

Twitter Sentiment Analysis

A Project Report

Submitted in Partial Fulfillment of Requirements for Sharpest Minds Mentorship
Program

Submitted By – Swaroop Todankar

Date – 14th November 2019

[1]



TABLE OF CONTENTS

	Page
List of Figures.....	iii
1. Introduction.....	1
2. Business Problem.....	1
3. People Interested in the Project – Target Audience	1
4. Data required.....	1
5. Methodology.....	2
5.1 Importing Necessary Libraries.....	2
5.2 Loading Twitter Credentials.....	2
5.3 Setting up Authorization and API objects.....	2
5.4 Defining the required functions.....	2
5.5 Fetching the tweets	4
5.6 Initialize the variables	4
5.7 Performing the total analysis.....	4
5.8 Obtaining the overall sentiments.....	4
5.9 Obtaining total amount of tweets.....	4
5.10 Visualization.....	5
6. Discussion.....	5
6.1 Learnings from the project	5
6.2 Problems faced during the analysis.....	5
6.3 Limitations	5
7. Conclusion.....	6
8. References.....	6

LIST OF FIGURES

	Page
Figure 1. Twitter Sentiment Analysis	i

1. Introduction

The social media platforms such as Facebook, Instagram, Twitter and Snapchat have become a place where people express their opinions regarding anything to which they appreciate or object.

Among these platforms, Twitter is of prime importance – being the easiest and most used platform to express your thoughts in real time. This project was intended to understand the sentiments of people expressing their views on Twitter.

The prime focus revolved around the Canadian Elections and to the way people reacted to some of the candidates. Natural Language Processing (NLP) Techniques were used to obtain the tweets and information regarding the users who expressed their views regarding a particular query. The tools provided by the NLP were used to assign a sentiment to the tweets obtained. The final step involved comparing the results obtained by the two libraries used for analysis.

2. Business Problem

The basic problem is divided into two questions:

1. How do people actually react to particular scandals or discoveries related to the elections?
2. Do the tweets put forth by people (politicians) or influential people bring about a change in the sentiments of people who follow them or among others who strictly don't support them?

This report aims to put forward an analysis of sentiments of twitter users from the tweets obtained from Twitter API –Tweepy, using NLP libraries – NLTK and TextBlob.

3. People interested in the Project – Target Audience

The following can be the target audience of the analysis:

1. Analytics companies who are solely interested to know the trends in the behavior of people.
2. The campaign managers of the politicians who are contesting the elections.

4. Data required

The data required to build a model to perform Sentiment Analysis:

Twitter Data: Data pertaining to the users is obtained using Tweepy API.

5. Methodology

The following steps were employed to obtain the required results:

5.1 Importing Necessary Libraries

The first step is to import the necessary libraries and packages.

- Numpy – For numerical calculations
- Matplotlib – plotting and visualization
- Pandas – Data manipulation
- NLTK – Natural Language Tool Kit library for Sentiment Analysis
- TextBlob – Another NLP library for Sentiment Analysis
- Re – Regular Expressions in order to clean the tweets of punctuations and special characters
- Tweepy – Twitter API library to fetch tweets

5.2 Loading Twitter Credentials

This step involves the loading of a python file consisting of private twitter developer account credentials. This is a required step to fetch the tweets.

5.3 Setting up Authorization and API objects

After the credentials are loaded the API object is created to be used for fetching the tweets depending upon the query and number of tweets required.

5.4 Defining the required functions

The following functions were created:

Fetch tweets – Using the API object to fetch tweets and filter retweets. Takes input as the search query and number of tweets required.

Clean tweets – The function takes the tweets stored in public tweets and removes the punctuations and special characters using regular expressions.

Sentiment NLTK – This function obtains the sentiment of the tweet passed using the NLTK library.

Sentiment TextBlob - This function obtains the sentiment of the tweet passed using the TextBlob library.

Display details – The function is used to display details related to the user who tweeted the tweet.

The details displayed are as follows:

- Tweet text , timestamp, user-id, screen name of the user, location, number of followers, number of tweets by the user, number of retweets on the tweeted tweet, sentiment predicted by NLTK and TextBlob and the user weights calculated by NLTK and TextBlob.

