Music Genre Classification
Interim- Project Report

priyanka singh	Manish	Aditya Singh
2018174	2018047	2018378
priyanka18174@iiitd.ac.in	manish18047@iiitd.ac.in	aditya18378@iiitd.ac.in

1. Introduction

Music forms are a very crucial part of our lives. Genre classification is an essential task with many real-world applications. Since the quantity by which music is getting released daily reaches the mountain peaks, the need for a proper and accurate genre classification rises in proportion. In today's era, internet services have a massive amount of multimedia exchanges and browsing. So the need for an effective way to organise and categorise these data rises in proportion. Searching a query song in a database containing millions of songs can become a cumbersome and time-consuming task. Categorising songs into different genres can reduce the search space for these query songs. The main objective of our project is to come up with an effective and accurate machine learning model that can automatically classify the music based on the genre.

2. Related Work

Music genre classification is not a new field. There has been work done in this field by many expert data scientists and ML enthusiasts. Some of the works that we looked into are [1] where they performed classification by extracting features like Timbral Texture Features, Rhythmic Content Features and Pitch content features and trained on the models GNN and KNN. In [2], a model is proposed that gives a human-like accuracy of 70 % in this the model used was CNN. In [4] they used SVM with K-L divergence to obtain the accuracy of 84%.With our research, we found out that [3] is the state-of-the-art current models perform with an accuracy of around 91% with the help of sparse representation-based classification. SVM , K-NN and CNN were the most popular techniques used in different papers.

3. Dataset and Evaluation

The dataset GTZAN used for building music genre classification has been taken from Kaggle . It consists of 1000 samples of songs. The dataset has 10 genre classes such as Jazz, Disco, Rock etc. Each of the 10 genre classes have 100 examples of songs. Thus the dataset is balanced. We used 20% of the dataset for testing and the remaining 80% of the data have been used for training and dev set.

3.1. Feature Extraction:

We used the songs examples present in the dataset to extract 63 features using the librosa module available in python. The extracted features includes mean and variance: 1. Mel Frequency Cepstral Coefficients 2. Spectral bandwidth 3. Spectral centroid 4. Zero crossing rate 5. Spectral roll-off 6. Beats location 7. Estimated global tempo 8. Chroma short time fourier transform 9. Root mean square energy for each frame 10. Harmonic component 11. Percussive component

3.2. Evaluation Technique:

Since, this was a classification task so evaluation metric like f1 score, recall, precision and accuracy comes in handy. But as shown above the data set was balanced with each of the 10 music genre classes containing 100 samples each .So evaluation metric -" accuracy " would be a right choice to evaluate our model performance.

4. Analysis And Evaluation

4.1. EXPLORATORY DATA ANALYSIS

In order to preprocess the data before the actual model can be built. We analysed several things which are mentioned below:

4.1.1 Missing value

The dataset has no NaN values so no elimination was required.

4.1.2 Label-Encoding

The target variable is the music genre which is categorical in nature. Hence label encoding was performed to obtain better results.

4.1.3 Standardisation

We analysed that the ranges of features differ by a significant amount. So in order to avoid the algorithm to give more importance to the feature with higher range, the feature set was standardised

4.1.4 Feature selection

Relevant features were selected using the insights drawn from variance and correlation.We checked for the variance of each of the features in the dataset. The feature with low variance in the data represents that it changes very less across the data . A feature with low variance doesn't have much predictive power and doesn't contribute much. So features like harmonic var(variance=1.357712e-04), percussive mean(variance=1.170194e-06), percussive var(variance=4.226615e-05) and harmonic mean(variance=2.835795e-06) were removed. Then we checked for the correlation among the features and correlation of different features with the target. If two features are highly correlated then the feature with highest correlation with the target label was kept in relevant features and the other with lower variance with the target variable was removed.So , we removed the features like 'tempo', 'mfcc19 mean', 'beats', 'spectral bandwidth variance'.So in this way we were able to select the 56 important features.

4.1.5 Outlier detection

We observed that there is a significant difference between mean, 75th percentile and maximum value for several features thus indicating presence of outliers. We removed the outliers using percentile capping method where anything greater than 99th percentile and lesser than 1st percentile has been removed. After removing the outliers and important feature selection the final dataset contains : 988 training examples and 56 feature sets. An example of such case analysis is shown in fig1: As an example the mean of spectral bandwidth variance is 13711.0238662 and 75 percent of the features have values equal to or below 182371.576801 and max value of this feature is is 694784.811549 Which indicates that there may be outliers present so we used a percentile capping approach to remove the outliers.

4.2. Design Choices

4.2.1 Learning method

By drawing conclusion from the Literature review we found that most popular classifiers that have been used so far are KNN,CNN and SVM. So we started with SVM as our baseline and then we looked for other classifiers Which includes KNN, NN, LDA, QDA, Naive Bayes and Logistic regression.

