iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.46004

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Predicting Covid-19 Misleading Information Using Sentiment Analysis

Mr. Venkatesh Prasad B S[1], Meghana H S[2], Nachiketh M N[3], Nishitha S[4], Sahana M S[5]
[1, 2, 3, 4, 5]*Dept. of Computer Science and Engg., Government Engineering College, Krishnarajpet, Karnataka, India*

***Abstract** During the go higher of grouping networking time, there has been a be moving as waves of user produced what is in. Millions of people having the same their ideas daily on microblogging site because of its quality of short and simple way of look. To mine the feeling from a having general approval now microblogging support, twitter, we offer and make observation of an example where user post true tweets about covid-19. By using both corpus based and dictionary based method we may clear an offspring of parts coming from different sorts way in, to come to a decision about the connotation adjustment of the emotion statements in tweets. To make clear the example or picture the use and good effects of the offered system an example learning process is presented.*

I. INTRODUCTION

Promise of continued growth in wide area network connectivity is highly enhanced opportunity for coaction and resource distribution. Currently, abundant social networking sites like YouTube, Myspace, Facebook, Twitter have captured popularity and we can't neglect them. Web 2.0 is becoming a more essential application. This permit users to make network connection with other users in a simple and appropriate manner. This permit users to share a variety of information and use network such as images sharing, Wiki, blog etc. It is Perceptible that the rice of these real time information servicing sites like twitter have built up an unrivaled public archive feelings about every universal system that is of interest. Despite the fact that twitter can furnish for an accomplishing channel for creating and presenting opinions, this presents new & divorce the process is incomplete without the efficient tools to analyse the challenges and those ideas accelerate their consumption

II. RELATED WORK

A. Various people are used to share their opinion about microblogging platform various connotation, therefore it is a precious source of public opinion.
B. Twitter has a huge numerous of text post and this develops day by day collected the aggregation can be informally giant.
C. Twitter listeners differ from formal user to VIP, company delegate and even the president of the country. So it is attainable to get text post of users from different communal and related groups.
D. Twitters listeners is described by people from several nations to naive bias the unigram model is implemented by Parikh and Movaset, one being naïve bias bigram models and the other is the maximum entropy model. These models are used for categorizing tweets. Naïve bias classifier works much better than maximum entropy pattern.

There are two approaches:

1) *Dictionary-based Approaches:* This approach aims to calculate the emotion based of the set of sentences from connotation based of lexicons. Connotation based include positive, negative or neutral.
2) *Corpus-based Approaches:* Aims to analyse whether the contents develop by users are negative or positive emotions about a particular topic.

III. LITERATURE SURVEY

Applying sentiment analysis to Twitter is an upcoming trend with researchers scientific testing and its potential applications. The challenges unique to this problem area are mainly attributed to the dominant informal tone of the microblogging. Pak and Paroubek [5] the rationale for using microblogging, and especially, Twitter as a fund rise for sentiment analysis. The above two approaches Naïve Bayes, Maxent and create model using Support Vector Machines (SVM). Their characters space abide of unigrams, bigrams and parts of speech reports that SVMs exceeded other models and unigram were more successful as facility. Pak and Paroubek [5] completed the same thing but classified the tweets as negative, positive, neutral & objective. To gather a collection of impartial terms, he obtained text information from of famous magazine and newspapers twitter account, where as "New York Times",

©IJRASET: All Rights are Reserved

"Washington posts". His classification is the multinational Naïve Bayes based on sorter that uses n-grams and parts of speech-tags as characteristics. Barbosa et al. [9] more classified tweets polarity or subjective and then subjective tweets were classified as negative, positive or neutral. The characters space applied include tweet attribute such as retweet, hashtags, links, punctuation, exclamation and are included exclamation mark with a combination of characters identical pre polarity of words and parts of speech of words.

The twitter user's emotional analysis tasks can be performed at multiple stages of functionality, i.e., feature stage, document stage, word stage and phrase or sentenced stage. Since Twitter permits the users to contribute small pieces of data known as "tweets" (limited up to 140 words), the word stage functionality suitably adapted to its setting. Survey among literature confirms that techniques at the word level automatically explanatory sense falls into 2 classes: (I) corpus-based & (II) dictionary-based approach. To self-operating analysis, separate methods have been applied to find the feelings of expressions, characters. These include Machine Learning (ML) & Natural Language Processing (NLP) algorithms. In this aim to mine sentiment from data in twitter, it offers a hybrid techniques that as well as benefits of the two corpus and dictionary-based methods combination of ML and NLP Built operations.

