International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr-2022 www.irjet.net p-ISSN: 2395-0072

Music Genre Classification using Machine Learning

Yukta Padgaonkar[1], Janhavi Gole[2], Bhanu Tekwani[3]

[1]Student, Department of Information Technology, Vidyalankar Institute of Technology, Maharashtra, India
[1]Student, Department of Information Technology, Vidyalankar Institute of Technology, Maharashtra, India
[2] Associate Professor, Department of Information Technology, Vidyalankar Institute of Technology, Maharashtra,India

---

Abstract - *Music is divided into different genres and sub genres. This is done by evaluating different characteristic of the music and putting them into categories/genre. Music genre classification is valuable not only for music journalist but also for artists, repertoire department and artists but the major reason for categorizing music is on an individual level to improve the listening experience. As the amount of music on the internet goes on increasing, it becomes very difficult to classify each and every song into a particular genre. Classifying music into genre helps in filtering through the songs or while giving suggestions and thus helps in music management. Machine Learning can help with this by classifying the music into different genres using a trained machine learning model.*

Key Words: machine learning, accuracy, music genre prediction, algorithms, precision

1.INTRODUCTION

Machine learning enables systems to learn from experience using data and take decisions without being explicitly programmed. It is a subset of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. The model is then tested on testing data and accuracy is obtained to evaluate the model. Supervised machine learning is an algorithm that learns from labeled training data to help you predict outcomes for unknown data. In supervised machine learning, you train the model using data that is well labeled. It means the data has correct outcomes. In Music Genre classification we are will use supervised Machine learning algorithms since we have a labelled dataset and we expect the model to give a certain kind of output. There are various supervised machine learning algorithms that we can use. In supervised machine learning, there are two kinds classification and regression. Since we need to classify the music files into different genres, we will use classification types of algorithms. But some of the algorithms can be used for both classification and regression like decision tree. Keeping that in mind we will be training various algorithms like Logistic Regression, Naive Bayes Classifier, K-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machines. Out of these the trained model that gives the highest accuracy will be selected.

2. Literature Survey

Nirmal Vekariya, Hemang Vyas, Nirav Dedhiya, in [1] "Music Information Retrieval And Genre Classification Using Machine Learning Techniques And Deep Learning", this research concludes that the maximum accuracy of 80% is obtained using Deep Neural Network for ten genre classes. They have also highlighted the facts on feature importance where features like rmse and chroma_stft stand out to be the most vital features.

Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia, Vidyullatha K S., in [2] "Music Genre Classification using Machine Learning Algorithms: A comparison", here, music genre classification is studied using the Free Music Archive small (fma_small) dataset. Two kinds of inputs were given to the models: Spectrogram images for CNN models and audio features stored in a csv for Logistic Regression and ANN model. Simple ANN was determined to be the best feature-based classifier amongst Logistic Regression and ANN models with a test accuracy of 64%. CNN model was determined to be the best spectrogram-based model amongst CNN, CRNN and CNN-RNN parallel models, with an accuracy of 88.5%. CRNN and CNN-RNN models are expected to perform well if the dataset is increased. Overall, image-based classification is seen to be performing better than feature based classification.

Muhammad Asim Ali, Zain Ahmed Siddiqui, "Automatic Music Genres Classification using Machine Learning", it was found out that the accuracy of classification by different genres and different machine learning algorithms is varied[3]. They extracted features and used k-NN and SVM Models. Overall, they found that SVM is more effective classifier which gave 77% accuracy for all features.

Vishnupriya S, K.Meenakshi, "Automatic Music Genres Classification using Machine Learning", this work shows provides a Convolution Neural Network based automatic music genre classification system[4]. The feature vectors are calculated using Mel Spectrum and MLCC. The python based librosa package helps in extracting the features and thus helps in providing good parameters for the network training. The learning accuracies are shown to be 76% and 47% for Mel Spec and MFCC feature vectors respectively.

Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R., "Music Genre Classification", in [5], they have extracted the feature vector from the audio files. They have used Convolutional neural network, where model is trained using the extracted feature vector. The model then gives the output. The training accuracy obtained here is 98% and the testing accuracy obtained is 73%.

[6] Derek A. Huang, Arianna A. Serafini, Eli J. Pugh, "Music Genre Classification", In this paper they have used K Nearest Neighbors, Simple Vector Machine, Feed forward neural network and convolutional neural network. All four models struggled with over fitting. The CNN performed the best, it took the longest time to train as well, but the increase in accuracy justifies the extra computation cost. However, there was similarity in accuracy between the KNN, SVM, and feed-forward neural network.

