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.**
    - Team Name: Grade Simps
    - Team Members: Yiting Zhao (yiting9), Matthew Do (mmdo2), Gautam Samudrala (gautams4), Ansh Bhalla(anshb2)
    - Captain: Matthew Do
2. **What topic have you chosen?**
    a. Free Topics - Reddit Subreddit/User Sentiment Analyzer Web App
3. **Why is it a problem?**
    a. It is time consuming for people to browse through streaming posts to find if they are positive or negative. Curious browsers of Reddit can get an sentiment analysis of their favorite communities and a specific user quickly.
    b. For example, there are a lot of reddits in UIUC_MCS, by tracking the community sentiments, we can easily understand if people are still satisfied with the program.
4. **How does it relate to the theme and to the class?**
    a. The second half of the course is related to text mining and covers sentiment analysis. We will use what we learned from course to inform feature building and sentiment prediction.
    b. We also plan to use the ideas in Lecture 12.6: Mining Topics with Social Networks. As Reddit is a large social network platform, the lecture information is helpful to our project.
5. **Briefly describe any datasets, algorithms or techniques you plan to use.**
    a. For dataset, we will be consuming data through the Reddit API
    b. For building features, we plan to use tools such as pattern discovery and topic mining. We will also try NLP pretrained models (e.g. BERT) to generate meaningful representations.
    c. For sentiment analysis, we will use ordinal logistic regression and research other models.
    d. For visualizing the sentiment trend, we will be using Plotly to create interactive figures.
6. **How will you demonstrate that your approach will work as expected?**
    a. We will present a live recorded demo of our application. We will guide the viewer on how our web application could be used. We will add and remove filters and parameters showing that our web application is responsive and robust. We will also demonstrate the value that our web application provides to Reddit users.
7. **Which programming language do you plan to use?**
    a. Python
8. **Main tasks to be completed, and the estimated time cost for each task.**
    a. Reddit API Investigation (15 hours)
        i. Endpoint Discovery and Testing API Capabilities (10 Hours)
        ii. Data Preprocessing - JSON to CSV (5 Hours)
    b. Sentiment Analysis Research (40 Hours)
        i. Text Preprocessing and Feature Engineering (15 Hours)
        ii. Comparison and Experimentation of different algorithms (20 Hours)
        iii. Visualization of Sentiment Trend (5 Hours)
    c. Front-End (35 hours)
        i. User Interface Design and Data Visualization (25 Hours)
        ii. Connecting Backend to Front End (10 Hours)