Project Final Report for CIS 419/519
Top Hit Songs - Genre and Decade Classifier

Christina Kim CHJK@SEAS.UPENN.EDU
Anmol Kathail ANMOLK@SEAS.UPENN.EDU
Inbar Tivon ITIVON@SEAS.UPENN.EDU

Abstract

For applications to the real world, we are creating a classifier that will identify songs by genre and decade of release. Initial research of music analysis led to 25 features to be used in creating a dataset to train this classifier with. These features include but are not limited to chroma frequency, spectral rolloff and 20 Mel Frequency Cepstral Coefficients (MFCC) features. By comparing several different classifiers, we deduced that using Gradient Boost Classifier was the best model to train our data and get the most accurate predictions. Through feature improvement which included variance threshold, and a chi-squared test, we were able to decrease our 25 features down to 4 significant ones. These 4 important features were spectral bandwidth and three Mel Frequency Cepstral Coefficients (MFCC) features. Using this information, we were able to create a classifier that is approximately seventy five percent accurate. We are know able to provide the classifier any song and it will label it as pop, rock, blues or metal. This success can be further improved and applied to decade classification.

1. Introduction

Music is an ever present art form that has always been changing over the years with the introductions to new styles and forms to alter its numerous features.(Wikipedia, 2019) It has evolved to not only be found in traditional places such as music halls, but applied in various fields outside of the conventional field, such as medical. For example, music has expanded to the medical field in form of therapy, meditation, or other research such as sleep or impact of music on the body. Similarly, business fields related to music such as production companies and movie music directors tend to study different attributes of music to create the next top hit. The impact of music on the body is explored based on different attributes both including and excluding effects of inducing different emotions or feelings. The ability to do this is not limited to one type of music. Today, music is often classified into different genres defined as groups per style.(Webster, 2019) Like all things sorted into groups, this is primarily done for easy identification. However, this classification, particularly for music, can be arbitrary and inconsistent. Due to the variability of music and each unique preferences, music is difficult to accurately classify without biases such as money.

As a result, we propose to use a machine learning classifier to identify the music genre and its associated decade of release based on technical features per song to remove bias. The technical features are found in the spectral, frequency and time domains such that they can also be used to find music preference more accurately which opens more applications for this classifier.

2. Approach

In order to create the music classifier, we have broken the project into 3 main components: dataset, classifier, and analysis.

2.1. Dataset

Like any research being conducted, we need data to train and test on. We created the dataset by importing different music from Blues, Rock, Pop, and Metal genres as AU files. These were then converted to WAV files with a lot of preprocessing. Although we originally stated our genres to be EDM, Classical, and R&B, we changed to the four listed above due to being constrained by availability of usable music data. Most of the music available online is before cleaning. While it is possible for us to do cleaning and labeling ourselves, we do not have anything to compare against to know if the resulting dataset would be accurate. Therefore, the resulting dataset has 100 songs

University of Pennsylvania, CIS 419/519 Course Project.

per genre such that each song is labeled in each genre category. We did not yet add the labels for release decade, but are planning to do this after creating a classifier that sorts music by genre in order to simplify and clarify our approach to the problem. The data was also ultimately normalized for better accuracy.

After importing and manipulating the data into a desired format, we extracted features from the WAV files. Firstly, the frequencies were visualized as shown figure 1 in form of a spectrogram for each song. The spectrogram is a basis on which we can start to define what our classifier will use in terms of common frequency (or notes) in certain patterns as features to accurately predict the genre and decade. We output the data as an audio file (waveform) and extracted specific features for a duration of 30 seconds within the song and saved this data as a CSV file.
We continued from this step to further find more features to end up with a total of 26 features such that the last column is genre labels , and the first was an index for the dataset. These features included the zero-crossing rate, spectral centroid, spectral roll-off, bandwidth, chroma frequency, and 20 different mel frequency cepstral coefficients resulting in a total initial dataset of 400x26 matrix.(McKay et al., 2005) The features listed will measure different aspects of music such as noisiness, balance point of spectrum, prominence, and rate of change in spectrum band (timbre) for classification of music.

