Project Proposal

1. 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.
   Matthew Walowski mlw6@illinois.edu Captain
   Zoheb Satta satta2@illinois.edu
   Jiacheng Guo jg17@illinois.edu

2. 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?

   Topic (Free Topic):
   Our application will have the user rank a set of movies/tv shows in their preferred order (favorite to least favorite). We will then use sentiment analysis on publicly available review data for the provided movies/tv shows to predict how the average person would rank these movies. From this, we can inform the user how unique their movie tastes are. I.E. we can tell the user if they ranked the movies similarly to the public.

   Interest:
   Making an analysis of the sentiment of these movie reviews could help people know the general category of these movies in advance. It will help people to judge if they would like to watch this movie and choose their own favorite movie.

   Approach\Tools:
   We will utilize The Movie Database API(https://developers.themoviedb.org/) to get review information about each of the movies. To perform sentiment analysis we will use an existing python library (NTLK https://www.nltk.org/howto/sentiment.html). We will experiment with different modes of feedback for the user (how do we express the similarity between their ranking and the general ranking).

3. Which programming language do you plan to use?
   Python backend to do all the data processing, javascript/html (react or angular) for frontend.

4. 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.

Workload:

Scraping API data - 10 hours

- Learn specifics about API usage
- Get movie ID from movie title
- Get reviews from movie ID
- Download and parse movie reviews

Sentiment Analysis - 15 hours

- Given a list of reviews, perform sentiment analysis on all
- Aggregate results
- Output ranking of movies

Similarity Measure - 10 hours

- Take a list of user rankings and list of generated rankings
- Output a numerical representation of similarity
- This requires a bit of experimenting so it will take extra time, we test multiple techniques to find a suitable solution:
    - Ex: Try MAP, gMAP, simple difference, VSM,...

Web Interface - 30 hours

- Build a web interface to interface with the user
- Problems to solve:
    - How to make it intuitive for users?
    - How to communicate between python backend and javascript frontend

Total Hours: 65 hours