Archit Rathore, Margaux Dorido, "Music Genre Classification", This project aimed to create an automated system for classification model for music genres[7]. They used a total of five features, namely MFCC vector, chroma frequencies, spectral roll-off, spectral centroid, zero-crossing rate were used for obtaining feature vectors for the classifiers from the GTZAN genre dataset. They used K Nearest Neighbors, Linear Kernel SVM, Radial Basis Function (RBF) Kernel SVM, Polynomial Kernel SVM, Sigmoid Kernel SVM, Decision Tree, Random Forest, Ada Boost, Naives Bayes, Linear Discriminant Analysis (LDA) classifier, Quadratic Discriminant Analysis (QDA) classifier, Logic Regression. Out of these, classifier that works best is SVM with Polynomial Kernel. This is followed by Ensemble classifier and QDA. It was also seen that some classifiers work well for some genres.

3. Dataset

We have used the GTZAN dataset from Kaggle. The GTZAN dataset is the most widely used dataset for evaluation in machine Learning research for music genre classification. The dataset comprises of 10 genres with 100 audio files, each file with a length of 30 seconds. It has 2 CSV files, containing features of the audio files. Both the files have same structure, except for the length of song. In one file the songs are split into 3 second audio files whereas in other the songs are of 30 seconds.

4. Methodology

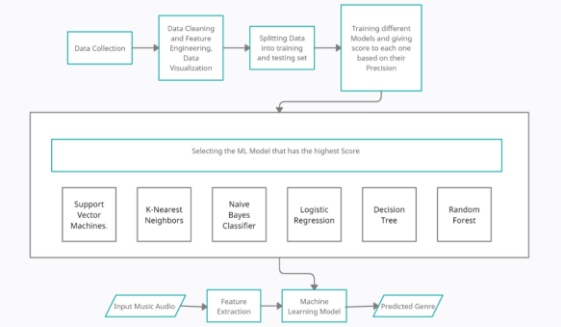

Figure 1: Architecture diagram

The architecture diagram explains the basic data flow of the entire methodology and explains each of the individual components of our approach. The different components are data collection, data cleaning and feature engineering, data visualization, splitting into training and testing set, selecting the model with highest accuracy, input music audio, feature extraction, giving features to selected machine learning model and predicting the genre.

Since we are using the csv file, we already have the data extracted in the required format. Feature selection also known as feature engineering helps the model to achieve a desired performance. While building a machine learning model, it is very important to select a robust set of features which will help us predict the correct outcome more effectively. In machine learning, the four categories of techniques that are used for feature selection are filter methods, wrapper methods, embedded methods and hybrid methods. We have used Random Forest Importance which is a technique that comes under embedded method. Using random forest importance, we obtained the top 20 features that have the most impact on the output label. We have also trained another set of classifiers with the all the features excluding name and length of audio and we have compared the results. We then split the dataset in the ratio 80:20, 80% will go to the training set and remaining 20% to the test set. After that we scale the data. The test and training data is scaled separately because we want our test data to be completely new to the model and we do not want any bias. When we scale the training data, we obtain the scaling parameters like mean and variance which are in turn used to scale the test data. The main reason to scale the data is that the model should not be biased to a particular feature of the dataset. Now we use this data to train our classifiers. We will use 6 classifiers namely, Support Vector Machine (with Linear, Polynomial, Sigmoid and RBF kernel), Decision Tree, Logical Regression, Gaussian Bayes, K Nearest Neighbors.

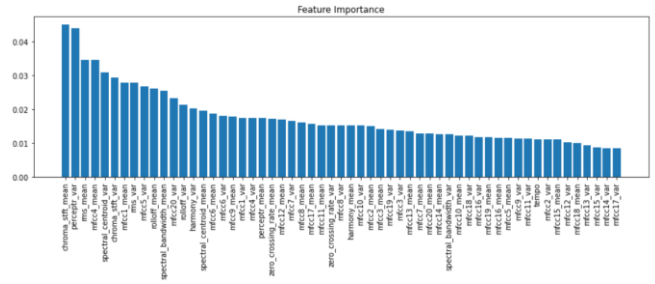

Figure 2: Feature Importance Plot

Classifiers

Support Vector Machine

Support Vector Machine is a supervised machine learning algorithm that which is used for both classification and regression. Each data item is plotted as a point in n-dimensional space where n denotes the number of features that are present. The value of each feature is the value of a particular coordinate on that dimension. After that, we find a hyper plane that will differentiate the two classes to perform classification. There can be multiple hyperplanes, we need to select the one that segregates the two classes better.

