Machine Learning in Music genre Classification

1st Jainam Patel
dept. CSE
Nirma University
Ahmedabad, India
20bce099@nirmauni.ac.in

2nd Chetan Jani
dept. CSE
Nirma University
Ahmedabad, India
20bce103@nirmauni.ac.in

3rd Madhav Kotak
dept. CSE
Nirma University
Ahmedabad, India
20bce133@nirmauni.ac.in

4th Hardi Joisar
dept. CSE
Nirma University
Ahmedabad, India
20bce112@nirmauni.ac.in

***Abstract*—As we have seen that when we listen to music on apps such as spotify, youtube, etc. we get recommendations based on what music we listen. This system is created by classifying the genres of music such as hip-hop, jazz, classical,etc. Music group categorization is very essential for recommendation systems of music. Prediction of genre of music is difficult job in the field of extraction of Musical Data. Machine Learning model is designed which automatically does the classification of genre of the music clip. Here, we directly classify the genres of music using dataset in which we have used three second of music of .wav format to train the model. There are approximately 55 features and 9990 files of music of different genres.**

I. Introduction

Music genre classification remains a challenging task within the domain of Music Data reclamation. Music genre classification is critical for music recommendation systems because music is heavily weighted in such algorithms and subsequent suggestions. A machine learning algorithm that autonomously identifies the type of the music track is created.Considering the present day innovative technology and easy access to massive data files over the web, it's critical to match each user's computing demands.Machine learning is a burgeoning field of artificial intelligence which has enabled these user expectations. Machine learning models are laying the foundation for classifiers such as music genre categorization, and therefore have proven to be highly effective in identifying classifications.

In present-day times,physical albums, discs and cassette are not the relevant sole resources of listening to songs , the net's introduction has exerted significant influence controlling the passage of data from the internet to every other individual on the planet who is linked to this connection. Anything from recordings to audiobooks, sound packets to audio clips may be downloaded in just a few seconds.The consequence is a vast collection of tracks and recordings strewn around in numerous files, which makes it difficult for a person to stay on top of the category of each track and order them accordingly.Categorizing recordings by labelling them to relevant genres is the most natural technique to managing such large quantities of audio files. An audio signal having attributes like amplitude, velocity, and frequency is used to represent sound. A machine can examine and interpret these acoustic input in a number of forms. Machine learning techniques can also be useful in audio assessment. Musical study utilizes a song's encrypted communications for a range of variables such as sound, dance ability, pace, intensity, etc. to determine the genre of songs a person might prefer listening to to listen to. Music is separated by genres, which are defined categories.

II. Music Genre

An audio signal having attributes like amplitude, velocity, and frequency is used to represent sound. A machine can examine and interpret these acoustic input in a number of forms. Machine learning techniques can also be useful in audio assessment. Musical study utilizes a song's encrypted communications for a range of variables such as sound, dance ability, pace, intensity, etc. to determine the genre of songs a person might prefer listening to to listen to. Music is separated by genres, which are defined categories.
In the field of Music Information Retrieval (MIR), that is concerned with examining, organising, and discovering large music collections, classifying is important. Sorting audio files to their appropriate genres is a challenging task. The main goal is to automatically categorise music into genres instead of manually entering it, using spectro-temporal characteristics, addressing an audio classification issue with a simple proposal. We want a model that can distinguish between several musical genres with reasonable accuracy.

III. Dataset

In machine learning research, the GTZAN is used to evaluate music genre recognition (MGR). Different recording settings were used, and data were acquired from several recording sources. It contains 1000 audio samples distributed evenly throughout 10 genres for just a total of 100 audio clips each genre. Each clip is around 30 seconds long. The structure for all tracks is as follows: Wav format, sampling rate of 22050Hz, mono 16-bit audio (.wav). This dataset includes 10 genres: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock.

IV. PROPOSED SYSTEM

Various ML techniques like Classifier Decision Tree,Classifier Support Vector Machine with Kernal,Classifier K-Nearest Neighbors,Classifier Naive Bayes and Classifier Multi-Layer-Perceptron are used and the technique with most accuracy is chosen and its confusion matrix is plotted.

