CSE 4/587 – Data Intensive Computing
Project - Phase 3

Thilak Reddy Dharam – thilakre – 50469154, Harsha Vardhan Bitra - hbirta - 50468952

Prediction of Songs Popularity on Spotify Data

Introduction:

The main aim of this project is to predict the popularity of the song when a new song is inserted into the dataset. The dataset we are using to solve the above is Spotify data: "SpotifyFeatures.csv" containing features such as "acousticness", "danceability", "instrumentalness", "liveness", "loudness", etc.

Dataset Description:

The dataset used in this project consists of songs from different artists on most popular music platform "Spotify" which has 232725 records and 18 columns.

Dataset Name: SpotifyFeatures.csv

Data Source: https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db

Attributes Description:

genre: Different genres of songs such as classical, jazz, hiphop, etc.

artist_name: Name of the artist who has composed the song.

track_name: Title of the song.
track_id: Unique ID generated for song by Spotify.

popularity: Popularity of song consisting of values between [0-100].
acousticness: Measures the acoustic of song , values consists between [0-1].

danceability: Describes if the song can be used to dance.
duration_ms: Duration of the song in milliseconds.
energy: Represents intensity and activity of the song.
instrumentalness: Represents the vocals of the song.
key: Overall key of the song, using standard pitch class notation. Ex-> 0-C, 1-C#, 2-D.

liveness: Represents the presence of audience in the song.
loudness: Overall loudness of the song in decibels.
mode: Indicates the modality of the song, representing “Major” for 1 and “Minor” for 0.

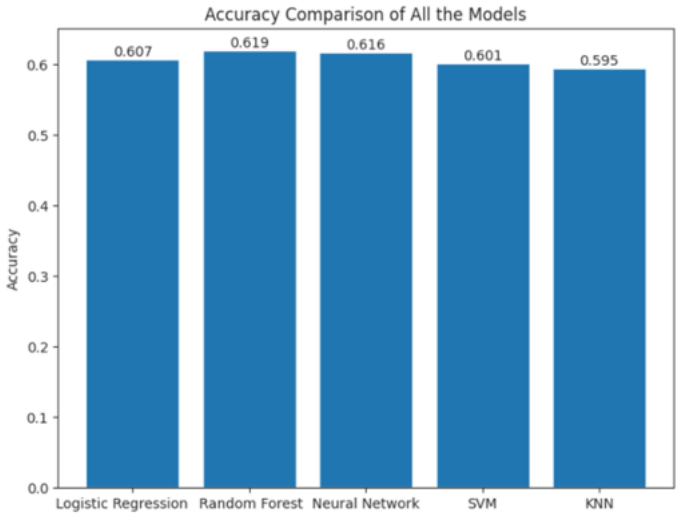
speechiness: Describes the measure and perfectness of spoken words in the song.

tempo: Beats for minute of the song.
time_signature: Tells the number of beats in each measure of the song.
valence: Musical positiveness delivered by the song.

Model:

Of the models from Phase 2, we are using the Random Forest model for predictions in Phase 3. The reason for picking the random forest model out of other models even though other models have a similar accuracy is its robustness to outliers and its less susceptibility to overfitting, versatility, scalability, and finally interpretability.

Models Input features: [‘Acousticness’, ‘Danceability’, ‘Energy’, ‘Loudness’, ‘Speechiness’]



Recommendations:

By predicting the popularity of songs from these features, it is possible to create a reliable prediction model that can help in a better recommendation of songs, marketing, and promotion, understanding modern trends, and generating revenue for the music industry.

We can even extend our project to provide a more comprehensive and personalized music recommendation system for users. Also, by implementing ideas of Collaborative Filtering, Genre Classification, and Mood Analysis, users can learn more about the songs they listen to, discover new music, and have a more engaging and enjoyable music-listening experience.

