SCHOOL OF COMPUTER SCIENCE & ENGINEERING

A Project Report on

Comparative Study on Various Algorithms used to perform Music Genre Classification

Team Members

Mona Reddy(19BCE0814)
Sri Pavani(20BCT0283)
Neha Ahirwar(20BCT0335)
Deepika Nuthi(20BDS0412)

Submitted to

Prof. Gunavathi C

Comparative Study on Various Techniques used to perform Music Genre Classification

Abstract: Music genre classification is one of the topics digital music processing is interested in. In this study, we extract the audio features using digital signal processing methods and then classify music using various models such as KNN, SVMs, Naive Bayes, Random Forests, XGBoost, Logistic Regression and so on. We compare the performances of the models based on the results obtained. In the study, GTZAN dataset has been used and the highest accuracy was obtained using the XGBooster algorithm.

Keywords: Music genre classification, GTZAN dataset, KNN, SVM, XGBoost, Naive Bayes, Random Forests

I. Problem Definition

Music is described as an art form and a cultural activity whose medium is sound. It is not only for entertainment or pleasure but is being used for a wide variety of purposes due its social and physiological effects. Today there are more than 35 million songs available online and thus it becomes very important to categorise them. Several music industries are making efforts to provide filtered and categorized content to their customers and the public, be it spotify, gaana, amazon music etc. Music genres are created in order to describe and categorise music but, there are no strict boundaries available. Thus the classification of music has become a wide area of research to understand patterns in each of these genres and classify them. This area also plays an important role in music retrieval systems and song suggestion systems. Efficient and accurate systems which can intelligently classify the vast amount of audio data available has become extremely important and there are many researches being done. This paper aims to understand patterns in the music in order to classify them accurately. Content and audio signal will be used as major parameters in order to perform classification. The aim is to compare various existing classification models which work using both supervised and unsupervised learning techniques in order to understand which model performs better. We have taken into consideration some widely used models such as SVM, KNN, XGBoost, Logistic Regression, Random Forest, Decision Trees, Naive Bayes and compared their accuracy to understand which one of them performs better. We have used GTZAN dataset for this purpose which has an extensive collection of music belonging to various categories like pop, blues, metal, country, classical, disco, jazz and hip-hop. Further we have proposed a model according to our research which gives a high accuracy of 90%.

II. Literature Survey

Music Genre classification has always remained a topic for research and improvement in the field of Artificial Intelligence. Giving the system capability to classify the given music or song based on the input that is audio signals, content etc. automatically was the aim for these systems. Today we have come far in this research field and following are some research work done

Tao Li, Mitsunori Ogihara and Qi Li [1] performed research on content based music genre classification. They compared the various models such as Support Vector Model and Linear Discriminant Analysis. This paper proposed a new method of feature extraction, DWCHs which can capture the local and global information of music signals simultaneously by computing histograms on their Daubechies wavelet coefficients.

Chetna Dabas, Adtya Agrawal [2] evaluated their proposed model for audio signal based music classification on various genres of music like pop, blues, metal, country, classical, disco, jazz and hip-hop. They considered different audio features like MFCC (Mel Frequency Spectral Coefficients), Delta, Delta-Delta and temporal aspects for processing the data. They found that their proposed model performed better than existing models in terms of accuracy by 95%.

Heerde, F., Vatolkin, I., & Rudolph, G [3] did research on the fuzz rules based music genre classification which offers advantage of understandability for end users, in particular in combination with carefully designed semantic features. They performed study on three approaches which operate on fuzzy rules: a complete search of primitive rules, an evolutionary approach, and fuzzy pattern trees. They found out that these methods have high error rates. Fuzzy Trees outperformed any other approach with a balanced validation error rate of 36.15% across all genres.

Bahuleyan, Hareesh [4] used various models to perform music genre classification. The first model used was CNN, that is trained end-to-end on an MEL spectrogram of the audio signal. He extracts the features in both, the time domain and the frequency domain of the input audio signal. These features are then used as an input to models such as Logistic Regression, Random Forests, XGBoost, Gradient Boosting and SVM CLassifiers. The dataset used was the 'Audio Set' Dataset. Finally, the CNN based model and XGBoost were determined to be the best feature-based classifiers along with reporting the most important features.

A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydin [5] extracted the acoustic features of music using digital signal processing methods and using various methods, performed music genre classification. The features extracted are used in classification in order to determine the most efficient algorithm. The algorithms used are KNN, Naive Bayes, Random Forests and SVM. They make use of the GTZAN dataset. According to the study, SVMs attained the highest performance accuracy as compared to the rest.

Costa, Y.M., Oliveira, L.S., Koerich, A.L., Gouyon, F. and Martins, J.G. [6] present an approach to music genre classification that converts the audio signal into spectrograms and then extracts various texture features from the time frequency images that is further used in modeling the various music genres in a classification system. The features are based on Local Binary Pattern (LBP), a structural texture operator used in image classification research. The two well known datasets used were Latin Music Dataset and the ISMIR 2004 dataset. On comparing the performances of the texture features with that of commonly used audio content based features, texture features always outperform the latter.

Chun Pui Tang, Ka Long Chui, Ying Kin Yu, Zhiliang Zeng, Kin Hong Wong [7] examined the application of Long Short Term Memory (LSTM) model in music genre classification. Two different approaches were explored in the paper: first being the single LSTM to directly classify 6 different genres of music and in the second approach, a hierarchical divide-and-conquer strategy was adopted to achieve 10 genres classification. Finally, the music is classified into strong and mild genre classes. Strong genre includes hip hop, metal, pop, rock and reggae and mild classes include jazz, disco, country, classic and blues.

Jeremy Reed, Chin-Hui Lee [8] provided a framework to give finer segmentation of hidden Markov models (HMMs) for automatic genre classification through acoustic segment modeling. They demonstrate that language-ignorant approaches provide results that are consistent with the current state-of-the-art for the genre classification problem. Modeling each one of these acoustic segments with the help an HMM builds a timbral dictionary in the same fashion that one would create a phonetic dictionary for speech.

Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing [9] proposed two ways to improve music genre classification with convolutional neural networks: first combining max- and average-pooling to provide more statistical information to higher level neural networks; second using shortcut connections to skip one or more layers, a method inspired by residual learning method. By comparing two different network topologies, the preliminary experimental results on the GTZAN dataset show that the two methods can effectively improve the classification accuracy, especially the second one.

III. Materials and Methods

A. Dataset

The **GTZAN dataset** is the most-used public dataset for evaluation in machine listening research in Music Genre Recognition (MGR). The files were collected in 2000-2001 from various sources that included personal CDs, radio, microphone recordings, in order to represent various recording conditions. It is a collection of 1, 000 music with 30-second, 22050 Hz sampling frequency and 16 bits. Following are the specifications of the dataset :

- Genres Original - Genres in the GTZAN include blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae and rock and all of these genres have 100 music files each.

- Images original - A visual representation for each audio file which helps as an input for the various neural network models.

- CSV files - The dataset contains 2 CSV files that contain features of the audio files.

B. Audio Features

The following is a detailed description about the features that were used from the dataset. Time and frequency domain digital signal processing was done. Also mean, median, standard deviation were used in order to find out useful features.

- **Zero Crossing Rate:** Zero-crossing rate is defined as the number of sign changes of a signal in a certain period of time. The transition of the signal between negative and positive values is called Sign change.
- **Harmonics and Perceptual :** Harmonics are known to be those characteristics that human ears cannot distinguish, it represents the sound color. Perceptual understanding of the wave represents the rhythm of sound and the emotion it carries.
- **Tempo BPM (beats per minute) :** A dynamic program based beat tracker.
- **Chroma frequencies :** Chroma features represent audio where the entire broad spectrum is divided into 12 bins where each one of them represent a distinct semitone/chroma of the musical octave.
- **EDA :** It performs analysis on the `features_30_sec.csv` file. This file contains the mean and variance for each audio file.
- **Spectral Centroid:** Spectral centroid is a feature used on a frequency domain and indicates the point of the center of gravity of the frequencies in the frequency bin.
- **Spectral Rolloff:** Spectral rolloff is the calculated normalized frequency at which the sum of the low frequency power values of the sound reaches a particular rate in the total power spectrum. In short, it could be defined as the frequency value that corresponds to a certain ratio of the distribution in the spectrum. This rate is usually considered to be 85%.
- **Mel Frequency Coefficient of Cepstrum-MFCC:** MFCCs are a set of features that describe the overall shape of the spectral envelope. It could be considered as a feature of timbre. The purpose of an MFCC is to adapt the cepstral coefficients to the human hearing system. Cepstral coefficients are linear scale.

