Using Machine Learning to Classify Music Genre

Rachaell Nihalaani
Computer Engineering Department, Thadomal Shahani Engineering College

Abstract: As Plato once rightfully said, 'Music gives a soul to the universe, wings to the mind, flight to the imagination and life to everything.' Music has always been an important art form, and more so in today's science-driven world. Music genre classification paves the way for other applications such as music recommender models. Several approaches could be used to classify music genres. In this literature, we aimed to build a machine learning model to classify the genre of an input audio file using 8 machine learning algorithms and determine which algorithm is the best suitable for genre classification. We have obtained an accuracy of 91% using the XGBoost algorithm.

Keywords: Machine Learning, Music Genre Classification, Decision Trees, K Nearest Neighbours, Logistic regression, Naïve Bayes, Neural Networks, Random Forest, Support Vector Machine, XGBoost

I. INTRODUCTION

Music is one of the most popular and widely loved art forms prevalent today. It is often a combination of instrumental and vocal sounds, expressing emotion, according to standards of rhythm, harmony, and melody. Music has been present in a number of styles throughout all periods of history all over the world. Music is a protean art that has and continues to permeate every human society in some way or the other, such as via ritual or drama. Today, popular culture incorporates music through film, musical theatre, radio, and obviously, the Internet. Music has the power to affect human behavior and has consequently been used in psychotherapy [1]. The emergence of various styles of music gave rise to a classification system called music genres. All music compositions that belong to the same genre are similar to each other in terms of form or style. Currently, there are roughly 1300 music genres in the world, each with its unique history and characteristics. Some of the most common and popular ones are classical, jazz, country, pop, and rock'n'roll [2]. With technological advancements, efforts have been made to delve into the research areas of music. Music genre classification is one such area where a considerable amount of research has been carried out, but there is always scope for more. Genre classification serves as a building block for other applications such as a music recommender model.

Machine Learning (ML) is a study of algorithms that learn from examples. Classification is a task that requires ML algorithms to learn how to assign class labels to examples. There are various types of classifications such as binary, multi-class, multi-label, etc., each with its own unique approach to problem-solving. Music genre classification is a problem best solved by machine learning. The machine learning model implemented in this paper uses multi-class classification, which is essentially a classification type wherein instances are classified into one among a set of more than two known classes [3]. The rest of this paper has the following organization. Section 2 discusses previous research in this area. Section 3 describes all algorithms necessary in the understanding of this paper. Section 4 discusses the dataset referenced and illustrates the methodology followed in building our model. Section 5 reveals the results obtained while Section 6 discusses the conclusion and further scope of the model.

II. LITERATURE REVIEW

This section reviews relevant past literature that has been published in the music genre classification domain. Paper [4] proposes two different approaches to solving the problem of music genre classification. It used a dataset of audio clips from Youtube. One approach comes under deep learning and trains a CNN model solely using its spectrogram. The spectrogram of the audio signal is treated as an image as VGG-16, which is an image classifier, is used to make predictions. The other approach trains traditional ML classifiers on time and frequency domain features from the audio signals. It was found that XGBoost is the best classifier. The former approach has been stated to outperform the latter approach. Finally, they observed an AUC value of 0.894 given by an ensemble classifier that combined both approaches. In [5], authors B. Liang and M. Gu have tackled the problem of audio-based classification using a transfer learning approach. They evaluated multiple models on a dataset of 1100 songs with 11 genres. They obtained an overall accuracy of 88% and the best model had 0.9799 ROC-AUC and 0.8938 PR-AUC. They faced the problem of misclassification and found that using thresholding instead of majority voting method would eliminate this problem.

Y. Yang et. al. have proposed in their paper [6], two BRNN-based models with serial and parallelized attention mechanisms. BRNN stands for Bidirectional Recurrent Neural network. They have compared both these mechanisms - serial and parallelized, and found that the latter is more flexible, relies less on BRNN, and gets better results in their experiment.