IV. EXISTING SYSTEM

Regardless of the accessibility of software to obtain information with respect to the sentiment of a particular individual service or product, organization and other data information workers still face problems with respect to data extraction.

V. PROPOSED SYSTEM

The work represented in this paper specifies a new method to analysis on Twitter data. A total of 16000 tweets have been made on covid different vaccines. To evoke emotion, we removed content based on polarity and subjectivity in tweets. Text blob method is used to obtain the connotation orientation of dictionary and a objective-based approach to obtain the connotation orientation of noun and verb. We identified vaccine-related positives, negative and neutral tweets.

VI. METHODOLOGY

This paper is classified into two phases. First, literature is studied, then system development includes studying on different emotions analysis approaches and methods presently in use. In step II, the functionalities and application requirements have been explained prior to its achievement. In addition, interface and architecture design identify the program and it interact. Twitter evolution for Analysis application, various tools are used, where as Notepad & Python Shell 2.7.2

1) The pre-processing of Tweet: A transaction file is created in which pre-processed opinion indicator.
2) Extracting Opinion intensifiers: The opinion intensifiers for tweets is calculated as in this type:
a) Tweet excerpt in caps: The sentence has 18 words in total which is in all caps. Here upon, Pc=1/18=0.055.
b) Length of repeated series, Ns=3.
c) Number of repeated series, Nx=3.

Opinion Vocabulary after a tweet has been pre-processed,

- The list of extracted Adjective Groups.
- The list of removed action Groups.
- Inscribing Module now that have their own adjective group and verb group, their connotation based has to be found. The count is based on the key.
- Adjective Group Score.

VII. DATA FLOW DIAGRAM

- *Level 0*

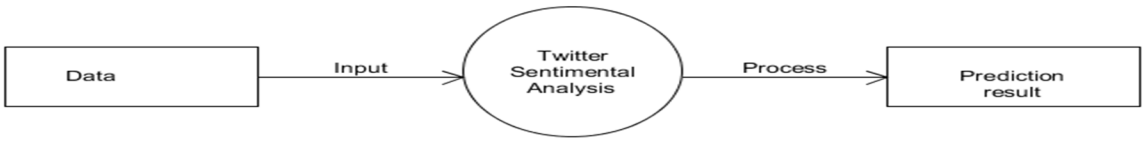


Figure 1. Dataflow diagram of Level 0

Describes the general method of this project. we tend to square measure passing data as an input the system can analyse and process the data and detects pretend tweets are true or not.

