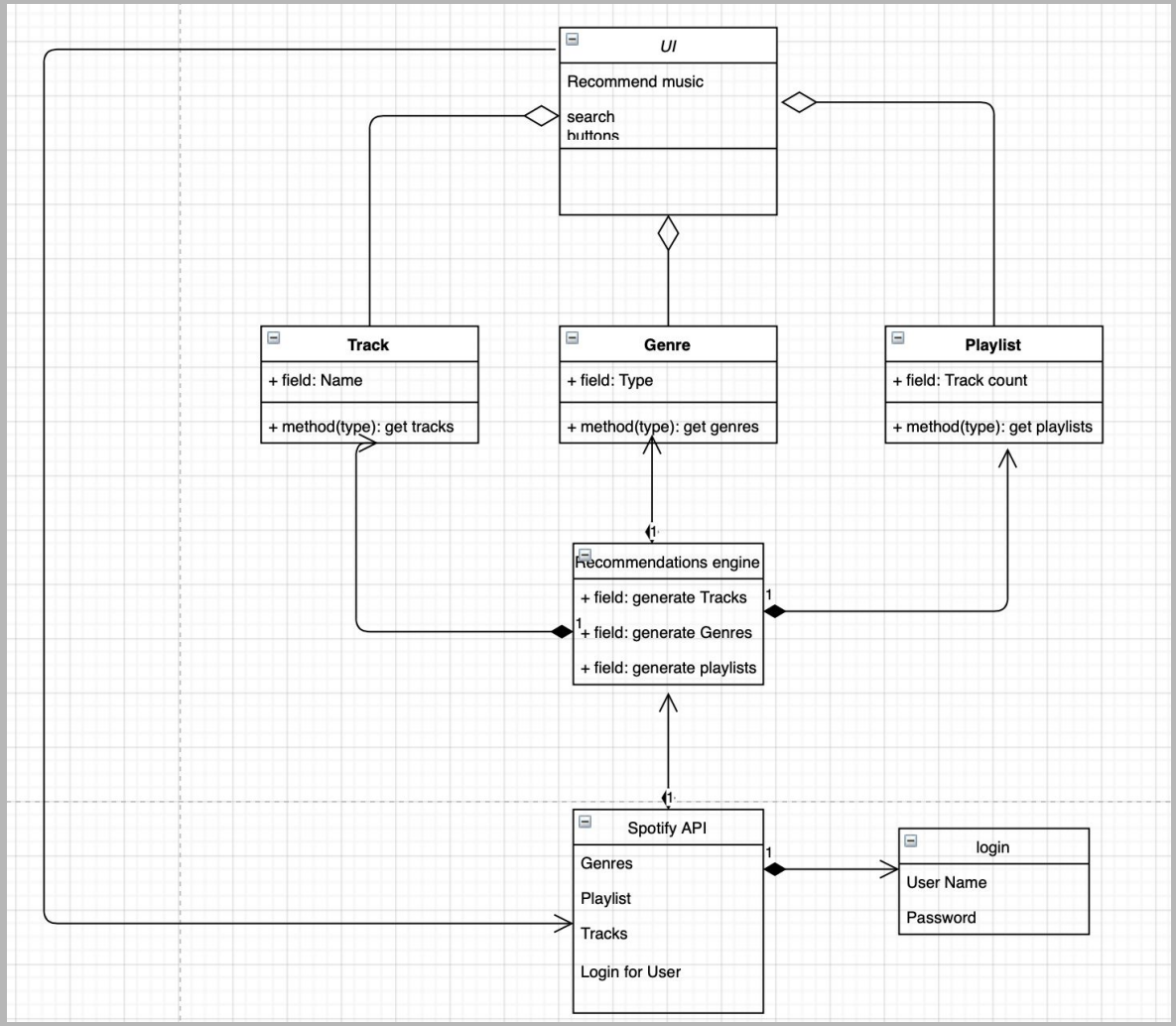
Final Project Presentation

Cameron McDavitt, Cameron Paul, Ola Nzeogu, Wenjie Ma, Deverell M

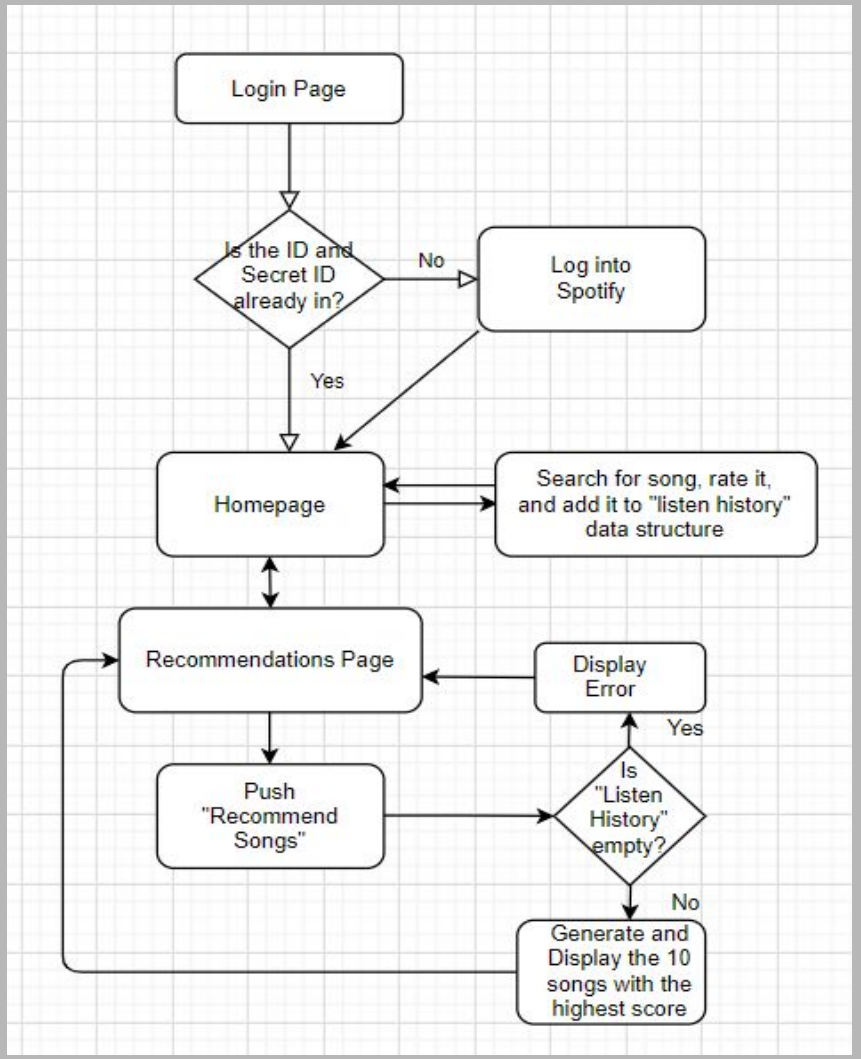
Purpose of system

The purpose of this system is to emulate the spotify app software. At the end of this project, a user should be able to search for a song of their choice, and the software should find the desired song from the used dataset and display it to the user. The machine learning aspect of the software should then be able to note when the user likes a song and should recommend more songs that are similar to the original song searched for by the user. The software uses Javascript, HTML/CSS, React, and Python for the machine learning.