They have concluded that BRNNs with attention mechanisms, especially parallelized CNN attention, achieve better results and outperform previous work. [7] has established a new and better model for CNN after training and comparing a few classification models on the GTZAN dataset. It included the following models - Artificial Neural Networks, Support Vector Machine, Multilayer Perceptron, Decision Tree, and Convolutional Neural Network and observed accuracies of 70%, 68.9%, 68.7%, 74.3%, and 91% respectively. They concluded that their proposed CNN model gave the best results and the highest accuracy, with only slight misclassification. This review of past literature in the music genre classification domain has given us insight and guidance to go about building and implementing our music genre classification model.

III. ALGORITHMS

All the algorithms described below are supervised learning techniques that are used for classification problems. Our model is also a classification problem.

A. Decision Trees

It has a tree-like structure with two types of nodes - decision and leaf. The former has multiple branches and makes decisions based on some dataset features, while the latter is the output of those decisions. It is used to get all possible solutions to a problem [8].

B. K Nearest Neighbours (KNN)

It stores all data available to it in categories and classifies new data points into these categories based on similarity [9].

C. Logistic Regression

It involves fitting a logistic function that indicates the likelihood of something by giving a value between 0 and 1 [10].

D. Naïve Bayes

It is based on Bayes' theorem and makes predictions on the basis of the probability of an object [11].

E. Neural Networks

It learns by processing labeled data. With practice, it learns the characteristics required to generate the correct output [12].

F. Random Forest

It is based on ensemble learning and contains multiple decision trees on various subsets of the given dataset and takes the average, instead of relying on a single decision tree. This prevents overfitting and leads to a more accurate model [13].

G. Support Vector Machine

It creates the best decision boundary, called hyperplane, with the help of extreme points or vectors, called support vectors. This boundary helps classify new data in the correct position i.e. above or below [14].

H. XGBoost

It is a kind of boosting algorithm. Boosting is an ensemble learning technique that uses several weak classifiers to build a strong one. XGBoost stands for extreme gradient boosting. It is an extension of gradient boosting, with improves speed and performance. It is highly scalable and consequently widely used [15].

IV. PROPOSED FRAMEWORK

A. Dataset

For the purpose of building this Music Genre Classification model, we have used the GTZAN dataset [16] which has been widely used for machine learning research in the domain of music genre recognition. This is an appropriate dataset for our model as it has a wide variety, with 10 genres collected from various sources. The genres included in this dataset are - blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. For each of these genres, the dataset contains 100 audio files each, bringing it to a total of 1000 audio files. Each of these files has a length of 30 seconds. The audio files are in waveform audio file format with '.wav' extension. In addition to the audio files, the dataset also contains the visual representation for each, in the form of mel spectrograms, for further applications. Further, the dataset also contains two CSV files with audio features. One of them has mean and variance for each song, while the other one has the same structure with the audio files split into files of length 3 seconds, to increase the input data.

B. Methodology

Our music genre classification model has been implemented in Python language, in the following steps. We have made use of the Librosa package that is used for music and audio analysis. The first and foremost step is to load the dataset, followed by performing data exploration and modification steps such as trim leading and trailing silences from audio files, to increase efficiency. We visualize data in the form of 2D sound waves and spectrograms. As our input audio files are in waveform audio file format, we can easily plot a 2D sound wave plot once it is in the form of a NumPy array. To plot a spectrogram, we first need to perform a Fourier transform, which is essentially the decomposition of the input time-domain signal into frequencies. A spectrogram is a visual representation of the spectrum of these frequencies as it varies with time.

Now that we have visualized our data, the next step is to familiarize ourselves with the following audio features that are present:

1) *Zero crossing rate*: It is the rate of change of audio signal from positive to negative or the other way round.
2) *Harmonics:* They are sound waves with an integer multiple of fundamental tone as frequency.
3) *Perceptual*: It helps us understand the shock wave that represents sound rhythm and emotion.
4) *Tempo*: It is the beats per minute i.e. the speed of the beat.
5) *Spectral Centroid*: It indicates where the center of mass for a sound is located.
6) *Spectral Rolloff*: It is the frequency below which some part (say 75%) of the total spectral energy lies.
7) *Mel Frequency Cepstral Coefficients*: They are a small set of features describing the shape of the spectral envelope and sometimes may need data to be scaled.
8) *Chroma Frequencies*: They are a representation of audio wherein the entire spectrum is projected onto 12 bins, each representing a distinct semitone (or chroma) of the musical octave.