One feature was chroma frequency. We computed a chromagram from the music. It takes the entire spectrum of audio and projects it into 12 bins representing different pitch classes. This feature captures harmonic and melodic characteristics of a song. Our second feature was spectral centroid. This indicates where the center of mass of the spectrum is located. Specifically it calculates the average of the frequencies in the song. Different genres tend to strongly differ in frequency range. For example, songs in the metal genre are expected to display a high spectral centroid. The third feature was spectral bandwidth. This computes the frequency bandwidth of the song. This gives the range of frequencies of the song rather than an average peak(spectral centroid). Our fourth feature was rolloff. This is a measure of the amount of the right-skewedness of the power spectrum. That is, it finds the frequency at which at least 85 percent of the song frequencies fall within. Our fifth feature was zero crossing rate. This indicates when the rate of the signal changes between positive and negative. This feature has been used heavily in both speech recognition and music information retrieval. Finally, we used 20 Mel Frequency Cepstral Coefficients (MFCC) features. MFCC features represent phonemes which are distinct units of sound.

To enhance the model, the next step was to improve feature selection. One such improvement was setting a variance threshold. We checked that none of the features were more than 80 percent identical across all training instances which resulted in 2 features being eliminated. This was done to ensure none of the features were not completely insignificant. That is, if a feature was identical across all instances then we would know it is not important in training the classifier. We also ran the features through feature ranking with recursive feature elimination and cross-validated selection of the best number of features (RFECV). This went through all possible combinations of all of the features to determine how many were actually needed to produce the highest accuracy. This told us that out of the 25 features, we just needed 4 optimal features for the same high accuracy. This led us to select the 4 best features through a chi-squared analysis. The chi-square test measures dependence between stochastic variables, so using this function weeds out the features that are the most likely to be independent of class and therefore irrelevant for classification.(**?**)

2.2. Classifier

To choose which classifier yielded the best accuracies for our purposes, we first took the dataset as described above, and randomly split it for different ratios. More simply, these were tested for different fold values for cross validation. These were used for six different known classifiers to test for accuracy levels for both training and testing. The six classifier types included: AdaBoost, Gradient Boost, Logistic Regression, Random Forest, Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM).
As previously mentioned, the data set was properly randomized, and through cross-validation, the data set was used to decide and train a classifier for each fold. The training data was analyzed to find the attributes of each feature per genre and decade. We are still trying to finish up analyzing the different attributes of each feature per label in order to initiate training and testing of the input data using average values.

2.3. Analysis

For analysis, all data is initially sorted and exported into a python interface as a text file. Once the classifier is trained using the average values, we will test the accuracy of running the classifier on a random song of one of the genres input in. Additionally, we will use the results to find the least number of features to maximize accuracy. This process will be repeated for release decade as well as a bank of songs. For further analysis, we downloaded a famous pop song, Ariana Grandas Break up with your girlfriend, and predicted its genre with our trained classifier. To do this we converted the song into our feature matrix and repeated the same feature selection steps described in our approach. This feature matrix was used as input for our classifier prediction. It correctly predicted that the song was pop!

3. Results and Discussion

The first result as discussed in the approach was a successful dataset of different songs with different features per genre. Extracted features were 26 total features such that the first column was data indices and the last was the label. Rather than stopping with just a couple features, we have been continually finding and testing more relevant features that we may want to include as an attribute that is a unique property of each genre. This was mainly due to the fluidity of music such that many different songs can be classified as more than one genre with overlap in certain properties. Contrary to our expectations, we have found that 4 features are optimal for this classifier to work. These features were spectral bandwidth and three Mel Frequency Cepstral Coefficients (MFCC) features.

For testing purposes, we retrieved audio waveforms of 30 seconds to get feature data. Using the methodology described in the approach, we were able to get different results of the feature as seen in figure 1.

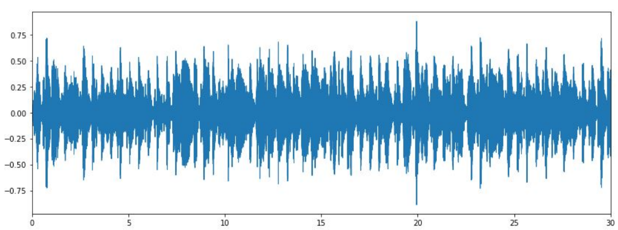


Figure 1. Audio waveform depicting amplitude over time=30s

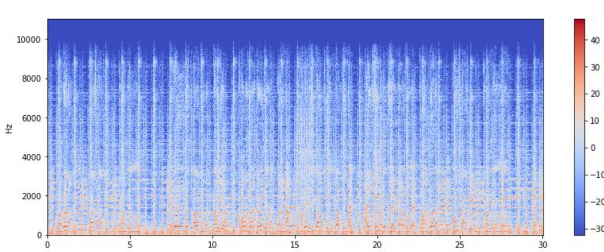


Figure 2. Spectrogram showing concentration of frequencies over time= 30s

Using the dataset, we have been testing the data to decide which classifier would best fit our needs. In order to see how accuracy was affected, and therefore, figure out a way to increase accuracy, we changed the ratios of how much data was left out. From the figures below, we were able to see that gradient boost classifier had the overall best accuracies such that it was 100