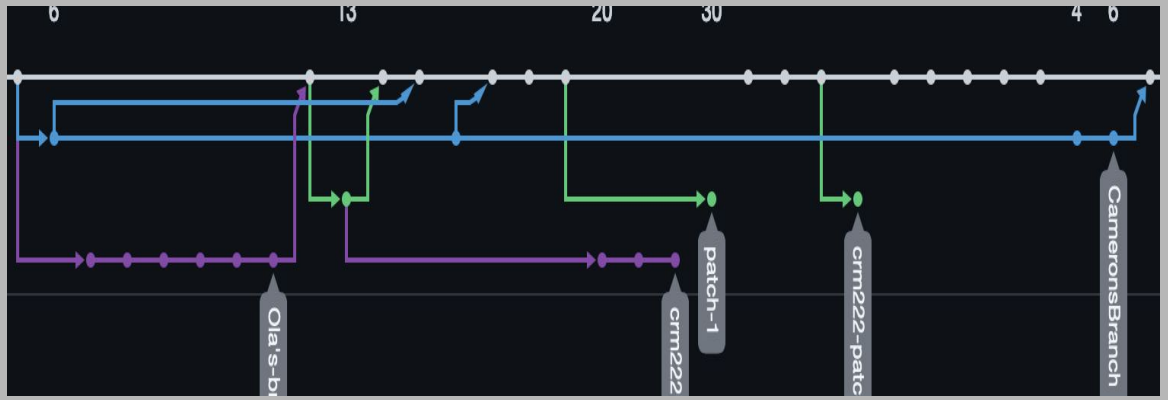
UML Class Diagram



System statechart diagrams



Github Flowchart



Machine Learning Program

- Uses a Kaggle dataset for spotify (https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks)
- Due to space limitations, the data was trimmed to one file of 1000 songs with song name, id, artists, and popularity (songs with 77 popularity or lower were trimmed)
- Runs a matrix algorithm known as a “TfidfVectorizer” through a sigmoid kernel, which compares artist name similarity for a given song to the artist names of the songs within the database.
- The function runs with the name of a song as a parameter and returns the list of songs in order of most similar artist names to least similar. Requires the song name to be contained within the dataset.
- Runs in python through pandas matrix manipulation and sklearn algorithms.
- Disconnected from the front end.

Data sample

	name	artists	id	popularity
2	drivers license	['Olivia Rodrigo']	7lPN2DXiMsVn7XUKtOW1CS	100
3	Mood (feat. iann dior)	['24kGoldn', 'iann dior']	3tjFYV6RSFtuktYl3ZtYcq	96
4	positions	['Ariana Grande']	35mvY5S1H3J2QZyna3TFe0	96
5	DÃKITI	['Bad Bunny', 'Jhay Cortez']	47EiUVwUp4C9fGccaPuUCS	95
6	BICHOTA	['KAROL G']	7vrJn5hDSXRmdXoR30KgF1	95
7	34+35	['Ariana Grande']	6Im9k8u9iIzKMrmV7BWtlF	94
8	Whoopty	['CJ']	5vGLcdRuSbUhD8ScwsGSdA	94
9	WITHOUT YOU	['The Kid LAROI']	27OeeYzk6klgBh83TSvGMA	94
10	Therefore I Am	['Billie Eilish']	54bFM56PmE4YLRnqpW6Tha	94
11	LA NOCHE DE ANOCHE	['Bad Bunny', 'ROSALÃA']	2XIc1pqjXV3Cr2BQUGNBck	94
12	What You Know Bout Love	['Pop Smoke']	1tkg4EHVoqnhR6iFEXb60y	93
13	you broke me first	['Tate McRae']	45bE4HXI0AwGZXfZtMp8JR	93
14	WAP (feat. Megan Thee Stallion)	['Cardi B', 'Megan Thee Stallion']	4Oun2ylbjFKMPTiaSbbCih	92
15	HOLIDAY	['Lil Nas X']	6zFMeegAMYQo0mt8rXtrli	92
16	Lonely (with benny blanco)	['Justin Bieber', 'benny blanco']	4y4spB9m0Q6026KfkAvy9Q	92
17	Anyone	['Justin Bieber']	31qCy5ZaophVA81wtlwLc4	92
18	The Business	['TiÃ«sto']	6f3Slt0GbA2bPZlz0aIFXN	92
19	Hecha Pa' Mi	['Boza']	3VvA1wSxukMLsvXoXtlwWx	92
20	Snowman	['Sia']	7uoFMmxln0GPXQ0AcCBXRq	91
21	Watermelon Sugar	['Harry Styles']	6UelLqGlWMcVH1E5c4H7lY	91
22	Good Days	['SZA']	3YJJjQPAbDT7mGpX3WtQ9A	91
23	For The Night (feat. Lil Baby & DaBaby)	['Pop Smoke', 'Lil Baby', 'DaBaby']	0PvFJmanyNQMseIFrU708S	91
24	ROCKSTAR (feat. Roddy Ricch)	['DaBaby', 'Roddy Ricch']	7ytR5pFWmSjzHJIeQkgog4	91
25	Heather	['Conan Gray']	4xqrdfXkTW4T0RauPLv3WA	91
26	Monster (Shawn Mendes & Justin Bieber)	['Shawn Mendes', 'Justin Bieber']	2Z8yfpFX0ZMavHkcIeHiO1	91
27	Life Goes On	['BTS']	249gnXrbfmV8NG6jTEMSwD	91
28	Dynamite	['BTS']	4saklk6nie3yiGePpBwUoc	91
29	DÃKITI	['Bad Bunny', 'Jhay Cortez']	4MzXwWMhyBbmu6hOcLVD49	91

```
929          Old Town Road
339    Old Town Road - Remix
325                   On Me
542                Errbody
110           Star Shopping
319                     F.N
398        Save That Shit
540             Sex Sounds
660            Beamer Boy
913                Losses
```

```python
import pandas as pd
from sklearn.metrics.pairwise import sigmoid_kernel
from sklearn.feature_extraction.text import TfidfVectorizer


df = pd.read_csv('data.csv')
df.dropna()
df_clean = df.drop(columns=['id', 'popularity'])
tf = TfidfVectorizer(analyzer='word', token_pattern=r'\w{1,}', ngram_range=(1, 3), stop_words='english')
matrix = tf.fit_transform(df_clean['artists'].astype('U'))
sigmoid = sigmoid_kernel(matrix, matrix)
song_title = df_clean['name']
indices = pd.Series(df_clean.index, index=df_clean['name'])


def song_recommendation(song_name):
    idx = indices[song_name]
    sim_scores = list(enumerate(sigmoid[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    sim_scores = sim_scores[1:11]
    song_indices = [i[0] for i in sim_scores]
    return df_clean['name'].iloc[song_indices]


print(song_recommendation('HOLIDAY').head(10))
```

