# Evaluation and Possible Improvements

**Avenues for Further Exploration**

* Tuning hyperparameters for Decision Tree and Random Forest took hours and days. If had access to more computational power, could refine the model even further.
* More feature engineering, PCA etc.
* Remove outliers to get normal distribution Naïve Bayes better, and to improve decision trees
* Try different distance measures for k-NN
* Limitation of having unbalanced classes, sometimes only 1 row in an album in the test set
* Handling of outliers, could be due to Attribute noise: refers to the corruptions in values of one or more attributes due to several causes, such as failures in sensor devices, irregularities in sampling or transcription errors (ref: <https://arxiv.org/pdf/1708.04321.pdf>)
* Combine with NLP to use lyrics for album prediction 🡪 even if songs might not be different in terms of audio attributes, themes explored in the lyrics might be different.
* Use GZTAN data (requires more domain expertise about music) 🡪 choose different features (frequency domain, mel-frequency cepstral coefficients, chroma features, peak centroid, peak smoothness). <https://ijirt.org/master/publishedpaper/IJIRT155461_PAPER.pdf>
* Use a Convolutional Neural Network (<https://ijirt.org/master/publishedpaper/IJIRT155461_PAPER.pdf> )
* Problem with selecting optimal audio features, using Discrete Cosine Transform: <https://ijcert.org/ems/ijcert_papers/V4I206.pdf>
* In favour of polynomial SVM: <https://ijcert.org/ems/ijcert_papers/V4I206.pdf>
* Imbalanced classes: [This](https://towardsdatascience.com/dealing-with-imbalanced-classes-in-machine-learning-d43d6fa19d2#:~:text=A%20simple%20way%20to%20fix,one%20class%20or%20the%20other.) article discusses some of the issues with imbalanced classes, such as diminished recall of the less-represented classes.
* Run K-Means clustering if time.
* Use Naïve-Bayes to predict genre including song lyrics and NLP as features: <https://towardsdatascience.com/i-built-a-naive-bayes-model-to-predict-genre-from-song-lyrics-and-it-went-ok-ish-639af0b0a078> Lyrics with multinomial Naive Bayes: <https://tmthyjames.github.io/2018/february/Predicting-Musical-Genres/>
* Great improvements in music genre classification have been achieved by combining Naïve Bayes with Support Vector Machines <https://projet.liris.cnrs.fr/imagine/pub/proceedings/ICPR-2010/data/4109e589.pdf>
* <https://www.irjet.net/archives/V9/i4/IRJET-V9I428.pdf> --> SVM definition: coordinates plotted in a n-dimensional space, finding hyperplane which best differentiates the classes.
* Naïve Bayes performs better with lyrics 🡪 <https://www.researchgate.net/publication/325645213_Lyrics_Classification_Using_Naive_Bayes>
* Logarithmic transformation for Naïve Bayes to achieve Gaussian distribution: <https://www.analyticsvidhya.com/blog/2021/05/how-to-transform-features-into-normal-gaussian-distribution/>

**k-NN**

An article by Liu and Wu addresses the challenges and limitations of the traditional k-NN classifier in the domain of music genre prediction. The first major obstacle hindering successful music genre classification is the algorithm’s assumption that each feature is equally important, thus ignoring the problem of collinearity or correlation between features. The second limitation is that the similarity between the test samples and ‘neighbors’ across different categories is not considered, only considering the quantity of neighbors in each category. To try to mitigate these challenges, these authors have suggested an improved version of the k-NN algorithm called *Double Weighted k-Nearest Neighbour*, which utilizes a complex notion developed in rough set theory (a branch of set theory dealing with uncertain knowledge) called ‘attribute dependence’, which acts as a metric for quantifying the influence of different features. This assigns weights to *features*, as well as neighbors based on their proximity to the test samples, which helps determine which audio attributes are better predictors of the genre label. Although understanding the concepts of rough set theory and the attribute-dependence calculation would require far more in-depth background reading and learning, it can result in much higher overall accuracy for music classification tasks (Wu & Liu, 2020).

**Summary of Confusion Matrix and How it Works for Evaluation**

“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.”

“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.”

“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.”

* <https://www.irjet.net/archives/V9/i4/IRJET-V9I428.pdf>

**Links about Metrics**

<https://www.edlitera.com/en/blog/posts/evaluating-classification-models#mcetoc_1ga47ipak18>

<https://towardsdatascience.com/6-metrics-to-evaluate-your-classification-algorithms-caddd65ecff5>

<https://www.analyticsvidhya.com/blog/2021/12/evaluation-of-classification-model/>

<https://www.topcoder.com/thrive/articles/metrics-to-evaluate-classification-algorithms>

<https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/#:~:text=Common%20metrics%20include%20accuracy%20(proportion,curve%20(AUC%2DROC>).