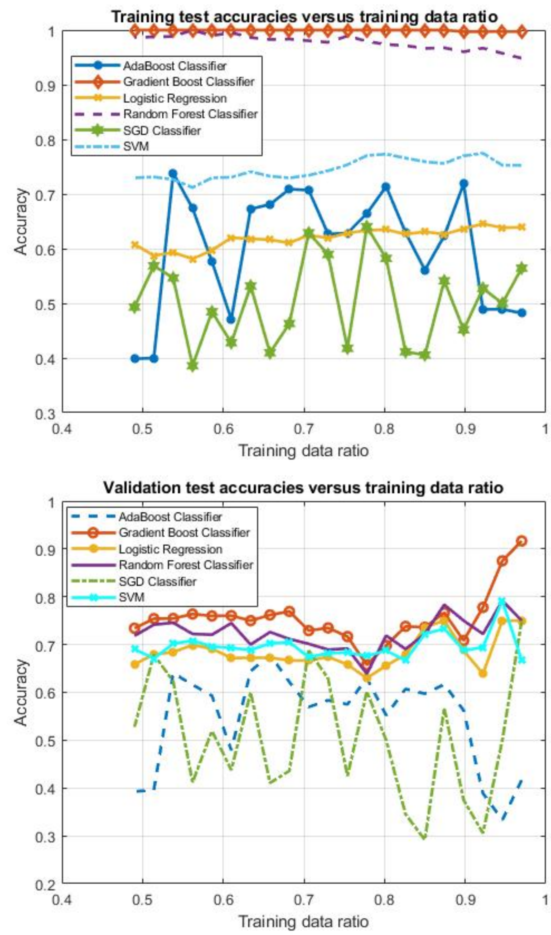


Figure 3. (top)Training accuracy per classifier (bottom)Validation accuracy per classifier

This graph represents varying the estimator parameter of our classifier. The depth was held constant at 3. It shows that 100 estimators was a good parameter to use while 500 would be worse. This graph represents varying the depth parameter of our classifier. The number of estimators was held constant at 100. It shows that a depth of 3 is the best option.

4. Conclusion

We have successfully created a classifier that is able to accurately label a random song by one of four specified genres. As specified in our objective, this was solely done based on technical features rather than features that could

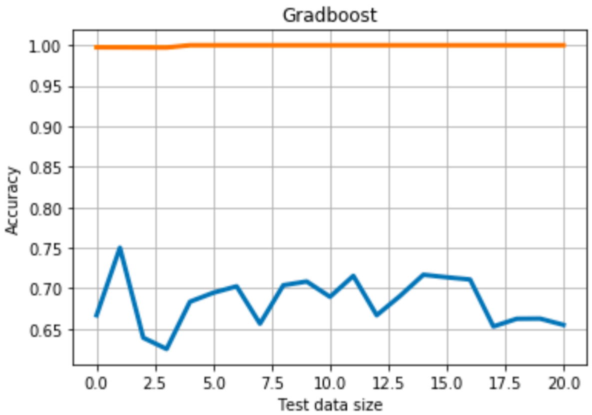


Figure 4. Gradient Boost Classifier using improved feature selection process(Legend : Orange = Accuracy on 100)

