CS 410 Project Proposal

Team: Working Alone v2

Member/Captain: William Shih (ws21)

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 about sentiment analysis on Twitter with data visualization. The Twitter API will be used to gather tweets and sentiment analysis will be performed. The user will be able to input a username or search topic as well as number of tweets (up to a maximum) and a range of dates. Then the results of the sentiment analysis of the tweets will be displayed on an interactive web page. The planned visualizations include an interactive line graph of the distribution of sentiment analysis scores and an interactive bar graph for most common words used in positive and negative tweets. Previous inputs will be remembered, and additional lines can be added to the line graph to compare the sentiment analysis of different inputs. Also, the sentiment analysis results will be able to be downloaded.

The task is to provide an interactive web page to visualize sentiment analysis on Twitter. It is important or interesting because it allows a user to compare the sentiment between different users, different topics, or the same user with different time periods with an interactive visualization.

The planned approach is to use a Python package and Twitter API to obtain the Tweets. Then, the NLTK package in Python can be used to clean the data and perform sentiment analysis. Then, the Flask package in Python and d3.js in JavaScript can create the data visualization.

The expected outcome is a website where a user can input a query to obtain the Tweets and a data visualization of the sentiment analysis results and download data button of the sentiment analysis will be displayed on the screen. NLTK package provides an annotated Twitter corpus to evaluate the training model on.

The programming languages used will be Python and JavaScript.

Workload (minimum 20 hours):

Learn how to use Twitter API: 5 hours

NLTK for Sentiment Analysis: >10 hours

Flask and d3 for Data Visualization >15 hours

Total: >30 hours