Project directory:

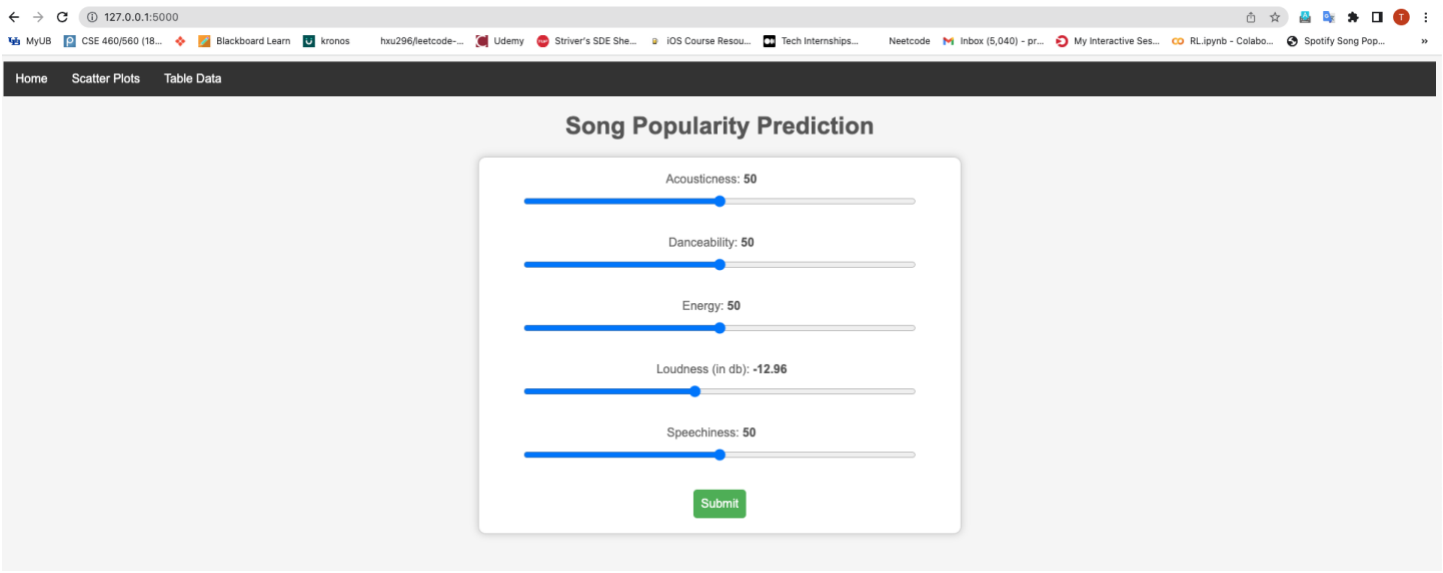
- Src:
    - Phase1
    - Phase2
    - Phase3
        - data => (this directory contains saved model pickle files and json files for visualizations)
        - static => (app.js file for handling dynamic visualizations)
        - templates => (html files)
        - app.py => (contains flask integration code)
        - requirements.txt
        - readme file

Working Instructions:

1. Prerequisites: Python3
2. Create a virtual environment for the project.
3. Install requirements using the below command
   pip install -r requirements.txt
4. Run the below command to start the web application
   flask --app app run

End Product Screenshots:

Home page:



Dynamic Scatter Plots:

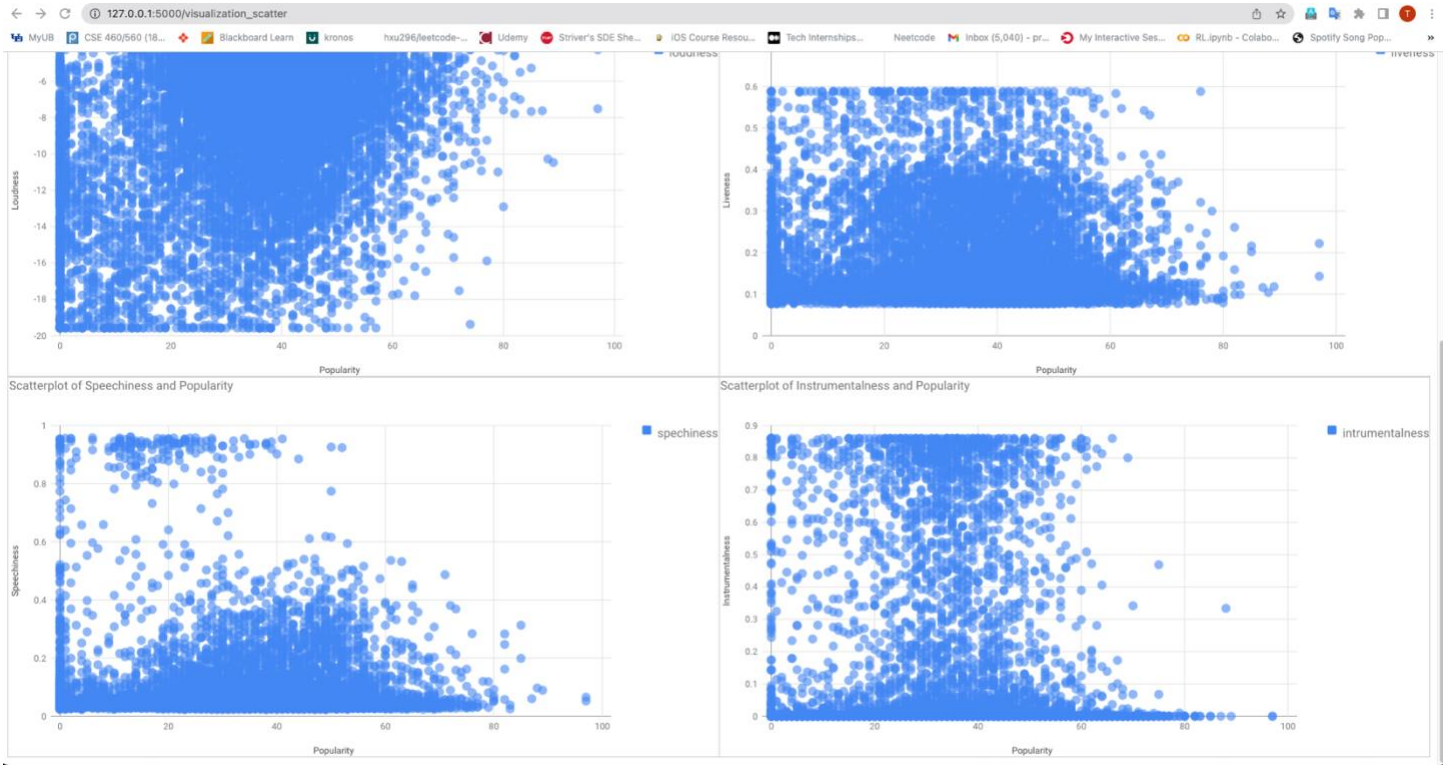
