

Predicting Song Popularity using Spotify's Music Dataset

Denisha Saviela

OVERVIEW

The Spotify music dataset contains numeric metrics generated by Spotify which measure the songs' danceability, mood, liveness, etc. The goal of this project was to predict the hit songs in future years from multiple artists based on the given dataset.

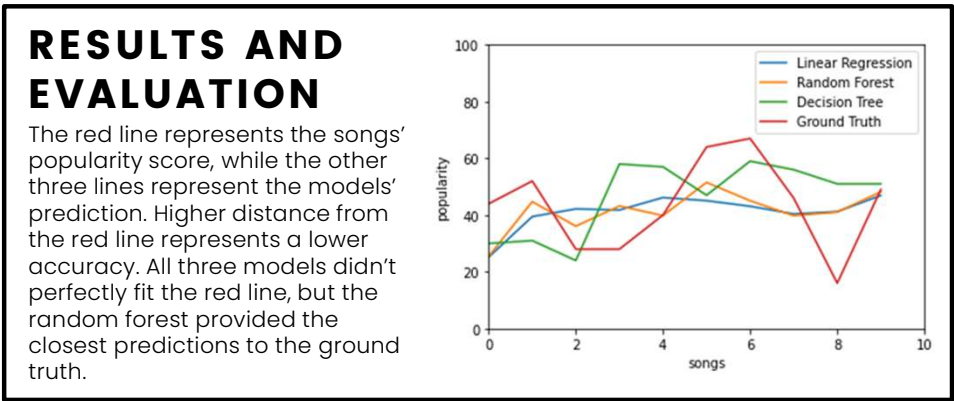
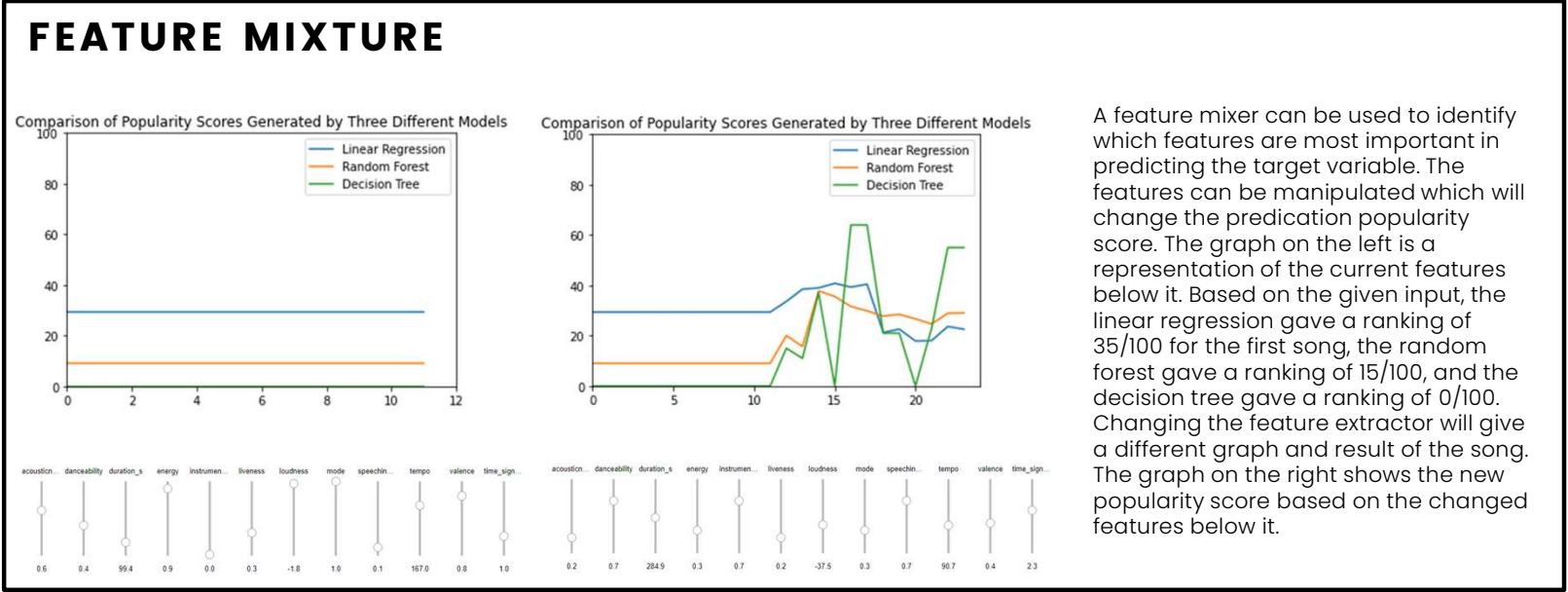
PRE-PROCESSING

REDUCING NUMBER OF SONGS

The database contains more than 232,000 songs. Following the data cleaning process, the number of songs was reduced to 176,000.

REMOVING FEATURES

Dropping duplicates of track_id and converted time_signature from type object to integer.



MODEL SELECTION

LINEAR REGRESSION	training score: 0.202 testing score: 0.199
RANDOM FOREST	training score: 0.906 testing score: 0.341
DECISION TREE	training score: 0.998 testing score: -0.350