Music Genre Classification and Hit Prediction

Aman Peshin (1225476655), Bitaan Chakrabarti (1225499119), Keval Shah (1225553979),
Rithik Kumar Goyal (1225572699), Shashank Karanth (1225314194), Vidushi Raturi (1225495622)

December 8, 2022

Abstract

In today's day and age, it has become very important to find what makes a piece of music a hit as well as classify music based on its specific genres. The current user base expects to get the music of their preferred genre(s) properly classified. Most popular music streaming services are expected to have this data correctly labeled and ordered in their databases as and when the user requires it.

For these classified genres, we want to establish the parameters that make a song a 'HIT'-song for each of the classified genres. The list of hit songs shall be based on the number of times a song has been 'listened to' to date that we shall be gathering from the 'features' dataset. Once we get the desirable features for each category, we shall try to predict if any song from our list shall be a 'Hit'-song or not.

Instead of going the usual route of taking tonal characteristics of a song that are produced by huge consortiums like Spotify and Soundcloud, we plan to take the actual musical (spectral features) of each song and use these to do both the genre classification and subsequent 'Hit-song' prediction. The reason for this is that the way the features are generated by these companies is highly confidential and most end-users only get an abstract view of the same. Thus we want to go to the root of the features that are used to generate these features and come up with an open source and open method of doing the same without any kind of black-box approach and also without the fear of keeping the whole power of changing the algorithm with only one entity.

We are using the spectral features extracted from actual music samples that are present in the Free Music Archive (FMA)'s 'metadata' dataset. Using these spectral features like MFCC, Spectral Roll-off, etc, we plan to create a preliminary model to Classify the Genre of music using multiple Statistical Machine Learning methods like Decision Trees, Naïve Bayes and so on. Having done this, we plan on creating a Hit prediction model for each genre and use this to classify whether a new song of that particular genre should be a hit or not solely based on the musical features. We shall use AUC and ROC plots in addition to the basic Model Accuracy scores, for each model to gain insight into the reliability of our models.

In this way, we believe we can remove the extra red tape that comes with marketing and distribution which may be intangibles that cause a song to fail; thus, the creator will have a clear idea as to whether their music is not good enough or if other reasons beyond their control caused it to not be as big of a hit.

1 Introduction

The music industry has grown from one where the best music is sold to one where other intangibles like Marketing, Distribution, Promotions, etc have become the driving factors for a piece of music to become popular, in addition to absolute chance events like random clips with that piece of music going viral in one of the many social media platforms or the music being used in movies and so on.

With this change in the landscape of the music industry, it has become more difficult for creators to come up with a way to understand why their fanbase likes their music and what are kinds of experiments their fans want to see from them. Thus, artists and creators are more at sea about their creative process than ever. They are completely at the mercy of abstract ways of judging music like charts and revenue numbers that are posted by organizations that have been proven time and again to have the singular aim of maximizing profit and ripping-off creators. Most consortiums who own the rights to music created by creators have little to no interest in actually being transparent about the creative experiments that creators often want to do, instead pushing the age-old money-making schemes of heavy beats and a lack of imagination. Thus, not only limiting the creativity of the artists but also making the audience musically 'dumber' and more tonally challenged than ever. In recent studies, it has been a common finding that music from previous decades was head and shoulders above now in terms of complexity and creativity. Somehow all genres nowadays have a tinge of the same top two or three genres because that is what the music labels and top execs believe sells and hence being fed the same thing over and over, the listeners are also resorting to listening to the same type of music just packaged in a different way. It's become less a game of creativity and more of one where the music that is promoted and packaged better sells more often than not.

This has prompted a lazy way of creating music, where even creators have taken to creating the same sort of music repeatedly to remain relevant and financially well-off. We, as music lovers, don't believe this is the correct way of dealing with music and feel that creators need to be in a position to take calls about their music based on tangible data. Hence in this day and age where data is king and we have multiple ways to gain insights about music and visualize the features that make a hit song for any particular genre, we have decided to make Data the 'Music Guru' who decides what should be a great music piece simply based on the spectral features of music produced and not be driven by sales numbers and some obsolete formula that has given rise to such drab music choices of late.

2 Motivation

Music is a moral law. It gives soul to the universe, wings to the mind, flight to the imagination, and charm and gaiety to life and everything" - Plato. Music is one of the most powerful forces which lets us communicate, feel and heal.

Humans have been using music for time immemorial to communicate, feel good and even create a sense of belonging. With time, music has evolved as a means to express our desires, our beliefs, our likes and dislikes, and most importantly, 'make a point. But what's alarming is that while the music itself has grown into a movement of expressing things, the creators and the listeners of music have been taken for granted repeatedly and have lost their power in the modern day. So much so that it has come to a point where most musicians don't have a say in what they are supposed to be producing and most audiences have music that is shoved down their throats based on what certain people feel is best for the current scenario.