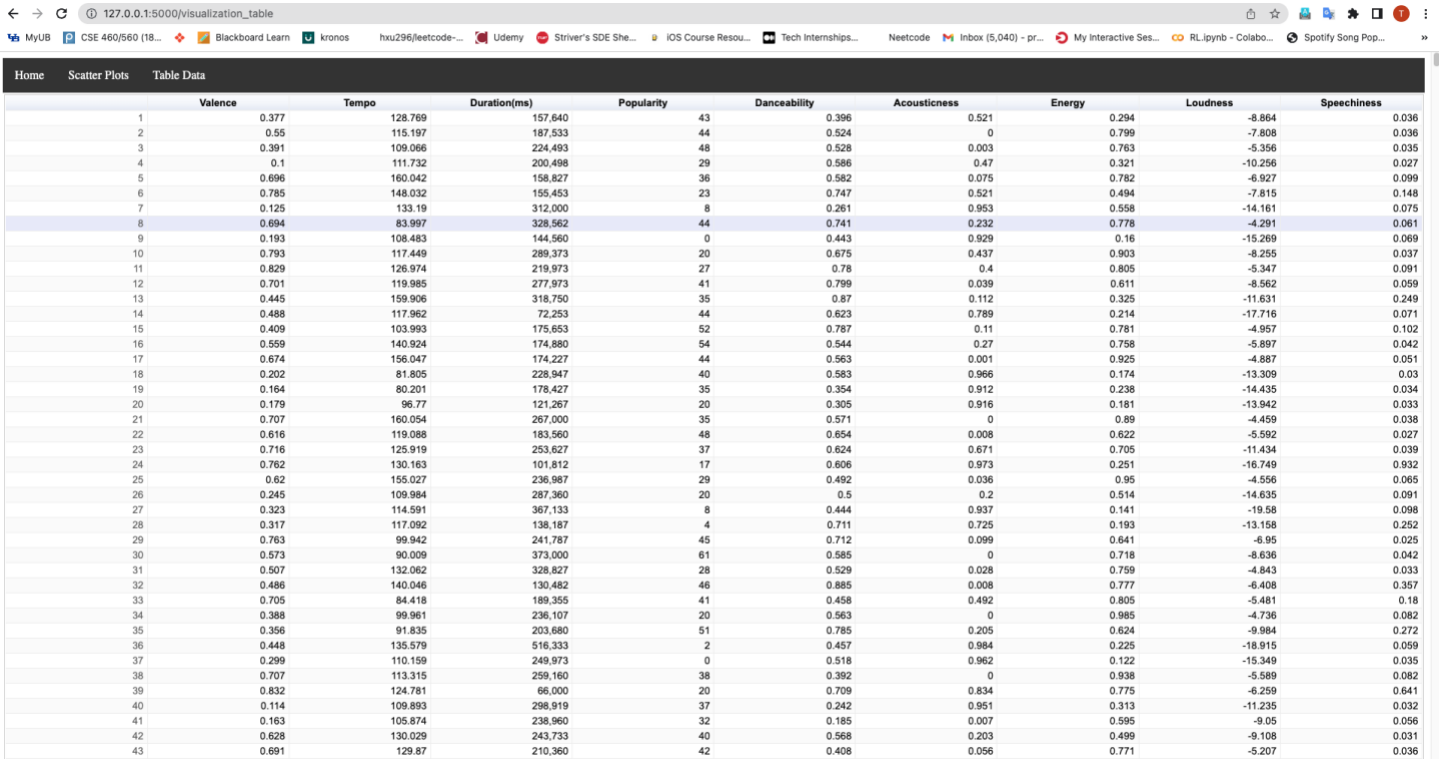


