Final Project Proposal for CS 410, Fall 2022

Project Github: https://github.com/rylandtikes/cs410-course-project

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 - Mountain Group
Binod Pandey   bpand3@illinois.edu
Charles Stolz (Captain)   cstolz2@illinois.edu
Michael Gambino   mgambin2@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?

We will perform Sentiment Analysis on Reddit comments in UkrainianConflict subreddit. Our plan is to create a Reddit application to facilitate the scraping of comments from UkrainianConflict and then connect to the application using the Python PRAW library. After we scrape the data we will use the NLTK Vader Sentiment Analyzer Python library to label it. Next we will preprocess the data using tokenization and stopword removal. After this we will perform additional analysis steps following Sentiment Analysis best practices until we have our final results. The results will be displayed on a Plotly Dash app deployed through Heroku. We will automate the pipeline and site creation displaying the results, including trends throughout the project. Tracking public opinion of a war is important, especially understanding how that opinion changes based on major events. In this project, we define the positive sentiment denoted by some threshold with NLTK Vader as support of actions pertaining to the war or a certain side, and negative sentiment as opposition of actions pertaining to the war or a certain side. With that defined, we will correlate both the magnitude of score and frequency over time with major events that had occured as a way to gauge the public opinion, i.e did the sentiment around this event correlate more for support of Ukraine (positive) or opposition to Russia (negative). For the result, we expect an inherent bias towards negative sentiment given that the data is collected from western social media in English mostly talking about the atrocities of war, but do also expect some rise in positive sentiment, mostly pertaining to tangible events of Ukrainian advances and victories. We will evaluate the work based on other similar projects, research papers, and feedback from our peers.

3. Which programming language do you plan to use? Python

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.

- 20hrs - Create a tool to pull and transform the data (Python)
- 20hrs - Create tool for Sentiment Analysis (Python)
- 20hrs - Create an app to display the results (Python with Plotly Dash and Heroku)