- *Level 1:*

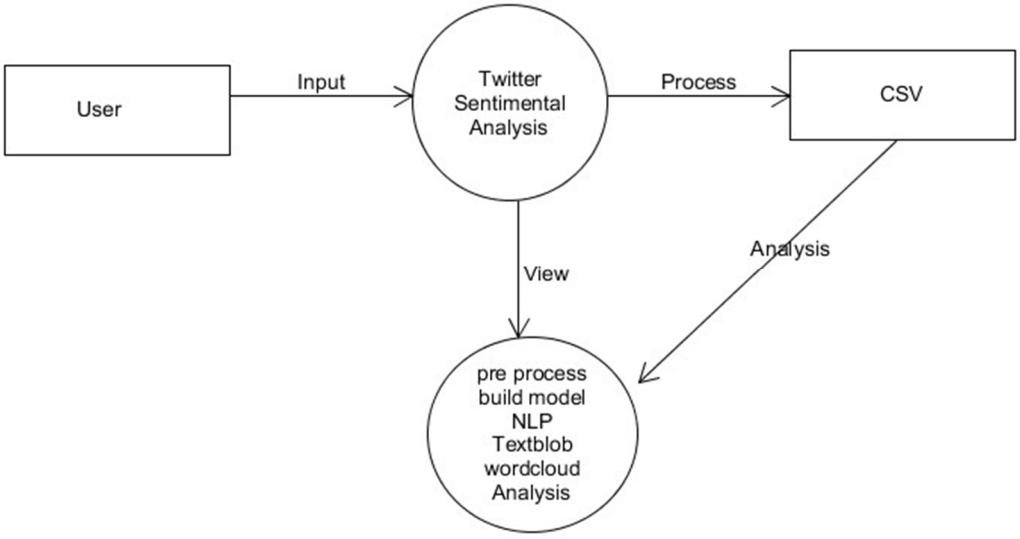


Figure 2. Dataflow diagram of Level 1

Describe the ultimate stage method of the project. User tweet the sentence as input the system can analyse and process the csv file and detect the result that is positive neutral or negative using NLP textblob method.

A. Usecase Diagram

This shows the overall activity of the project and in this more commonly using csv file to store the dataset as input.

1) Working: By doing this below steps system detect whether the tweets are real or fake news.

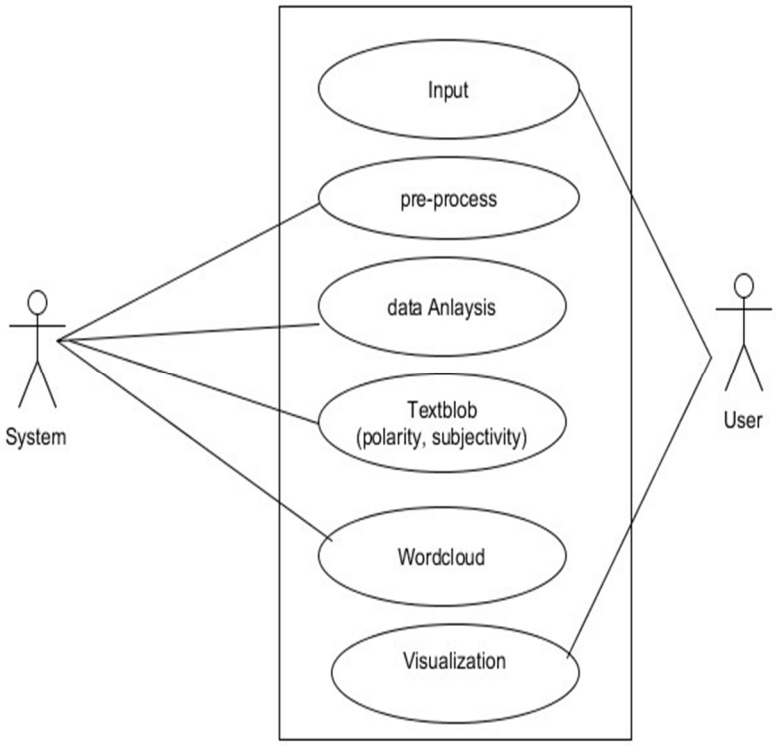


Figure 3. Usecase diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

VIII. ACTIVITY DIAGRAM

This is basically a flowchart to present flow from one activity to another. The activity can be described as an operation of the system. In this below activity data collected from Csv file where we store dataset of the project. taking user tweets for processing the data only if its true users of tweeter. After the processing using Nlp techniques output will be shown to users.