Table Data:

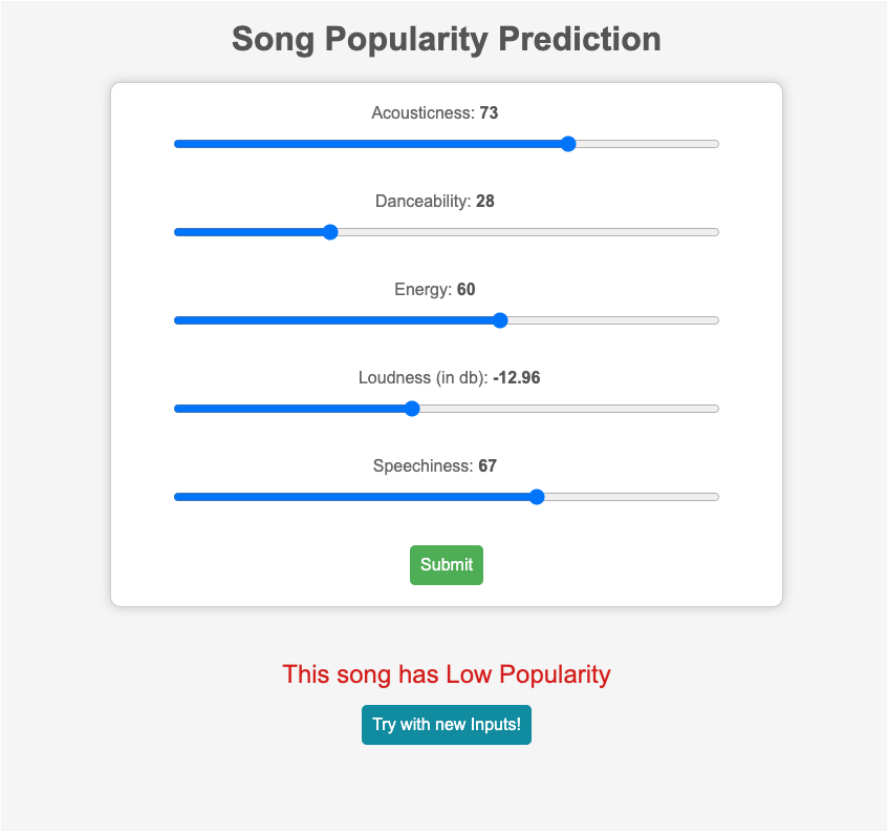


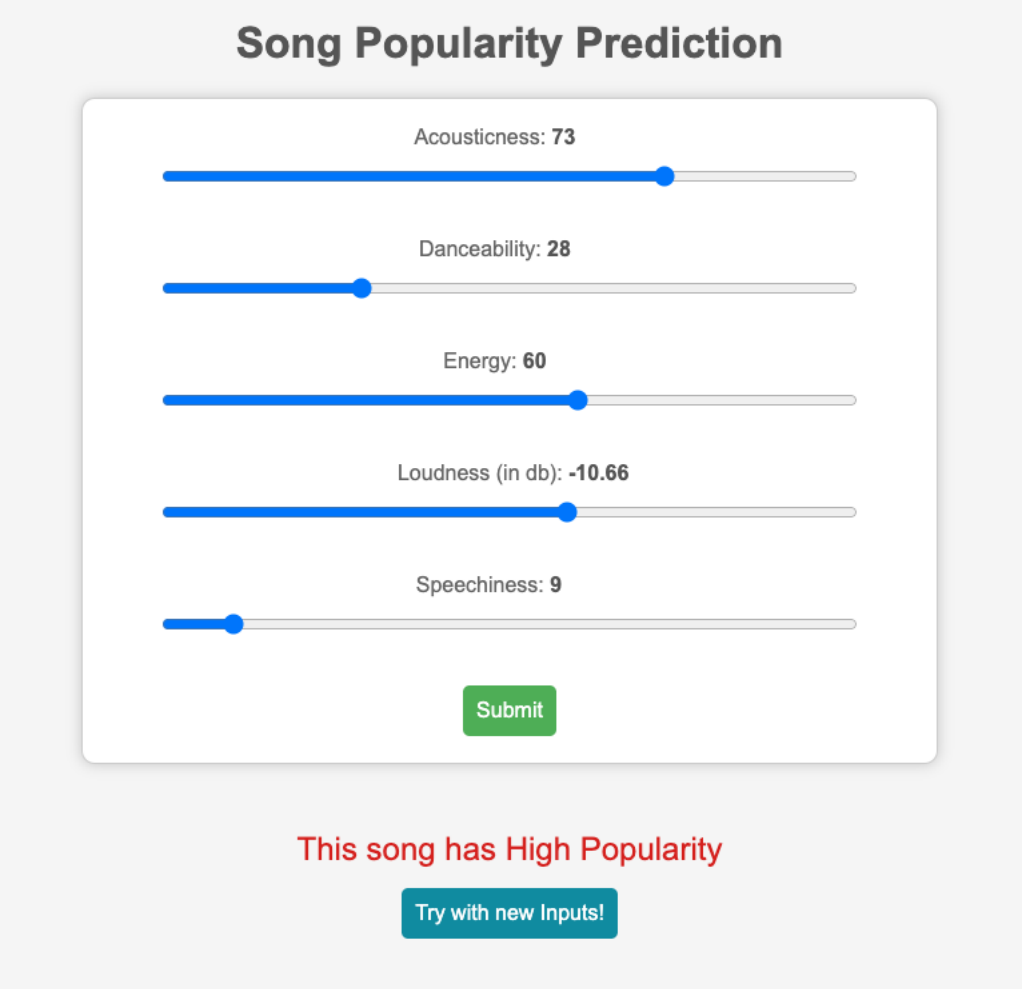
127.0.0.1:5000/visualization_table

Home | Scatter Plots | Table Data

	Valence	Tempo	Duration(ms)	Popularity	Danceability	Acousticness	Energy	Loudness	Speechiness
1	0.377	128.769	157,640	43	0.396	0.521	0.294	-8.864	0.036
2	0.55	115.197	187,533	44	0.524	0	0.799	-7.808	0.036
3	0.391	109.066	224,493	48	0.528	0.003	0.763	-5.356	0.035
4	0.1	111.732	200,498	29	0.586	0.47	0.321	-10.256	0.027
5	0.696	160.042	158,827	36	0.582	0.075	0.782	-6.927	0.099
6	0.785	148.032	155,453	23	0.747	0.521	0.494	-7.815	0.148
7	0.125	133.19	312,000	8	0.261	0.953	0.558	-14.161	0.075
8	0.694	83.997	328,562	44	0.741	0.232	0.778	-4.291	0.061
9	0.193	108.483	144,560	0	0.443	0.929	0.16	-15.269	0.069
10	0.793	117.449	289,373	20	0.675	0.437	0.903	-8.255	0.037
11	0.829	126.974	219,973	27	0.78	0.4	0.805	-5.347	0.091
12	0.701	119.985	277,973	41	0.799	0.039	0.611	-8.562	0.059
13	0.445	159.906	318,750	35	0.87	0.112	0.325	-11.631	0.249
14	0.488	117.962	72,253	44	0.623	0.789	0.214	-17.716	0.071
15	0.409	103.993	175,653	52	0.787	0.11	0.781	-4.957	0.102
16	0.559	140.924	174,880	54	0.544	0.27	0.758	-5.897	0.042
17	0.674	156.047	174,227	44	0.563	0.001	0.925	-4.887	0.051
18	0.202	81.805	228,947	40	0.583	0.966	0.174	-13.309	0.03
19	0.164	80.201	178,427	35	0.354	0.912	0.238	-14.435	0.034
20	0.179	96.77	121,267	20	0.305	0.916	0.181	-13.942	0.033
21	0.707	160.054	267,000	35	0.571	0	0.89	-4.459	0.038
22	0.616	119.088	183,560	48	0.654	0.008	0.622	-5.592	0.027