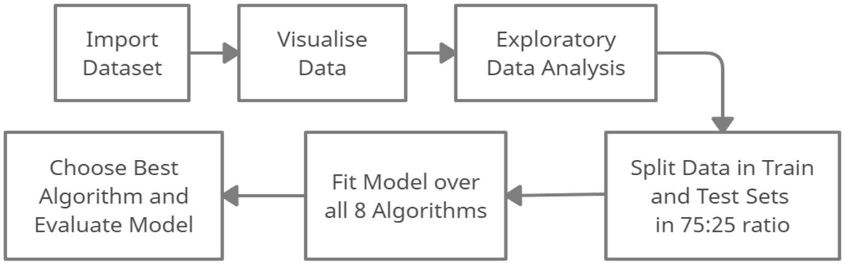
Import Dataset
Visualise Data
Exploratory Data Analysis
Split Data in Train and Test Sets in 75:25 ratio
Fit Model over all 8 Algorithms
Choose Best Algorithm and Evaluate Model

Fig 1. Proposed Framework for Music Genre Classification

Our dataset has another part that needs to be explored - the CSV file containing the mean and variance for each audio file. We visualize the means of these 10 genres with 100 audio files each in the form of a correlation heatmap. Additionally, the genre distribution is visualized as a box plot. Such extensive exploration and visualization are necessary due to the input files being audios. Now, we are fully equipped to begin building our model. We normalize the data as and when necessary and create target and feature variables. A target variable is a particular feature of the dataset that we want to understand better. Our target variable is the column containing the genres of the audio files and our feature variables are all of the columns containing audio features. The next step is to split the data into two sets, the first with 75% of the data will be used to train our model while the remaining 25% will be used to test it. We have used 8 classification models - Decision Trees, K-Nearest Neighbour, Logistic Regression, Naive Bayes, Neural Networks, Random Forest, Support Vector Machine, and XGBoost. Each of these has been explained briefly in Section 3. The preliminary model is fit over all 8 algorithms, using the training set and the accuracies are found. Whichever algorithm shows the highest accuracy is then improved upon, and a final model is created. This final model is evaluated over the testing set and a confusion matrix is returned. This confusion matrix is used to find values for the performance measure indices such as precision, recall, and F1 score.

V. RESULTS AND ANALYSIS

This section reveals the results we obtained on the implementation of our music genre classification model. First, we performed data visualization to get familiar with our data. For visualization purposes, we considered one instance of our dataset - the 11th audio file belonging to the Pop genre. Fig. 2 shows the 2D sound waves for this file. This wave plot has been generated using the Librosa library. Fig. 3 shows the spectrogram for the same file. A spectrogram is a visual representation of the spectrum of these frequencies as it varies with time.

