

## Identifying Analytics of Sentiment Analysis on Twitter Data

Ms. Puja M. Dadhe<sup>1</sup>, Dr. R.N. Jugele<sup>2</sup> and Dr. D.S. Sadhankar<sup>3</sup>

<sup>1</sup> Research Scholar, Department of Computer Science,  
Shivaji Science College, Nagpur.

<sup>2</sup> Associate Professor, Department of Computer Science,  
Shivaji Science College, Nagpur.

<sup>3</sup> Assistant Professor, Department of Computer Science,  
SFS College, Nagpur.

**Abstract:** Social Media becoming the most reliable platform for people to express their views, Sentiment Analysis has gained major popularity in research Area. Twitter is the most popular social media, people tend to put up their views very often. Corona Pandemic has raised various issues in human life, that is why social media has become a most convenient way to spread a word about opinions. This paper analyses corona tweets for Sentiments using Textblob and Vader libraries of Python. This work is based on real time Sentence level Sentiment Analysis (SLSA) which predicts the sentiment of Sentences mentioned in the Tweets.

**Keywords:** Corona, Sentiments, Lexicon Approach, Textblob, Vader, Tweets.

### 1. INTRODUCTION:

Sentiment analysis is a problem in natural language processing (NLP) that involves machine learning to train a system to interpret human opinion derived from various sources. Sentiment analysis, often known as opinion mining, is a text analytic technique that studies people's feelings, emotions, sentiments and attitudes about various items or entities using written language[1]. Sentiment analysis is a technique for analyzing the opinions of individuals or groups, such as a subset of a brand's followers or a single consumer in contact with a customer service agent. Sentiment analysis is a technique for determining whether or not articulation is positive, negative or neutral and to what extent[2]. Sentiment Analysis involves mechanisms to categories text in to positive, negative and neutral polarity. levels of sentiment analysis are

- **Document level:** This classifies the whole document text into positive or negative polarity.
- **Sentence level:** This extracts the polarity of each sentence of a document into positive or negative polarity.
- **Aspect/entity level:** This classifies the sentiment polarity of each entity's aspect or feature of a sentence/document[3].

Social networking connects people. Multiple social networking sites are available for users to post their views. Twitter, one of the most popular social networking sites, has recently garnered investigation due to its rapid expansion. Twitter began as an online microblogging service in March 2006, allowing users to create status messages known as tweets.

A user can also follow other users and watch their tweets, as well as retweet them to their followers. Twitter's user-generated content covers a wide range of themes, including products, events, people, and current events. It can be beneficial in the decision-making process of businesses and other communities[4].

To answer the challenge of sentiment analysis, Twitter is widely used and tweets are readily available. These tweets can be fed into a variety of sentiment analysis algorithms and labelled as positive, negative or neutral. Because of the following factors, Twitter messages are a valuable source for sentiment analysis:

1. Tweets are 140-character long and have a more abstract tone.
2. Real-time analysis on tweets is possible.
3. Access to a large quantity of tweets for analysis.

The goal of sentiment analysis is to find and extract opinions from user-generated content. From review sites to microblogs, there has been progress in the field of sentiment analysis. It can be beneficial in the decision-making process of businesses and other communities.

### **Two basic techniques for sentiment analysis**

**Supervised :** Machine Learning (ML) based sentiment analysis. Supervised learning is the process of inferring a function from labelled training data. The training data consist of a set of training examples. Supervised learning is a widely used solution for classification purpose and is been used in most of the sentiment classification techniques. Techniques for sentiment classification include SVM, Neural Network and Decision tree Classifiers. Other commonly used algorithms include K-Nearest Neighbour, Bayesian Network.

**Unsupervised :** Rule-based sentiment analysis, uses a dictionary of words labelled by sentiment to determine the sentiment of a sentence. Unsupervised learning is used to find inferences from datasets consisting of input data without labelled responses. In most of the classification techniques, especially text data it is very difficult to create training labelled data and it requires much of the human effort. Making use of unsupervised techniques can help overcome the disadvantage.

In this paper Lexicon based techniques are employed to determine the polarity of 5000 corona tweets.