23	0.716	125.919	253,627	37	0.624	0.671	0.705	-11.434	0.039
24	0.762	130.163	101,812	17	0.606	0.973	0.251	-16.749	0.932
25	0.62	155.027	236,987	29	0.492	0.036	0.95	-4.556	0.065
26	0.245	109.984	287,360	20	0.5	0.2	0.514	-14.635	0.091
27	0.323	114.591	367,133	8	0.444	0.937	0.141	-19.58	0.098
28	0.317	117.092	138,187	4	0.711	0.725	0.193	-13.158	0.252
29	0.763	99.942	241,787	45	0.712	0.099	0.641	-6.95	0.025
30	0.573	90.009	373,000	61	0.585	0	0.718	-8.636	0.042
31	0.507	132.062	328,827	28	0.529	0.028	0.759	-4.843	0.033
32	0.486	140.046	130,482	46	0.885	0.008	0.777	-6.408	0.357
33	0.705	84.418	189,355	41	0.458	0.492	0.805	-5.481	0.18
34	0.388	99.961	236,107	20	0.563	0	0.985	-4.736	0.082
35	0.356	91.835	203,680	51	0.785	0.205	0.624	-9.984	0.272
36	0.448	135.579	516,333	2	0.457	0.984	0.225	-18.915	0.059
37	0.299	110.159	249,973	0	0.518	0.962	0.122	-15.349	0.035
38	0.707	113.315	259,160	38	0.392	0	0.938	-5.589	0.082
39	0.832	124.781	66,000	20	0.709	0.834	0.775	-6.259	0.641
40	0.114	109.893	298,919	37	0.242	0.951	0.313	-11.235	0.032
41	0.163	105.874	238,960	32	0.185	0.007	0.595	-9.05	0.056
42	0.628	130.029	243,733	40	0.568	0.203	0.499	-9.108	0.031
43	0.691	129.87	210,360	42	0.408	0.056	0.771	-5.207	0.036

Giving Inputs and Prediction:





References:

1. https://flask.palletsprojects.com/en/2.2.x/
2. https://developers.google.com/chart/interactive/docs
3. https://developer.mozilla.org/en-US/docs/Web/HTML
4. https://developer.mozilla.org/en-US/docs/Web/CSS
5. https://developer.mozilla.org/en-US/docs/Web/JavaScript