



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 12, December 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)



# Opinion Mining using Twitter Platform by Natural Language Processing

Mr. Vikram Singh Kardam<sup>1</sup>, Nishank Vadhera<sup>2</sup>, Raghav Girdhar<sup>3</sup>, Ridhi Gupta<sup>4</sup>, Saksham Singh<sup>5</sup>

Assistant Professor, Department of Electronics and Communication Engineering, HMR Institute of Technology and Management, Delhi, India<sup>1</sup>

U.G. Student, Department of Electronics and Communication Engineering, HMR Institute of Technology and Management, Delhi, India<sup>2</sup>

U.G. Student, Department of Electronics and Communication Engineering, HMR Institute of Technology and Management, Delhi, India<sup>3</sup>

U.G. Student, Department of Electronics and Communication Engineering, HMR Institute of Technology and Management, Delhi, India<sup>4</sup>

U.G. Student, Department of Electronics and Communication Engineering, HMR Institute of Technology and Management, Delhi, India<sup>5</sup>

**ABSTRACT:**In the modern digital era, most of the population is using online networking on the Internet. Opinion mining (sentimental analysis) has turned into a primary way of investigating such a considerable amount of information. Different applications turn up in a wide variety of modern areas. Meanwhile, opinions have profuse pronunciations which mobilize investigative challenges. The research challenges recently made opinion mining an unsteady research region. In this paper, Natural Language Processing (NLP) techniques for opinion mining are evaluated. At first, NLP is reviewed then briefed about its common and useful pre-processing step. In the paper, opinion mining for numerous levels is viewed, analysed and the result is showcased. In the end, issues are identified, and some recommendations are suggested for sentiment analysis.

This project discusses the obstacles of opinion analysis in Twitter that classify tweets based on the sentiment expressed in them: strongly positive, weakly positive, positive, strongly negative, weakly negative, negative, and neutral. Twitter is online blogging and social networking platform that allows users to present their opinions on various issues, whether on poll discussion, the stock market, or any other social injustice toward anyone but within a word, limit or say the length of not more than 280 characters.

It is an exponentially expanding service with over 396 million registered users and still increasing. Out of which, around 206 million are active users. By saying active users, they log in to their Twitter account daily and generate nearly 500 million tweets per day. Because of this substantial usage, we expect to achieve hindsight of public sentiments by analysing the sentiments expressed in their tweets. Analyzation of public sentiments is foremost for many applications, such as firms trying to find out the response of their products in the market, predicting election results, and predicting socio-economic phenomena like the stock exchange.

The objective of this project is to have a classifier for accurate and automatic classification of an unknown tweet stream based on sentiments.

**KEYWORDS:**Opinion Mining, Sentiment Analysis, Natural Language Processing, Python, and Natural Language Toolkit (NLTK).

## I. INTRODUCTION

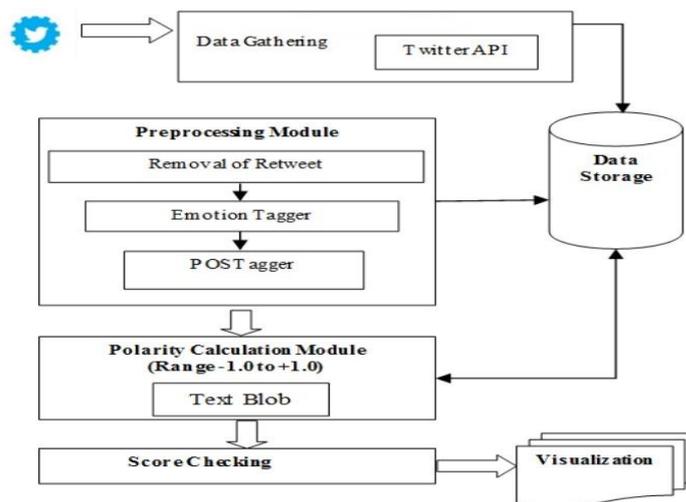
In the past few decades, many populations have shown their inclination towards social networking sites like Facebook, Instagram, and Twitter. These sites bring out their views towards the society, emotions quotient and the opinions about the society, personalities (Politicians, Actors/Actresses, etc.), and the places which enable other users another way to look at the situations.

There are several methods used for sentiment analysis. These methods are categorized primarily as Natural language processing, artificial intelligence, statistical, and knowledge base are based on different methods. It isn't easy to analyse sentiments and opinions computationally [1]. Therefore, it extracts the information from the available data through the Twitter account for marketing, electoral poll results, business analysis, communication, research, and educational solutions.

