

Anamitra Bhattacharyya
PREDICT-452 Sect. 55
Individual Assignment 4, Option 1
Sentiment Analysis

Executive Summary

The aim of this project was to extract tweets from Twitter related to the US Federal Special Prosecutor search term 'Robert Mueller'. Using the Twitter authentication credentials used in Assignment 2, the Twitter stream was used to automate data collection and extract tweets referencing the 'Robert Mueller' search term. The extracted information was parsed for positive, negative and neutral sentiment tweets yielding summary statistics of their relative abundance, and an output of the five examples of positive and negative tweets displayed.

Research design and methods

The strategy used for this exercise was to develop a Python program (using the *tweepy* package) designed to access Twitter data related to mentions of US Special Prosecutor 'Robert Mueller'. Specifically, the Python program (**MineTwitter.py**), which has 3 main sequential steps:

- Authorize twitter API client (using *tweepy* package)
- Make a GET request to Twitter API to fetch tweets for a particular query 'Robert Mueller'
- Parse the tweets, and classify each tweet as positive, negative or neutral (uses *textblob* package)

The Twitter API gets 200 tweet results. The *tweepy* package uses the Twitter API to access and get the tweet data, while the *textblob* package is used mainly to derive the tweet sentiment. It does this using a sentiment classifier to produce a sentiment polarity on a scale from -1 to 1, so that positive, neutral and negative sentiments have scores of 1, 0 and -1, respectively, using the classifier.

Implementation and programming

The **MineTweets.py** program creates and instantiates an object of class `TwitterClient`, which encodes several helper methods to first access the Twitter API (passes authentication credentials), perform a search query (e.g. 'Robert Mueller') passed to it by the main method, then parses the resulting 200

tweets collected by cleaning them of links and special characters (using the *re* regular expressions package), and then categorizes the sentiment using the *textblob* package into positive, negative and neutral. Some summary statistics of the relative percentage of positive, negative and neutral tweets are calculated and displayed, and finally five examples of positive and negative tweets are displayed in the screen output (see below).

Output:

```
In [218]: %run "/Users/anamitrabhattacharyya/Desktop/MSPA/452-DL-55/Assignments/Assignment04/MineTweets.py"
Positive tweets percentage: 34 %
Negative tweets percentage: 10 %
Neutral tweets percentage: 55 %

Positive tweets:
Expect more 'surprises' from Mueller probe, former crusading prosecutor says https://t.co/0pQEqUfkng via @washtimes
RT @NewYorker: The special counsel Robert Mueller has lately been moving so fast that it is becoming difficult to keep track of the intricacies...
RT @nxtompson: I found some interesting news about Rob Goldman, the FB executive who was retweeted by Trump—including his apology today to...
Ken Star Special Counsel Robert Mueller Shows Nothing of Trump Obstruction of Justice https://t.co/J4i2jGlnr6
RT @ABC: The 58 stores in Trump Tower have been home to some controversial residents—including several Russia-linked figures who were later...
Democracy in action to expose the truth that our nation do have checks and balance a true patriot think you Robert... https://t.co/LeuJP24LW3
Gojo doing more than Robert Mueller to bring down the Trump administration https://t.co/G4b10BiSsJ
RT @axios "RT @alayanatreene: Joining WTOP in the next few minutes to discuss the latest developments in Robert Mueller's Russia investigation"
RT @aroseblush: Robert Mueller issues new indictment against Paul Manafort—asks for forfeiture of homes and accounts https://t.co/Z0m...
I liked a @YouTube video https://t.co/y0g9pmsilj Bruce Springsteen Sings - Robert Mueller's Comin' to Town

Negative tweets:
RT @funder: These people are on Schiff's side: Obama, George W, Robert Mueller, our entire intelligence community, the American military &...
RT @stopthenutjob: Reminder: the only reason the GOP produced the fake Nunes memo to begin with was to discredit the FBI in order to discred...
Now the left will call Starr crazy. Wait and see. https://t.co/nlpwHDCUW
After months and months — given the absence of evidence showing Trump collusion with Russia — what could Mueller's... https://t.co/XBMRwORFLX
@FoxNews @jonathanvswan @realDonaldTrump This is what happens when your ego becomes maniacal. @realDonaldTrump thin... https://t.co/7npjumwqT
```

Conclusions

Automated data acquisition of 200 tweets related to the “Robert Mueller” was successfully obtained in the project, using the Twitter API. The extracted information from the Twitter feed was parsed, cleaned and classified for type of sentiment (*e.g.* either positive, negative or neutral). A summary of the relative proportions of sentiment for the search term, ‘Robert Mueller’ was, positive (34%), negative (10%) and neutral (55%).