Proceedings of the International Conference on Research in Management & Technovation pp. 63–67
DOI: 10.15439/2020KM20
ISSN 2300-5963 ACSIS, Vol. 24

Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing

Ditiman Hazarika[1], Gopal Konwar[1], Shuvam Deb[1], Dr.Dibya Jyoti Bora[2]
[1]*Student, School of Computing Sciences(IT), The Assam Kaziranga University, Jorhat, Assam, India*
[2]*Assistant Professor, School of Computing Sciences(IT), The Assam Kaziranga University Jorhat, Assam, India*
E-mail: cs19mscit003@kazirangauniversity.in, cs19mscit005@kazirangauniversity.in,
cs19mscit001@kazirangauniversity.in, dibyajyotibora@kazirangauniversity.in

***Abstract*—The Internet has become an innovative platform regarding online learning, exchanging content, sharing views. In this paper, we will use Twitter as our social networking platform. Sentiment analysis on Twitter is based on opinion mining on posts to obtain the user's point of view. The leading goal deals with how opinion mining techniques can be accessed to analyze some of the tweets in many reports involving various types of tweet languages on Twitter and classify its polarity. Practical implication shows that the proposed machine learning classifiers are efficient and highly accurate.**

***Index Terms*—sentiment, opinion, polarity, classification techniques, TextBlob.**

I. Introduction

SENTIMENT analysis, one of the key emerging technologies provides people the freedom to analyze a huge amount of user-generated content available on the web[6]. Sentiment analysis is a type of NLP tool for tracking the mood of the public about a specific product or service. It deals with constructing a system to gather and analyze reviews about a certain product [4].

Few fields of research involve:-

- **Sentiment classification**:
  Handles the classification of complete documents based on opinions towards a particular object.
- **Opinion Summarization**:
  In this task, only the important characteristics of the product are extracted on which the customers have given their viewpoints.

II. Challenges in Sentiment Analysis

Some of the issues faced by sentiment analysis are:-

- A word having positive and negative sentiment has opposite orientations in various application domains.
  E.g. - "This movie sucks" implies a different meaning compared to "This vacuum cleaner really sucks''.
- Words with or without sentiment in certain sarcastic sentences are hard to examine.
- Sometimes words in a sentence without sentiment simply opinion.

III. Sentiment Analysis Tasks

The collected tweets are pre-processed for performing the data cleaning.

By using any feature selection methods important features are taken out from the clean text.

The data is divided, manually labeled as negative, positive & neutral Tweets for building a training set.

Features extracted and the training set labeled are given as an input to the classifier built for creating the test set [4].

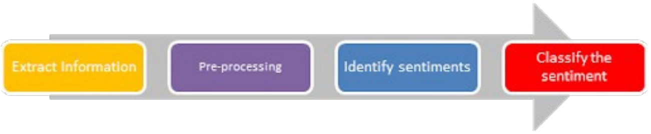


Fig. 1. Steps in sentiment analysis
Source: www.quora.com

A. *Data Sources*

Opinion mining plays a major role in selecting the data sources. Twitter, a micro-blogging, social networking site with fixed content size and generally accessible information has to gain lots of popularity [3].

B. *Methods for collecting Tweets*

The tweets for research can be collected by following methods given below –

- Data repositories- UCI, SNAP & Tweepy.
- APIs- Two types of APIs provided by Twitter are- search API and stream API. Search API - Tweets are collected with respect to hashtags and stream API -It involves streaming real-time data
- Automated tools - classified into premium tools like -Sysomos, Radian6, Simplify360, Lithium & non-premium tools like -Topsy, Keyhole, Social Mention & Tagboard.

C. *Preprocessing of Data*

- Twitter data mining is a difficult process as it involves raw data and it is essential to clean it by the following methods-

Hashtags (#), retweets (RT), and account Id (@) needs to be removed.

Symbols, URLs, hyperlinks, non-letter data & emoticon are removed as only text data is needed.