Execution of sentimental analysis involves specific steps, namely - data collection, pre-processing of data, feature extraction, selecting baseline features, sentiment detection, and performing classification either using simple computation or else machine learning techniques.

To take out sentiments from tweets, opinion mining is performed, also known as sentiment analysis. In this paper, a pre-processed data set is formulated using NLP to filter out the tweets where we embraced Bag of Word (BoW). NLP comprises several techniques, tokenization, stemming, lemmatization, POS, named entity recognition, and chunking. The agenda of this project is to recognize the sentiments of the tweet by the usage of polarity that is dividing it into seven classes – strongly negative, negative, weakly negative, neutral, weakly positive, positive, strongly positive where tweets are gathered by using Twitter streaming API from Twitter. We employ these tweets as unaltered data.

The rest of the paper is assembled as follows. Section II provides the literary survey to dissect the previous work. Section III discusses the proposed work's methodology and the related discussions. In section IV, detailed experimental result have been provided with their respective figures for the better understanding of the proposed system. Section V provides the conclusion to the paper.



**Fig. 1. Framework for The Sentiment Analysis on Twitter Accounts**

Sentiment analysis is a fast-growing region of Natural Language Processing with research starting from document-level classification and extending to learning the polarity of words and phrases. Because of the limitation of characters on tweets sorting the sentiments of tweets is parallel to sentence-level sentiment analysis however it remains a difficult task to accomplish because of the usage of informal and special language in tweets, and due to the very nature of the microblogging domain.

In the past few years, several research has taken place in this field. Nowadays there had been numerous papers looking for Twitter sentiment and buzz. Many researchers have begun to investigate the use of POS features, but the results remain varied. Researchers are also on the way to finding various methods of automatically collecting the training data. Some researchers also use hashtags for creating training data, but they bound their research to sentiment/non-sentiment classification instead of 2-way polarity classification, as we do. We use text extraction methods and apply NLP algorithms for the classification to obtain the best result:

➤ Natural Language Processing:

Natural Language Processing (NLP) is a part of computer science and artificial intelligence in which computers analyse, understand, and develop meaning from human language in a smart and useful way. By utilizing NLP, several tasks can be performed in an organized and structured way such as sentiment analysis, chatbot, speech recognition, machine translation, spell checking, keyword searching, information extraction, advertisement matching.



NLP treats text as a hierarchical structure of language. By analysing language for its meaning, NLP systems have long filled useful roles, such as correcting grammar, converting speech to text, and automatically translating between languages[2].

NLP is categorized as a challenging problem in computer science. Human language is hardly precise or simply spoken. To comprehend human language is to understand not only words but the idea behind the word and how they are linked with each other to establish meaning. Understanding language is the easiest thing for the human mind, but the abstruseness of language makes natural language processing problematic for computers to master.

## II. METHODOLOGY

This section displays our proposed work and concentrates on a strategy to sentiment analysis on Twitter data. The architectural overview describing an overall process design of sentiment analysis which is shown in Fig. 1. There are three important parts to the developed method that are Extraction of Data from a particular project or product, pre-processing of the extracted Tweet using Natural language toolkit (NLTK) and Classifier model that evaluates the sentiment of each Tweet.

Performing sentiment analysis using Twitter platform using natural language processing is investigative research which itself is a problematic as well as tough job. There are certain steps involved for performing sentiment analysis of tweets. These steps are as follow:

### 1. Preparing the Text

Once the twitter developer account has been registered, the following steps are performed:

- a. Using the API credentials, authenticate our Python script.
- b. Implement a keyword search to download tweets
- c. After registering with Twitter, we will get our authentication credentials.

Following that, coding part will begin. In the third step, we will download tweets based on the term we are attempting to retrieve.



### 2. Pre-Processing of Tweets

Data preprocessing is an essential step as it determines the efficiency of later stages. It entails correcting the tweets to ensure their syntactic accuracy.

Below are a few steps involved in preprocessing tweets:

- a. In Opinion mining, words matter most. Punctuation, however, cannot give you a sense of sentiment. Therefore, it doesn't matter what you say.
- b. From the perspective of future processing, usernames, and URLs do not matter, so their presence is futile. Usernames and URLs are converted into generic tags or removed.
- c. Stemming – Using roots to replace words and reducing similar words in different forms. This provides a more dimensional set of features.  
For example, word like “Affection, Affecting, Affected, Affectation” would yield to the word “Affect” after removing prefixes and suffixes.
- d. Lemmatization – Similar to stemming. It groups together different inflected form of a word called “LEMMA”.  
It again several word into its one common root and gives output as a proper word.