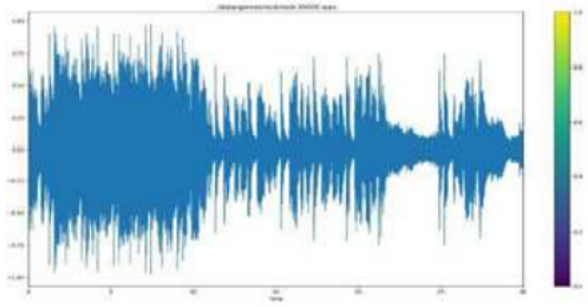

Fig. 1. Wave plot of an audio clip (Rock) in the dataset.

A. Classifier Decision Tree

A decision tree is a technique for extracting information from a data collection that is highly essential for data gathering. Typically, a TDIDT (Top-Down Induction Decision Tree) technique is used to extract information.It generates a tree structure through dividing data onto increasingly homogenous categories depending on the divide and conquer approach This splitting procedure comes to an end once the subset comprises only a category or if no further enhancement is conceivable. A pre-determined halted condition can also be used to interrupt the split process. A split can be represented mathematically by a test function t:-→ X Rt that translates instances into split outcomes, ensuring that each potential consequence of a node's split is connected with a different exiting branch. The C4.5 method is an ID3 evolution that use Gain Ratio as a classification technique. The component with the greatest Gain Ratio is chosen as the root node, and the data set is divided depending on the values of the root element. The Information Gain for each sub-node is determined, and the procedure is continued till the estimate is complete. CART (Classification and Regression Tree) and C4.5 algorithms employ an intensive research approach to identify the thresholds that will be employed by the nodes to partition the continuum characteristics.The phase 2, pruning, is in charge of reducing the tree's intricacy. It entails lowering the size of trees that are overly huge and rooted. The issue of distortion and overfitting lowers data efficiency and accuracy. First and foremost, the Gini Index must be introduced. The Gini Index is a criterion based on impurities that assesses the divergences in the probability distributions of the target attribute values. in the presence of a training set S and a goal characteristic y The signifies the selection of tuples, such as y=cj. S represents the selection of occurrences of the attribute y belonging to cj class, given the dataset S.

$$Gini(y,S) = 1 - \sum_{c_j \in dom(y)} (\frac{\sigma_{y=c_j} S}{|S|})^2.$$

As a result, the analysis factor for choosing the attribute ai is specified as As a splitting criterion, the C4.5 algorithm

$$GiniGain(a_i,S) = Gini(y,S) +$$

$$- \sum_{v_{i,j} \in dom(a_i)} (\frac{\sigma_{a_i=v_{i,j}} S}{|S|}) \times Gini(y, \sigma_{a_i=v_{i,j}} S).$$

employs the Gain Ratio. When the amount of instances to be divided falls below a particular limit, the splitting process is terminated. The algorithm can deal with numerical characteristics. To illustrate the Gain Ratio, we first must establish the notions of Information Gain and Entropy. A sample's Entropy reveals how mixed the class values are. It is made up of a purity measurement device. A number of 0 indicated that the data is homogenous, whereas a value of 1 suggests that it is disordered. The definition of entropy is as follows: As the algorithm picks which feature to split on,

$$Entropy(y,S) = \sum_{c_j \in dom(y)} -\frac{|\sigma_{y=c_j} S|}{|S|} \times log_2 \frac{|\sigma_{y=c_j} S|}{|S|}$$

Information Gain employs the Entropy metric to compute the shift in homogeneity coming from just a split on every feasible feature.

$$InfoGain(a_i,S) = Entropy(y,S) +$$

$$- \sum_{v_{i,j} \in dom(a_i)} \frac{|\sigma_{a_i=v_{i,j}} S|}{|S|} \times Entropy(y, \sigma_{a_i=v_{i,j}} S)$$