Stop words like am, is, was etc. do not show any emotions. So these are removed for decompressing the data set.

Compress extra letter words like ‘Funnyyy’ to ‘Funny’.

Slag words like c8, g9 are decompressed which are adjectives or nouns signifying the highest sentiment level. Removal of these words is essential.

D. *Feature Extraction*

-It involves the extraction of various aspects like adjectives, verbs, nouns which are identified as positive or negative to detect the polarity of the whole sentence. Some popular methods of Feature extraction are-

1) **Terms Frequency and Presence**: Individual and unique words and counting of their repetition are given these features.
2) **Negative Phrases:** Negative words changes meaning or orientation regarding a particular opinion.
3) **Parts Of Speech (POS):** Nouns, adjectives, verbs, etc. are founded as they are the important figures of thoughts.

E. *Techniques related to Sentiment Classification*

For identification of text, two techniques are used which are given below-

- **Technique based on knowledge** (Lexicon based technique)

  -It relies on extracting the opinion-based lexicons from the text and then the polarity of those lexicons is identified.
- **Machine Learning**

  - The major objective of this method is to develop an algorithm for optimizing performance by using training data like examples, past knowledge & experiences. This method provides the solution in two steps [5]:

  1) The model is developed and trained using already labeled data.
  2) Dividing of data into unlabeled or unclassified, which is based on the trained model.

Supervised and unsupervised techniques are two broad categories of ML techniques. Subjective data is used for supervised machine learning techniques which especially depend upon labeled training data. Textblob will be used in our project.

IV. Project Design

A. *Flow diagram*

Figure 2.

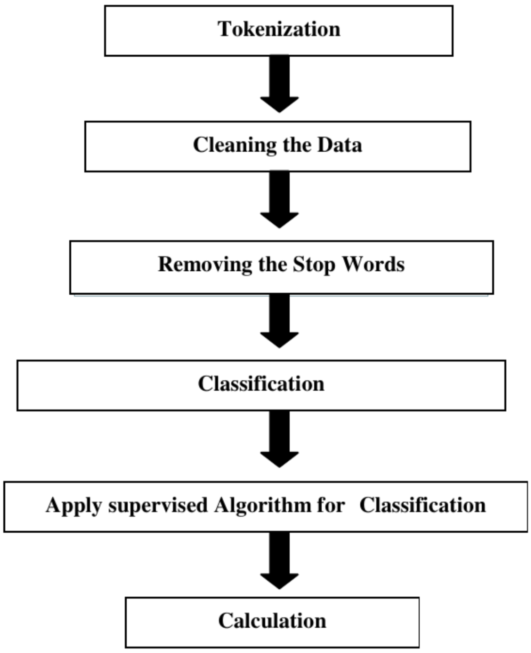


Fig. 2. Flow diagram of our project

B. *Steps involved*

- **Tokenization**- Dividing a paragraph into different types of a statement or dividing a statement into different types of words.

  Example- Let us say a simple statement “The movie was great.”

  *The *movie *was *great *. (After tokenization)
- **Cleaning the Data** - It involves removing special characters or any other words which do not add any value to the analysis part. Punctuations like the comma, full stop, exclamation are removed.

  Example- From the above Statement

  “The movie was great.”

  . (Full stop) is removed.
- **Removing the stop words -** Remove all

  Those stop words, do not add any value to the analysis part. Words as ‘a’, ‘is’, ’was’, ‘the’; do not indicate any sentiment.

  Example- From the above Statement

  “The movie was great.”

  “The”, “was” is removed.
- **Classification-** The step classifies where the statement is positive, negative, or neutral. For positive words, we assign sentiment scores as “+**1”,** for the negative words we assign **“-1”,** and for neutral “0”.

  Example- From the above Statement

  “The movie was great.”

