**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 and most popular music streaming services are expected to have this data correctly labelled 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’ till 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 are highly confidential and most end-users only get an abstracted view of the same, thus we want to go to 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 about 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 which cause a song to fail; thus, the creator will have a clear idea as to whether their music is not good enough or if there were other reasons beyond their control that caused it to not be as big of a hit.

**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 to 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 one’s desires, our beliefs, our likes and dislikes, and most importantly ‘make a point’. But what’s alarming is that while 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 really have a say into what they are supposed to be producing and most audiences have music that is shoved down their throat 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 is what differentiates 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 really 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 actually create the music. Hence the onus to create good music that the audience is actually interested in, goes back to the creators and the industry moves away from the lazy method of creating the same kinds of music with lack of tonal differences and a total lack of experimentation.

**Introduction**

The music is industry has grown from one where the best music 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 actually likes their music and what are the kinds of experiments that their fans actually want to see from them. Thus, artists and creators are more at sea about their own 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 who 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 ‘dummer’ and more tonally challenged than ever. In recent studies it has been a common finding that music from previous decades were 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 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 so as 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 own music based of 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.