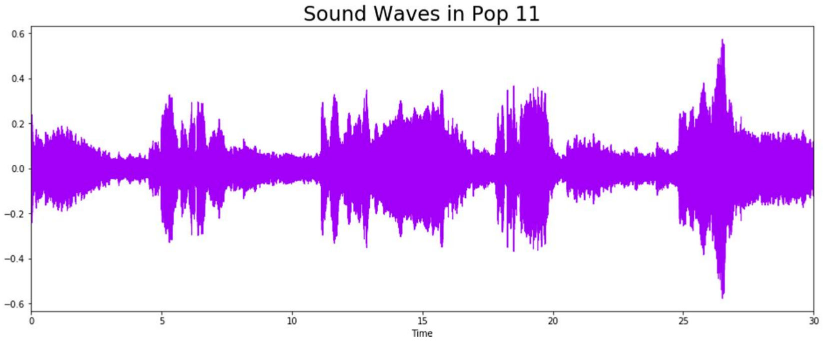

Fig. 2 2D Sound Waves for Pop 11

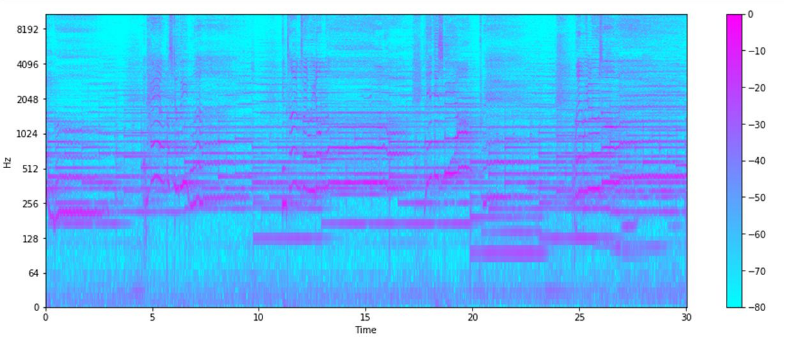

Fig. 3 Spectrogram for Pop 11

Second, we looked deeper into the audio features. Continuing with the same file belonging to the Pop genre, we have found the following values or graphs. This particular audio file had a Zero Crossing Rate of 57691, which means that the audio signal changed from positive to negative or vice versa 57691 times. Fig. 4 shows the Harmonics and Perceptual of the audio file. The portion in blue (the lighter shade in the case of grayscale) represents the harmonics and the portion in pink (the darker shade in the case of grayscale) represents the perceptual. This file had a Tempo of 107.666015625 which indicates that the speed was approximately 108 beats per minute. Fig. 5 and Fig. 6 show the wave plot of the Spectral Centroid and Rolloff of this file respectively. For both, the portion in pink indicates the normalized version, to help with visualization. Both the Spectral Centroid and the Rolloff (the darker shade in the case of grayscale) have been described in the previous section. The Mel Frequency Cepstral Coefficients have a shape of (20, 1293) this means that 20 MFCC's are calculated on 1293 frames. Fig. 7 shows the Chrome Frequencies, with a (12, 133) shaped chromagram.