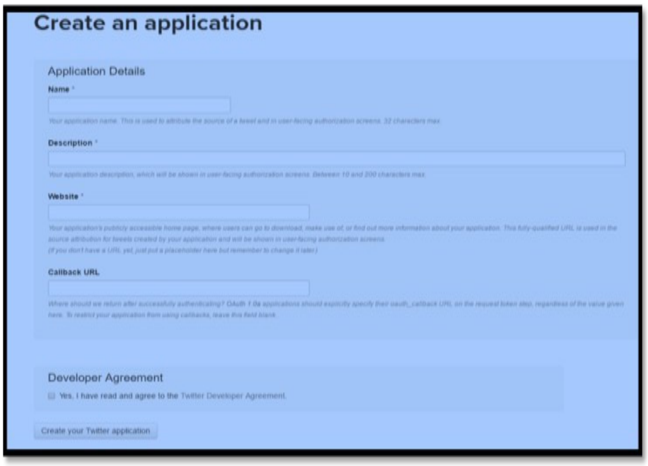
  movie – 0 great - 1
- **Apply supervised Algorithm for Classification** - This is the part where we train our model with a bag of words or lexicons and test it on the analysis statement. Once the model has been trained, we can perform a test on the analysis statement to test its accuracy for classification.

V. Feature Implementation

A. *A Twitter developer account*

To access the Twitter API, we will need to sign up for the Twitter Developer website and then perform some steps for the creation of an application.

After complete approval, an access token is generated and we will have our Consumer Secret key, consumer key, Access token secret & Access token from the Keys and Access Tokens tab.



Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

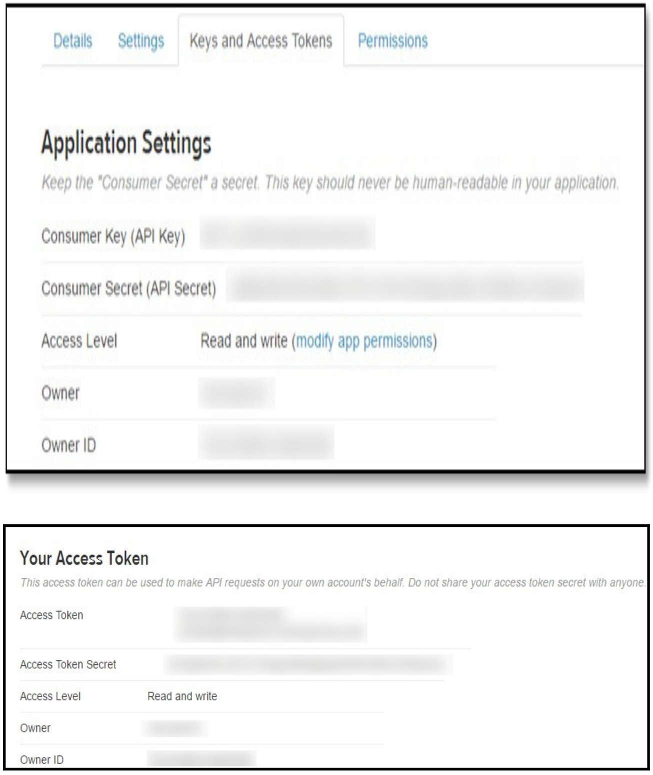
Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Developer Agreement

□ Yes, I have read and agree to the Twitter Developer Agreement.

Fig. 3. Twitter Developer interface.
Source: https://docs.inboundnow.com/guide/create-twitter-application



Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	[redacted]
Consumer Secret (API Secret)	[redacted]
Access Level	Read and write (modify app permissions)
Owner	[redacted]
Owner ID	[redacted]

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	[redacted]
Access Token Secret	[redacted]
Access Level	Read and write
Owner	[redacted]
Owner ID	[redacted]

Fig. 4. Twitter access token and customer key interfaces.
Source:https://docs.inboundnow.com/guide/create-twitter-application

B. *Anaconda navigator Installation*

-It is a GUI that launches applications, manages conda packages, environments & channels without the need for accessing command-line commands.

