

Shao-Min Yeh

Project Proposal

CS 410: Text Information Systems

Project Progress Report

1. Which tasks have been completed (pictures listed at the bottom of the report):
 - a. The functionality of the three models works and is implemented correctly. The models were implemented similarly to how the proposal stated (models, libraries, etc.).
 - b. The Flask application is implemented completely and incorporates the above models. I now have paths for each of the three models, so the website works along with outputting results from the models.
 - c. HTML pages have been implemented according to the Flask application. I now have each of the models with their respective HTML pages, along with functionality with Flask and the models themselves.
 - d. I also had to add a song search method along with the HTML page and Flask routes. This allowed the similar song and sentiment analysis models to work.
2. Which tasks are pending:
 - a. The models still need parameter tuning and accuracy improvements as listed:
 - i. Search Engine: The parameters need tuning because the results can falter at times.
 - ii. Similar Songs: The combined features need normalization, as musical features weigh much more than TF-IDF features.

- iii. Sentiment Analysis: The results with query tokenization are bad, which suggests that the non-tokenization model works due to overfitting only.

Also, the model takes a long time to run, which can be improved upon.

- b. The HTML pages need further customization. The current pages have the same format as a past project, but this theme doesn't work well for this webpage.

3. Challenges faced:

- a. Modeling Challenges:

- i. Search Engine:

- 1. The main challenge was how to incorporate ranking with title/artist along with lyrics - I found that the former was still very useful to ranking and should be separated. Thus, there had to be further utility for separated/combined rankings.

- ii. Similar Songs:

- 1. TF-IDF, musical, and combined all had different Python types, so each had to be implemented separately. Furthermore, this had to be the case, as TF-IDF and musical features are both helpful for finding similar songs.
 - 2. Another challenge was how to balance these two feature types. Right now, musical features are much more relevant, but this is not completely the case with similarities in songs.

- iii. Sentiment Analysis:

- 1. Due to the lyrics dataset not being too extensive, there was overfitting without tokenizing each query. Furthermore,

incorporating Metapy along with Sklearn was a big challenge for this model.

2. The other main challenge revolved around the time it takes to run; the model is somewhat slow (it can take minutes with tokenization). This was already mitigated somewhat, but is still a challenge.

iv. Utility Methods:

1. Preprocessing was challenging due to the Kaggle dataset containing faulty or non-English data. Furthermore, the sheer size of a lyrics dataset yielded problems with efficiency.
2. Song searching was hard due to data being stored in dat/csv datasets, so I had to make a separate class for easier utility.

b. Flask/HTML Challenges:

- i. Flask initially had issues with Metapy, which led to changes with threading (modeling is slower as a result).
- ii. Web pages are relatively cluttered and could be improved upon; however, I found it easier to make many separate web pages rather than combining multiple pages.

4. Pictures of the website and results:

Home Page

The Spotify Feature App

We have three different functionalities revolving around Spotify songs:

Search Engine

Sentiment Analysis

Similar Songs

Search Engine Query Page

Query Search

Type in a query name to retrieve the top 25 songs

If you do not want separated rankings, leave the weights empty
If any weight is left empty, rankings will default to combined rankings
Please round your weights to the hundredth place

Query:

Separated Ranking Weights:

Search Engine Results

Top Songs

Here are the top 50 songs that match your query

Song Name	Artist Name	Computed Score
Full Circle	Five Finger Death Punch	9.929620742797852
Breathe	The Prodigy	9.274965286254883
Russian Roulette	Rihanna	8.50214958190918
BLACK EFFECT	The Carters	8.313254356384277
Reach (NBC Olympic Version)	Gloria Estefan	7.985852241516113
Just Say Yes	Snow Patrol	7.884729385375977
Work	Chris Lorenzo	7.649186611175537

Sentiment Analysis Search Page

Sentiment Analysis Song Search

First, type in your query and choose title or artist to indicate your search choice

Query:

Search Choice:

Sentiment Analysis Song Selection

Songs List

Here are the songs that match your query

you want to tokenize the lyrics dataset (Takes longer but less prone to overfit). Furthermore, pick a random state longer than a minute to run.

Song Name	Artist Name	Song Selection
ELEMENT.	Kendrick Lamar	<div>No Tokenization Random Model State</div> <div>Select Song</div>
King Kunta	Kendrick Lamar	<div>No Tokenization Random Model State</div> <div>Select Song</div>
Money Trees	Kendrick Lamar	<div>No Tokenization Random Model State</div> <div>Select Song</div>
Swimming Pools (Drank)	Kendrick Lamar	<div>No Tokenization Random Model State</div> <div>Select Song</div>

Sentiment Analysis Results

Sentiment Analysis Results

contains the predicted sentiment score versus the actual score. This also contains the random state and model

Predicted Score	Actual Score	MSE	Random State
0.40500011191361085	0.405	8.77167263557751	70203

This table contains the most significant words in your song.

Word Lyric	Significance Score
cold	0.24459773751142017
there	0.21627248023650797
outside	0.2075890304351032

Similar Song Search

Similar Song Search

First, type in your query and choose title or artist to indicate your search choice

Query:

Search Choice:

Submit

Back

Similar Song Selection

Songs List

Here are the songs that match your query

g a song, choose a weighting option within that row for the similar song search (Note: this may take i

Song Name	Artist Name	Song Selection
Echoes Of Silence	The Weeknd	<div>Only TF-IDF ▾</div> <div>Select Song</div>
Same Old Song	The Weeknd	<div>Only TF-IDF ▾</div> <div>Select Song</div>
Tell Your Friends	The Weeknd	<div>Only TF-IDF ▾</div> <div>Select Song</div>

Similar Song Results

Top Songs

Here are the 25 most similar songs

The first song listed is your chosen song

Song Name	Artist Name	Computed Distance
Can't Feel My Face	The Weeknd	0.0
Most Girls	Hailee Steinfeld	7.686529190920055e-11
Cold Water	Major Lazer	8.258005390615608e-11
Never Go Back - Robin Schulz Remix	Dennis Lloyd	9.440681569827802e-11
Let You Love Me	Rita Ora	1.2903023094423816e-