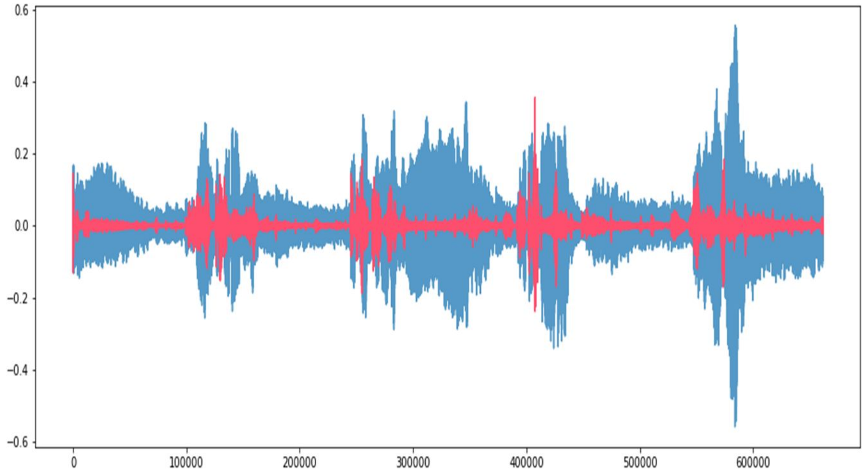

Fig. 4 Harmonics and Perceptual of Pop 11

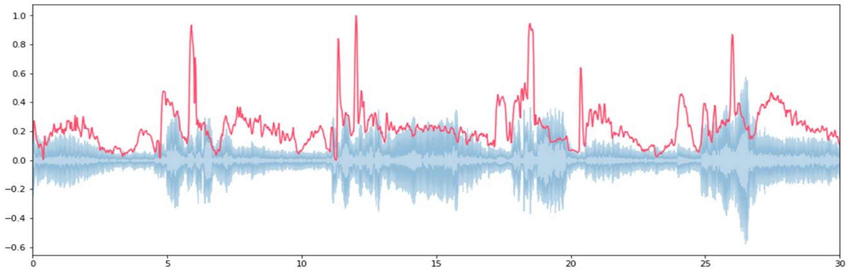

Fig. 5 Spectral Centroid of Pop 11

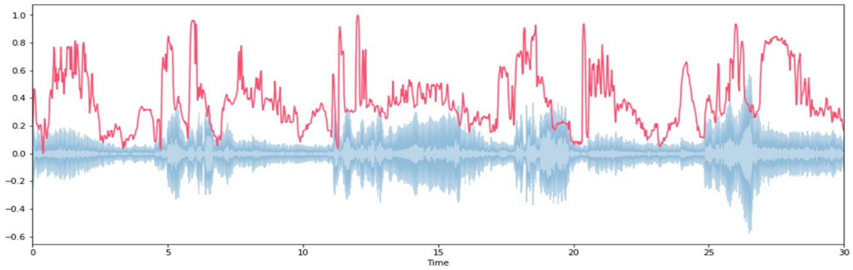

Fig. 6 Spectral Rolloff of Pop 11

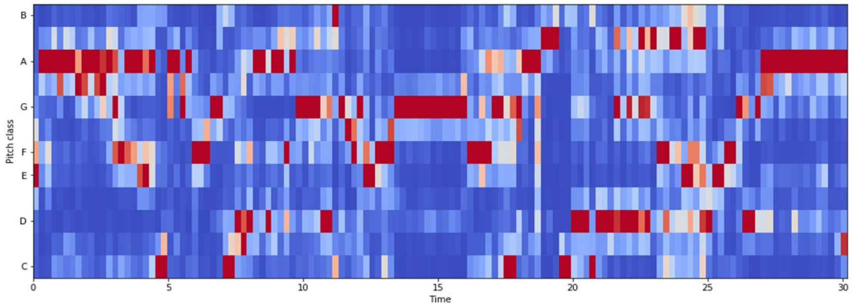

Fig. 7 Chroma Frequencies of Pop 11

Next, we fit our preliminary model over all of the 8 algorithms. Table 1 shows the accuracies observed. As is clear, XGBoost showed the highest accuracy of 91% followed by K Nearest Neighbours and Random Forest showing accuracies of 81%. We moved forward with XGBoost and evaluated it over the testing data set.

TABLE I
Algorithms and Accuracies

Algorithm	Accuracy	Algorithm	Accuracy
Decision Trees	0.63	Neural Networks	0.68
K Nearest Neighbours	0.81	Random Forest	0.81
Logistic Regression	0.69	Support Vector Machine	0.75
Naïve Bayes	0.51	XGBoost	0.91

Finally, we obtained the Confusion Matrix, given by Fig. 8. This multi-class confusion matrix is to be read as follows. The vertical labels are the actual class values and the horizontal labels are the predicted class values. For example, jazz has been wrongly predicted as classical 10 times, and country has been wrongly predicted as blues 8 times.

