MUSIC POPULARITY PREDICTOR

Sean McCormick, Ryan Power, Sam Springer, and Jenna Dean

DESCRIPTION

Create a model that can successfully predict the popularity of a song and/or artist based on various quantitative and qualitative characteristics mined from the artist/song.

INTERESTING QUESTIONS

1. Can a model be created to predict the popularity of an artist or song using Spotify?
2. Can an artist's popularity be predicted based on the acoustic attributes of their songs and albums?
3. Is it possible to correlate a set of song characteristics that will effectively predict an artist's and/or song's future popularity and success?
4. Can a combination of genres predict which artists will be the most popular?

Prior work

- Spotify API query framework that retrieves current data (returned to API user in JSON) about a song or artist.
- Pandora Music Genome Project: analyzes each song using up to 400 distinct musical attributes that are relevant to understanding the interests of a listener. (https://www.pandora.com/corporate/mgp.shtml)
- Spotify Predictive Model: a predictive model that determines whether a listener likes or dislikes a song.
- Spotify Predictor of Song Popularity: a model based on how often a song is added to a playlist. (https://developer.spotify.com/documentation/web-api/reference/#object-trackobject)

DATASETS

- Queried data from Spotify's API which includes artist and song characteristics for just under 30,000 artists.
- A compiled dataset from Kaggle which can be found with the URL: (https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks)
- Our data is installed on our machines locally and can also be found on our github project repository: (https://github.com/jede4829/Data-Mining-Team-Project)

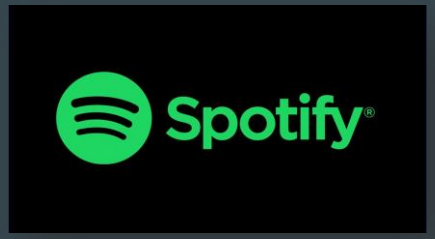
artists	id_artists	release_date	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalne	liveness	valence	tempo
['Uli']	['45tIt06XoI0Iio4LBEVpls']	2/22/1922	0.645	0.445	0	-13.338	1	0.451	0.674	0.744	0.151	0.127	104.851
['Fernando Pessoa']	['14jtPCOoNZwquk5wd9DxrY']	6/1/1922	0.695	0.263	0	-22.136	1	0.957	0.797	0	0.148	0.655	102.009
['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	3/21/1922	0.434	0.177	1	-21.18	1	0.0512	0.994	0.0218	0.212	0.457	130.418
['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fUm3X2']	3/21/1922	0.321	0.0946	7	-27.961	1	0.0504	0.995	0.918	0.104	0.397	169.98
['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	1922	0.402	0.158	3	-16.9	0	0.039	0.989	0.13	0.311	0.196	103.22
['Dick Haymes']	['3BiJGZsyX9sJchTqcSA7Su']	1922	0.227	0.261	5	-12.343	1	0.0382	0.994	0.247	0.0977	0.0539	118.891
['Francis Marty']	['2nuMRGzeJ5jJEKlfS7rZ0W']	1922	0.51	0.355	4	-12.833	1	0.124	0.965	0	0.155	0.727	85.754
['Mistinguett']	['4AxgXfD7ISvJSTObqm4aIE']	1922	0.563	0.184	4	-13.757	1	0.0512	0.993	1.55E-05	0.325	0.654	133.088
['Greg Fieler']	['5nWlsH5RDgFuRAiDeOFVmf']	1922	0.488	0.475	0	-16.222	0	0.0399	0.62	0.00645	0.107	0.544	139.952

PROPOSED WORK

- Capture Data
  - Query Spotify API for various data sets to compare and contrast against local data set to confirm preciseness and accuracy.
- Data Cleaning
  - Remove outlier attributes
  - Fill in missing attributes as needed
    - Utilize regression techniques to define missing values
- Data Preprocessing
  - Dimensionality reduction
    - Find correlations between attributes
  - Normalization
- Data Integration
  - Call Spotify API using Python script
  - Using Python Dataframes > csv
  - Upload to Tableau to analyze data

LIST OF TOOLS

- Spotify API
- MLXTEND – Apriori
- Numpy
- Pandas
- Sys
- Python
- Tableau



EVALUATION

- Compare modeled popularity results to what Spotify deems as popular based off of current streaming trends and charts.
- Compare Spotify recommendations to results generated by our predictive models.