

CSCI 5408, Summer 2018
Assignment 2: Sentiment Analysis of Twitter Tweet Data using ETL
Process and Elasticsearch

Submission date 07-06-2018

Lakshmi Maligireddy
Manojha Panneerselvam

B00773792
B00783628

GitHub Link: <https://github.com/Manojha18/Assignment2CSCI5408.git>

Table of Contents

S. No.	Title	Page No
1	Objective	2
2	Task Description	2
3	Twitter Tweet Extraction	2
4	Sentiment Analysis Algorithm	3

OBJECTIVE:

- To learn extracting data from social media platforms.
- To learn extract, transform and loading process.
- To learn advanced data management process using NoSQL DB.
- To learn sentiment analysis and basic machine learning techniques.

TASK DESCRIPTION:

The objective is achieved by creating a personal twitter account and twitter app. Twitter developer API is used to extract the tweets. In order to achieve this, we need to have customer key, customer_secret_key, access_key, access_secret_key. We also created new EC2 instance and installed Elasticsearch as in assignment 1. Once the tweets are extracted from twitter using twitter API it is preprocessed, tokenised, tagged and compared with sentiwordNet. Later the tweets are classified into positive,negative and neutral.

TWITTER TWEET EXTRACTION:

- Creating a new App in Twitter
- While creating the App, collect Consumer Key and Consumer secret and create access token.
- Thus, Access token and Access token secret are generated.
- Python and tweepy are imported in Ubuntu. Later a python script is written to extract President Donald Trump tweets
- Tweets are preprocessed and kept ready for tokenization. In preprocessing we actually removed special characters and was trying to clean the tweets.

Figure (a) realDonadTrump tweets

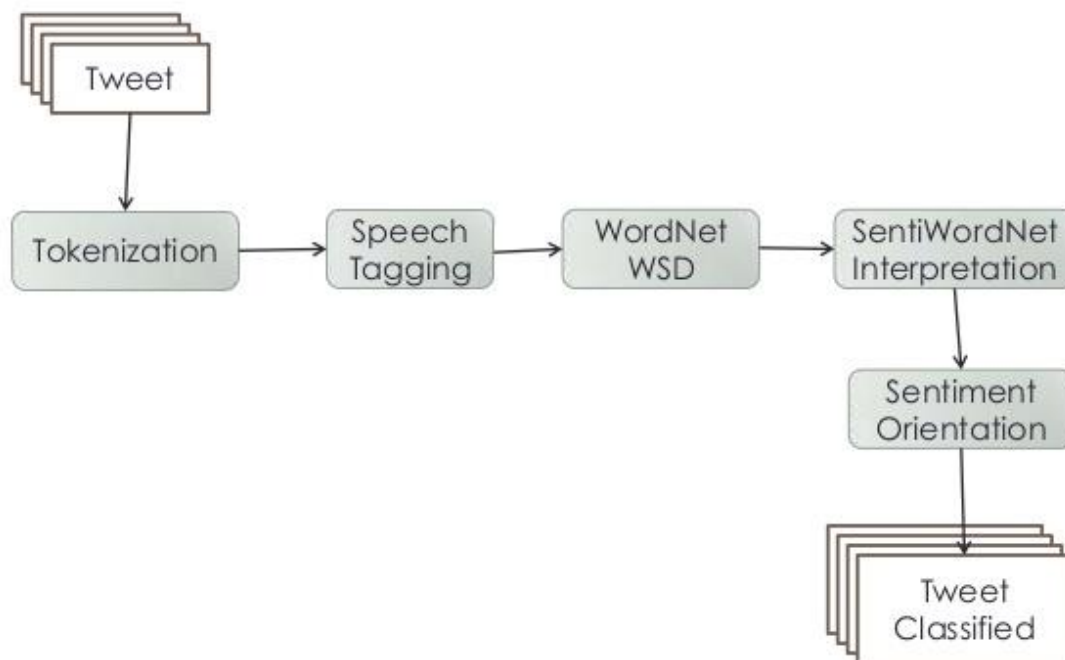
Figure (b) preprocessed tweets

```

If word == not
    T_positive = T_positive + 0
    T_negative = T_negative + 16
    Count = count + 1
Else
    T_positive = T_positive + 16
    T_negative = T_negative + 0
    Count = count + 1
Step - 4: After calculating the score of positive and negative
If T_positive > T_negative
    Print "Positive"
Elif T_positive < T_negative
    Print "negative"
Else
    Print "neutral"
Step - 5: Print the result as a column in csv file

```

Sentiment Classification Phases



In order to perform sentiment analysis, we imported NLTK a natural language processing toolkit.

We performed tokenization and separated the tweets with commas, later after extracting the tweets, it is tokenized and compared with sentiwordnet and then the tweets are classified. For word comparison we used SentiWordNet to perform analysis.

Tweets	Classification	Confidence
b'RT @realDonaldTrump: The HISTORIC Rescissions Package we\	neutral	0.436
b'RT @PluckyQ416: Karma has a new name.\n@realDonaldTrump wants to test	neutral	0.525
b"@realDonaldTrump You need to cut the budget for Congress. It's wildly out of	neutral	0.502
b'RT @realDonaldTrump: The HISTORIC Rescissions Package we\	neutral	0.436
b'RT @realDonaldTrump: Dennis Rodman was either drunk or on drugs (delusio	neutral	0.387
b'@ManInTheHoody @MyMysticMuse @realDonaldTrump He is a freaking trait	negative	0.445
b'@debra12862 @catherinemccoin @TomGardnerIII @realDonaldTrump Bottis	neutral	0.434
b'@real_farmacist @realDonaldTrump Hahahha you lost !!!!!'	negative	0.642
b'RT @TheLastRefuge2: 2. Mitch McConnell is simply trying, yet again, to avoid	neutral	0.522
b'@realDonaldTrump I heard there was a guy who took a knee to your mandato	neutral	0.538
b'RT @stopthenutjob: @realDonaldTrump The Cost of Bigotry: tourism to US in	negative	0.436

Figure (c) After the tweets are classified

References

- [1]“Tutorial of Sentiment Analysis”, *Slideshare.net*, 2018. [Online]. Available: <https://www.slideshare.net/faigg/tutotial-of-sentiment-analysis>. [Accessed: 07- Jun- 2018].
- [2]“sanju1920/Sentiment-Analysis-Using-Sentiwordnet-in-Python”, GitHub, 2018. [Online]. Available: <https://github.com/sanju1920/Sentiment-Analysis-Using-Sentiwordnet-in-Python/blob/master/idoctor.py> [Accessed: 07- Jun- 2018]
- [3]“SentiWordNet”, *Sentiwordnet.isti.cnr.it*, 2018. [Online]. Available: <http://sentiwordnet.isti.cnr.it>. [Accessed: 07- Jun- 2018].