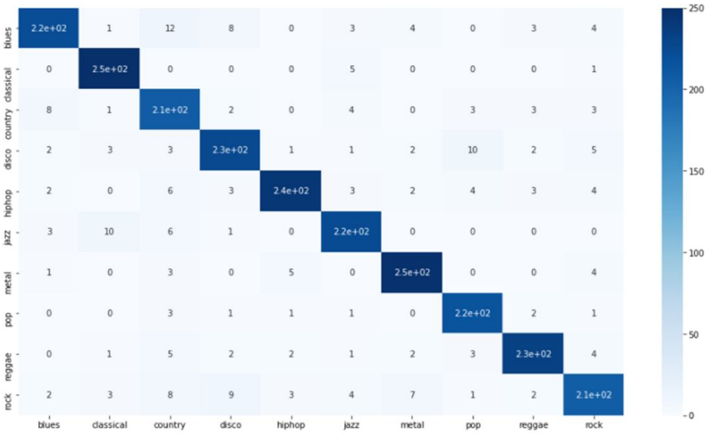

Fig. 8 Multi-class Confusion Matrix

VI. CONCLUSION AND FUTURE SCOPE

Music has played an important role in the world in the past and in the present, and it is very likely to continue to do so in the future. Significant advancements in the machine learning domain are also in the near future. Analysis of music via machine learning has opened up many possibilities in the field, many of these ideas have already been quite established. In this literature, we discussed a machine learning model that classifies music audio files based on their genre. We explored the dataset in detail, and fit our model on 8 algorithms - Decision Trees, K-Nearest Neighbour, Logistic Regression, Naive Bayes, Neural Networks, Random Forest, Support Vector Machine, and XGBoost. We found XGBoost to perform the best and give us an accuracy of 91%, which is the maximum of the 8 algorithms. This is a satisfactory starting point to build on. This model could be improved by considering numerous other appropriate machine learning algorithms or even using a different dataset. The future scope of this model could be using it to build a music recommender system, or even a music generation system.

VII. ACKNOWLEDGMENT

I would like to express my gratitude to Professor Juhi Janjua for her guidance and to my parents for their encouraging words.

REFERENCES

[1] G. Epperson. "music". Britannica.com. https://www.britannica.com/art/music. (accessed Sep. 30, 2021).

[2] “What is a Music Genre? Definition, Explanation & Examples.” promusicianhub.com. https://promusicianhub.com/what-is-music-genre/ (accessed Sep. 30, 2021).

[3] J. Brownlee. “4 Types of Classification Tasks in Machine Learning.” machinelearningmastery.com. https://machinelearningmastery.com/types-of-classification-in-machine-learning/ (accessed Sep. 30, 2021).

[4] H. Bahuleyan. (2018). Music Genre Classification using Machine Learning Techniques. [Online]. Available: https://arxiv.org/pdf/1804.01149.pdf

[5] B. Liang and M. Gu. (Aug. 2020). Music Genre Classification Using Transfer Learning. Presented at MIPR 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9175547

[6] Y. Yang, S. Luo, S. Liu, H. Qiao, Y. Liu and L. Feng. (2019). Deep attention based music genre classification. [Online] Available: https://www.sciencedirect.com/science/article/abs/pii/S0925231219313220

[7] A. Ghildiyal, K. Singh and S. Sharma. (2020). Music Genre Classification using Machine Learning. Presented at 2020 4th ICECA. [Online] Available: https://ieeexplore.ieee.org/abstract/document/9297444

[8] “Decision Tree Classification Algorithm”. javatpoint.com https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm (accessed Sep. 30, 2021).

[9] “K-Nearest Neighbor(KNN) Algorithm for Machine Learning”. javatpoint.com. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning (accessed Sep. 30, 2021).

[10] “Logistic regression in Machine Learning”. javatpoint.com. https://www.javatpoint.com/logistic-regression-in-machine-learning (accessed Sep. 30, 2021).

[11] “Naive Bayes Classifier Algorithm”. Javatpoint.com. https://www.javatpoint.com/machine-learning-naive-bayes-classifier (accessed Sep. 30, 2021).

[12] “Neural Network”. Deepai.com. https://deepai.org/machine-learning-glossary-and-terms/neural-network (accessed Sep. 30, 2021).

[13] “Random Forest Algorithm”. javatpoint.com. https://www.javatpoint.com/machine-learning-random-forest-algorithm (accessed Sep. 30, 2021).

[14] “Support Vector Machine Algorithm”. javatpoint.com https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm (accessed Sep. 30, 2021).

[15] “Understanding XGBoost Algorithm | What is XGBoost Algorithm?”. mygreatlearning.com. https://www.mygreatlearning.com/blog/xgboost-algorithm/ (accessed Sep. 30, 2021).

[16] GTZAN Dataset, Kaggle.com. [Online]. Available: https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification