Will My Song Be Spotify Famous?

STATS 418 - Final Project

Wesley Lau
Austin Haynesworth
Daehyun Kim

Table of Contents

- Introduction
- Overview of Data Set
- Exploratory Data Analysis and Data Transformations
- Model Selection and Performance
- Flask App Production
- Future Directions

Using audio features to predict song popularity

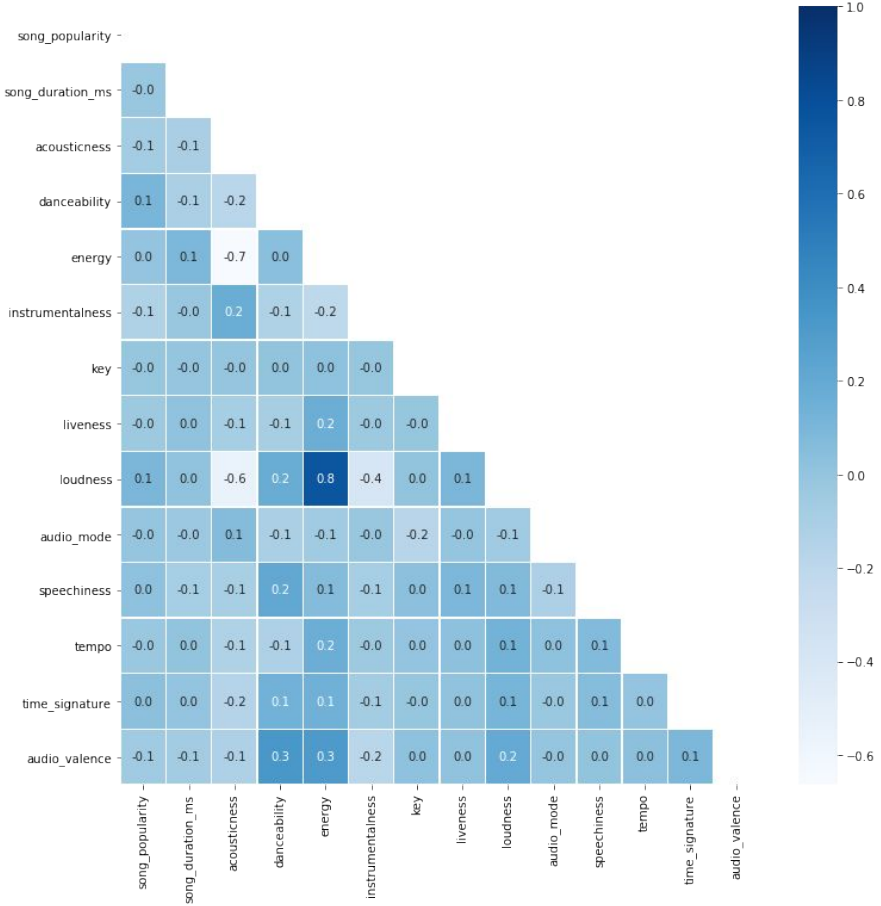
- Determining the popularity of a song is useful to singers, songwriters, and labels, and can help them decide the direction of their music.
- What makes a song popular is a complex question. Undoubtedly, external social factors such as the level of fame of an artist or a songs social significance can impact its popularity.
- Our goal is to see how musical features themselves (i.e. the key signature, beats per minute, minor/major tonality, etc.) can predict a songs popularity.
- Using data audio features (i.e. tempo, danceability, etc.) from 19000 songs on spotify, we will build an ML model to predict level of popularity.
- We will also build a Flask App with user interface such that a user can manually input the audio features of their own song, such to predict its popularity