  Steps included in an MFCC are Frame Blocking, Windowing, Fast Fourier Transform, Mel Frequency Wrapping and Spectrum, respectively.

C. Music Genre Classification using various Algorithms

1. **SVM :** Support Vector Machine is a popular Supervised Learning algorithm, that is used for various Classification as well as Regression problems. The goal of the algorithm is to create the best line or a decision boundary that segregates an n-dimensional space into classes such that a new data point can be easily put into the correct category in future. This best decision boundary is termed as hyperplane.

An SVM chooses the extreme points/vectors that help in generating the hyperplane. These extreme cases are known as support vectors, and hence the algorithm is termed as Support Vector Machine.

2. **K- Nearest Neighbours:** It is one of the most used algorithms for classification and regression predictive problems. Its performance depends on three factors: the distance metrics, the distance rule, and the value of K. The distance metrics give the measure to locate the nearest neighbors of any incoming data point. The distance rule helps us classify the new data point into a class by comparing its features with that of data points in its neighborhood. And the value of K decides the number of neighbors to be compared with. In order to obtain an optimal value of K, the training and validation are segregated from the initial dataset. Now a graph based on the validation error curve is plotted to achieve the value of K. For all predictions K value will be used.

3. **Convolutional Neural Networks :** CNN is an efficient recognition algorithm which is widely used in pattern recognition and image processing. It is a deep learning algorithm that takes an image input and assigns importance (weights/ biases) to various objects in the input image such that it can differentiate one from another. The role of CNN is to reduce the RGB images of an object into a form that is easier to process without losing any feature that is critical to obtain a good prediction. Additionally, the preprocessing required in CNN is much lower than that of other algorithms.

4. **Stochastic Gradient Descent :** Stochastic Gradient descent is a variation to gradient descent which overcomes the drawback of slow performance on very large dataset. Slope of a function is known as Gradient and it is used to measure the change of one variable with respect to the changes in another variable. Gradient Descent is a convex function for which the output is taken as the partial derivative for the set of parameters of its inputs. A slope which is steeper will always have a greater gradient. In stochastic gradient descent some random samples will be selected instead of taking the whole dataset. In each iteration gradient will be calculated. Here updates to the coefficient is performed for each training instance rather than at the end of the batch. This algorithm requires only a small number of passes through the dataset to get the coefficients.

5. **Logistic Regression :** Logistic regression is a statistical model that makes use of a function that models a binary dependent variable, even though many more complex extensions exist. Despite the name being regression, LR is used for classification problems for predicting binomial and multinomial outcomes, having the goal of estimating the values of the parameter's coefficients using the sigmoid function. Logistic regression is used for clustering and when a transaction is ongoing it examines the values of its attributes and tells whether the transaction should proceed or not.

6. **Recurrent Neural Networks:** RNN is a class of artificial neural networks where the connection between multiple nodes helps in forming a directed graph along a given temporal sequence. Thus, it exhibits a temporal and dynamic behavior.

Since RNN is derived from feedforward neural networks, they can make use of their internal state memory to process the variable length sequences of the given inputs. This allows them to be applicable to tasks related to applications such as unsegmented, connected handwriting recognition or speech recognition. These neural networks are designed for analyzing streams of data by means of hidden units. Since RNNs deal with sequential data, they are well suited for the health informatics domain where enormous amounts of sequential data are available to process.

7. **Random Forest :** The Random Forest is a popular algorithm that belongs to supervised learning techniques. Random Forest is a great model which can be used for dual purpose. Classification and Regression Problems in ML. It is based on the concept of ensemble learning**,** which is a process of *combining multiple classifiers in order to solve a complex problem and to improve the performance of the model.* A random forest algorithm creates decision trees on data samples and then gets the prediction from each one of them. Finally, it selects the best solution by the means of voting. It is an ensemble method that is better than a single decision tree because it reduces the over-fitting by averaging the result.

8. **Decision Trees :** The Decision Tree Algorithm is a data mining induction technique that recursively shares a set of records. This algorithm is used to solve regression and classification problems using tree representation. A decision tree contains one root node, internal nodes, and leaf nodes. The nodes are labeled using attribute names, edges are labeled using attribute values. In order to predict a class label for a record, we start from the root of the tree, compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node, comparing our record’s attribute values continuously with other tree’s internal nodes and this is done till some predicted value is reached.

9. **K - means clustering:** K-means clustering is an unsupervised learning algorithm where the model divides or classifies the data into various clusters or classes based on certain parameters. This is an iterative algorithm and the data will be partitioned into K predefined non overlapping clusters. This algorithm begins by defining K number of clusters and then initializing it with K randomly chosen data points as centroids without replacement. The algorithm is iterated till the point there is no change to the centroids, that is the centroid data points remain fixed during the iterations. The data points are assigned a cluster during the algorithm. The centroid for a particular data point is found out by computing the sum of the squared distance between data points and all centroids. The data points will be assigned to the closest cluster (centroid). Finally the centroids will be computed for the clusters by taking the average of the all data points that belong to each cluster.

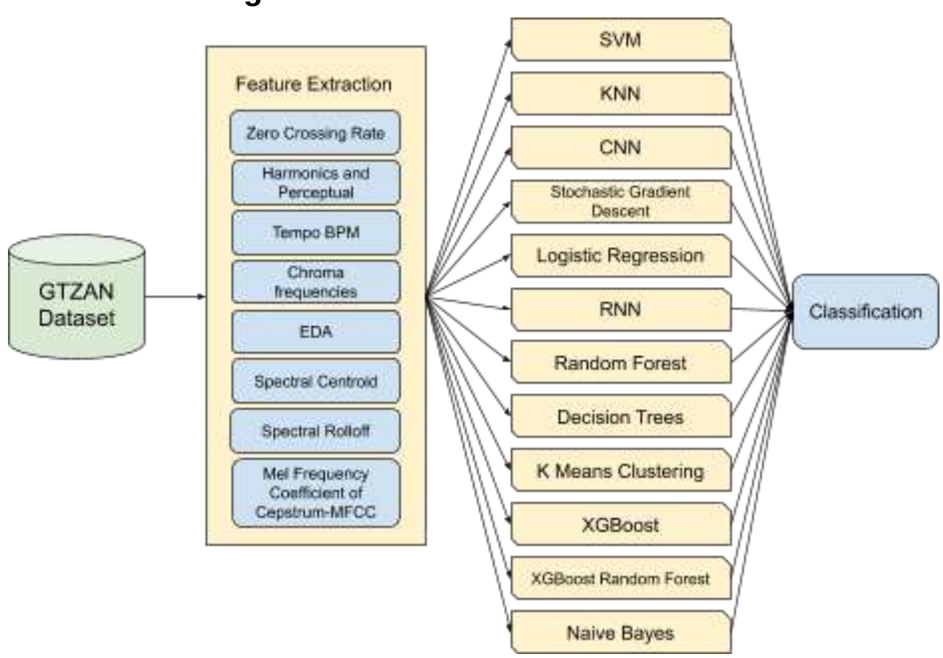
10. **Cross Gradient Booster :** Cross Gradient Boost is an ensemble of decision trees and uses gradient boosting as a framework. Cross Gradient boost is a variation to the basic gradient boost model and has better computational efficiency and often the model performs better. This algorithm is highly scalable.

