Harry Neal

Spotify MoodGrid

11th April 2023

OVERVIEW

Spotify Moodgrid Playlist Subsetter is a tool to combine your multiple playlists, give them a happy score and an energy score, and then output a new personal playlist centred around your chosen Happy/Energy mood.

Spotify playlists have become a staple in the lives of music lovers worldwide. They provide a personalised and curated listening experience that can reflect our unique tastes, preferences, and moods. A third of Spotify listening time happens on Spotify curated playlists, with another third happening on user-generated playlists.

People love making playlists because it allows them to express themselves, showcase their favorite songs and artists, and share their musical tastes with others. Similarly, listening to playlists offers a convenient and effortless way to discover new music and enjoy familiar favorites. Playlisting can be a quick way of changing or enhancing a mood with music, but currently, there is no way in Spotify of outputting one or more playlists to a more specific subset based on mood. That's where the Spotify MoodGrid Playlist Subsetter comes in. By using multiple playlists and analyzing their happy/energy scores, the app allows users to create a new playlist centered on their desired mood, providing a more tailored and fulfilling listening experience. It's a tool for anyone who wants to enjoy their favorite songs in a way that matches their current mood or to discover new music that will enhance their listening experience.

Data Download, Cleaning & Exploratory Data Analysis

A dataset of 160,000 tracks was downloaded using Spotify's Web API. Playlists were searched using queries that fit under one of four moods: Happy, Sad, Energetic, Chilled. 53 different queries were used to gain a broad spread of feelings across each mood. Some examples include:

- Happy: blissful, contented, ecstasy, euphoric, happy, positive
- Sad: angry, crying, depress, grief, sad
- Energetic: adrenaline, energetic, heavy, high octane, pumped
- Chilled: calm, chill, easy, mellow, relax

1881 playlists were downloaded, with 20 features for each track, including artist and track details, genre, and audio features.

Before any processes were applied to the data an 80/20% train/test split was carried out

Data cleaning involved removing irrelevant playlists to their given mood, removing/filling NaNs, removing duplicates for the same mood, and removing duplicates for an opposite mood. I.e. duplicates in Happy & Sad were dropped, but duplicates in Happy & Energetic were retained.

Feature engineering involved grouping niche Spotify genres to more general groups - reducing down from 3620 to 9. Time signature was converted from numerical beats per measure into a binary 'common time' feature. Track duration was Winsorized to a maximum of 12 minutes in order to reduce the influence of > 1 hour outliers. Categorical columns were converted into dummies and the 'explicit' column was binarised.

EDA highlighted the features that were likely to be good predictors of Happy vs Sad songs, or Energetic vs Chilled songs.

The same pre-processing was applied to the train and test sets, after which the total cleaned dataset consisted of 134,042 tracks.

Modelling

The dataset was split into two parts: a Happy/Sad (HS) dataset and an Energetic/Chilled (EC) dataset.

These datasets were well balanced, with a 50.5/49.5% Happy/Sad split, and a 50.2/49.8% Energetic/Chilled split.

A logistic regression was refined separately for each dataset, via grid search to optimise hyperparameters. A standard scaler and SelectKBest was used. In order to obtain both a happy score and an energy score for every track, the models were required to be applied to the opposite dataset. I.e. The EC model was fitted on the EC data and scored on both the EC and HS data. Conversely the HS model was fitted on the HS data and scored on both the EC and HS data. To achieve this, the feature sets of the two models were made equivalent. A total of 26 features were included in the final model, and the following optimised model accuracy scores were attained:

- Energetic/Chilled: 82.1% train accuracy, 80.1% test accuracy
- Happy/Sad: 68.9% train accuracy, 67.6% test accuracy

The most influential features for predicting an Energetic song were 'energy', *'acousticness'*, *'popularity'*, 'genre_metal', and 'loudness'. (*italics* indicates a negative coefficient).

The most influential features for predicting a Sad song were 'metal', 'loudness', *'danceability'*, *'valence'*, and *'genre_electronic'.* (*italics* indicates a negative coefficient).

MoodGrid

The models were fitted on the full dataset, and happy/energy scores were generated for every track. This gave every track a MoodGrid coordinate. The fitted model could then be used to score any new song read in with the Spotify API. An interactive Streamlit application was created to allow a user to input multiple Spotify URLs, display the songs on their own MoodGrid, and select a number of songs to output to a new playlist, centred around their chosen Happy/Energy score.