CS 410 Group Project Proposal

What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

Team Captain: Maya Subramanian - mayass3
Meng Mu - mmu3
Tayo Amuneke - amuneke2
Jessica Nwaogbe - nwaogbe2

What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?
We want to perform sentiment analysis in the Spotify music library using Spotify/Genius API in order to create "smart" playlists for listeners.
We plan to get song information from the Spotify API and lyrics from the Genius API and store this data in a key, value NoSQL database (i.e. MongoDB). We will process and clean these lyrics to remove stopwords. We will then perform a sentiment analysis, likely using the scikit-learn sentiment analysis library in Python, on the lyrics to give each lyric a sentiment score (positive vs. neutral vs. negative).

DB Example with fake data: **Key:** 'Single Ladies' **Value:** {'single': 0.5, 'ladies': .08}

In the frontend, we will prompt the user to enter their current mood/feelings. Then, we will run sentiment analysis just as we did on our song lyrics on the user-entered phrase, which would assign sentiment scores to each of the words in their phrase. We will then compare our song lyrics' previously stored sentiments with the user-entered phrase's sentiment scores in order to give the user a playlist relevant to their mood of songs with the closest sentiment score match.

We expect to give the user a 10-song playlist that is well-matched to their mood. Our goal is to return this playlist so the user can click on an embedded Spotify player and listen directly from the web app.

We will utilize explicit and implicit feedback to evaluate our web app. We will use explicit feedback by giving the user checkboxes after each song in the playlist with the prompt "Does this song fit your mood?" and "Yes" or "No" as options. Based on this feedback, we will tweak our sentiment analysis parameters to improve the results.

Which programming language do you plan to use?
Python, JavaScript

Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.

We have 4 team members, so our expected workload time is ~80 hours.

Tasks include:

Creating the song database: ~20h

- Fetching song data from the Spotify API (limited number of songs, limited to English & other parameters)
- Fetching lyric data for these songs from the Genius API
- Clean the lyrics (i.e. remove stopwords, spaces, etc)
- Perform sentiment analysis on the song lyrics and store these as key, value pairs

Handling user input: ~20h

- Perform sentiment analysis on the user-entered phrase and put the words as keys and sentiment scores as values in a nested dictionary
- Pass this dictionary to a function which will match the phrase's sentiment scores with the top-ten best matched songs from our song database

Handle user feedback: ~20h

- Based on user feedback, if the user clicks "No, the song does not fit my mood" for a particular song, we can adjust parameters and re-run our sentiment analysis for that particular song in our database and store this updated value

User interface: ~20h

- Create the input box for user to enter phrase
- Either embed Spotify web player or create our own web player on web app with returned playlist
- Implement user feedback interface with checkboxes next to each song in playlist