Project Proposal: CS 410
Topic: Sentiment Analysis
Team Canyon

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.

- Amrutha Gujjar   agujjar2@illinois.edu
- Balakumar Balasubramaniam   bbalasub@illinois.edu
- Himanshu Pandey   hpandey3@illinois.edu

    Himanshu Pandey is the Captain

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 Name: Sentiment Analysis

The task is to capture the sentiments on the data. For eg: In current Presidential Campaign, we read the news through different channels including social media. Streaming the social media data like Twitter, can give a dataset on the topic "Presidential Campaign" and then we will process the data to find and evaluate the sentiments of the users on both the Presidential Campaign.

Dataset: We will be using the twitter data for this project.

Tools: Scrapping/Streaming the twitter data can be done through R or Python in Real Time.

Evaluation: Precision/Recall Measures

Which programming language do you plan to use?

Python

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.

The task is divided into following sub tasks:

- Data streaming and cleaning. We are planning on stream the data at different times from the Twitter and once we have 30-40k rows, we can start cleaning the data.
  8 hours

- Data cleaning and Analysis
  6 - 8 hours

- Writing the code to train the data
  20 hours

- Test and refine
  8 - 10 hours

- Improving the model accuracy
  6 – 8 hours

- Evaluation of the test results
  6 – 8 hours

- Generating charts
  6 – 8 hours