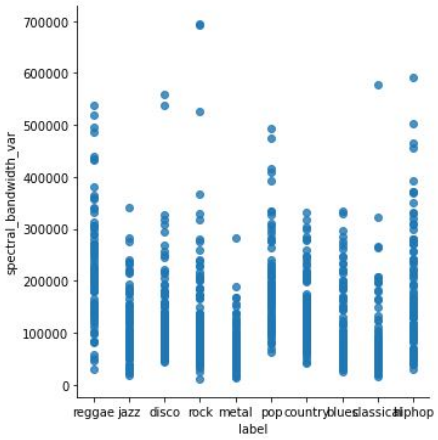

Figure 1. Example showing the presence of outlier

4.2.2 Model selection strategy

As shown above our dataset was balanced. so we used accuracy as a evaluation metric to analyse the performance of different model. We have achieved maximum of 73.4 %accuracy. We will be also trying for ensemble learning method and CNN and then the model giving best accuracy on unseen test data with lower bias and variance will be chosen finally to build this classification system.

4.2.3 Hyperparameter setting

We have used Grid Search method along with k- fold cross validation resampling technique to choose the best parameter from given set of parameters. The chosen parameters gives best performance on the dev set and hence these parameters were used to finally estimate the genre of test data.In future we will be looking for Random search method and other resampling techniques like leave one out cross - validation.

4.2.4 Illustrations And Graph Plot

1.The dataset is not linearly separable as we observed that the train accuracy obtained on SVM model was not 100% on the best hyperparameter evaluated but was 0.73422. This shows that the data is not linearly separable.The same has been illustrated by taking example of two features in figure3
2. We checked the training and testing error of our models and following were the results received- As we can see that there is not a case where training error is too low and test error is too high means our model is not overfitting and also there is not a case where both training and test error are high means our model is not underfitting as well.So it shows that the final tuned hyperparameter is right.

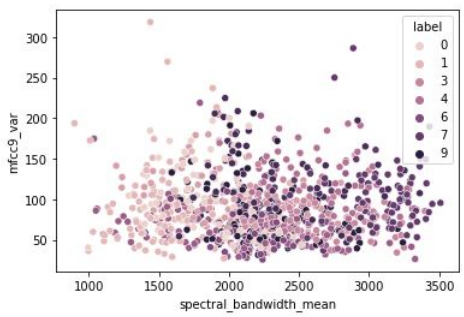

Figure 2. An example to show that dataset is not linearly separable

Model Name	Test Error	Training Error
KNN	0.27	0.20
Linear Kernel SVM	0.28	0.19
Neural Network	0.28	0.17
Poly Kernel SVM	0.26	0.16
Sigmoid kernel SVM	0.29	0.25
RBF Kernel SVM	0.27	0.25
LDA	0.28	0.24
Random Forest	0.27	0.20

Figure 3. Error Analysis

5. Result

The learning model we have explored so far and the corresponding accuracies obtained have been mentioned in Table 1. The highest accuracy obtained is by using Poly Kernel SVM model with accuracy as 73.4%. Now we will shift our focus on models with accuracy 70% or above as human accuracy to classify is 70% [2] We got to know that SVM and neural networks performed well and 8 of our 10 models have accuracy 70% however random forest also performed well with accuracy of 72% it means that all these models similar in performance since difference in accuracy is not much. Gap between between our best model and state of the art model is about 20% where our poly kernel SVM gives about 73% accuracy the state of the art model is at 91%

6. Future Work

Since the results we have obtained through different model has slight different accuracy. So we will be trying to use ensemble learning techniques with majority voting scheme. We will also be using confusion matrix to draw insight which models classifies most of the genre with correct accuracy.If time permits than we will look for convolution neural network techniques to attain more accurate result.Random Search for hyperparameter tuning and K-stratified cross validation technique will be explored.

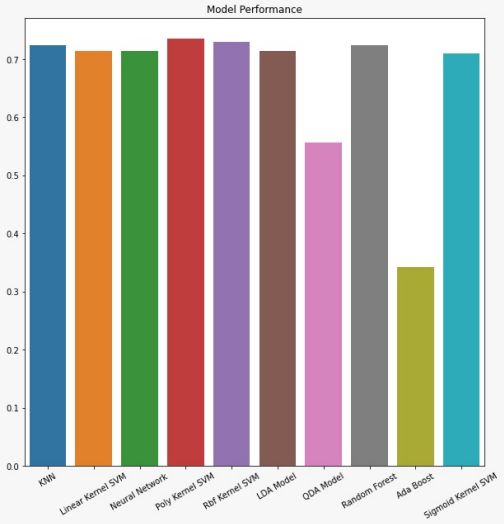

Figure 4. Accuracy obtained using different classifiers

7. Individual Contribution

Priyanka singh: Feature selection, EDA, Outlier Detection, Application of SVM , KNN and Neural Network,Hyperparameter Tuning.
Manish:Hyperparameter tuning,Application of Random forest, LDA,QDA,Ada Boost, SVM and model comparison.
Aditya singh:Feature extraction and application of SVM, KNN and Neural Network

8. Refrences

[1]G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5):293–302, July 2002.G. Tzanetakis and P. Cook. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5):293–302, July 2002.

[2] Mingwen Dong. Convolutional neural network achieves human-level accuracy in music genre classification. CoRR, abs/1802.09697, 2018.

[3] Y. Panagakis, C. Kotropoulos, and G. R. Arce. Music genre classification via sparse representations of auditory temporal modulations. In 2009 17th European Signal Processing Conference, pages 1–5, Aug 2009.

[4] Changsheng Xu, N. C. Maddage, Xi Shao, Fang Cao, and Qi Tian. Musical genre classification using support vector machines. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., volume 5, pages V–429, April 2003.