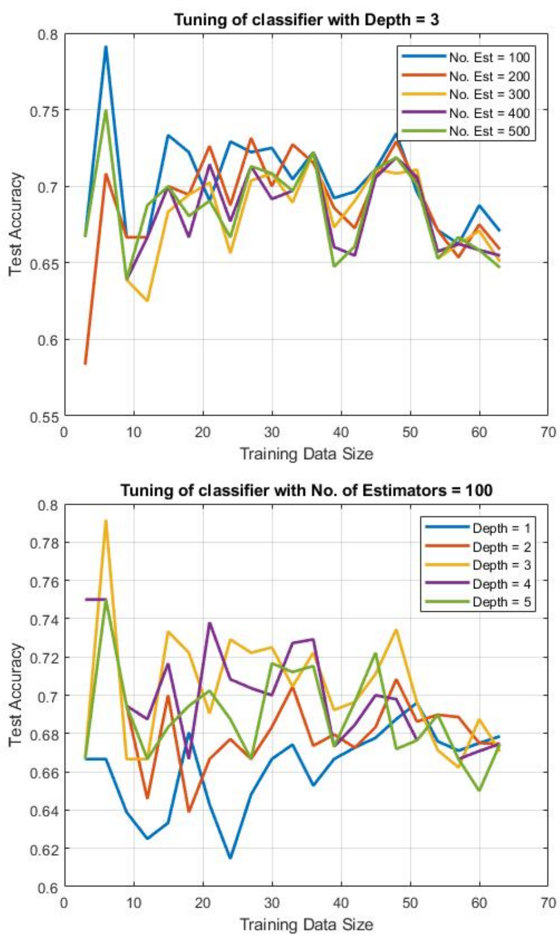


Figure 5. (top)Estimator (bottom) Depth

be externally biased. As a quantifiable method to categorize music, it will be able to used in a wide variety of applications such as predicting new trends, probability of success, or even creation of new music. Similarly, it can be used to identify certain features that are more conducive to relaxation and/or stress in music therapy. However, future uses of this application will be limited in that the implementation itself would be biased due to training on previous/already existing music.

In the duration of the project, it quickly became evident that the most challenging portion of the project was preprocessing the data rather than training and testing the data. There were many limitations that impeded the progress of our project. One of the major impediments was finding available music we could easily access in a desired format. While we had originally strived to create a classifier that would identify genre and decade, we were only able to completely produce results for a classifier that accurately identifies the genre. This was mainly due to time limitations. Next steps for this project would be to complete this portion of the project.

Acknowledgments

We would like to acknowledge all resources cited to find and implement information and data.

References

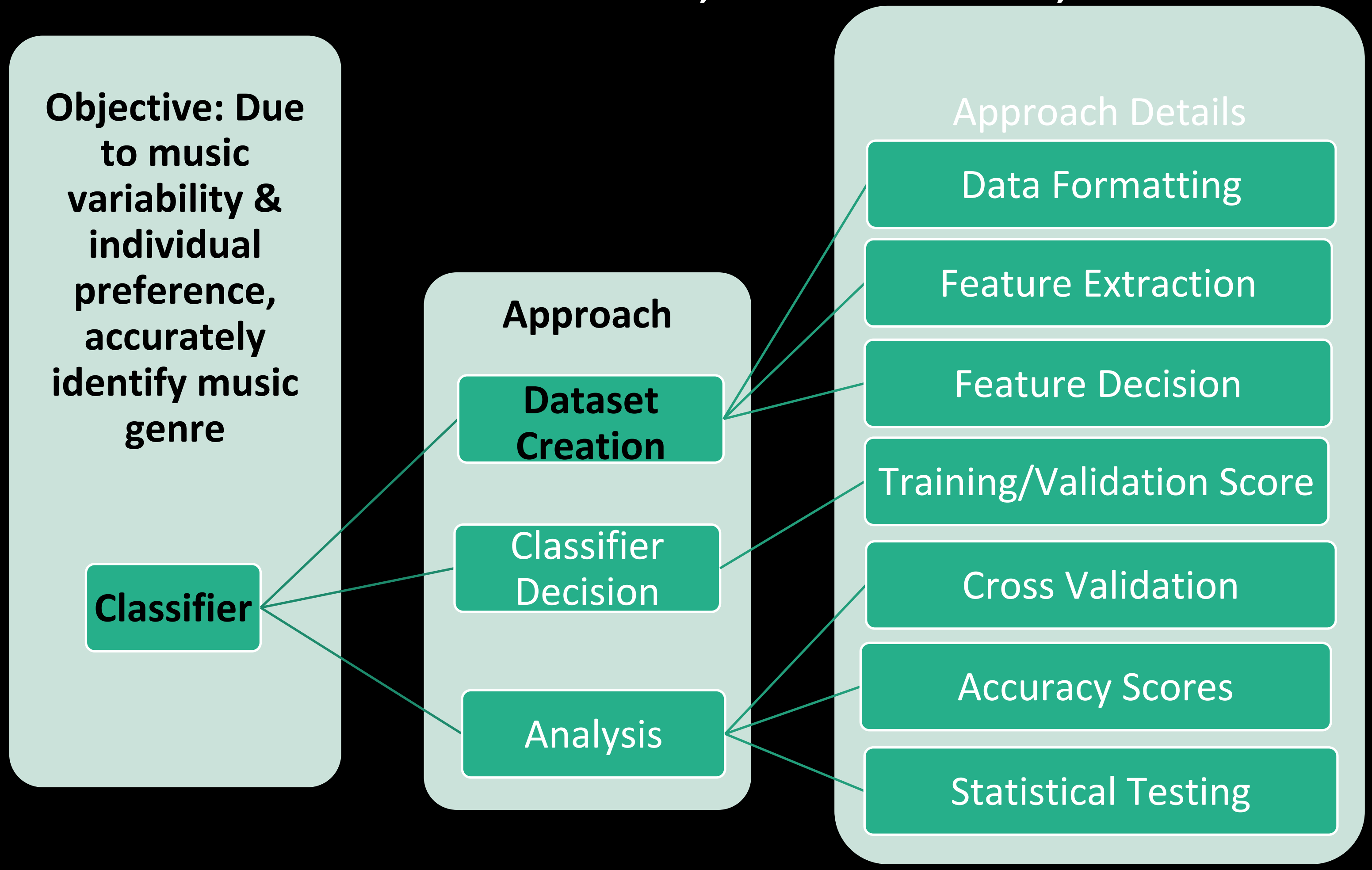
McKay, Cory, Fujinaga, Ichiro, and Depalle, Philippe. jaudio: A feature extraction library. In *Proceedings of the International Conference on Music Information Retrieval*, pp. 600–3, 2005.