While it is not a new thing for producers and record labels to dictate which kind of music actually 'sells', it has become a huge monopoly of late. And the so-called 'Music Gurus' are taking away the power of creation from the actual creators and simultaneously making the people listen to the same type of beats and tempos that they believe will be commercially successful.

What these people fail to realize is that music has always been about feelings and less about a sure-shot formula to make millions. Subtle nuances in tone, pitch and other features in a song are what differentiate how one perceives it. Thus, different types of music make us feel differently, hence there are multiple genres and hence it is important to categorize music into different genres before we even start to tackle the problem of creating a hit song.

The music industry in our opinion, is a highly closed environment where new and upcoming and independent artists find it difficult to get their music noticed and it takes them years to come up with their first 'hit' song. With this project, we are trying to democratize the industry and give power back to the people who create the music. Hence the onus to create good music that the audience is interested in goes back to the creators and the industry moves away from the lazy method of creating the same kinds of music with a lack of tonal differences and a total lack of experimentation.

3 Problem Description

This research aims to map the bridge of how spectral features are used by online platforms to predict upcoming hit songs across each genre. We have attempted to make this process simplified and accessible to artists. In this paper, we have developed and compared approaches that help us extract spectral features from song snippets, select features that best represent the task at hand and predict which genre the song falls under and whether it would be a hit in that particular genre. We have also come up with a simplified solution that would use STFT and ISTFT on selected features to make the song more appealing to users. Our approach is to potentially make an end-to-end solution that could be used as a tool in the industry.

4 Methodology

The dataset used in the experiment is the Free Music Archive (FMA) dataset. It is a collection of 106,574 tracks encoded as mp3s from multiple artists and genres. The dataset has audio files as well as metadata related to the tracks such as artist name, date released, listens, publisher, etc.

Feature extraction is a critical part of both genre classification and hits classification. The process of feature extraction entails the evaluation and extraction of significant information from audio in order to produce a succinct, machine-readable description. Features are typically chosen in the context of a particular task and domain.

For our experiments, we used the librosa library to extract features from audio files. A total of nine audio features were extracted and summarized with seven statistics - mean, standard deviation, skew, kurtosis, median, minimum and maximum. The features are - Chroma, Tonnetz, Mel Frequency Cepstral Coefficient (MFCC), Spectral centroid, Spectral bandwidth, Spectral contrast, Spectral roll-off, Root Mean Square energy, and Zero-crossing rate.

A brief description of each -

- Chroma - It is a powerful tool for analyzing music features whose pitches can be meaningfully categorized. They capture harmonic and melodic characteristics of music while being robust to changes in timbre and instrumentation.
- Tonnetz - It is the tonal centroid features or the central tones. These are features that help in detecting harmonic changes in music or variances due to tones in audio.
- MFCC - It is based on a logarithmic scale and is able to estimate human auditory response in a better way than the other cepstral feature extraction techniques.
- Spectral centroid - The spectral centroid indicates at which frequency the energy of a spectrum is centered. This is like a weighted mean.
- Spectral bandwidth - It is the measure of the extent of the spectrum.
- Spectral contrast - Spectral contrast considers the spectral peak, the spectral valley, and their difference in each frequency subband.
- Spectral roll-off - Spectral Rolloff is the frequency below which a specified percentage of the total spectral energy.
- Root Mean Square - The root mean square in this context refers to the signal's overall magnitude, which is equivalent to the audio file's loudness or energy parameter.
- Zero crossing rate - Zero-crossing rate is a measure of the number of times in a given time interval/frame that the amplitude of the speech signals passes through a value of zero. It is a key feature in classifying percussive sounds.

There are two parts to our project. Given an audio file, we first extract features from it. Based on these features, the genre of the music is found. Subsequently, based on the predicted genre, we also classify the song as a hit or not. We train multiple hit classification models using data from each genre.

For genre classification, the feature used is MFCC. Most of the existing work with genre classification uses this feature only and that is the reason behind this. Once the features are extracted, we explore the data based on the genre field in the metadata. We found that the data is highly imbalanced. A