Machine Learning Test Cases

```
929                Old Town Road
339     Old Town Road - Remix
325                       On Me
542                     Errbody
110               Star Shopping
319                         F.N
398             Save That Shit
540                 Sex Sounds
660                 Beamer Boy
913                     Losses
```

Case 1: Input a song that is in the dataset

Result: Generated list of 10 songs

Case 2: Input a song that is not in the dataset

Result: Key error, unhandled

```
Traceback (most recent call last):
  File "C:\Users\Camerons Desktop\PycharmProjects\SpotifyAPIMachineLearning\venv\lib\site-packages\pandas\core\indexes\base.py", line 3080, in get_loc
    return self._engine.get_loc(casted_key)
  File "pandas\_libs\index.pyx", line 70, in pandas._libs.index.IndexEngine.get_loc
  File "pandas\_libs\index.pyx", line 96, in pandas._libs.index.IndexEngine.get_loc
  File "pandas\_libs\index.pyx", line 126, in pandas._libs.index.IndexEngine._get_loc_duplicates
  File "pandas\_libs\index.pyx", line 133, in pandas._libs.index.IndexEngine._maybe_get_bool_indexer
  File "pandas\_libs\index.pyx", line 150, in pandas._libs.index.IndexEngine._unpack_bool_indexer
KeyError: 'dfhhhfd'

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "C:/Users/Camerons Desktop/PycharmProjects/SpotifyAPIMachineLearning/main.py", line 25, in <module>
    print(song_recommendation('dfhhhfd').head(10))
  File "C:/Users/Camerons Desktop/PycharmProjects/SpotifyAPIMachineLearning/main.py", line 17, in song_recommendation
    idx = indices[song_name]
  File "C:\Users\Camerons Desktop\PycharmProjects\SpotifyAPIMachineLearning\venv\lib\site-packages\pandas\core\series.py", line 853, in __getitem__
    return self._get_value(key)
  File "C:\Users\Camerons Desktop\PycharmProjects\SpotifyAPIMachineLearning\venv\lib\site-packages\pandas\core\series.py", line 961, in _get_value
    loc = self.index.get_loc(label)
  File "C:\Users\Camerons Desktop\PycharmProjects\SpotifyAPIMachineLearning\venv\lib\site-packages\pandas\core\indexes\base.py", line 3082, in get_loc
    raise KeyError(key) from err
KeyError: 'dfhhhfd'

Process finished with exit code 1
```

Login System