1, means the nearest neighbour, it gives maximum accuracy while classifying music genre. [[knn](https://github.com/Sk-singla/Music-Genre-Classification/blob/master/KNN.ipynb)]

**2. Naïve Bayes:** Naïve Bayes is one of the highly practical Bayesian learning methods. This classifier is based on applying Bayes’ theorem with strong independence assumptions between features. Bayes Optimal Classifier calculates the probability of a class from each group of existing attributes, and determines which class is the most optimal.

**3. Decision trees:** Decision tree is a technique very useful in data mining to extract information of a data set, normally using a TDIDT (Top-Down Induction Decision Tree) algorithm. It is a tree shaped (if-else) classifier in which internal nodes represents features, branches represent decision rules and leaf nodes represent outcome.

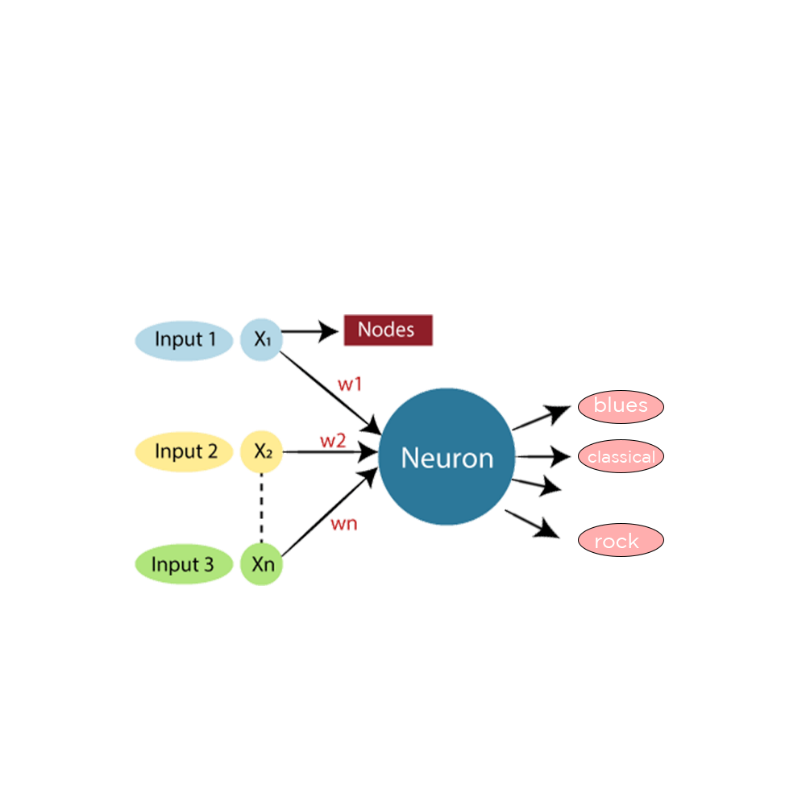
**4. Random Forest:** A forest contains multiple trees. Similarly random forest algorithms create given number of decision trees and get the best result by means of voting. It reduces the over-fitting by averaging the result. Larger the number of trees, more accurate is the result.

**5. SVM:** A machine learning technique which is based on the principle of structure risk minimization is support vector machines. SVMs transform the original input data into a high dimensional space using a kernel trick. The transformed data is linearly separated using a hyperplane.

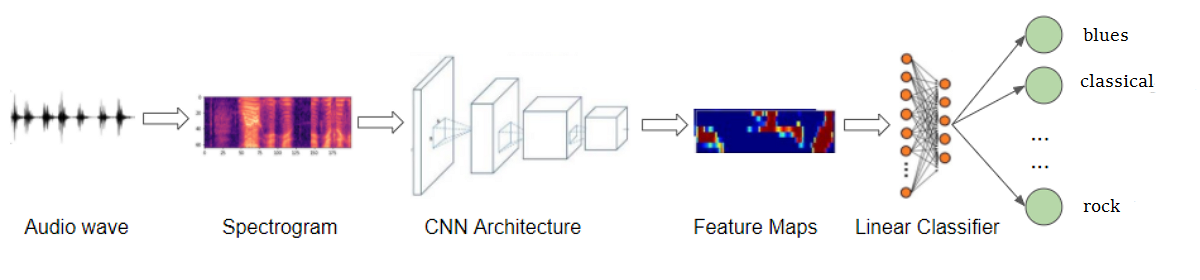


**6. Logistic Regression:** This linear classiﬁer is generally used for binary classiﬁcation tasks. For this multi-class classiﬁcation task, the LR is implemented as a one-vs-rest (OvR) method and uses the cross-entropy loss.

**7. ANN:** Neural Networks are the ones that develop structure of our brain. Artificial Neural network is derived from that. Like the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks.



**8. CNN:** Convolutional Neural Networks is mainly used to process multidimensional vectors such as images. Understanding Images is done by CNN. So here we are using Spectrogram of audio file to classify genres. But it is found that it don’t works so good in this case.

****

1. **Evaluation**

**4.1 Model Evaluation**

To Evaluate the models described in section 4, the following metrics will be used.

* **Accuracy:** It refers to the percentage of correctly classiﬁed test samples.
* **F-score:** Based on the confusion matrix, it is possible to calculate the precision and re-call. F-score is then computed as the harmonic mean between precision and recall.

|  |  |  |
| --- | --- | --- |
|  | **Accuracy** | **F-Score** |
| KNN | 0.68 | 0.7 |
| SVM | 0.66 | 0.67 |
| Logistic Regression | **0.72** | **0.73** |
| Naïve Bayes’ | 0.5 | 0.51 |
| Decision Tree | 0.45 | 0.45 |
| Random forest | 0.63 | 0.64 |
| ANN | **0.68** | **0.72** |
| CNN | 0.33 | 0.35 |

* 1. **Feature Evaluation**

Here we will evaluate 20 most Important features to classify music.

Chart

Description automatically generated

From this chart, it is clear that MFCC is most important feature of audio. Other than that, Zero Crossing Rate, Chroma Frequencies, Harmony and Perceptrual are also among top 20 features.

1. **Conclusion**

In this work, the task of music genre classification is proposed using 8 different machine learning Algorithms. The techniques KNN, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, SVM and Artificial Neural Networks used hand-crafted audio features to predict music genre. But CNN used Spectrograms for that. Logistic Regression is found to be the best performing model. But if we talk about Neural Networks, Simple ANN technique also performed well, which can even be further improved. Also, MFCC is found to be the Most Important feature for music classification. It has been found that the audio clips used in this study are bit noisy. Future studies can find a way to pre-process noisy data before feeding it into a machine learning model, in order to achieve better performance.

**References**