Cross Gradient Boost provides a huge advantage since it is an ensemble of decision tree models. One model may not be sufficient to rely upon and thus cross gradient boost gives the power of multiple decision trees. The multiple decision trees made aim to reduce errors of each other. In each consecutive step, the error will be reduced.

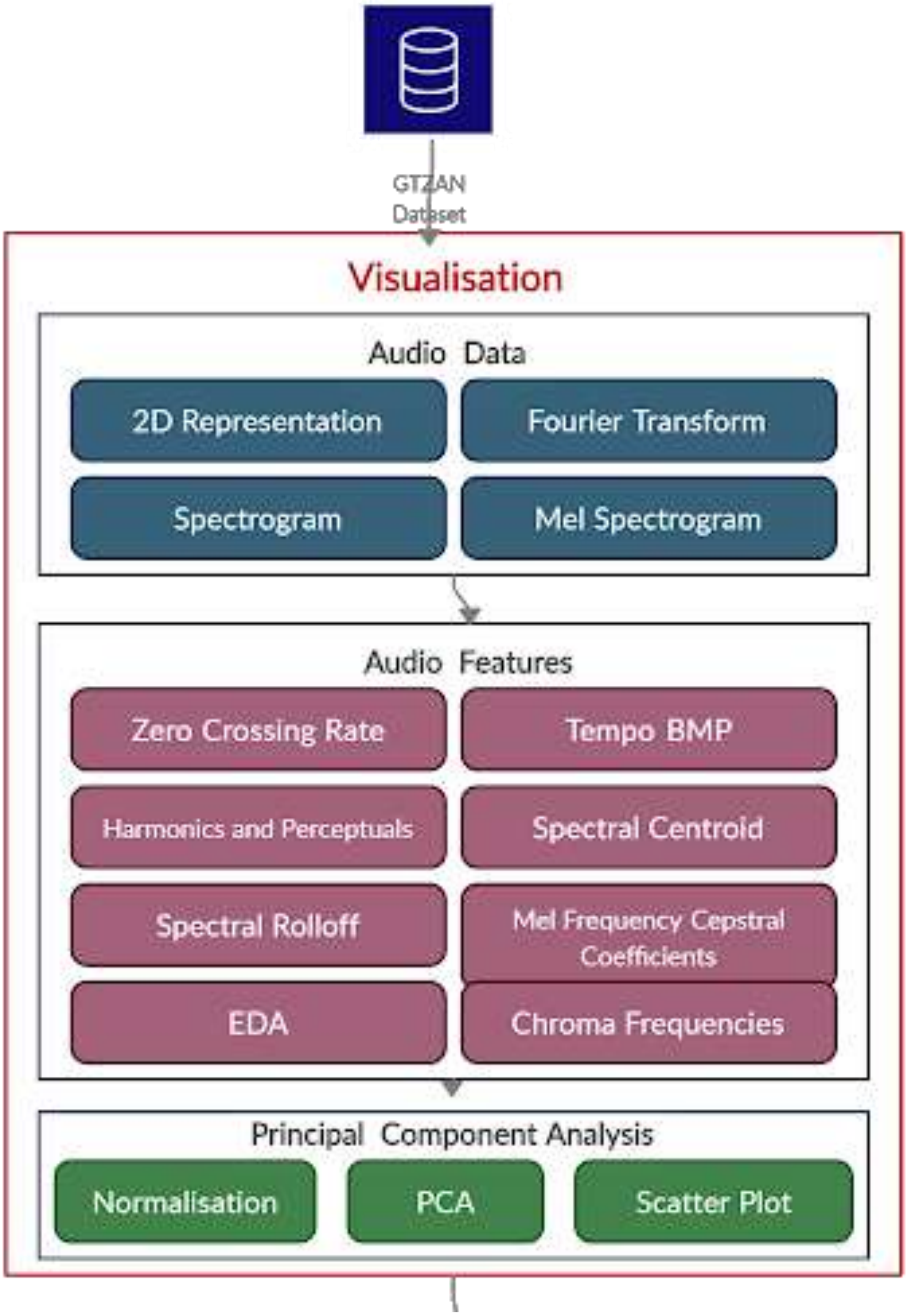
11. **Cross Gradient Booster (Random Forest) :** Cross Gradient Boost as discussed is used for training Gradient Boosted Decision Trees and other Gradient Gradient Boosted Models. Random Forest is a different training algorithm but it also uses the same inference and model representation as decision trees. Thus, both this algorithm can be combined to make Cross Gradient Booster Random Forest Model where one can use Cross Gradient Boost to train standalone random forest or random forest is used for gradient boosting as a base model.

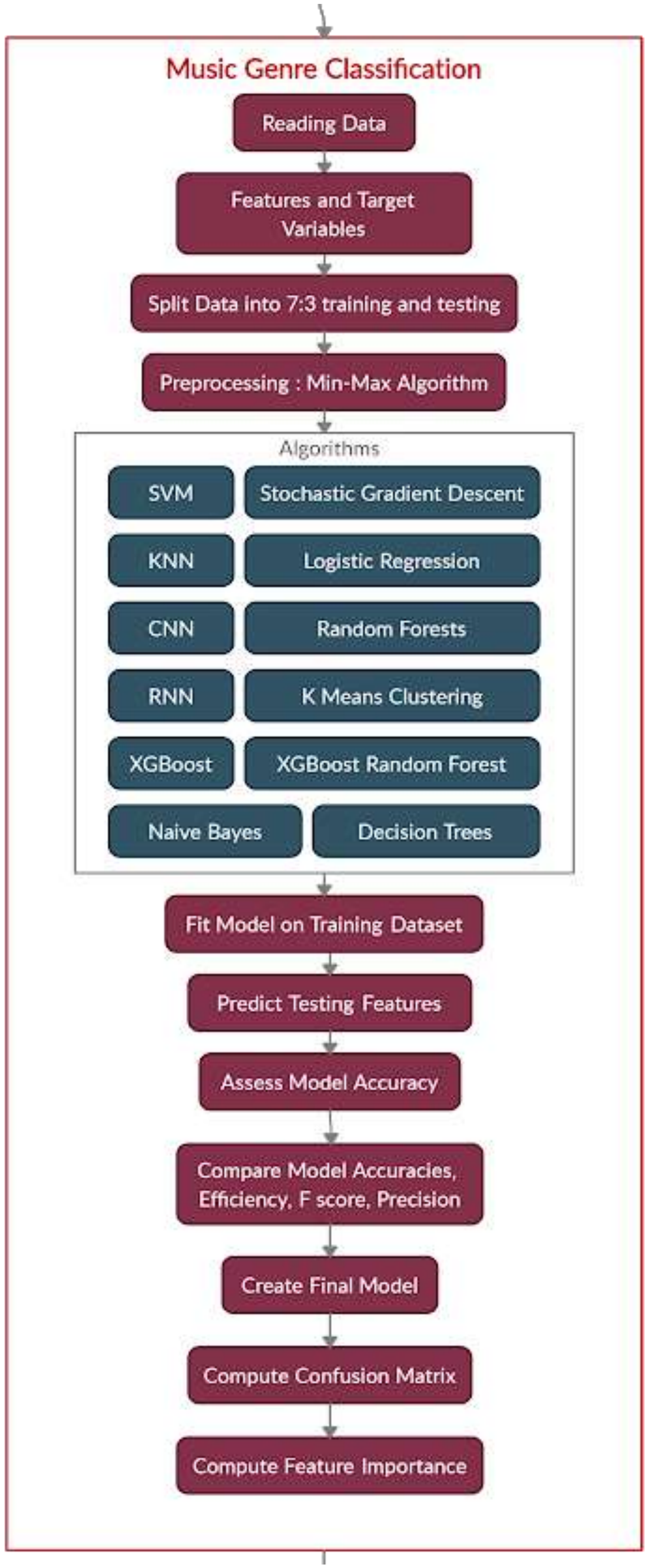
12. **Naive Bayes:** Naive Bayes classifier works on the Bayes Theorem. This is a great classification algorithm which is very useful in case of millions of records and even real time data. This model specifically gives great results in terms of textual data and thus can be a good model in case of content based music genre classification. It works on conditional probability. Given that something else has already occurred, the probability of happening something is known as Conditional probability. We can calculate the probability of an event using the beforehand knowledge with the help of conditional probability. This algorithm helps in classifying and predicts classes or membership probability that given record or data point can belong to. The class which has the highest probability will be considered as the appropriate class and will be known as Maximum A Posteriori (MAP).

