

## CS 410 Final Project Documentation

Here's an **overview of the function of the code** (what it does and what it can be used for). Our project is a terminal based application that performs sentiment analysis of Twitter users as positive, negative, or neutral. This type of code can be used to give information to people about the perception of a public 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.

Here is **how the software is implemented**. The user can input a public Twitter handle and type of analysis to be used for sentiment analysis. In our application, there are 4 options: "personal", "mentions", "liked", "overall". If "personal" is picked, the inputted user's tweets will be collected. If "mentions" is picked, tweets that mention the inputted twitter handle will be collected. If "liked" is picked, tweets that the user liked are collected. If "overall" is picked, all 3 of the categories will be collected. For the first three categories, 200 of the most recent tweets will be collected. For the "overall" category, 200 each of the other categories will be collected for a total of 600 tweets. These tweets are represented as a list of strings.

Once the tweets are collected, we clean up the tweets by removing special characters. After that, we got the sentiment for each tweet using the textblob library. When we pass a tweet into the TextBlob to create a TextBlob object, it cleans the data further by tokenizing the tweets, removing the stop words, and does POS tagging. We then check the sentiment polarity to classify tweets as positive, negative, or neutral. TextBlob uses a dataset that has positive and negative labels for the data. Internally, the Naive Bayes Classifier trains the data. At the end, we calculate the percentage of positive, negative, and neutral tweets collected for the user and output these percentages and the overall sentiment (positive, negative, or neutral) of the user.

Here are some **instructions on how to run this code**. These directions assume you have python 3 installed on your computer:

- 1) Clone the repo from Github by running **git clone** <https://github.com/ckesan2/CS410CourseProject.git> in the terminal
- 2) Run **pip3 install tweepy** in the terminal to use the Twitter API
- 3) Run **pip3 install textblob** in the terminal to be able to use the textblob library
- 4) Run **python3 -m textblob.download\_corpora** in the terminal
- 5) **cd into the directory** that contains the **project.py** file using the terminal or open the directory in **VS Code**.
- 6) Run **python3 project.py** on the terminal
- 7) Enter a valid **public** Twitter handle, analysis type. Analysis type options are "personal", "mentions", "liked", and "overall".

An example would be: **@KDTrey5, personal**

At this point if you get an error message that says forbidden access or token has expired, these steps must be followed to create a Twitter Developer account and access the tweets:

- 1) Go to this link: <https://apps.twitter.com/> and press “Create New App”
  - 2) Then fill out the application details describing the sentiment analysis usage.
  - 3) Then go to the Project in the dashboard and **request for elevated access** and fill out the details to the questions.
  - 4) Then go to the “Keys and tokens” section of the project.
  - 5) Press Regenerate for the Bearer Token section. Copy the bearer token
  - 6) Paste the bearer token into the **BEARER\_TOKEN variable** in the project.py file.
  - 7) The code will now run properly once the Bearer Token is updated.
- 4) Step 3 can be repeated as many times as the user wants.

If these steps are followed, the overall sentiment of the twitter handle will be shown on the terminal in addition to the percentages of positive, negative, and neutral tweets included in the collected tweets.

Here is a **brief description of the contribution of each team member**. Each of us researched different ways of collecting tweets and training models for sentiment analysis and went through multiple tutorials. We also each went through the optional sentiment analysis lectures in Coursera. Muhil (muhila2) worked on processing the input from the user and gathering tweets for the inputted public Twitter handle and the inputted analysis type. Chandrachur (ckesan2) worked on training the model and calculating the sentiment of each tweet and outputting the calculations of negative, positive, and neutral tweets and overall sentiment to the terminal. Lloyd (quadros2) worked on standardizing and cleaning each of the tweets to prepare them for classification and worked on error handling in case of exceptions. Each of us spent over 20 hours on the project.