-We can access the following applications through the navigator:

Jupyter Lab, Jupyter Notebook, Spyder, Pycharm.

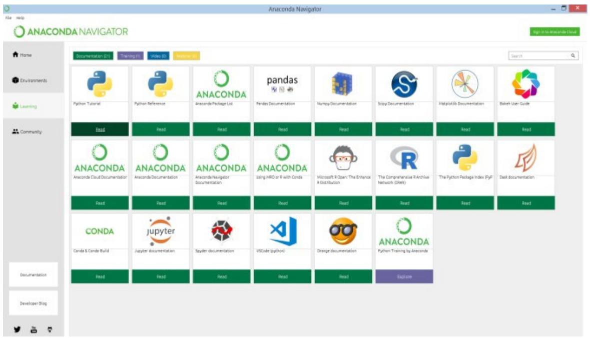


Fig. 5. Anaconda Navigator interface
Source: Project handling device

C. *Installing Python Libraries*

Python library is a collection of core modules that contains a reusable chunk of code that can be used subsequently.

The Following are some important libraries that we will be using in our project.

- **TextBlob :**

TextBlob, a Python library which deals with textual information which gives a simple API for accessing NLP activity like Noun Phrase extraction, PoS tagging, sentiment analysis, etc.

```
pip install -U textblob
```

- Install TextBlob: It can be installed by the following command through the anaconda prompt.

D. *Features of TextBlob:*

1) *Tokenization:*

It involves a large division of a large paragraph into words and phrases. A token simply means a word. Initially, we import the TextBlob object, TextBlob module & pass it the sentence for tokenization.

2) *Lemmatization:*

It refers to the tracing of a word to its origin as given in a dictionary. For using it through TextBlob, we need to access the Word object from the TextBlob module, pass the word & call the lemmatize method.

Example-

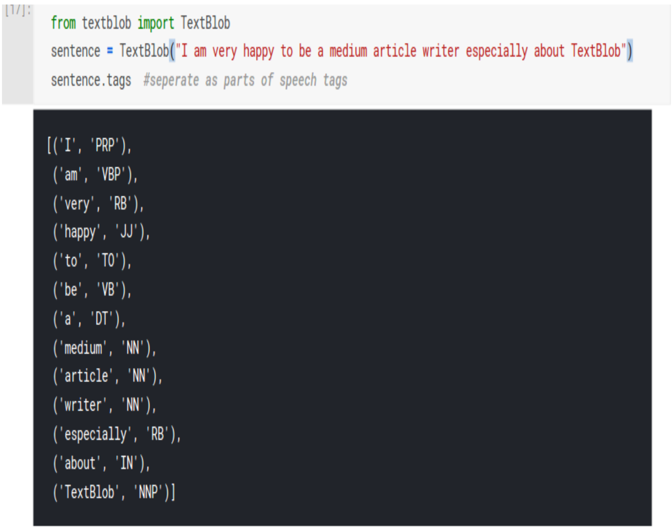
```
from textblob import Text
text1 = Text("oranges")
 print("oranges:", text1.lemmatize())
text2 = Text("radii")
print("radii:", text2.lemmatize())
text3 = Text("smaller")
print("smaller:", text3.lemmatize())
```

We perform lemmatization on the words "apples", "media", and "greater". In output get “apple”, “medium” and “great”.

3) Part of Speech tagging

It is the process of identifying the structural elements of a text document, such as verbs, nouns, adjectives, and adverbs.

