Spotif-why

scho31, kagarwa5, nlee26, awang167

Goal

Given the many song features provided by Spotify and the wide selection of songs on Spotify, we were curious if a song's popularity could be predicted based on its features. While this was a prediction project, we feel we were also testing the hypothesis that song popularity can be predicted, using musical features.

Data

We obtained data from a Kaggle dataset which gave us songs from 1921-2020. We also scraped songs from Spotify's U.S. weekly top 200 songs from 2017-present to get a sample of popular songs. This gave us a total of 160,341 songs. This dataset had song features like valence, danceability, acousticness, loudness, energy etc.

Model+Evaluation Setup

We want to predict song popularity using song attributes. We played around with different attributes and narrowed down on ones that gave the best results. We also used our correlation heatmap to narrow down the attributes. The attributes used to predict popularity were: valence, danceability, energy, loudness, mode, speechiness, acousticness, instrumentalness, and liveness. We augmented the data by using min-max scaling and threw away duplicates- found by the same song ID or the same track and artist name. We did a 80-20 train-test split to minimize any bias and to evaluate the predictive ability of the models. Our primary accuracy metric is MSE
We first set up a multiple linear regression. Not satisfied with the result obtained, we decided to use Support Vector Regression to unravel more complex relationships. The SVR didn't work as expected (see Challenges section), so we decided to make a Neural Net to improve performance.

Challenges

Initially, our data only consisted of songs scraped from Spotify's U.S. weekly top 200 songs from 2017-present. This dataset had a weird distribution of popularity scores, which severely limited model performance (as presented in the analysis deliverable). As a result, we expanded our dataset to include the Kaggle dataset, to get a much more even distribution. This new dataset almost halved our testing MSE for linear regression (0.0678 -> 0.0350).
However, as a result of the much bigger dataset, the SVR, given its complex nature, did not finish running (we tried Google Colab and GCP). Hence, we do not have metrics for the SVR for the new dataset but we did compare it to Linear Regression on the original dataset and obtained test MSE of 0.0535 and 0.0520 respectively.

Results and Analysis

Claim #1: Popularity can't be exactly predicted but can be approximated using musical features
Support for Claim #1: We saw an MSE of 0.0276 for the Neural Net and 0.0350 for Linear Regression. This means at best, we'll be ~16.5 (scale of 100) points off from our actual popularity metric, on average. The Linear Regression is ~18.7 (scale of 100) points off, on

average. This implies that predicting exact or near exact popularity, using our approach, seems impossible but we can probably predict popularity within a range i.e. the model will not predict a popular song as unpopular and vice-versa.

Claim #2: More complex models aren't the key to predicting musical popularity
Support for Claim #2: As can be seen from the graphs below, the neural net is only slightly better at predicting popularity than linear regression. This, we believe, is evidence that we need another approach to modelling musical popularity, than just more complex models.
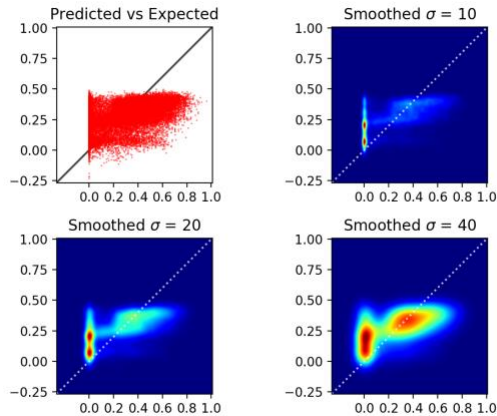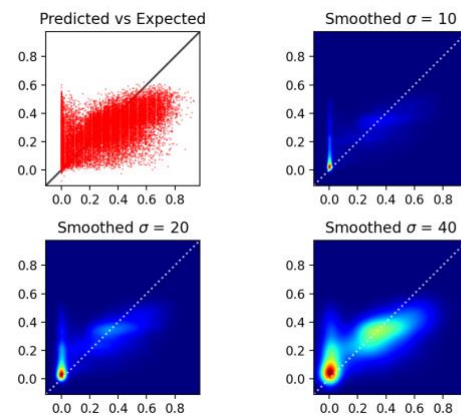

Figure 1: Linear Regression

Figure 2: Neural Net

Claim #3: Acousticness, Instrumentalness, Loudness, and Energy are the most significant contributors to musical popularity.
Support for Claim #3: Looking at the correlation heatmap of features generated, these 4 attributes have the highest values influencing popularity (both negative and positive), as seen in the bottom row below.
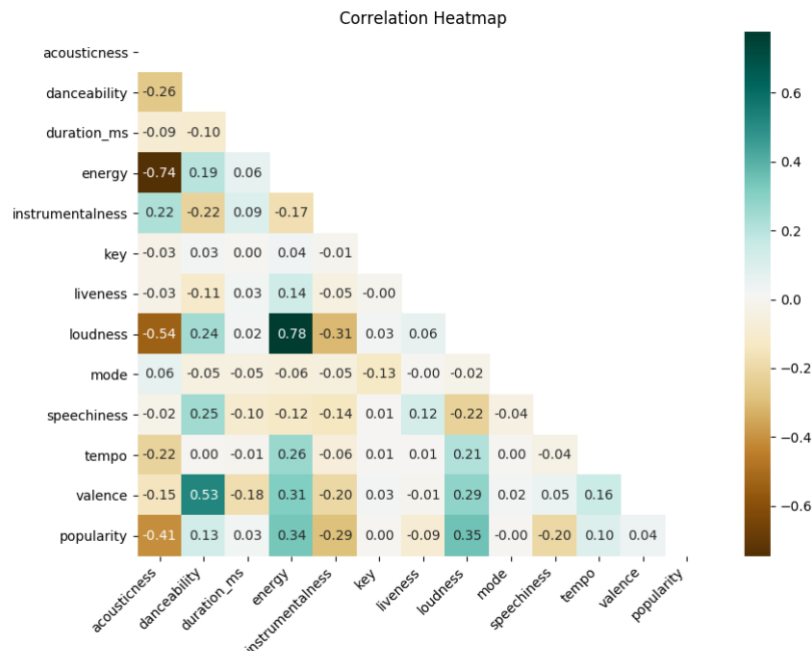
Moving Forward

On the basis of the claims above, we do not think that solely using musical features to predict popularity is the best way forward. We have two ideas for models that could have better performance:

1. Segregating by genre, region: There might be different combinations of attributes, in different genres and regions, that contribute to musical popularity (as we talked about in the ethics report). Calculating separate popularity scores region-wise, and training genre specific models might lead to better performance.

2. Segregating by popularity: From our analysis deliverable, we found that the skew in data between less popular and more popular songs is hurting our performance. A naïve idea we tried was removing outliers. We thresholded on popularity i.e. only ran models which had popularity scores above 50, and then for below 50. Both times, performance improved significantly (MSE ~ 0.0045). This made us think that perhaps there are different combinations of attributes that work for popular vs unpopular songs. Training separate models for them, and then training a simple classifier that predicts if the song will be 'popular' or 'unpopular' to decide which model to run the song through, might lead to much better performance.