Webster, Merriam. Definition of genre, 2019. URL `https://www.merriam-webster.com/dictionary/genre?u` $sdutm_medium = serputm_source = jsonld$.

Wikipedia. Music, 2019. URL `https://en.wikipedia.org/wiki/Music`.

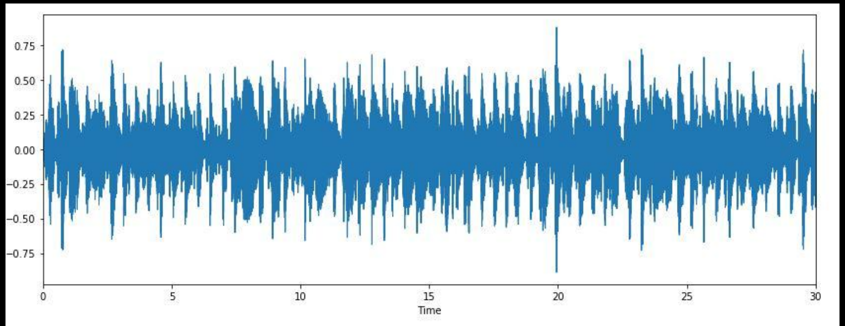
Top Hit Songs: Genre & Decade Classifier

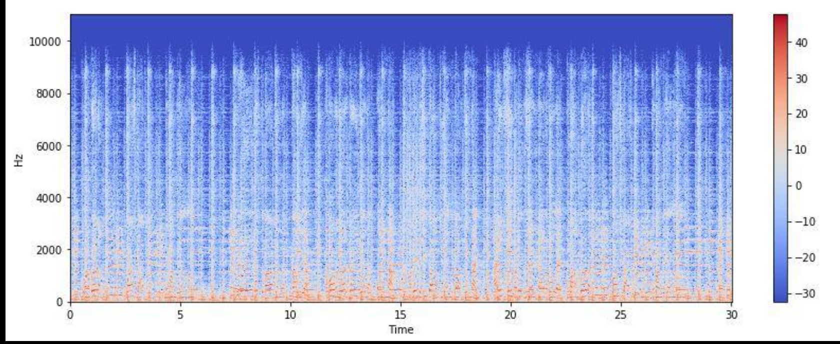
Group Members: Christina Kim, Inbar Tivon, & Anmol Kathail