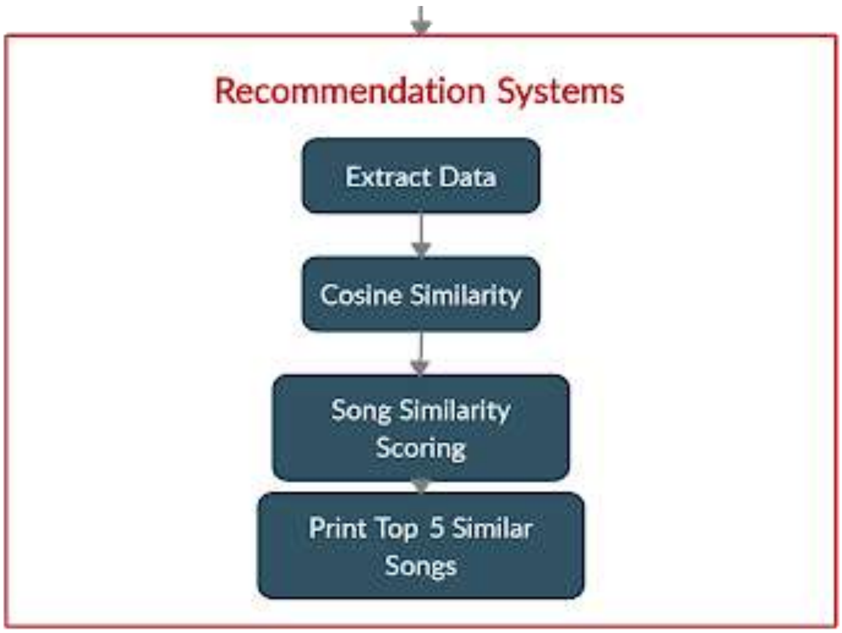
D. Architecture Diagram



E. Detailed Architecture Diagram



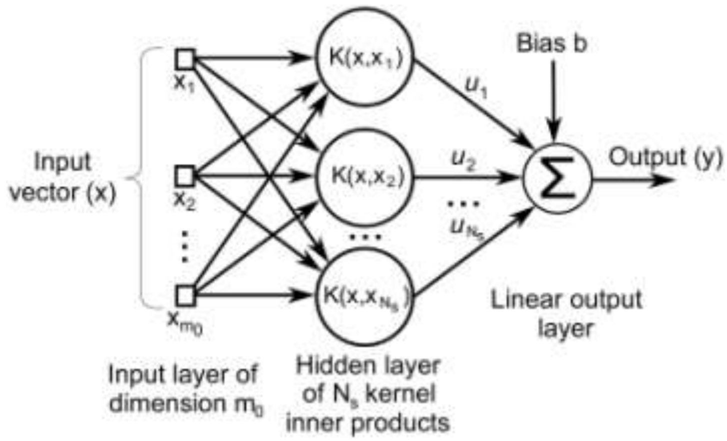

Music Genre Classification
Reading Data
Features and Target Variables
Split Data into 7:3 training and testing
Preprocessing : Min-Max Algorithm
Algorithms
SVM
Stochastic Gradient Descent
KNN
Logistic Regression
CNN
Random Forests
RNN
K Means Clustering
XGBoost
XGBoost Random Forest
Naive Bayes
Decision Trees
Fit Model on Training Dataset
Predict Testing Features
Assess Model Accuracy
Compare Model Accuracies, Efficiency, F score, Precision
Create Final Model
Compute Confusion Matrix
Compute Feature Importance



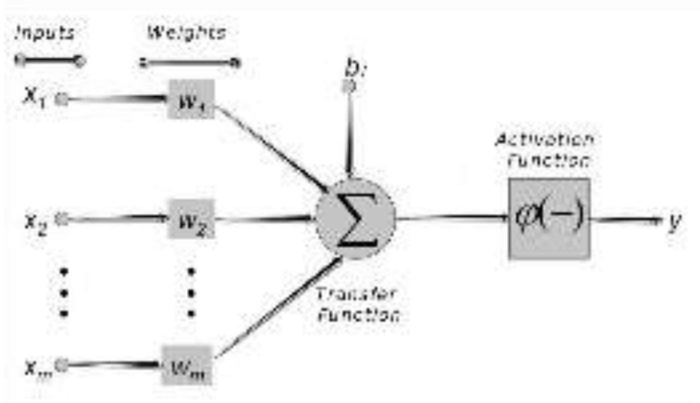
The above diagram depicts the detailed architecture diagram which covers the detailed flow/architecture of - 1. Visualisation, 2. Music Genre Classification, 3. Recommendation Systems.

F. Detail Model Design

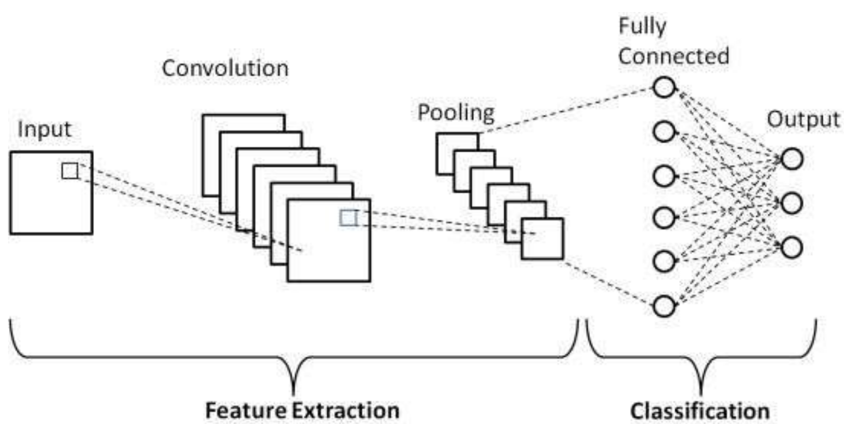
1. Support Vector Machines (SVM)



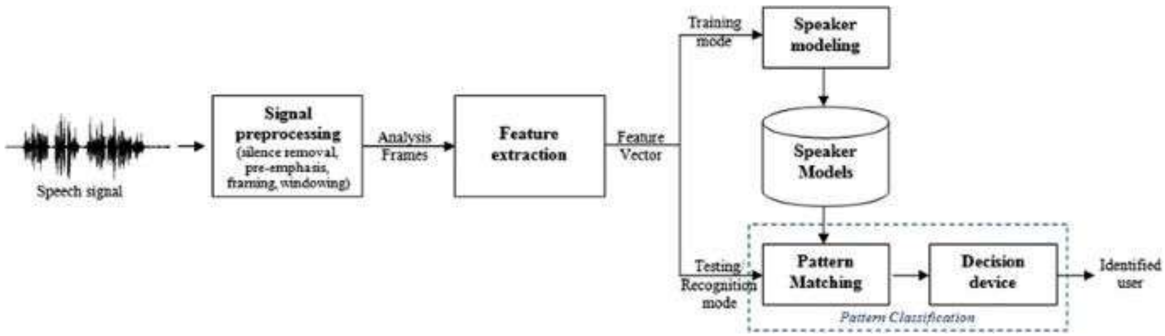
2. K- Nearest Neighbours (KNN)