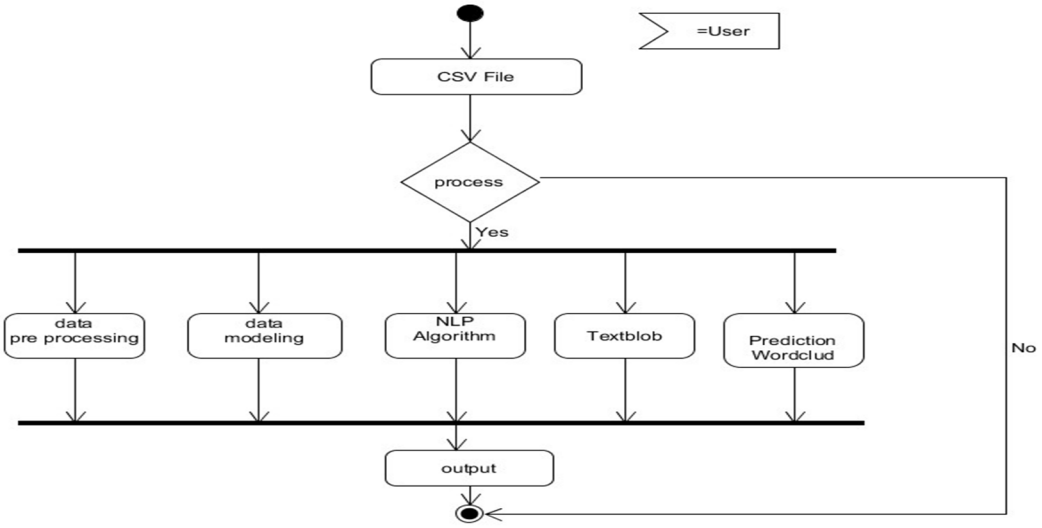


Figure 4. Activity diagram

IX. IMPLEMENTATION

At first, data was made to go through pre-processing. All things that constitute towards making the data clean and trim for it to be able to pass through a model is pre-processing. Treatment for inaccurate data-types, handling missing values, and removing outliers was done in this stage.

The following steps are involving in the pre-processing

1) *Stage I:* For every tweet record, sixteen features and one class label and also include tweet, data, retweet etc.

2) *Stage II:*

a) Data Cleaning: Data have a more inapplicable words and absent part, to make this part in relevant data cleaning is done. It includes handling absent part, analysis of data, feature engineering, noisy data etc.

b) Missing Data: It causes when some data is absent in the dataset. It can be maintained in several ways.

- *Ignore the Tuples:* This may implement when we have large or more numbers of values are absent within a tuple of the dataset.
- *Fill the Missing Values:* In this we have several ways to this approach. We can select this approach manually by attribute mean or believable value.

3) *Stage III*

The obtained dataset from stages are taken into consideration then data is trained using the classification algorithm and result is analysed and represent in the graph using python library. The obtained data is also trained using Machin Learning Algorithms nlp. The obtained result is compared for better Accuracy.

X. CONCLUSION

Covid-19 misleading information prediction is a big issue in machine learning. But still is there to figure out regarding to the emotion or opinion of texts accurate exactly is difficulty within the English language. In the project we concentrate to particularized in fake and real fact for covid vaccine tweets. There are total 16000 tweets made n covid different vaccine. These tweets are analysed using natural language processing techniques and first we have identified the positive and subjectivity sentence in the tweets using text blob which gives the better result. The results are passed as input to each covid vaccine data to identify the positive, negative and neutral.