Decision Tree

Decision Tree is a supervised machine learning algorithm that which is used for both classification and regression. A decision tree is a flowchart-like structure in which each node represents a decision on a feature, each leaf node represents a class label or a decision taken after computing all features. Branches represent conjunctions or the possible values of the features that lead to those class labels.

Random Forest

Random Forest is a supervised machine learning algorithm that which is used for both classification and regression. Instead of using one model, it uses an ensemble of decision trees. All the different models are trained independently. The models are rained on different subsets of data. Some models may predict the correct result while others may not. After the results from all the models are taken, the final output is based on majority voting. This helps in predicting the correct outcome.

Logical Regression

Logical Regression is a supervised machine learning algorithm that which is used for classification where data is to be classified into two or more classes. In Logistic regression, we compute the probability of an outcome which lies between 1 and 0. If the probability lies beyond a certain value, the outcome is A or else the Based on whether the probability lies above or below that value, outcome is decided. A Sigmoid function is used to map the predicted values to probabilities.

Gaussian Bayes

Gaussian Naive Bayes is a variant of Naive Bayes that follows Gaussian normal distribution. It supports continuous data. Naive Bayes Classifiers are based on the Bayes Theorem. Here, an assumption is made that the value of a particular feature is independent of the value of any other feature. Here an assumption is made, while working with continuous data that the continuous values associated with each class are distributed according to a normal distribution.

K Nearest Neighbors

K Nearest Neighbors is a supervised machine learning algorithm that which is used for both classification and regression. It uses similarity to predict outcome of the new data. The distance between the feature of the new. We need minimum distance in features space from our new data point whose outcome is to be predicted. Then based on the minimum distance k points are selected. The class to which most number of points selected belong is obtained as the result.

5. Results and Discussion

We found that Support Vector Machine with RBF kernel performs the best with an accuracy of 74% followed by Support Vector Machine with polynomial kernel and Logistic Regression. Support Vector Machine with polynomial kernel and Logistic Regression obtained an accuracy of 69%. Naive Bayes performed the worst with an accuracy of 45.5%.

Table 1. Comparison of accuracy of different classifiers

SN.	Algorithm	Accuracy (with all features)	Accuracy (with top 20 features)
1	SVM with linear kernel	68.0	67.0
2	SVM with polynomial kernel	69.0	66.0
3	SVM with RBF kernel	74.0	65.5
4	SVM with sigmoid kernel	53.5	48.5
5	Random Forest	67.5	65.0
6	K-Nearest Neighbors	65.0	60.0
7	Decision Tree	49.0	47.5
8	Logistic Regression	69.0	66.5
9	Naive Bayes	45.5	53.5

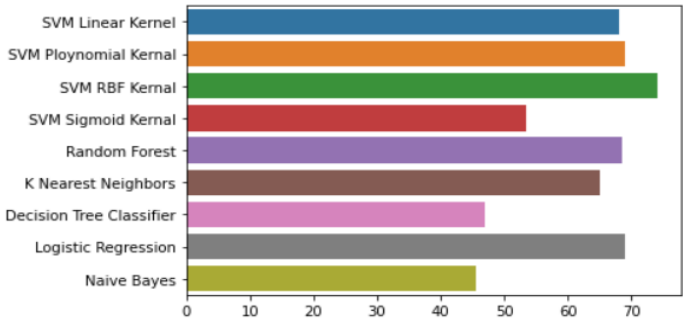

Figure 3: Classifier(trained with all features) performance comparison graph

We saw that as the number of features are reduced, the accuracy of the most of the classifiers goes on decreasing. This does not hold true for Decision Tree and Naive Bayes.

Confusion matrix

Confusion matrix provides information about the performance of the model created. It also gives information about which class/label is being predicted correctly and which label is being predicted incorrectly and as what label. It helps to know what errors are made and where errors are made. It is a summary of the predicted results, it tells you which results are calculated correctly and which ones are incorrect, using true positive (when a positive result is predicted correctly), true negative(when a negative result is predicted correctly), false positive(when a negative result is predicted as positive) and false negative (when a negative result is predicted incorrectly as a negative one). From the confusion matrix, we can see that SVM with RBF kernel trained on all features works best with blues, jazz, classical, disco and country genres. From the confusion matrix of other classifiers, we can see that most of the classifiers performed well on classical and blues genre.