Example-The interface for using the PoS tagging feature.



```
from textblob import TextBlob
sentence = TextBlob("I am very happy to be a medium article writer especially about TextBlob")
sentence.tags  #seperate as parts of speech tags
```

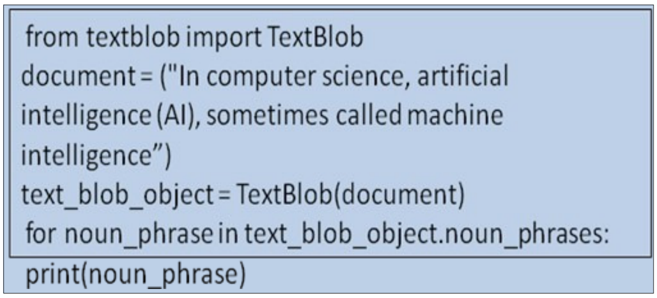
```
[('I', 'PRP'),
 ('am', 'VBP'),
 ('very', 'RB'),
 ('happy', 'JJ'),
 ('to', 'TO'),
 ('be', 'VB'),
 ('a', 'DT'),
 ('medium', 'NN'),
 ('article', 'NN'),
 ('writer', 'NN'),
 ('especially', 'RB'),
 ('about', 'IN'),
 ('TextBlob', 'NNP')]
```

Fig .6. TextBlob implementation
Source: https://www.analyticsvidhya.com/blog/natural-language-processing-for-beginners-using-textblob

4) Noun Phrase Extraction

It involves the extraction of phrases from a given context that contains nouns.

Let’s understand this feature with the following example.



```
from textblob import TextBlob
document = ("In computer science, artificial
intelligence (AI), sometimes called machine
intelligence”)
text_blob_object = TextBlob(document)
for noun_phrase in text_blob_object.noun_phrases:
print(noun_phrase)
```

The output will give all the nouns in the document.

```
computer science
artificial intelligence
AI
machine intelligence
```

5) Tweepy

An open-source repository on python for communicating with the Twitter platform and use its API.

Tweepy has many features like:

1. Get tweets.
2. Make and delete tweets.
3. Follow and unfollow users.

VI. Research Scope in Opinion Mining and Sentiment Analysis

The main areas for future sentiment analysis are:

- Spam Detection Sentiment;
- Sentiment Analysis on short Sentence like short text;
- Improvement in sentiment word identification;
- Development of the automatic analyzing tool;
- Accurate Analysis of policy-related content;
- Proper classification of bipolar sentiments.

VII. Limitations

The major limitations related to Sentiment analysis or opinion mining are as follows:-

- Spam and fake reviews detection-
  The spam contents on the web can be removed by identifying duplicates, removing outliers & consider the reputation of the reviewer.
- Classification Filtering limitation-
  There is a limitation in determining popular thought or concept. For better sentiment analysis this limitation needs to be reduced.
- Availability of opinion mining software-
  Great quality Opinion mining software -very expensive and only can be afforded by big organizations and government.
- Domain Independence-

The sentiment words are domain-dependent i.e. good performance of one feature set in one domain but at the same time, it may perform very poorly in another domain.

VIII. Conclusions

Sentiment analysis or opinion mining covers a wide area of real-time applications, meanwhile, it has suffered from many research limitations.

Since the fast growth of the internet and internet-related applications, Sentiment Analysis becomes the most interesting research area among the natural language processing community. In this research paper, we have analyzed the sentiment on the Tweets, extracted from Twitter and classify them according to their polarities.

REFERENCES

[1] https://www.lexalytics.com/news/press-releases/lexalytics-unveils-sentiment-analysis-of-emoticons-acronyms.

[2] https://www.digitalvidya.com/blog/twitter-sentiment-analysis-introduction-and-techniques

[3] https://www.quora.com/What-is- Twitter-sentiment-analysis

[4] Kiruthika, Sanjana, & Giri, " Sentiment analysis of Twitter Data", *International Journal of Innovation in Engineering and Technology*, Vol No.6, April 2016

[5] Mitali Desai, Mayuri A. Mehta, "Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey", *International Conference on Computing, Communication, and Automation* (ICCCA2016)

[6] Sanjeev Kumar Sharma, "Sentiment Analysis: An analysis on its past, present and future scope", *International Journal of advanced research in Science and Engineering*, Vol.No.6, Issue No.07, July 2017.