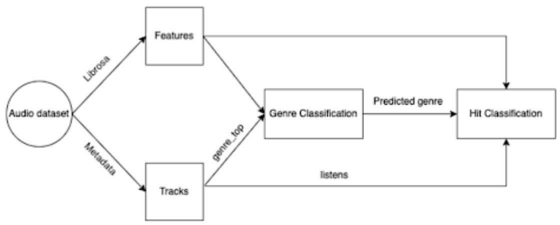

Figure 1: Flow Chart

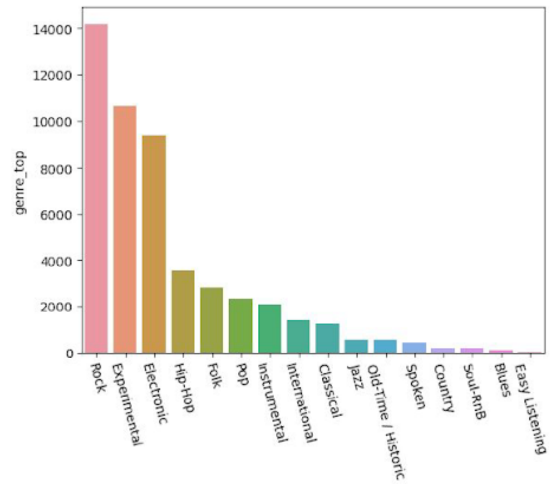

Figure 2: Number of tracks in each Genre

large number of tracks, around 14,000 out of the 100,000, belonged to the 'Rock' genre. Almost half of the samples belonged to the top 4 genres. To avoid the issues prone to models trained on imbalanced data, we take a balanced subset of 2000 songs per genre from the top 7 genres. MFCC features consist of 140 columns in total, which include the various statistics such as mean, standard deviation, skew, kurtosis, median, minimum and maximum. Since the values are in a wide range, we normalize the data using the MinMaxScaler. The genre target label is obtained from the metadata provided in the dataset.

Multiple models were trained and tested on the above data. Keeping the scope within the topics taught in class, we have tested Support Vector Machine, Gaussian Naive Bayes, K Nearest Neighbours, and Random Forest. To get better results, we have also tried using some ensemble classifiers such as AdaBoost and Extreme Gradient Boosting (XGB). Accuracy and confusion matrices were used to evaluate the performance of each model. From the experimental results, XGB performed the best with an accuracy of 60The predicted genre is used to select the hit classification model. Since a hit song's characteristics vary depending on its genre, a separate model is trained exclusively on songs from each genre. Due to the nuances of what makes a song a hit, additional features such as Chroma, Tonnetz, Spectral centroid, Spectral bandwidth, Spectral contrast, Spectral roll-off, Root Mean Square energy, and Zero-crossing rate, along with MFCC, are used in order to capture more features of the track. Hit prediction is a binary classification task. Since it is supervised learning, we use the number of listens available in the metadata to determine each track's target label. For our task, we classify a song as a hit if the number of listens is more than the 60th percentile for that genre.

The algorithms used for hit prediction were Naive Bayes, K Nearest Neighbours, Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression. From the results, we found that Random Forest showed the best results, with an average of around 65% accuracy for all the genres. Since the labels are of a 60-40 split, we have also used a confusion matrix and ROC AUC scores to evaluate the models.

5 Results

For Genre Classification, we used 6 different machine learning models namely, Support Vector Machine, Gaussian Naive Bayes, K Nearest Neighbours, Random Forest and XGB. We trained these models and compared their accuracies. Out of all the models, XGB gave the best accuracy, close to 60%. Below is the visual representation depicting the performance of each of the aforementioned models. We implemented hyperparameter tuning to improve the accuracy of the models for both of the use cases.

For Hit Prediction, we used Naive Bayes, K Nearest Neighbours, Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression.

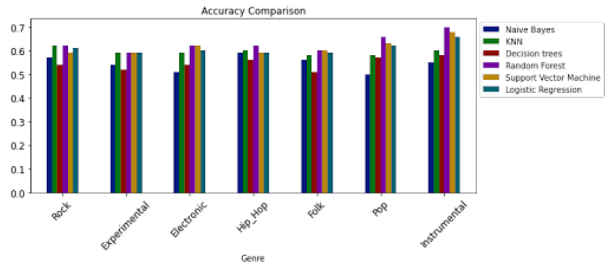

Figure 3: Hit Prediction - Accuracy for each algorithm per Genre

These models were hyperparameter tuned like in the case of SVM where we used One over Rest, and used a high value of K to find the maximum number of neighbors.
Using the above techniques, we found that we get a high classification accuracy of 70% using the Random Forest classifier albeit it overfits the data a little.
Due to the inherent problem with two classes classification problem where data is highly skewed to one class, we have used an AUC-ROC plot to determine the true number of positive and negative classifications and created a confusion matrix used to visualize important predictive analytics like recall, specificity, accuracy, and precision