The Gain Ratio normalizes the Information Gain and is represented as:

$$GainRatio(a_i,S) = \frac{InfoGain(a_i,S)}{Entropy(a_i,S)}$$

Evaluation: The Confusion Matrix is a tool for determining the qualities of a categorization rule. It shows the amount of items categorised properly or inaccurately for each class. The primary horizontal axis displays the amount of data successfully categorised for every class. The off-diagonal elements indicate the number of occurrences that were misclassified. The Confusion Matrix M(Ci,Cj), for I j = 1,...,k classifications, is an important sign of categorization efficiency because it displays the number of variables properly identified and predicted for every class. Let T represent a set of instances and h represent a hypothesis. The Confusion Matrix of h is generated by:

$$M(C_i, C_j) = \sum_{\forall (x,y) \in T : y = c_i} \| h(x) = C_j \|$$

B. K-Nearest Classification

K-Nearest Neighbors produces estimations or predictions centered on/focusing on the data items that support their resemblance on metrics or units of measurement such as distance between them. K-nearest neighbours may soon become a obvious regression and classification machine learning rule. We do this through, by importing library from sklearn, "from sklearn.neighbours import KNeighborsClassifier". The cost of item is the outcome of K-Nearest Neighbors Regression. It depicts the average of k closest or most nearest neighbors' values. This is a classification technique in which each and every mathematical computations are delayed until the evaluation of function isn't accomplished. To get accurate target is every crucial, in-order to achieve highest possible accuracy, the training data is normalised as it is based on categorization of distance and accuracy will increase if alternatives reflect entirely distinct units or are accessible on broadly variety of scales. The neighbors are chosen after the completion of collecting the objects to which class or article value has been known. This is done due to algorithm's training.

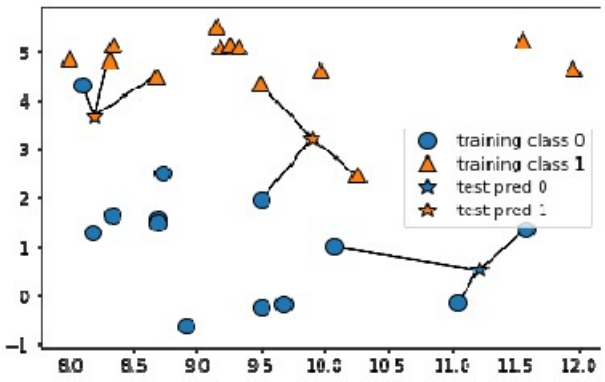

Fig. 2. Example of KNN.

C. SVC

we tried to use Support Vector Machines, after using both extraction procedures. To begin, the original data set (which had 9990 pieces) was divided into two groups of 6993 (training) and 2997 (test) pieces. In order to achieve this goal we used the train_test_split routine. Then we instantiated an SVC object, with all its parameters but probability set to their default values: we made this choice because we will run a Grid Search to choose the most promising ones.
We give the results of an SVM classifier run in this section. The deeper the blue on the diagonal cells in the confusion matrix, the more correct classifications there are.

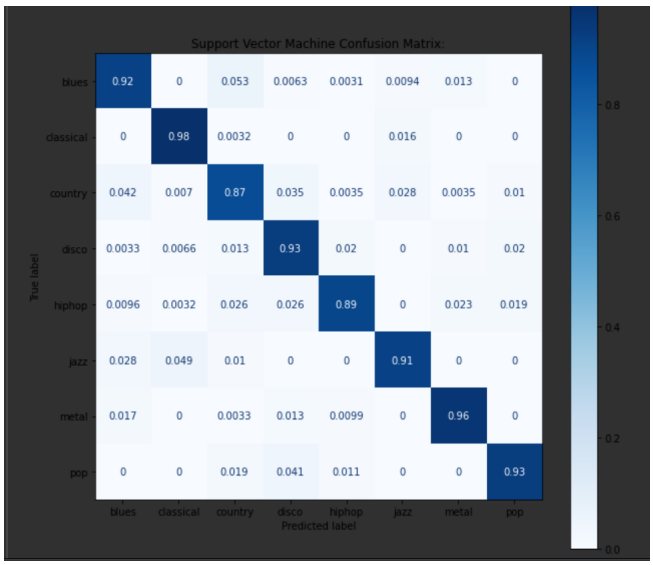

Fig. 3. Support Vector Machine Confusion Matrix

Support Vector Machine Classification Report:

