

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 still have the following **remaining tasks**. We need to collect a training dataset of tweets using the tweepy API, train the Naive Bayes Classifier model, and collect tweets that mention the inputted Twitter user, tweets that the inputted Twitter user liked, and tweets that the inputted Twitter user tweeted online to score sentiments about the user.

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.