Unfortunately, the above-mentioned approaches did not make any significant improvement. We deduced that the reasons for the same could be the following: Firstly, the variance of values of features for different classes is very low. The values after Min-Max scaling our features were so close to each other that essentially it's a coin toss as to which class it's going to end up in. This led to overlapping clusters making it difficult for the models to classify the songs and predict hits. We have used dimensionality reduction to get to the root cause and see how apart are the clusters of each class, we have used two techniques principal component analysis and T-SNE plot. The given plot shows the feature space after using PCA and reducing the dimensionality to two, we can see the heavy overlap and a wide scattering of data points from different classes. We have used T-SNE to proportionally decrease pairwise distances and separate the clusters from having smaller overlapping. We have used a perplexity of 100 to get this result. Even with

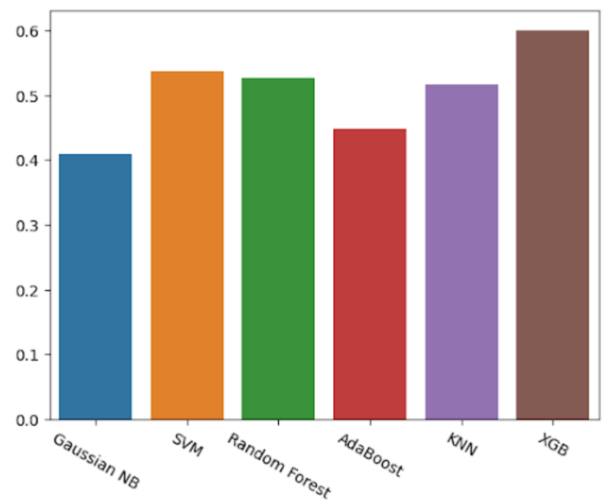

Figure 4: Genre Classification - Accuracy for each algorithm

these approaches, we see a heavy overlap of data points which essentially shows us that the given features are not good measures to predict genres or hits. Lastly, it could also be inferred from the performance of these statistical models that using a deep neural network might lead to better results and could save us the task of selecting different features for the two different tasks at hand.

6 Related Work

There have been various research done where spectral features are used to predict the genre of a given song. We have referred to the research paper "Automatic Music Genre Classification and Its Relation with Music Education" written by Hasan Can Ceylan et al. have discussed how spectral features and Neural networks can be used to classify songs. Although their research used only MFCC features from the GTZAN dataset, convolutional neural networks and very small segments of songs. We have tried to improve on this by selecting longer segments, using a more extensive dataset called free music archive and trying to use simpler models to replicate the same problem statement. We have gone a step forward by using the same dataset to even predict whether the song is going to be hit and how a song can be synthesized

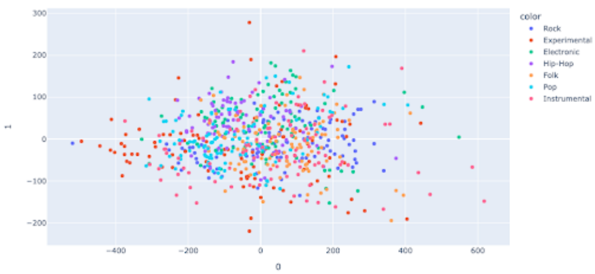

Figure 5: PCA for dimensionality reduction

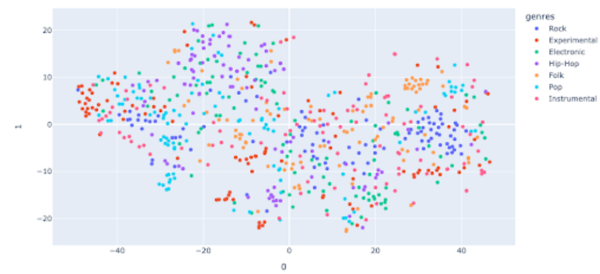

Figure 6: t-SNE for dimensionality reduction

to achieve this.

7 Conclusion

Classifying music and recommending new music in music listening apps and platforms is an essential and current concern. The work done in this field is still relatively new and has the potential to improve. Following the tasks that were performed during our research on this project, we feel that in order to mimic leading platforms like Spotify, we would need to use deep learning models which would learn how they combine multiple features and come up with features like tempo, danceability etc. We could come up with this mapping for our base statistical models but we can be sure even with knowing such mappings and tuning our models it would not improve the accuracy to such an extent that deep learning models could be benched.
We have worked on the stretch goal of using the Griffin-Lim algorithm to synthesize audio features and regenerate the audio snippet using ISTFT and STFT on magnitude spectrograms. This algorithm primarily depends on phase reconstruction and is independent of the prior audio.

References

[1] Song Hit Prediction: Predicting Billboard Hits Using Spotify Data

[2] HITPREDICT: Predicting HIT Songs using Spotify Data

[3] Genre Classification of Spotify Songs

[4] Music Genre Classification using Spectral feature

[5] Multi-Label Music Genre Classification from Audio

[6] FMA: A Dataset For Music Analysis