## **2. METHODOLOGY :**

The two main areas of research in sentiment classification are lexical and machine learning approaches. For the lexical approach, a dictionary is prepared to store the polarity values of lexicons. For calculating polarity of a text, polarity score of each word of the text, if present in the dictionary, is added to get an 'overall polarity score'. For example, if a lexicon matches a word marked as positive in the dictionary, then the total polarity score of the text is increased. If the overall polarity score of a text is positive, then that text is classified as positive, otherwise it is classified as negative. Though this approach seems very basic, variants of this lexical approach have been reported to have considerably high accuracy. Since the polarity of the text depends on the score given to each lexicon, there has been a large volume of work dedicated to discovering which lexical information is most efficient[5].

This paper aims to analyze the performance of Textblob and Vader of Python for polarity detection (positive, negative and neutral) of textual data. 5000 corona tweets are fetched using rest Api of Twitter and reports the implementation of the Twitter sentiment analysis. Tweepy is a library of Twitter API for fetching the tweets directly from Twitter that are tweeted by the different Twitter user. The real time



Fig 1: Libraries imported and Secret Keys for Twitter API

Once the connection is successful, 5000 tweets are fetched and saved in corona.csv file.

```
In [8]: tweets_df = tweet_search(key_word)
tweets_df

Tweets downloaded: 4999 / 5000253 / 5000 496 / 50001277 / 50005000 1872 / 5000 2428 / 5000 3217 / 5000 3506 / 5000 3888 / 5000
4308 / 5000 4392 / 5000 / 5000
```

```
Out[8]:
```

Datetime	Tweet	Username	Retweets	Followers	CleanTweet
2022-02-23 11:12:15	@JoeSmithSDK @2NJoyMore @Matth_4America also n...	gregoryient	0	4285	also, note, five-eyes, is, having, the, most, ...
2022-02-23 11:12:14	RT @VarunKrRana: Corona report of Nawab Malik ...	KapilSGalav	65	823	corona, repoot, nawab, malik, came, negative, ...
2022-02-23 11:11:19	@bent_shamo Corona Outlet 🗨️	ClassicalQ8	0	704	corona, outlet, 🗨️
2022-02-23 11:11:03	RT @EverythingEnch1: #Newpost 🇮🇳 My Journey fr...	DeliciouslySavv	18	33768	🇮🇳, 🇮🇳, my, journey, from, london, uk, to, india, ...
2022-02-23 11:10:55	RT @INC_akhter: To compensate for the loss in ...	kushal_gehlot	14	3009	to, compensate, for, the, loss, in, education, ...
...	...	...	...	...	...
2022-02-21 21:59:32	RT @naturalphoton: Dear @UNGeneva - you should...	Amrita1224	16	532	dear, -, you, should, go, one, step, back, and, ...
2022-02-21 21:58:23	RT @LawrenceSellin: Stolen elections have cons...	GoneWit59335088	19	739	stolen, elections, have, consequences, the, us, ...

Fig 2: Tweets fetched.

Tweets fetched are then preprocessed and then Textblob and vader libraries are used to classify these tweets as positive negative and neutral.

```
In [9]: def vader_compound_score(tweet):
vader = SentimentIntensityAnalyzer()
if vader.polarity_scores(tweet)['compound'] >= 0.05:
return 'Positive'
elif vader.polarity_scores(tweet)['compound'] <= -0.05:
return 'Negative'
else:
return 'Neutral'

def textblob_sentiment(tweet):
analysis = TextBlob(tweet)
if analysis.sentiment.polarity > 0:
return 'Positive'
elif analysis.sentiment.polarity == 0:
return 'Neutral'
else:
return 'Negative'
```

Fig 3: Python code for sentiment Analysis using Textblob and Vader.

### 3. RESULTS:

The result is visualized by different plot methods using matplotlib that is the most popular library in python to the visualization of a result like a bar chart, histogram, pie chart etc.

The result of tweets for 'coronavirus' based on 5000 tweets from Twitter.

Textblob shows Positive tweets percentage 40.3% Negative tweets percentage 30.6% Neutral tweets percentage 29.1%.

Vader shows Positive tweets percentage 40.2 % Negative tweets percentage 21.1% Neutral tweets percentage 38.7%.

Following Table shows number of positive, negative and neutral tweets evaluated by Textblob and vader.