Tweet scores – This function assigns a user particular score depending upon the number of tweets put forth by the user.

Follower scores – This function assigns a user particular score depending upon the number of followers the user has.

Retweet scores – This function assigns a score to the user depending upon the number of retweets the tweet tweeted by the user obtained.

Weight Assignment Blob – This function calculates the user weights for Text Blob analysis.

This function has 3 parts depending upon the polarity –

- Calculate the number of positive, negative and neutral tweets according to TextBlob
- Calculate the user weight = (Tweet score + Follower score + Retweet score)/3
- Calculate the total weight = Calculate the total weight of all users for the analysis

Weight Assignment NLTK – This function calculates the user weights for NLTK analysis.

This function has 3 parts depending upon the scores –

- Calculate the number of positive, negative and neutral tweets according to NLTK
- Calculate the user weight = (Tweet score + Follower score + Retweet score)/3
- Calculate the total weight = Calculate the total weight of all users for the analysis

Overall sentiment Blob – This function calculates the overall sentiment according to TextBlob.

This function has 2 parts –

- Display the results of Text Blob analysis – Total positive and negative weights and the number of positive, negative and neutral tweets
- Overall sentiment – To display the overall sentiment according to Text Blob analysis

Overall sentiment NLTK– This function calculates the overall sentiment according to NLTK.

This function has 2 parts –

- Display the results of NLTK analysis – Total positive and negative weights and the number of positive, negative and neutral tweets
- Overall sentiment – To display the overall sentiment according to NLTK analysis

Total Tweets – This function calculates the total tweets analyzed for a particular query.

Visualize Blob – This function is used to visualize the entire analysis using Text Blob in the form of a Pie Chart.

Visualize NLTK – This function is used to visualize the entire analysis using NLTK in the form of a Pie Chart.

5.5 Fetching the tweets –

This step involves the input of the query to be searched and the number of tweets to fetch.

5.6 Initialize the variables –

This step involves assigning counters to zero. The counters in this case are positive tweets, negative tweets, neutral tweets, user weight, and total weight negative and positive for both NLTK and TextBlob.

5.7 Performing the total analysis –

This step involves calling the above defined functions, obtaining the tweets and the details regarding it.

5.8 Obtaining the overall sentiments –

This step involves calling the overall sentiments function and obtaining the sentiment based on TextBlob and NLTK results.

5.9 Obtaining total amount of tweets –

This step involves calling the total tweets function and obtaining the total number of tweets.

5.10 Visualization –

This step involves calling the visualize functions and obtaining the pie charts for the analysis using both the libraries.

6. Discussion

6.1. Learnings from the project –

The following tools and things were learnt during the entire analysis –

1. Twitter API – Tweepy hands on experience
2. Google Collab – Creation of credential files and loading python files
3. Creation of functions and calling them – In order to simplify the code

6.2 Problems faced during the analysis –

1. The initial problem faced was inclusion of twitter credentials for the developer account. It was required that the credentials remain private to prevent misuse when uploaded on github.

This problem was solved by creating a python file consisting only credentials and then loading it into google collab. The credentials are read in the back end and are not mentioned anywhere in the script.

2. Initially the analysis was done using just TextBlob, but as the project progressed, analysis using NLTK was included.

Both of these libraries give the sentiment of the tweet passed to it but the scoring system is different. The difference is that TextBlob provides polarity while NLTK provides score. In order to find a common scoring system the score functions were defined and the analysis was completed.

6.3 Limitations -

Although it is not possible to manually label each and every tweet with its associated sentiment i.e. Positive, Negative and Neutral, a small sample of the tweet was selected. It was found that most of the tweets were labelled according to the values provided by lexicons of each libraries. But the libraries failed to take in to consideration sarcasm. This presented a problem statement for future work where the problem statement could be framed like –

How can these libraries take in to consideration the effect of sarcasm on sentiment?

7. Conclusion

Using Tweepy, tweets related to the query were fetched. Required functions were created and called. The details for the tweets were obtained and the analysis was performed.

8. References

1. Analyticsvidhya, "Twitter Sentiment Analysis," [Online]. Available: <https://www.analyticsvidhya.com/blog/2018/07/hands-on-sentiment-analysis-dataset-python/>