CSE343 Machine Learning - Project Proposal

Manav Mittal	Utkarsh Venaik	Akash Kushwaha	Lakshay Kumar
2021538	2021570	2021514	2021061
B.Tech CSAI	B.Tech CSAI	B.Tech CSAI	B.Tech CSE

1. Motivation

The project's primary motivation is to find whether there exists a relationship between music popularity and data science. It questions whether specific songs' popularity has some underlying data-driven factors or is just coincidental. The main objective of this project is to leverage attributes of a song to predict its popularity and the degree of acceptance by listeners. An analysis of this kind could greatly benefit music creators, composers, singers, and instrumentalists. It offers them insights for adapting their creative output based on anticipated audience preferences.

We went through multiple journals, research articles, and websites for project ideas and came across the fact that it is possible to predict the popularity of music using its characteristics of it. We all find this topic captivating and innovative, with substantial room for exploration, so we decided to proceed with it only.

2. Related Work

1. In this paper, they presented a methodology for predicting if a song will appear on Spotify's Top 50 Globalranking after a certain amount of time. [1]

2. In this paper they evaluated different classification and regression algorithms on their ability to predict popularity and determined the types of features that hold the most predictive power. [2]

3. This project tries to answer three key questions: Are there certain characteristics for hit songs, what are the largest influencers on a song's success, and can old songs even predict the popularity of new songs? [3]

3. Timeline

1. Dataset formation using the Spotify API [1 week]
2. Pre-processing of the dataset and feature engineering [1 week]
3. Exploratory data analysis, visualisation and planning [1 week]
4. Model selection [2 weeks]
5. Training of machine learning models [2 weeks]
6. Model evaluation [2 week]
7. Documenting and presenting the entire project and all steps carried out [1 week]

4. Individual Task

1. Extraction of the dataset using the Spotify API - **Akash**
2. Preprocessing Dataset and feature engineering - **Utkarsh, Lakshay**
3. Exploratory data analysis, visualisation and planning - **Manav, Utkarsh**
4. Researching and selecting appropriate machine learning models and setting up a baseline model - **Manav, Lakshay**
5. Hyper-parameter tuning, Selecting the best model based on validation results - **Manav, Akash**
6. Model Evaluation - **Everyone**
7. Documenting the entire project - **Everyone**

5. Final Outcome

We will gain insights into how characteristics of the sound waves corresponding to a song decide the track's popularity. We aim to create multiple visualisations and run multiple trials and tests with different models and hyperparameters to predict a song's popularity with a high accuracy finally.

We aim to develop a machine learning model that, given a song, can classify it as whether or not a song will be famous. A lot of current studies aim at binarily classifying whether or not a particular song would be popular. We would aim to iterate over it by also classifying the degree of popularity that the song would achieve using multi-classification or regression (will be decided in due course after the model selection and data analysis phase).

This machine learning model can be of great use to music composers and artists to know whether or not their songs will be popular and can also guide them on how they can improve their songs to make them more popular among the audience.