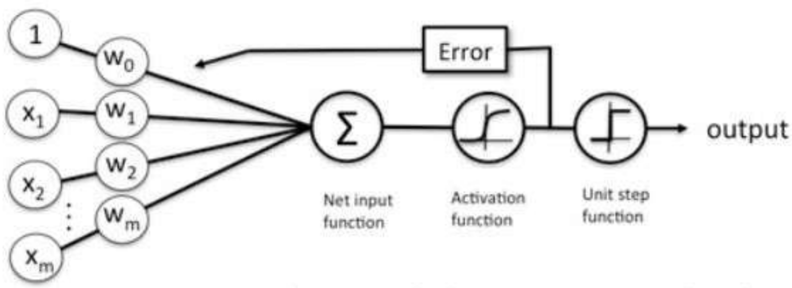
3. Convolutional Neural Networks (CNN)



4. Stochastic Gradient Descent

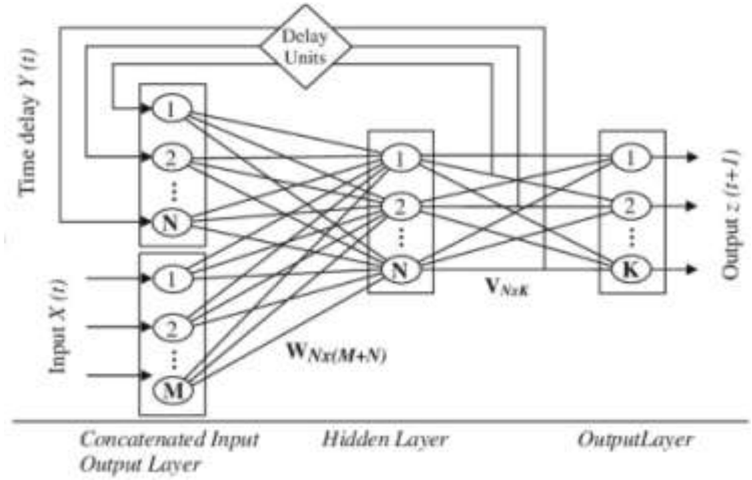


5. Logistic Regression

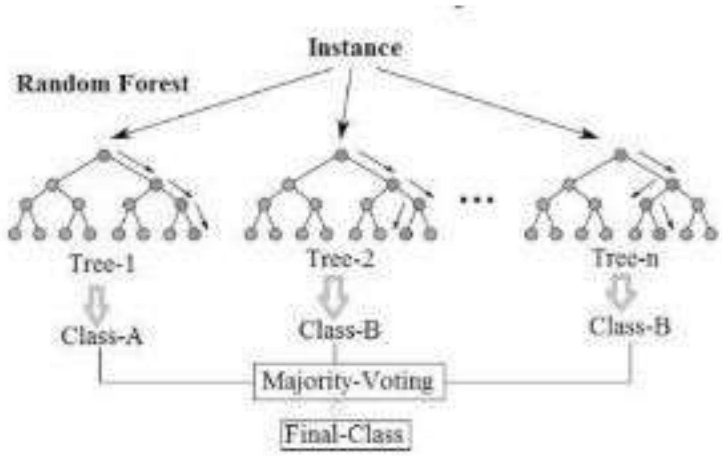


Schematic of a logistic regression classifier.

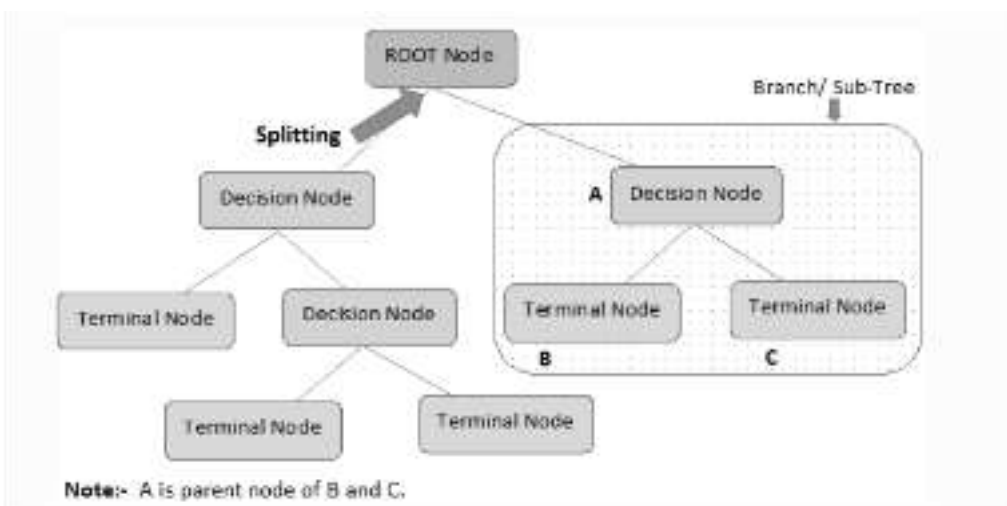
6. Recurrent Neural Network (RNN)



