CS 410 Project Proposal

There are three people in our group. The members are Chandrachur Kesana (netid: ckesan2), Lloyd Quadros (netid: quadros2), and Muhil Arumugam (netid: muhila2). The team captain will be Chandrachur Kesana (ckesan2).

Our free topic is sentiment analysis for Twitter users. We plan to collect a number of tweets using Python's tweepy API and then use the nltk library to perform sentiment analysis on twitter users based on these tweets. The project will be a terminal based application where the user can run the script and input a twitter user's username as an argument. The output would be deciding whether the perception of the twitter user is positive, negative, or neutral.

We are planning on using the NaiveBayesClassifier to train our model on tweets. To train the model, we will pull a sample of pre-classified tweets as our training data set. We would run our trained model on a sample of Tweets that mention the Twitter user that was inputted into the terminal. From there, every time the sentiment analysis is finished running, we can display the tweets that are classified as positive, negative, or neutral for that Twitter user who was mentioned. From there, we can evaluate how well the model is classifying tweets about a given username since application users can decide whether a tweet about a Twitter user should have been positive, negative, or neutral after seeing the results.

This topic is important because it gives information to people about the perception of a Twitter user. They can decide whether that person is a credible source or not or if they are someone who is beneficial/detrimental to be associated with in the public.

We have 3 people in our group so N = 3. These are the tasks that need to be completed for the project: **Go through Sentiment Analysis Lectures (3 hours), Research about tweepy, nltk, and different classifiers to train a model (10 hours), Collect training data set of tweets (15 hours), Clean up data (normalizing/tokenizing data) (15 hours), Setting up a classifier to model as positive, negative, or neutral (5 hours), Grab tweets that mention the username inputted into the terminal (10 hours), Run model on these sample tweets and output sentiment of the user (5 hours).** For all of these tasks combined, we expect to spend a total of 63 hours on the project, which is greater than 20*3 = 60.