

Abstract

The goal of this project was to create a web app where a user would be able to input the characteristics of a particular song, and then get a predicted value for the popularity of that song. I would use the Spotify API to access the data and then put the data into a SQL database. Once it is in the database, I would clean it and train a linear regression model. I would then use this model in my web app and allow the user to toggle the features of the song.

Design

I came up with the idea of this project by thinking about another way that record labels can gauge the popularity of songs that they release. Beyond being able to determine the popularity of a specific song, this tool could also be used to see what features are important in determining the popularity, which can allow the record label to make specific decisions about the music such as making the tempo faster, or releasing the song in a different month.

Data

I have pulled 100,423 rows and 22 columns off of the Spotify API. Each row of the dataset represents a song. I pulled data from two main feeds, the basic song information and the audio features of the song. There were no missing values in the dataset, and I did not use any categorical datapoints in the final model.

Algorithms

API Pull

The spotify API requires the user to pull a maximum of 50 songs at a time based on the specified search criteria. I built functions to iterate through these search lists, pulling each song and adding it to a pandas dataframe. Once the data frame was complete, I loaded the data into a SQL database.

Feature Engineering

I created new variables for dates so that I could see if seasonality has an impact on the popularity. I also added polynomial features.

Models

I used basic linear regression, lasso regression, and ridge regression. I used 5-fold cross-validation to determine which model had the highest R^2 score and then moved forward with that model for the web app.

Streamlit

I built a streamlit app with slider and number inputs so that users can inputs their own values.

Tools

Numpy and Pandas for data manipulation

Scikit-learn for modeling

Pickle for model storage

SQL for Data Storage

Spotipy for Data Acquisition

Streamlit for Web App

Communication

In addition to the slides, the web address for the web app is

https://share.streamlit.io/liammoran13/spotify_app/main/streamlit_code.py