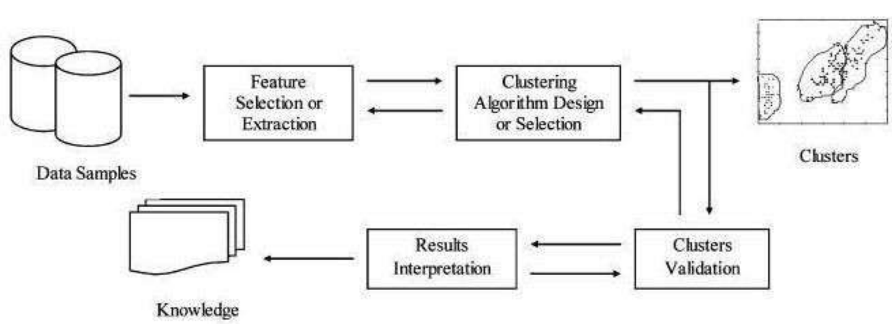
7. Random Forest



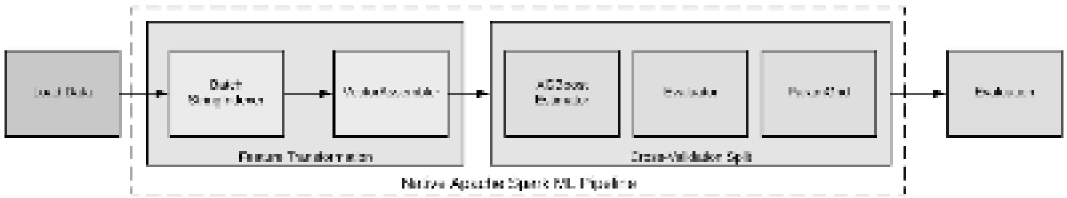
8. Decision Trees



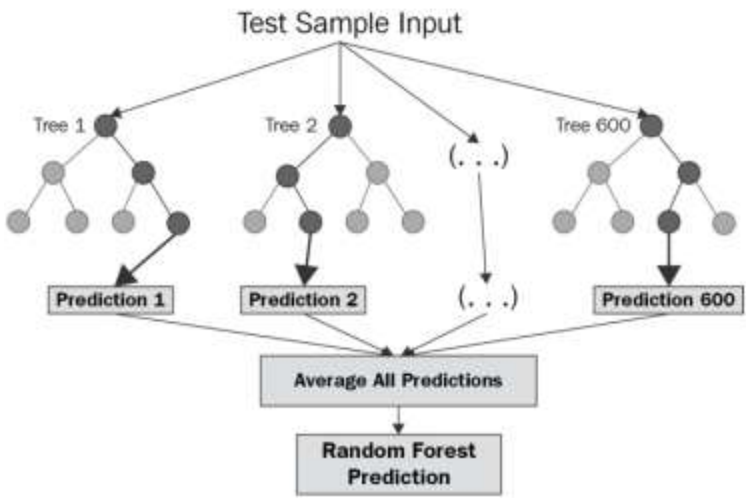
9. K-Means clustering



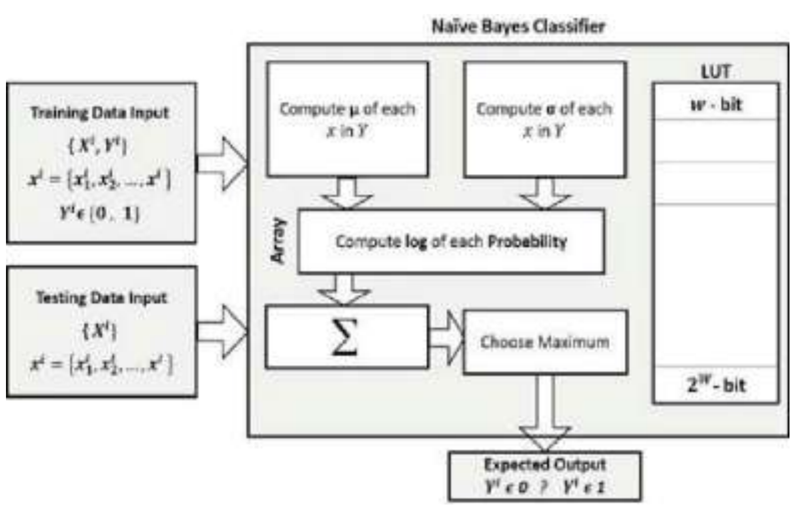
10. XG Boost



11. XG Boost with Random Forest



12. Naive Bayes



G. Work Done

We use librosa, the mother of audio files.

1. Understanding Audio

Initially, we explore the audio data. We use `sample.wav` file to understand the features and visualise them. We first calculate the value of sound and sample rate. We then trim the audio file such that the leading and trailing silence is eliminated.

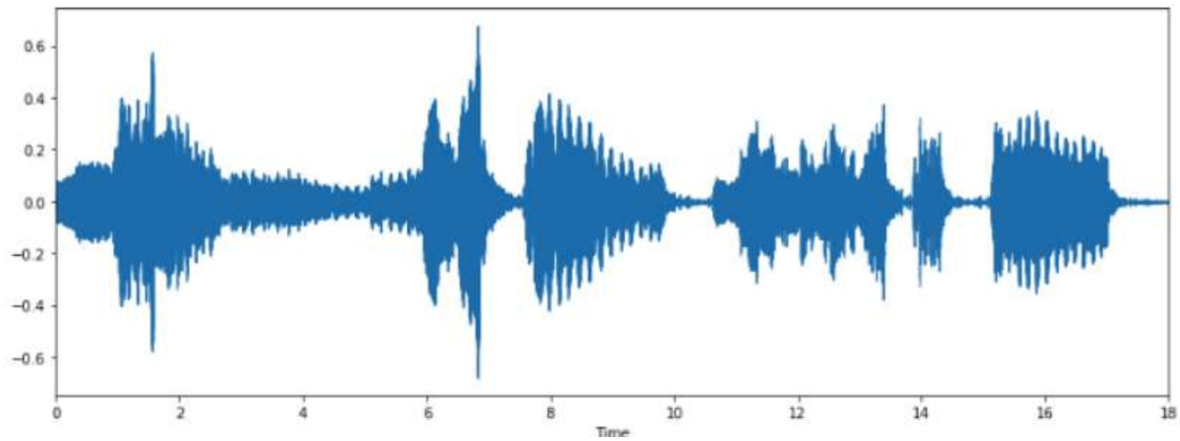


Figure 1. Visualising sound waves in 2D space

Moving on, we visualise the spectrogram of the audio file, which is a representation of the spectrum of frequencies of a signal with respect to time. Given below, is an amplitude spectrogram converted into a decibel scale spectrogram. The audio file used is *reggae.00036.wav*.

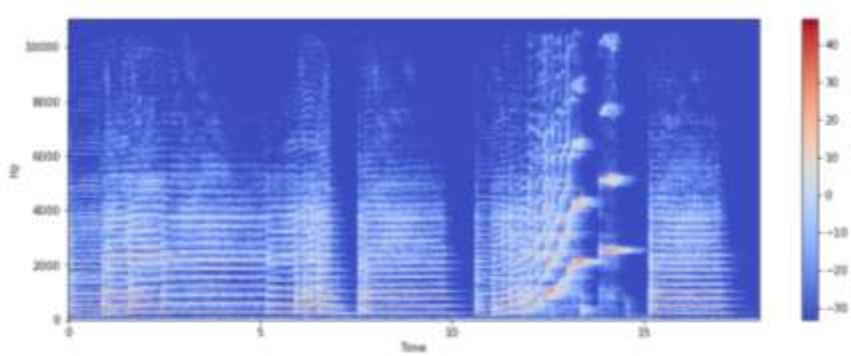


Figure 2. Decibel Scale Spectrogram -sample.wav

Next, we visualise the Mel Spectrogram. The Mel scale is a result of a non linear transformation of the frequency scale. It is a regular spectrogram but with the Mel scale on the y axis. The audio file used for this is *sample.wav.*

The same visualisation on *sample.wav* gives the following result.

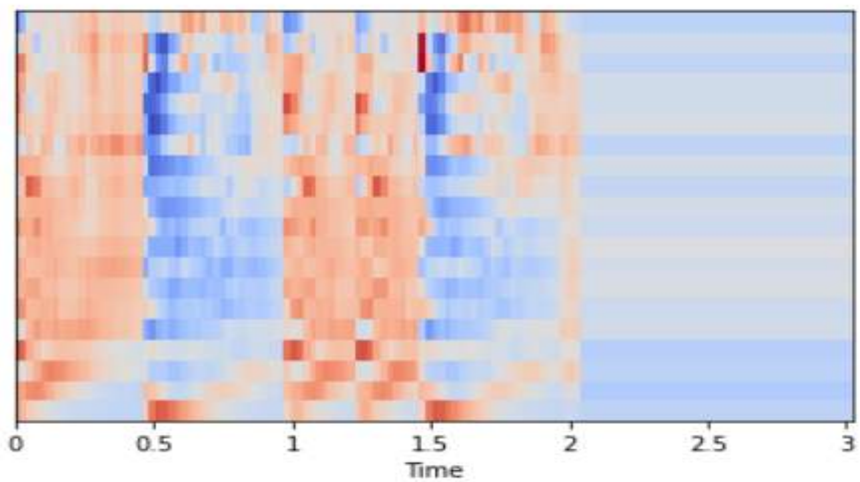


Figure 3. Classical Mel Spectrogram - sample.wav

We then visualise the features of the audio file such as **Zero Crossing Rate, Harmonics and Perceptual, Tempo BPM, Spectral Centroid, EDA Spectral Rolloff and theMel-Frequency Cepstral Coefficients.**