	No. of Positive tweets	No. of Negative tweets	No. of Neutral tweets
Textblob	2010	1055	1935
Vader	2015	1530	1455

Table 1: Result

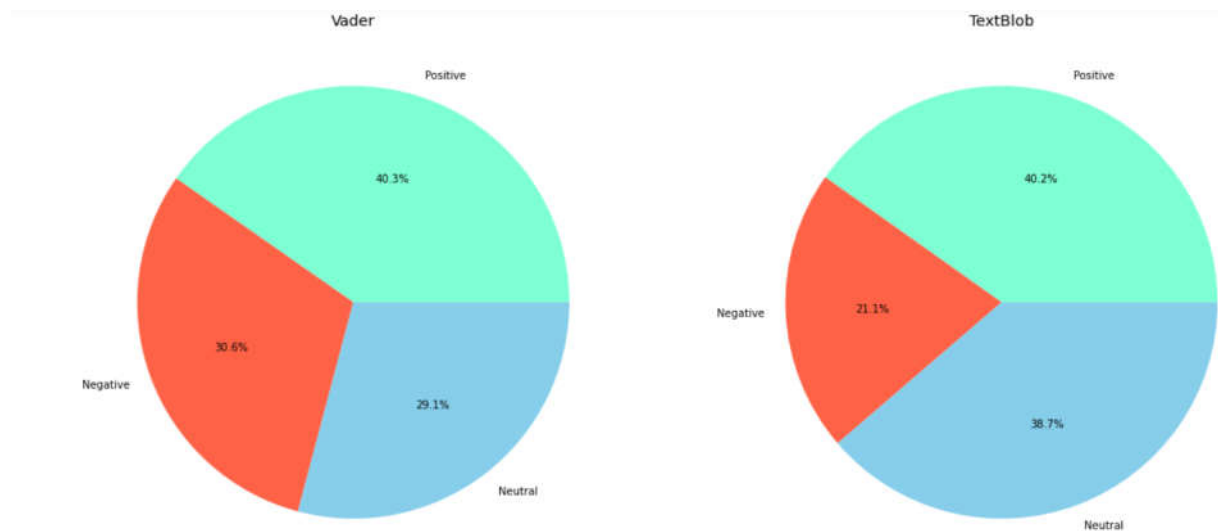


Fig 4: Percentage of Polarized Tweets

#### 4. CONCLUSION:

With the Advancement in the field of Natural language processing algorithms, understanding and managing text based data is now easy. For the analysis of attitudes with data, the algorithm has a greater accuracy rate. Both Textblob and Vader have a plethora of features and functionalities — plotting the graph for the comparison revealed that Vader is primarily developed for social media platforms and can provide superior results as well as intensity when used with data from social media platforms such as Twitter. The rule-based methodology for sentiment analysis sometimes has a flaw in that it focuses just on individual words while completely ignoring the context in which they are used. The result indicated that VADER performs well than Text blob in sentiment analysis of tools.

#### 5. REFERENCES:

1. Shubham V. Pandey, A. V. Deorankar, "A Study of Sentiment Analysis Task and It's Challenges", 978-1-5386-8158-9/19/\$31.00©2019IEEE.
2. Abdullah Alsaedi1 , Mohammad Zubair Khan, "A Study on Sentiment Analysis Techniques of Twitter Data", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 10, No. 2, 2019.
3. Ms. Puja M. Dadhe, Dr. R.N. Jugele, "Inspection of Retrospection - Challenges in Sentiment Analysis", Compliance Engineering Journal, Volume 11, Issue 1, 2020.

4. K. Revathy, Dr. B. Sathiyabhama, “A Hybrid Approach for Supervised Twitter Sentiment Classification”, International Journal of Computer Science and Business Informatics, Vol. 7, No. 1. NOVEMBER 2013.
5. Chetashri Bhadanea ,Hardi Dalalb , Heenal Doshi, “Sentiment analysis: Measuring opinions” International Conference on Advanced Computing Technologies and Applications (ICACTA2015) volume 45, 2015, Pages 808-814.
6. <https://textblob.readthedocs.io/en/dev/>
7. <https://www.geeksforgeeks.org/python-sentiment-analysis-using-vader/>