CS 410 Progress Report

Our project is related to sentiment analysis of twitter users. As for **progress**, we have finished going through the sentiment analysis lectures on Coursera to get a better understanding of what methods are available for opinion mining/sentiment analysis. Furthermore, we have created an outline of what we want our program to look like. We have compiled a document that discusses sample input, sample output, and which methods we need to implement for our program to work. We have researched how to use the Python tweepy API for our specific use case, and we have read online tutorials for sample sentiment analysis projects. Additionally, we have read some articles about different ways to train a classifier for sentiment analysis, specifically about the Naive Bayes Classifier.

We created messages that will be outputted in the terminal to direct the user on how to use the program. The program currently has the functionality to take a twitter handle and analysis type (personal, mentions, liked, and overall) arguments. We have also made logic to collect the tweets tweeted by the inputted twitter username (personal), the user's liked tweets (liked), and tweets that mention the user (mention) depending on the analysis type the program user wants to run for sentiment analysis. "Overall" means it uses all three types of tweets to analyze the sentiment of the inputted twitter handle.

We still have the following **remaining tasks**. We still need to preprocess the raw tweets to standardize and tokenize them. Some tweets could contain unwanted symbols or emojis that need to be properly handled for accurate sentiment analysis. We need to collect a training dataset of tweets using the tweepy API and train the Naive Bayes Classifier model. Additionally, we need to run the Naive Bayes Classifier model on the collected tweets for the inputted twitter handle and classify them as positive or negative. Then, we can use the number of positive and negative labels to give a score out of 100 for the inputted twitter username and output that in the terminal.

We have faced some **challenges/issues**. It has taken us some time to gather useful documentation for tokenizing and normalizing tweets to make sure accuracy can be increased when classifying tweets as positive or negative. It has been taking time to read through all of these tutorials to determine which ones would be the best to follow without it being too complicated.