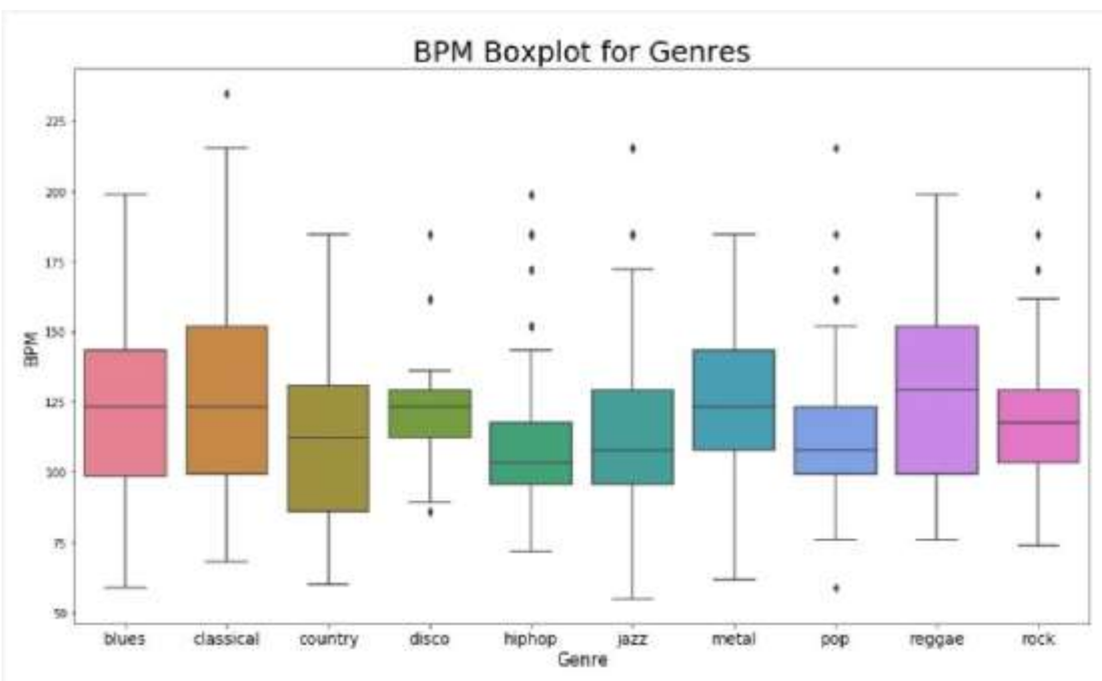


Figure 4. Boxplot depicting distribution of Genres

We perform Principal Component Analysis (PCA) in order to visualise the possiblegroups of genres.

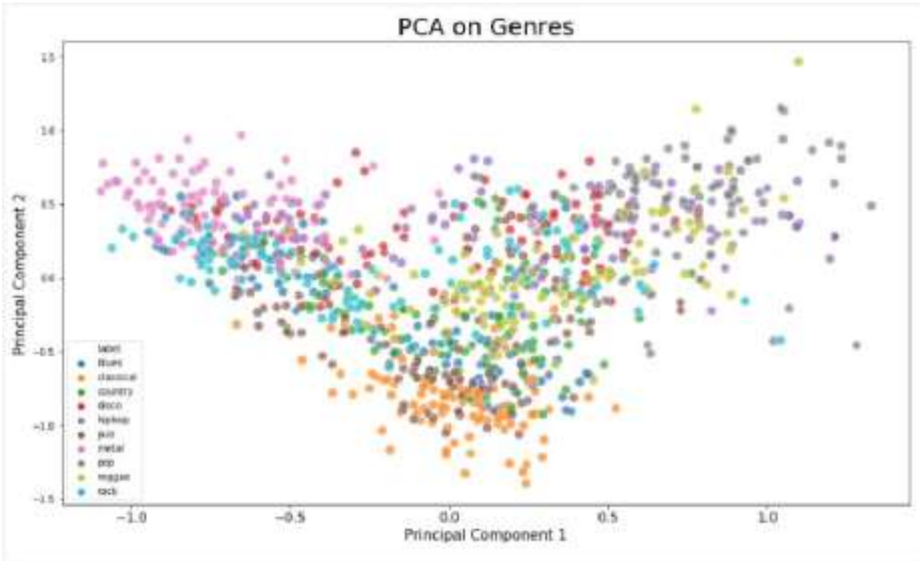


Figure 6. Scatter plot depicting PCA on Genres

Preprocessing (Min-Max algorithm)

The data is normalised before being used. Min-Max Normalisation algorithm was used in this research to obtain normalised data set. Min-Max Normalisation is one of the most common methods used for performing normalisation. This algorithm normalises the values such that the most minimum value becomes 0 and the most maximum value is transformed to 1. Other values that occur are transformed into decimal values between 0 and 1. Min-Max Algorithm is not capable of handling outliers well and this is one of the drawbacks of this algorithm. Following is the formula that is used to perform this normalisation:

$$\frac{value - min}{max - min}$$

Music Genre Classification

We read the *features_3_sec.csv* file to build a classifier that accurately predicts the genre for any audio file input. We create the target and feature variables, normalise the data and then split it into 70% to 30% ratio as training : test data. On executing the algorithms and assessing the accuracy, we found that **XGBooster was the best** performing model.

```
Accuracy Naive Bayes : 0.51952

Accuracy Stochastic Gradient Descent : 0.65532

Accuracy KNN : 0.80581

Accuracy Decision trees : 0.64631

Accuracy Random Forest : 0.81415

Accuracy Support Vector Machine : 0.75409

Accuracy Logistic Regression : 0.6977

Accuracy Neural Nets : 0.67734

Accuracy Cross Gradient Booster : 0.90224

Accuracy Cross Gradient Booster (Random Forest) : 0.74875
```

We find XGBooster to be the model with highest accuracy - 90% (approx)

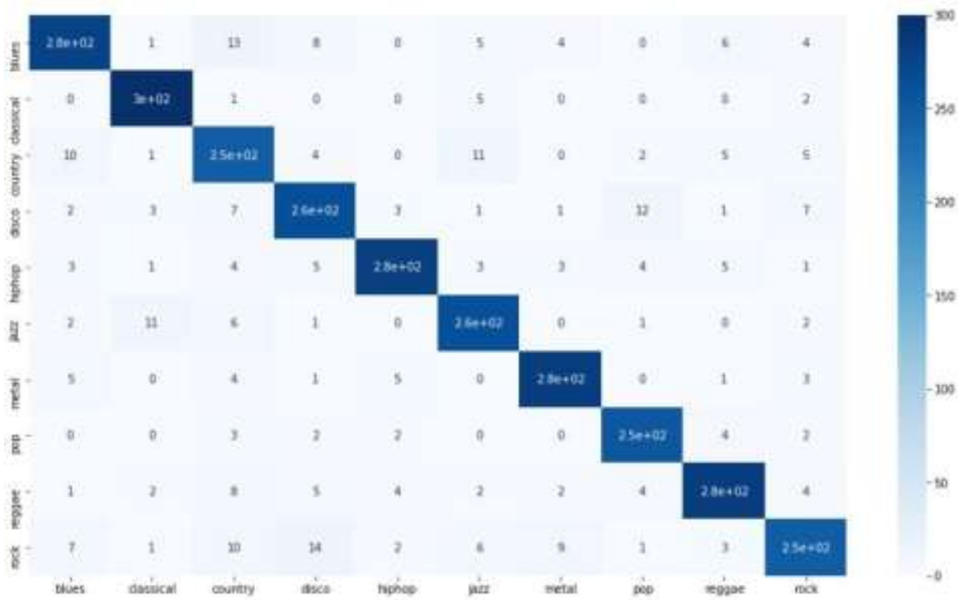


Figure 7.Confusion Matrix for XGBooster Algorithm

Weight	Feature
0.1205 ± 0.0095	perceptr_var
0.0416 ± 0.0031	perceptr_mean
0.0390 ± 0.0049	mfcc4_mean
0.0345 ± 0.0044	chroma_stft_mean
0.0339 ± 0.0062	harmony_mean
0.0280 ± 0.0065	harmony_var
0.0228 ± 0.0049	mfcc9_mean
0.0208 ± 0.0049	mfcc6_mean
0.0181 ± 0.0024	rms_var
0.0174 ± 0.0026	mfcc3_mean
0.0148 ± 0.0031	spectral_bandwidth_mean
0.0147 ± 0.0056	mfcc11_mean
0.0137 ± 0.0046	tempo
0.0116 ± 0.0036	chroma_stft_var
0.0113 ± 0.0026	mfcc7_mean
0.0109 ± 0.0038	mfcc1_var
0.0101 ± 0.0029	mfcc3_var
0.0089 ± 0.0057	mfcc8_mean
0.0089 ± 0.0020	mfcc5_mean
0.0072 ± 0.0038	mfcc18_mean

Figure 8. Feature importance along with their weights-Screenshot

Sl no.	**Weight**	**Feature**
1.	0.1205 ± 0.0095	Perceptron Variance
2.	0.0416 ± 0.0031	Perceptron Mean
3.	0.0390 ± 0.0049	MFCC Mean
4.	0.0345 ± 0.0044	Chroma Mean
5.	0.0339 ± 0.0062	Harmony Mean

Table III.1. Top 5 Most Important Features that determine the Music Genre

2. Recommender Systems

Firstly, we scale the input data: *features_30_sec.csv* file. Using *cosine_similarity* library, we find the best similar matches ranked in descending order for any given vector input. It calculates the similarity of two audio files, pairwise and generates a 1000 x 1000 matrix where every cell depicts the similarity of the corresponding audio files. We use the *features_30_sec.csv* file to predict the same. We write a function named *find_similar_songs* that takes the name of a song and returns the top five best matches for the input.