	precision	recall	f1-score	support
blues	0.70	0.92	0.80	319
classical	0.91	0.98	0.95	308
country	0.62	0.87	0.72	286
disco	0.67	0.93	0.78	301
hiphop	0.66	0.89	0.76	311
jazz	0.83	0.91	0.87	286
metal	0.79	0.96	0.87	303
pop	0.75	0.93	0.83	267
reggae	0.00	0.00	0.00	316
rock	0.00	0.00	0.00	300
accuracy			0.73	2997
macro avg	0.59	0.74	0.66	2997
weighted avg	0.59	0.73	0.65	2997

Fig. 4. Support Vector Machine Classification Report

D. Naive Bayes Classification

The Bayes Theorem is the foundation of the supervised machine learning technique known as Naive Bayes Classification, which is used to categorise data. In general, it makes predictions based on the object's or data item's probability. Naive Bayes classification comes in 3 flavours:

- 1. Gaussian Naive Bayes - The Gaussian model presupposes that feature distributions are normal. This indicates that the model thinks that predictor values are samples from the Gaussian distribution if they take continuous values rather than discrete ones.
- 2. Multinomial Naive Bayes - When the dataset is multinomially distributed, the Multinomial Naive Bayes classifier is employed. It is typically used for documen-

tation classification issues, indicating which category a specific document falls under, which including sports, government, academia, etc.

- 3. Bernoulli Naive Bayes - Identical to the Multinomial classifier, the Bernoulli classifier uses independent Boolean variables as predictor variables. such as determining whether a word is used or not by a document.

One of the quick and simple machine learning techniques to predict a dataset class is naive bayes. It can be used for multiclass classifications along with binary classifications.

Steps to take are:

- Pre-processing of data - set
- Training set is fit by Naive Bayes
- Prediction of the result of testing data
- Creating Confusion Matrix of get the result of test data set accuracy
- Visualization of the result of the testing dataset

E. Multilayer Perceptron Classification

Also known as MLP, Multilayer Perceptron is a neural network that has multiple layers which are fully connected dense layers which transform any input dimension into another desired dimension. Neural Network is created by combining neurons which produces some desired output and that output becomes input for other neurons. In Multilayer Perceptron, we have 100 hidden layers and have used Logistic Activation Function and used Stochastic Gradient Descent as it is faster and it is iterated 1000 times.

Fig. 5. Multilayer Perceptron Confusion Matrix

V. Discussion

Here we have tried to apply different classifiers is Machine Learning of Music Genre Classification to classify different genres of music and found the optimum algorithm that is Support Vector Classifier. We compared Accuracies of different classifiers and displayed the Confusion Matrix of some Classifiers.

VI. Conclusion

As per our approach, the accuracy of different classifiers are as follows:

- Decision Tree - 58 percent
- K Nearest Neighbor - 68 percent
- Multilayer Perceptron - 66 percent
- Naive Bayes(Gaussian Distribution) - 52 percent
- Support Vector Classifier - 79 percent

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[2] T. Li and M. Ogihara. Music genre classification with taxonomy. IEEE, 5:197–200, (2005).

[3] J. J. Aucouturier and F. Pachet. Representing musical genre: A state of the art. Jornal of New Music Research, 32(1):83– 93, (2003).

[4] D. Cast´an, A. Ortega, and E. Lleida. Speech/music classification by using the C4.5 decision tree algorithm. FALA 2010 - VI Jornadas en Tecnolog´ıa del Habla and II Iberian SLTech Workshop, pages 197–200, (2010)

[5]] T. Li and M. Ogihara. Music genre classification with taxonomy. IEEE, 5:197–200, (2005).

[6] N. M. Norowi, S. Doraisamy, and R. Wirza. Factors affecting automatic genre classification: An investigation incorporating non-western musical forms. International Conference on Music Information Retrieval (ISMIR), pages 13–20, (2005).

[7] R. Timofeev. Classification and Regression Trees (CART) Theory and applications - Master Thesis. Humboldt University - Berlim, (2004)