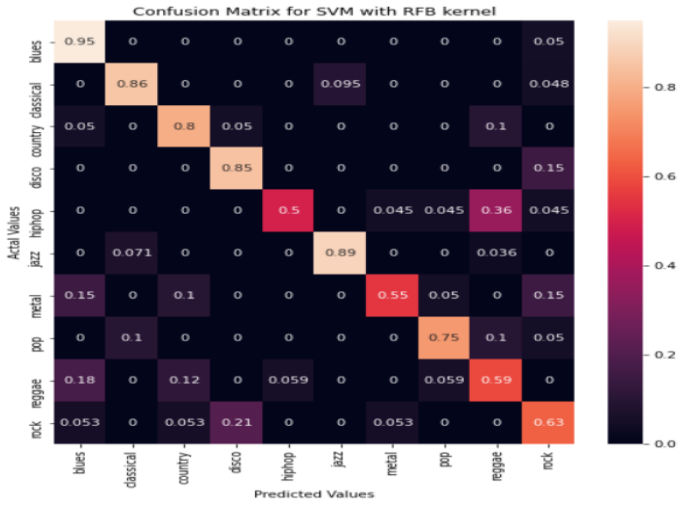

Figure 4: Confusion matrix for SVM Classifier with RBF Kernel

Precision, recall, f1-score, support

Precision indicated the quality of positive prediction made by a Machine learning model. Precision is equal to the number of true positives divided by the total number of positive predictions which means sum of true positives and false positives.

Recall indicates the sum of number of true positives that have been made and the number of positives that could have been made. It is equal to the number of true positives divided by the total number of true positives and false negatives.

F1-score is the harmonic mean of precision and recall. It is used to compare the performance of classifiers and it is used to measure model's accuracy on a dataset.

Support is equal to the number of times the class occurs in the dataset.

Table 2. Precision, recall, f1-score, support for all genres using SVM with RBF Kernel

SN.	Genre	Precision	Recall	F1-score	Support
1	blues	0.70	0.95	0.81	20
2	classical	0.82	0.86	0.84	21
3	country	0.76	0.80	0.78	20
4	disco	0.69	0.85	0.76	13
5	hiphop	0.92	0.50	0.65	22
6	jazz	0.93	0.89	0.91	28
7	metal	0.85	0.55	0.67	20
8	pop	0.83	0.75	0.79	20
9	reggae	0.43	0.59	0.50	17
10	rock	0.57	0.63	0.60	19
Accuracy	0.74				200

6. CONCLUSION

The classifier that works the best when trained with all the features is SVM with RBF kernel. SVM with Polynomial kernel and logistic regression also did well with an accuracy of 69%. When trained with the top 20 features is SVM with linear kernel with an accuracy of 67% followed by Logistic Regression and SVM with Polynomial kernel having with 66.5% and 66.0% accuracy. The accuracy of the classifiers increases as the number of features used for training the classifiers increases. The accuracy of the model can be increased by experimenting with other advanced machine learning algorithms. It can also be improved by finetuning the hyperparameters.

REFERENCES

[1] Nirmal Vekariya, Hemang Vyas, Nirav Dedhiya", Music Information Retrieval And Genre Classification Using Machine Learning Techniques And Deep Learning", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 07, July 2020

[2] Snigdha Chillara, Kavitha A S, Shwetha A Neginhal, Shreya Haldia, Vidyullatha K S., "Music Genre Classification using Machine Learning Algorithms: A comparison", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 05, May 2019.

[3] Muhammad Asim Ali, Zain Ahmed Siddiqui, "Automatic Music Genres Classification using Machine Learning", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 8, 2017.

[4] Vishnupriya S, K.Meenakshi, "Automatic Music Genres Classification using Machine Learning", 2018 International Conference on Computer Communication and Informatics (ICCCI -2017), Jan. 04 – 06, 2018.

[5] Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R., "Music Genre Classification", SSRG-IJCMS, Jan-April 2020.

[6] Music Genre Classification Report by Derek A. Huang, Arianna A. Serafini and Eli J. Pugh from Stanford University.

[7] - http://cs229.stanford.edu/proj2018/report/21.pdf

[8] Archit Rathore, Margaux Dorido, "Music Genre Classification", Indian Institute of Technology, Kanpur Department of Computer Science and Engineering, 2017.