Predicting Spotify Charting Success

Group Members

Alexis Bloor, Jasmine Jones, Hannah Hill

Overview

This project aims to develop machine learning models which are able to predict two outcomes related to the popularity of a song on the platform Spotify: 1) whether the song in question will chart in the Spotify Top 200 rankings, and 2) if so, for how many weeks will it remain in the Top 200 chart. These models will use features of songs collected from the Spotify API including tempo, loudness, key, instrumentalness, and speechiness to predict successful outcomes. These models could be used by music artists and producers to gauge how likely their newest records are to achieve Top 200 status.

Data

Randomly Selected Database of Spotify songs (10,000 per 26 genres)
https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db?select=SpotifyFeatures.csv
Dimensions: (232725, 18)

Top 200 Spotify Chart 2020-2021
https://www.kaggle.com/sashankpillai/spotify-top-200-charts-20202021?select=spotify_dataset.csv
Dimensions: (1556, 23)

Deliverables

1) Classifier that uses musical features of a song to predict whether or not it will reach the modern Top 200 chart on Spotify.
    a) Strategies:
        i) Logistic Regression
        ii) K-Nearest Neighbors clustering
        iii) Random Forest Decision Trees
        iv) Neural network
2) Regression model that uses musical features of songs predicted to chart to anticipate how long the song will be in the Top 200.
    a) Strategy:
        i) Linear Regression
3) HTML/Javascript Flask-based application that allows users to input music features and outputs the predictions of the models (Will Chart? Yes/No, How long? # of weeks)

Goal

The goal is to optimize the models to achieve a minimum of 75% accuracy.