Overview of Dataset

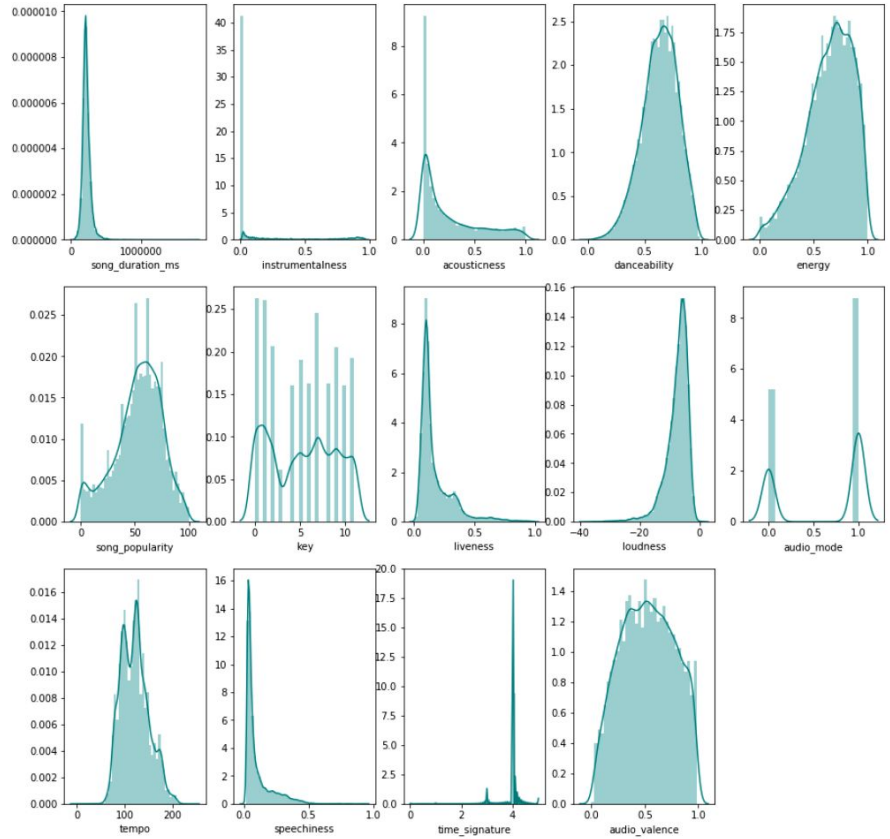
- Dataset includes ~19k songs, each with 14 variables:

Feature	Explanation
duration_ms	The duration of the track in milliseconds.
key	The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
audio_mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
time_signature	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
acousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.
instrumentalness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
loudness	A categorial variable representing the loudness of a track. Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values that range between -40 and 0 db are binned into levels 1-4 with 4 being the loudest value.
speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music.Values below 0.33 most likely represent music and other non-speech-like tracks.
audio_valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
song_popularity	Song ratings of spotify audience.

Correlation Chart for Predictor Variables



EDA and Data Transformations



Distributions of Variables

- In our model, the only input variable transformed is loudness
- Loudness is changed from units of dB's to a categorical bins 1-4 (of 10dB each)
- Our dependent variable, song popularity is binned into 5 categories from the 100-point scale
- However, we will see for interpretability and other reasons, our data and requested inputs may be scaled in the flask app for UI purposes

Model Selection and Performance

Introduction to data

Response variable: Popularity, as determined by number of spotify plays. The 19000 songs in our data set were binned into 5 categories, 1-5, which 5 being most popular.

Predictor variables: 13 audio features of a given song

Model Selection: Our goal was to find the most accurate model for predicting popularity

- Data was split into a 75:25, test:train sets.
- A grid search was run on the train data set using Linear SVC, K Nearest Neighbors, Perceptron, Logistic Regression, Multi-layer Perceptron (MLP) neural networks, and random forest models with various parameters.
- A total of 27 models were created via the grid search, and the model with highest accuracy was chosen.

Model Parameters and Performance

Grid search produced 27 models, and the model with highest accuracy was selected

	mean_fit_time	std_fit_time	mean_score_time	std_score_time	param_classifier	mean_test_score	std_test_score	rank_test_score
21	2.974273	0.221245	0.052843	0.004034	MLPClassifier(activation='tanh', alpha=0.0001,...	0.333693	0.011125	16
22	7.793696	2.549174	0.060719	0.000354	MLPClassifier(activation='tanh', alpha=0.0001,...	0.325789	0.000055	19
23	7.085579	1.688945	0.081027	0.035185	MLPClassifier(activation='tanh', alpha=0.0001,...	0.325789	0.000055	19
24	0.787318	0.070880	0.013565	0.000384	MLPClassifier(activation='tanh', alpha=0.0001,...	0.284832	0.091468	23
25	0.855330	0.179652	0.015349	0.001421	MLPClassifier(activation='tanh', alpha=0.0001,...	0.341597	0.011231	11
26	2.715604	0.005287	0.140488	0.000879	(DecisionTreeClassifier(class_weight=None, cri...	0.542646	0.008196	2
27	5.211217	0.327196	0.277498	0.028202	(DecisionTreeClassifier(class_weight=None, cri...	0.542933	0.009285	1