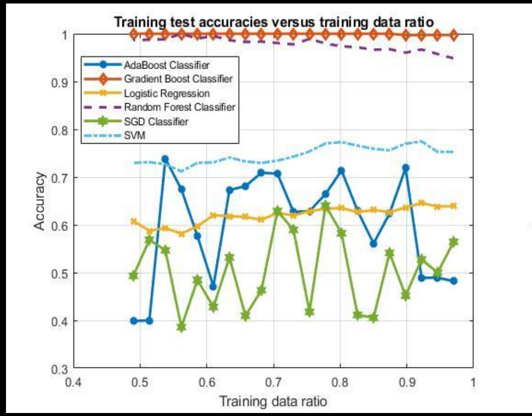


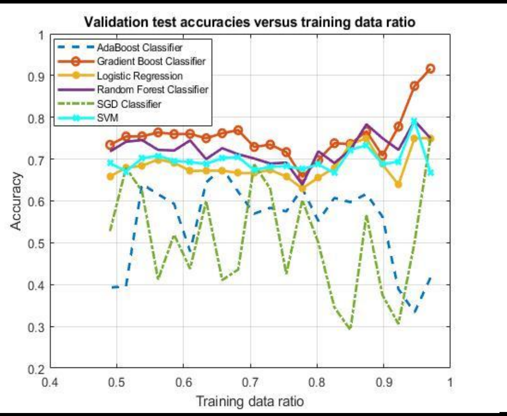
Who cares: Entertainment industry, music apps (i.e. Spotify), doctors, anyone interested in music

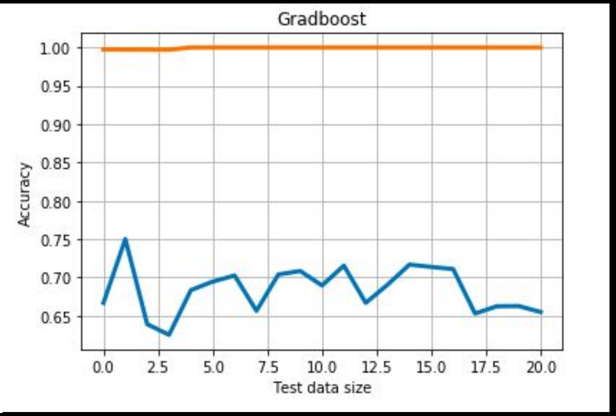
Results



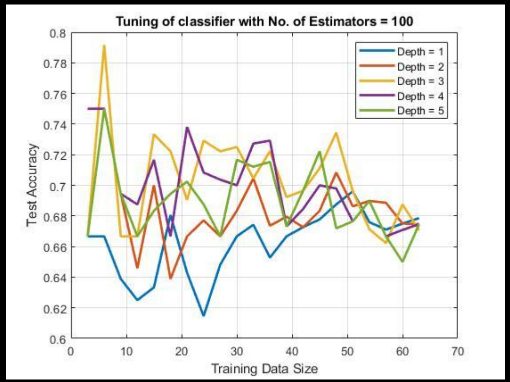


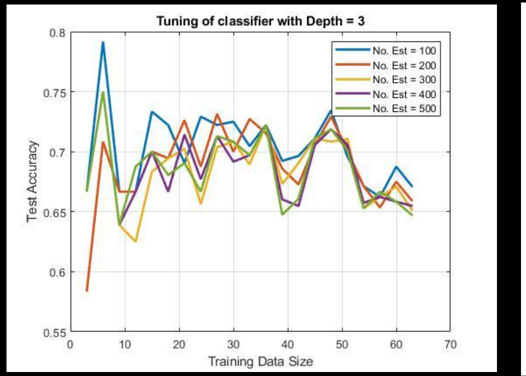


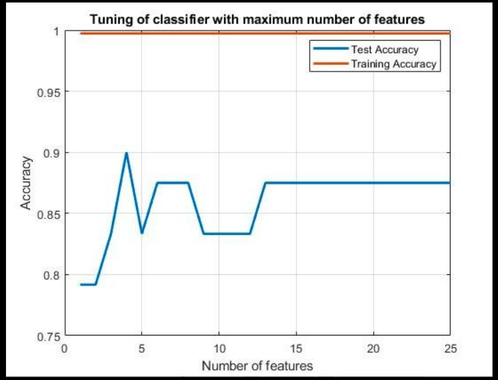




- Training vs. Validation Accuracies per ratio of data show Gradient Boost Classifier best choice
- Gradboost Graph shows classifier using improved feature selection process (Legend : Orange = training data, Blue = Cross Validation on training data)







- 25 total features were found per song
- Each was tested for relevance
- Amplitude of song for duration of t=30s
- Spectrogram to see frequency of notes for duration of t=30s

- varying number of estimators, depth, and number of features to improve classifier accuracy