CSE343 : Machine Learning Project Proposal
OK Computer : Employing Machine Learning to analyse the impact of different parameters in foretelling the next earworm

Ansh Arora	Jishnu Raj Parashar	Nandika Jain	Tushar Mohan
2019022	2019048	2019064	2019393

1 Motivation

Music, like other forms of art, is something that has historically been subjective to the listener's ears. However, in the age of TikTok and Reels, it is claimed that the songs regularly reaching the top of the charts are getting repetitive, almost following a pattern that guarantees its success. Some artists have even gone on to make rather braggadocious claims that they know well before the release of their song whether it's going to be a hit or not. Are we, as consumers of these tracks, so naive? Is there in fact a way to find out whether a song would be hit or not purely based on its characteristics?

In this project, we wish to answer these burning questions - What is it, after all, that seems to make a song more popular than others, and can we accurately predict the popularity of a track. Modern machine learning techniques might help us identify the features in a song that make it a hit, as well as accurately tell if a song, based solely on the selected characteristics, would be popular or not.

2 Related Work

Araujo et al in *Predicting Music Popularity on Streaming Platforms* make use of SVM, Random Forest and Gaussian Naïve-Bayes models to classify whether a song is successful or not based on its appearance on the Spotify Top 50 Global charts [1]. Ni et al in *Hit Song Science Once Again a Science* make use of features such as duration, loudness and tempo to binarily classify a song as hit or not hit, employing a shifting perceptron [2]. Raza et al in *Predicting a hit song with ML* make use of Logistic Regression, Decision Trees, Naïve Bayes and Random Forest methods to find out if there is a secret formula to predict which song will be on the top of charts on release [3].

3 Final Outcome

The objective of this project is to find what are the features that may drive the popularity of a song and make it more likeable among the masses. We aim to find features of a track that correlate the most with its popularity and create a gradient scale to judge the same based on various performance measures like chart performances, sales and YouTube video views. Other than basic characteristics of sound like tempo and loudness, we also plan on using modelling techniques like Text vectorization, Topic modelling, etc. for song lyrics and compare all these features using various Machine Learning models under multiple metrics.

Our main goal is to find which features affect the popularity of the song the most and create a model that is able to accurately predict how popular a song can get based on these characteristics, thus eventually helping budding artists find sounds that appease to the wider public.

4 Timeline

Week	Task
Week 1-2	Data Collection
Week 3	Data Preprocessing and visualization
Week 4	Feature Extraction
Week 5	Feature Analysis, Selection, Correlation
Week 6	Linear regression, SVMs
Week 7	Logistic Regression, K-Means, NB
Week 8	RF, Decision Trees, NNs
Week 9	Analysis of the outcomes
Week 10	Hyperparameter tuning
Week 11-12	Writing Report

5 Individual Tasks

Tasks	Team Member(s)
Data Collection	Tushar and Nandika
Data Preprocessing and Visualization	Jishnu and Ansh
Feature Extraction	Ansh and Tushar
Feature Analysis	Jishnu and Nandika
Linear Regression, SVMs	Ansh and Nandika
Logistic Regression, K-Means, NB	Tushar, Jishnu and Ansh
RF, Decision Trees, NNs	All
Analysis of the outcomes	Ansh and Tushar
Hyperparameter tuning	Jishnu and Nandika
Writing Report	All