Rakesh Singh

ISTM 6200

Professor Yixin Lu

Spring 2021

April-20, 2021

**Team Assignment 1: Documentation**

**Project Objective:**

The program will analyze the data extracted from Twitter leveraging Tweepy API and will implement analysis to find out the result where a stock/crypto is trending positively, negatively or neutrally in the public domain. This will allow the customer to take a decision to buy, sell or hold according to the program of python.

**TweePy:**

This project involves implementing the TweePy RESTful API class. The goal of using this API class is to provide a centralized, parameter-based approach to call and scrape tweets off of twitter.

**Initialization:**

After importing the required modules, the “OAuth” handler is called and assigned to a variable object. This handler is required for user account authentication. The handler takes in two arguments as part of its function: “consumer key ID” and “consumer secret key”. These two go hand in hand in performing a full-fledged authentication over a secure channel. The program will take the user to the twitter developed website for authentication and the 6 digit code generated from Twitter app needs to be keyed into the program for final authentication.

**Retrieval of access token:**

On passing the two arguments along with the ‘OAuthHandler’, we receive an access token to access Twitter, programmatically. This redirects a user to a ‘session’ which happens after the user inputs a secret pin. This step is carried out in the backend in line with the OAuth protocol.

**OAuth protocol:**

OAuth is a standardized authentication protocol used for third-party client access over HTTP channel to retrieve access tokens as part of user authentication. These access tokens can be retrieved to access the desired information from an endpoint, in this case, Twitter. Generally, OAuth is preferred for accessing live streams of feeds such as RSS and Twitter due to its dynamic nature of browser based URL redirection.

**Reading Tweets:**

After gaining access to the tweets’ stream, a “Cursor” method is called by the API class. Cursor is the tool to perform pagination on web pages or twitter feeds in this case. Pagination involves parsing and processing of text based information and selecting objects. Cursor helps us “get” all items present in that specific twitter page. The text is displayed on the program to perform sentiment analysis and appended to a text file in the local system for creating visual dashboards.

**Sentiment Analysis:**

User can enter a keyword to search the stream of text obtained by the cursor. The word is stored in terms of “positive” or “negative” for assigning it the appropriate sentiment. Positive and negative in this case are empty arrays which take in this word and a method called “Prediction” is called, which classifies the word and detects its intent. To detect its intent, the word’s polarity is measured by assigning another empty array, in addition to “positive” and “negative” attributes. There is a python dictionary used to perform the analysis and the words for “positive” and “negative” are determined by the dictionary.

**Polarity Scores:**

The negative, positive and polarity values are determined by calling the “Polarity\_Scores” method. This method assigns scores for these attributes by filling the pre discussed empty arrays. The tweets are checked against positive, negative or neutral attributes by the measure of percentages and are weighed accordingly.