IV. Result Inference and Analysis

Algorithm	Accuracy	Efficiency	F-score	Precision
SVM	0.75409	0.75410	0.75210	0.75209
K-Nearest Neighbours	0.80581	0.80567	0.79432	0.79338
Convolution Neural Networks	0.67734	0.67723	0.66478	0.66473
Stochastic Gradient Descent	0.65532	0.65527	0.64213	0.64205
Logistic Regression	0.69970	0.69966	0.68778	0.68764
Recurrent Neural Networks	0.69012	0.69034	0.68982	0.68991
Random Forest	0.81415	0.81427	0.80126	0.80113
Decision Trees	0.64631	0.64660	0.63990	0.63984
K-Means Clustering	0.72146	0.72123	0.71291	0.71278
Cross Gradient Booster	0.90224	0.90222	0.90121	0.90125
Cross Gradient Booster (Random Forest)	0.74875	0.74863	0.73245	0.73249
Naive Bayes	0.51952	0.51940	0.50349	0.50412

Accuracy, efficiency, f-score and precision are few of the most important parameters that are used to evaluate the performance of the algorithms used and decide which one to finally incorporate in the project. We can deduce from the above table that the algorithm with the highest accuracy, efficiency, f-score and precision is the *Cross Gradient Booster Algorithm* and the one with the lowest is *Naive Bayes Algorithm.* Hence it is ideal to choose the **Cross Gradient Booster** for building our model.

V. Applications

Music Genre classification has wide areas of applications. This is the basis for many applications that can be made using this classifier. Genre of the music is a very useful knowledge that can be used for a wide variety of manipulations as it provides some great insight.

1. Music Genre Classification can be used to build a music Recommender System. A music Recommender system understands your choice of music on the basis of your selections and then recommends suitable music that you can listen to. Being able to classify the

type of music that the user listens to will enable the model to suggest even more music in the same category.

2. Music Genre Classification can be used in Sentimental analysis systems. Based on the genre of music the user is listening to or would like to listen to, the system can understand the genre of music and then predict the mood of the person. For example - Person listening to pop music can be happy.
3. Music Genre classification can be used to automatically make playlists for songs. This is a system that many famous applications use in order to make playlists and suggest them to users.
4. Music Genre classification enables systems to manage, classify and store the large amount of data we have in audio format. Genre classification will help systems to automatically classify them making everything very organised.
5. Good classification and categorisation of data will also help making retrieval systems much more efficient. Retrieval system is something that enables you to retrieve based on some provided data like the lyrics, category, music itself etc. Thus music genre classification will also play a major role in this field.

Thus we can see that music genre classification is the basis for a wide variety of problems and it can provide very useful knowledge about any music.

VI. New Learning Experience

We were completely new to the domain of Artificial Intelligence. By successfully implementing this project, we learnt various concepts and algorithms that enhanced our knowledge in the area.

We used Kaggle notebook due to the ease and efficiency.

1. Firstly, we performed a literature survey on the existing methods to analyse the various algorithms used to classify music.

We learnt the working of these algorithms such as Naive Bayes, XGBoost, Support Vector Machines and so on.

2. We thoroughly analysed how each algorithm processes the data on taking the audio file input and provides the respective music genre as the output.

3. We further learnt more on various libraries such as `pandas, numpy, seaborn, matplotlib.pyplot, sklearn` and the mother of a6ll audio files, `librosa.`

4. We began by trying to understand the features of an audio file. We visualised the files by performing fourier transforms on the sin waves and also learnt more on various kinds of spectrograms and their significance.

5. We performed a Principal Component Analysis on the dataset and studied the scatter chart to visualise possible groups of genres.

6. Then we performed the classification using the algorithms mentioned above and assessed the model accuracy to compare their performances.

7. On finding **XGBoost** to be the best model, we created the final model and displayed the confusion matrix to visualise the accuracy.

8. On performing feature importance using XGBoost, we learnt about the top features that play an important role in classifying the music.

9. After implementing the classification system, we were further interested in implementing more functionalities into the project. This led us to learning more on recommendation systems.

10. Using cosine similarity we could successfully implement the recommendation system that outputs top 5 similar songs based on the audio input.

    Overall, it was a great learning experience and we are looking forward to experimenting more in this domain.

VII. Innovation

1. **Optimal Approach** - After going through quite a good number of research papers, we discovered an efficient approach and an optimal flow of building a music genre classification model which had around 90% accuracy that had not really been used before.

2. **Features** - Discovered not only a few but around ten important features from the dataset that determine the genre of music and affect the classification criteria extensively. Few of those being Chroma frequencies, zero crossing rate and spectral rolloff. These features play a vital part of being the distinguishing factors among hundreds of different existinggenres.

3. **Recommendation** - We managed to create a recommender system that suggests similar songs to the user based on the given input. The algorithm considers the input given by the user and finds similar genre songs and lists them to the user based on his/her likings. Again, feature extraction comes into the picture to perform this function.

4. **Comparison** - We took the 12 of the best existing algorithms that are being used today in similar technologies and individually observed their performance. This was followed by an advanced comparison mechanism of all the algorithms together by assessing their efficiency and accuracy percentage to find the best one which was found to be : XGBoost.

VIII. Conclusion

As per the accuracy results, we find that **XGBooster has the highest accuracy of 90%**, and then Random Forests and KNN have an accuracy of 81% and 80% respectively. We build our final model using the XGBooster algorithm and perform feature importance. We find *perceptr_var, perceptr_mean and mfcc4_mean* to be features that have the highest weights of importance. Finally, using the cosine_similarity library we find the best similar matches to the given input audio.

IX. References

[1] Li, T., Ogihara, M., & Li, Q. (2003, July). A comparative study on content-based music genre classification. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 282-289).

[2] Dabas, C., Agarwal, A., Gupta, N., Jain, V., & Pathak, S. (2020). Machine Learning Evaluation for Music Genre Classification of Audio Signals. International Journal of Grid and High Performance Computing (IJGHPC), 12(3), 57-67.

[3] Heerde, F., Vatolkin, I., & Rudolph, G. (2020, April). Comparing Fuzzy Rule Based Approaches for Music Genre Classification. In International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar) (pp. 35-48). Springer, Cham.

[4] Bahuleyan, Hareesh. (2018). Music Genre Classification using Machine Learning Techniques.

[5] A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydin, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Adana, 2018, pp. 1-5, doi: 10.1109/ASYU.2018.8554016.

[6] Costa, Y.M., Oliveira, L.S., Koerich, A.L., Gouyon, F. and Martins, J.G., 2012. Music genre classification using LBP textural features. *Signal Processing*, *92*(11), pp.2723-2737.

[7] Chun Pui Tang, Ka Long Chui, Ying Kin Yu, Zhiliang Zeng, Kin Hong Wong (2018). Music Genre classification using a hierarchical Long Short Term Memory (LSTM) model. ICMR 18, 11-14 June 2018, ,Yokohama,Japan

[8] Jeremy Reed, Chin-Hui Lee (2006). A Study on Music Genre Classification Based on Universal Acoustic Models.

[9] Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing, 2016. Improved Music Genre Classification with Convolutional Neural Networks. INTERSPEECH 2016, San Francisco, USA