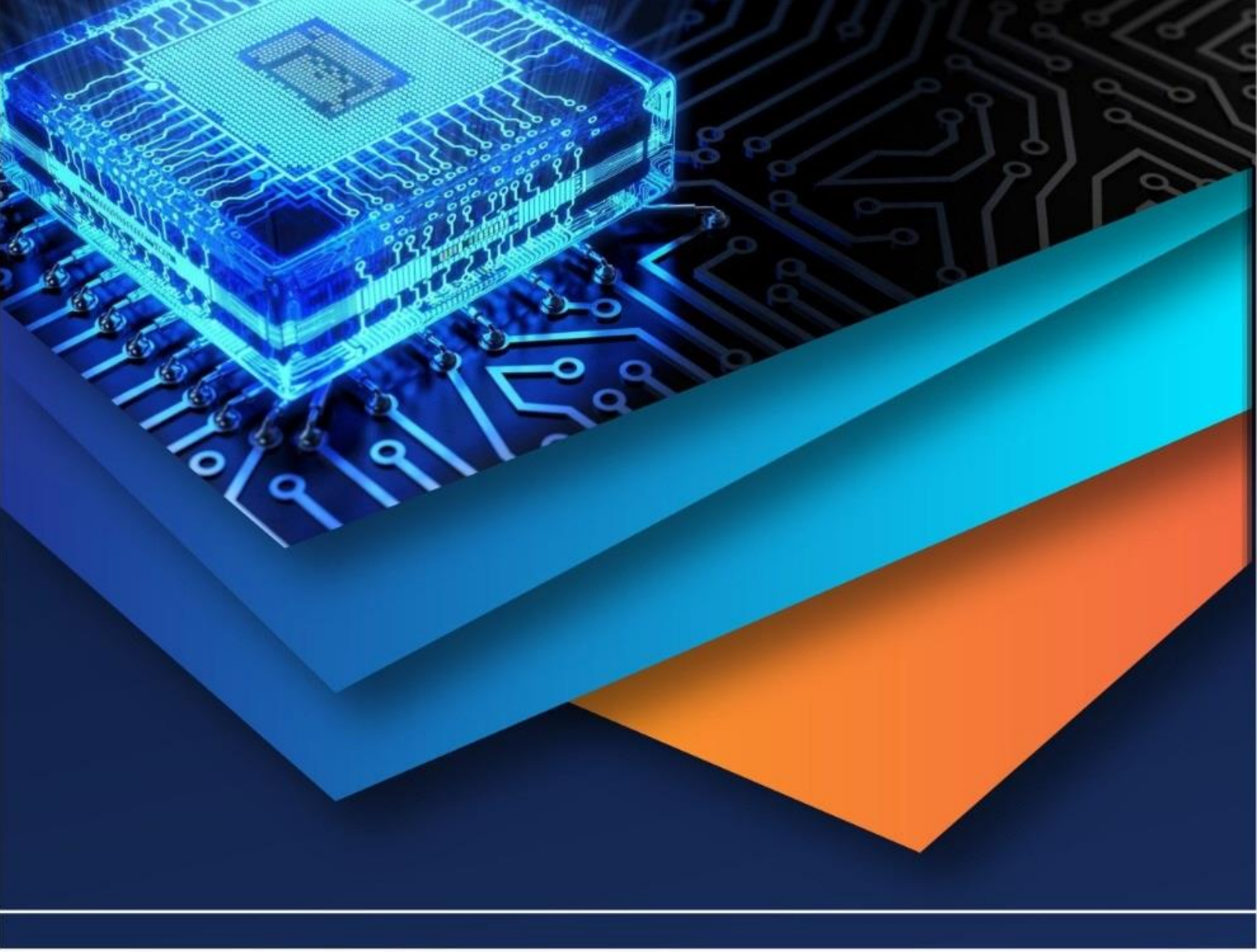
©IJRASET: All Rights are Reserved

A. Future Enhancements

In future result we can go check for different vaccine datasets and try to build more accurate result compared to present study. Other NLP methods can be used to identify the results and better performance.

REFERENCES

[1] Pooja kumari, Shikha singh, Devika more, and Dakshata Talpade, "Sentimental Analysis of Tweets", IJSSN:2349-784X, Volume:1, Issue:10, pp: 130-134,2015
[2] Asst Prof.A Kowcika, Aditi Gupta, Karthik Sondhi, Nishit Shivhre, and Raunaq kumar, "Sentimental Analysis for social media", International Journal of Advanced research in computer science and software engineering, ISSN :2277 128x, volume:3, issue:7,2013.
[3] Ali Hasan, Sana Moin, Ahmad Karim, and Shahaboddin Shamshirband, "machine learning based sentiment analysis for twitter accounts", Mathematical and computational applications, ISSN:2297-8747, Volume:21, Issue:1,2016.
[4] Rasika wagh, and payal punde, "survey on sentiment analysis using twitter dataset" 2nd international conference on electronics, communication and aerospace technologies IEEE conference, ISBN:978-1-5386-0965-1,2018.
[5] Pak and Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining", proceedings of the international conference, LREC 2010,17-23 may 2010, Valletta, Malta.
[6] Mohammed H.Abd El-jawad hodhod, and Yasser M. K.Omar, "sentiment analysis of social media network using machine learning",2018 14th international computer engineering conference, ISBN:9781-5386-5117-9,2018.
[7] Ajitkumar Shitole and Manoj Devare, "Optimization of person prediction using sensor data analysis of IOT enabled physical location monitoring, "Journal of advanced research in dynamical and control system", ISSN:1943-023X, Volume:10, Issue:9, pp,2800-2812, dec 2018.
[8] Ellen Riloff, Ashequl Qadir, "Sarcasm as contrast between a positive sentiment and negative situation", 2013 conference on emperial method in natural language processing, pp:704714,2013.
[9] Barbosa and Feng, "Robust sentiment detection on tweeter from biased and noisy data", coling 2010,23rd International conference on computational linguistics, posters volume,23-27 august 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)