The most accurate model was a random forest model. Test data was then run through the model, with the following results:

```
Accuracy of model is 0.5723
Precision of model is 0.6717
Recall of model is 0.5047
F1 score of model is  0.5155
```

Flask App - Features

- On the first page, a user can extract audio features for a reference song of their choice using a song title and artist name
- This will help a user understand what values are appropriate to use for their own song since these metrics can be something unfamiliar

Please input track name and artist name to extract Spotify audio features for your reference track

Artist Backstreet Boys

Track I want it that way

Submit

Spotify Audio features for your reference track :

Acousticness: 2.2

Danceability: 6.9

Duration(s): 213.6

Energy: 7

Instrumentalness: 0

Key: 6

Liveness: 1.6

Loudness: 4

Mode: 0

Acousticness: A confidence measure from 0.0 to 10.0 of whether the track is acoustic. 10.0 represents high confidence the track is acoustic.

Danceability: describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 10.0 is most danceable.

Duration: The duration of the track in seconds.

Energy: Energy is a measure from 0.0 to 10.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.

Instrumentalness: Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 10.0, the greater likelihood the track contains no vocal content. Values above 5.0 are intended to represent instrumental tracks, but confidence is higher as the value approaches 10.0.

Key: The key the track is in. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. Liveness: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 8.0 provides strong likelihood that the track is live.

Loudness: The overall loudness of a track binned into a dB range. Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). The 4 levels of loudness range from 1 (least loud) to 4 (loudest).

Mode: Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.

Speechiness: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 10.0 the attribute value. Values above 6.6 describe tracks that are probably made entirely of spoken words. Values between 3.3 and 6.6 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 3.3 most likely represent music and other non-speech-like tracks.

Tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

Time Signature: An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).

Valence: A measure from 0.0 to 10.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

Now predict the popularity of your own song

link to prediction

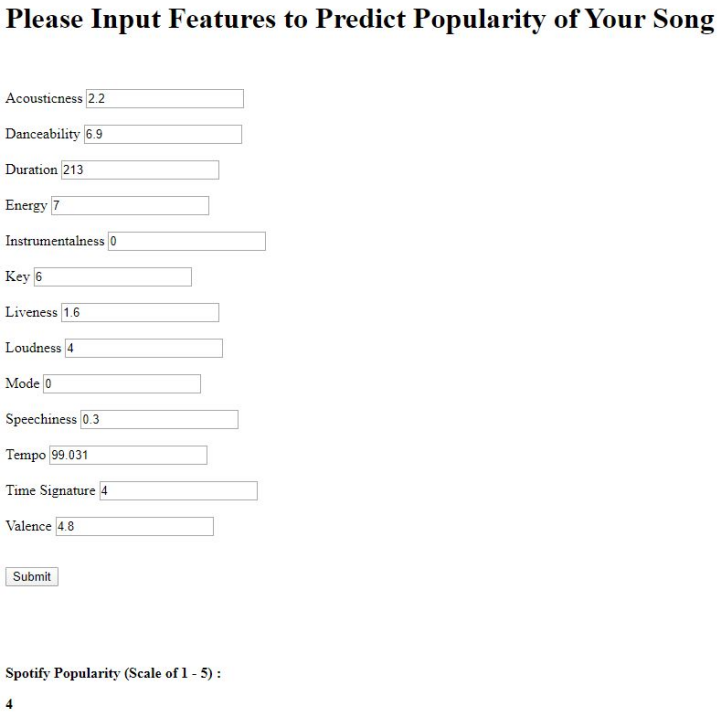
Flask App - Features

- Users can navigate to the second page to input audio features

  of their own song that is not on Spotify

  In order to predict its Spotify Popularity

DEMO



Insights and Future Directions

- Currently, users must estimate the values of numerical audio features, using their understanding of these numerical values based on reference songs.
- It would be more useful if the user could instead upload audio files, and the song would be parsed for these audio features, like Spotify currently does.
- Unfortunately, Spotify’s parsing software is proprietary and not open source.
- For future directions, we would like to improve input methods for users, such that the numerical values they input for audio features are more precise, hence leading to a more precise popularity prediction.