CS 410: Final Project Proposal

Team Stanley (Fall 2021)

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.

   Name: Stanley Ho
   NetID: smho2@illinois.edu
   Team: Stanley

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?

   My free topic is **Brand Sentiment on Twitter using Sentiment Analysis**. More specifically, sentiment analysis can capture the market or customer sentiment towards a brand. Companies can use this information to better understand their audiences' reactions to the brand's news or marketing campaigns, and to further enhance the brand. Investors or traders can also leverage this information to determine whether they should long or short their positions in the stock behind the brand.

   The task is to perform sentiment analysis on the Twitter tweets related to a given brand over a period of time, and create a sentiment trend graph which visualizes the sentiment towards the brand.

   My approach is as follows:
   a. Generate a dataset using the Twitter API to collect the tweets related to a given brand over a period of time.
   b. Perform sentiment analysis on the dataset using one of the public cloud's NLP services (e.g. AWS Comprehend, Google's Natural Language API, etc.).
   c. Based on the result from the sentimental analysis, create a sentiment trend graph for the brand.
   d. To evaluate the result, I plan to pick a brand which had some significant sentiment shift over the past few months. I also plan to collect significant news or events related to the brand over the same period of time. By analyzing the sentiment trend graph, we should be able to link any significant shift in the graph with the news or events.

3. Which programming language do you plan to use?

   Java or Golang.

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.

   a. Write a program to generate a dataset by collecting the tweets from Twitter API for a given brand over a period of time. (5+ hours)
   b. Write a program to submit the dataset to one of the public cloud's NLP service for sentimental analysis. (7+ hours)
   c. Create a sentiment trend graph based on the result of the sentimental analysis for data visualization. (4+ hours)
   d. Collect recent brand's news or events. (1+ hours)
   e. Analyze the sentiment trend graph, and link any sentiment shift with the recent news or events. (3+ hours)