For example, a lemmatizer maps the words like “Gone, Going, Went” to “Go” which stemming is unable to give.

Algorithm used:

```
def cleanTweet (self, tweet):  
  
    #Remove Links, Special Characters etc from tweet  
  
    ps-PorterStemmer()  
  
    processed_tweet=[]  
  
    dataset=pd.read_csv ('result.csv', names = ['tweet'])  
    for i in range (50):  
  
        self.cleanTweet=re.sub(@[w] , dataset ['tweet'][i])  
  
        self.cleanTweet=re.sub('[^a-zA-Z#]',",",tweet)  
  
        self.cleanTweet=tweet.lower()  
  
        self.cleanTweet=tweet.split()  
  
        self.cleanTweet=[ps.stem(token) for token in tweet if not token in stopwords.words('english')]  
  
        self.cleanTweet-' '.join(tweet)  
  
    processed_tweet.append(self.cleanTweet)
```

### 3. Function Creation

An important aspect of opinion mining is analysing body of text for understanding the opinion it reveals. In general, we quantify this sentiment by giving it a positive or negative value, called polarity. Depending on the sign of the polarity score, the overall sentiment can be interpreted as positive, neutral, or negative. So, polarity can be calculated using the natural language processing toolkit. The Natural Language Toolkit, sometimes called NLTK, is a set of Python programs and libraries designed for statistical and symbolic natural language processing (NLP). There are many ways NLTK can be used as a teaching tool, as a study tool, as a platform to build prototypes, and as a tool for research.

### 4. Testing a Model

In the output, we will be asked the keyword on which we want to search the recent tweets, along with the number of tweets we want to search at the same time.

A pie chart is used to show a detailed report of various opinions.

## III. EXPERIMENTAL RESULT

After pre-processing of the datasets taken from Tweepy API and implementation of above algorithm has evaluated the last 25 tweets with the searchTerm equal to 25 and the general report conveys it as weakly positive stats –

```
Enter how many tweets to search: 25  
How people are reacting on MS Dhoni by analyzing 25 tweets.  
  
General Report:  
Weakly Positive
```



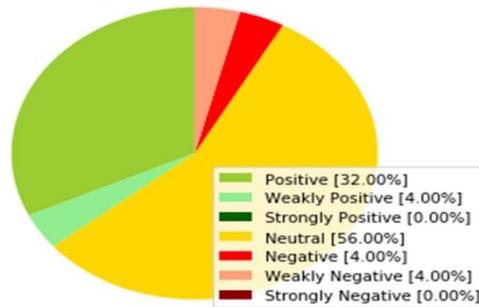
Subsequently a detailed report is generated that is classified into 7 above mentioned classes as follow –

Detailed Report:  
32.00% people thought it was positive  
4.00% people thought it was weakly positive  
0.00% people thought it was strongly positive  
4.00% people thought it was negative  
4.00% people thought it was weakly negative  
0.00% people thought it was strongly negative  
56.00% people thought it was neutral

At last, a graphical representation in the form of pie chart is obtained which is constructed based on detailed report generated earlier.

PIE CHART:

How people are reacting on MS Dhoni by analyzing 25 Tweets.



IV. CONCLUSION

Twitter sentiment analysis falls under the category of opinion mining and sentiment analysis. Businesses use opinion mining as a powerful tool for measuring attitudes, feelings, and emotions about their brand. These brands can gain insights into consumer behaviour by investigating and analysing customer sentiment and, ultimately, offer better products, services, and experiences to their customers. It contains steps like text pre-processing, function creation, data collection, training, and testing the model.

REFERENCES

- [1] Jagdale, O.; Harmalkar, V.; Chavan, S.; Sharma, N. Twitter mining using R. Int. J. Eng. Res. Adv. Tech. 2017, 3,252–256.
- [2] John Rehling “How natural language processing helps uncover social media sentiment”, Available: <https://mashable.com/archive/natural-language-processing-social-media#VbWC8PySNqyy>, November 8, 2011.
- [3] <https://www.monkeylearn.com/sentiment-analysis/>
- [4] <https://brandwatch.com/blog/understanding-sentiment-analysis/>
- [5] Vadivukarassi, M. (2017). Sentimental Analysis of Tweets Using Naïve Bayes Algorithm, Tamil Nadu: idios.
- [6] NLP: The Essential Guide to Neuro-Linguistic Programming Book by Susan Sanders, TomDotz , and Tom Hoobyar



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details