

# Visualization of Real Time Big Data Through Sentiment Analysis

Vijaya Lakshmi Saragadam

M.E. Big Data Analytics Karpaga Vinayaga College of  
Engineering and Technology  
(Affiliated to Anna University, Chennai, Tamilnadu)

Dr. S. Parasuraman,

Under the guidance and support of  
M.E., Ph.D., Mentor, Supervisor  
Head of the Department, Karpaga Vinayaga College of  
Engineering and Technology

**Abstract**—Sentiment Analysis and Opinion Mining is a prominent field for analyzing and extracting information from text data from many sources such as Facebook, Twitter, and Amazon, among others. It entails a computational analysis of an individual's purchasing behavior, followed by the mining of his ideas on a company's corporate entity. User-generated content, such as reviews, ratings, and comments, can be analyzed for better business insights. E-commerce was investigated cognitively in this study. The twitter dataset used in this research was obtained from a dataset source. Then we must put the pre-processing strategies into action. The system is then constructed using NLP approaches. Then we must put the machine learning method, such as Logistic regression, into action. The accuracy is demonstrated by the experimental findings

**Keywords**— *Big Data, Machine Learning, Sentiment Analysis, NLP, Logistic Regression, Opinion Mining*

## I. INTRODUCTION

Big Data is a buzz word these days, but what is the difference between data and big data?. From few decades, there is a huge increase of software programs which resulted in huge amounts of data generation in this digital age. Big data analytics, a complex process that involves analysing massive volumes of data in order to identify hidden patterns, customer preferences, correlations and market trends that can aid businesses in making better decision.

Sentiment analysis aids in the examination of this opinionated data and the extraction of key insights that will aid other users in making decisions. Sentiment analysis systematically discovers, extracts, and subjective info analysis from textual data using text analysis, machine learning(ML) and natural language processing(NLP). Reviews, survey findings, healthcare media, online media, and other bases to determine customer mood or opinion. The purpose of this research using sentimental analysis is to evaluate the insolence of a speaker, writer, or other topic in respect to a particular problem or contextual polarity to a certain event, discussion, forum, interaction, or among other things. The fundamental use of sentimental analysis in this current world is to determine the polarity of text at the feature, phrase, and document levels.

Because of growing internet usage, each user is more interested in expressing his or her viewpoint on the internet through various channels, resulting in the development of opinionated data on the internet. Social media data includes product reviews, movie reviews, airline reviews, cricket

reviews, hotel reviews, employee involvement, healthcare reviews, news and stories, etc. The technique of extracting and comprehending the sentiments defined in a written document is known as sentiment analysis. Consumers now have new opportunities to express their thoughts about products, people, and places thanks to the explosion of data on social network platforms like Twitter, Facebook, and LinkedIn.

The user's viewpoint is always expressed as textual data. Using social media and online commerce platforms, millions of text messages are sent every day. Plagiarized Investigating and evaluating public opinion is a critical responsibility. To identify whether the sentiment of an opinion is good, negative, or neutral, NLP with artificial intelligence capacity and text analytics are applied. Opinion mining and sentiment analysis aren't limited to a single platform or domain. It has expanded to all social media networks, healthcare, management, the economics, and many other areas, and it's proven to be extremely beneficial to the growth of many businesses and organization.

Sentiment analysis and sentiment categorization are two approaches used in opinion mining. Both have their own distinct characteristics, but they can be used as interchangeable polarity at times. By providing class labels to the document or segment, sentiment classification reflects the sentiment orientation. Sentiment orientation is a text classification method that categorizes text data according to the sentiment orientation of the opinions expressed. Sentiment orientation denotes polarity of a true or incorrect view based on subjectivity. The method of determining whether a particular text or review data is subjective or objective is known as subjective analysis. Several sentiment analysis approaches are discussed in this study. Despite the fact that a lot of scholars have proposed articles in this field, there is still a need to increase the comprehension of sentiment analysis and the accuracy.

Sentiment analysis comes in handy in a variety of scenarios. However, due to the intricacy of human language, this is a highly challenging task. There are various variations, such as grammatical, cultural, and so on. "My order has been delayed," for example, can be easily misinterpreted by humans. "Excellent." Similarly, the word "thin" can be construed positively with respect to a laptop but negatively in the context of an apartment wall. As a result, in order to make the best decision, sentiment research

must be more company relevant. In this modern age, the utilization of E-commerce has skyrocketed. People are preferring to shop online rather than shopping in stores. In e-commerce, data created by customers from product ratings and reviews can help to verify the product's validity and publicize it.

These reviews and rating always help consumers to decide and confirm whether to buy any particular product or not. Such content could include good or negative feedback from customers who have used the product before. E-commerce companies can benefit from an precise analysis of this user- generated information to get insights and better understand their customers' needs and intensions. Machine Learning Algorithms can assist us in creating realistic visual representations of this type of consumer behaviour. Such visual representations would aid in a more in-depth analysis of the dataset in order to realize and draw exact conclusions about consumer behaviour on the e-commerce platform. Natural Language Processing (NLP), a machine learning technique, can evaluate text data and detect positive or negative customer feedbacks. This approach is also called Sentiment Analysis.

As a result, customer behaviour, such as purchasing intentions, influenced by data presented by ratings. With this smart analysis, real-time recommendations may be generated to give a tailored shopping experience, encouraging customers to buy more and so increasing the total profit of these businesses. The goal of this research is to look into characteristics that aid in fine-tuning organizational directions using bulk amounts of user-generated content.

## II. RELATED WORK

T. Ahmad, A. Ramsay, and H. Ahmed [12] mentions that at first glance, assigning sentiment labels to documents appears to be a simple multi-label classification operation. Various strategies have been employed for this goal, but the current state-of-the-art systems utilize deep neural networks (DNNs). As a result, it appears that typical machine learning techniques, such as these, will be a viable option. We present an alternate method that involves constructing a weighted vocabulary of sentiment phrases using probabilities, then changing the lexicon and finding appropriate thresholds for each class. This technique outperforms DNNs and other conventional strategies, as demonstrated. DNNs aren't a panacea, and we believe that focusing on type of data you're attempting to learn from is more important than experimenting with increasingly powerful general-purpose machine learning algorithms. We show that the strategy of deliberately generating a lexicon outperforms both the DNN algorithms described above, and we speculate on why this is the case.

A. Bandi and A. Fellah [2] describes that the use of social media is growing at a rapid pace. The impact of societal changes is bending in the direction of people's expressed thoughts on social media. Because of its real-time nature, Twitter has gotten a lot of attention. By constructing Socio- Analyser, we study current societal trends in the me-too movement. Socio-Analyser was

implemented using our four- phase approach. The data world website yielded a total of 393,869 static and stream data, which was analyzed using a classifier. The data is identified and classified into three groups by the classifiers (positive, neutral, and negative). The majority of people's opinions are neutral, according to our findings. The next highest number of persons disagreed, and the results were compared with TextBlob. We validate 765 weather-related tweets and extrapolate the results to Me-too data. When considering neutral tweets as positive, the precision values of Socio-Analyser and TextBlob are 70.74 percent and 72.92 percent, respectively. We've seen these movements turn into social media protests or rallying calls at times. Interviews, surveys, and opinion polls are used by sociologists to investigate people's reactions to social movements. However, this will be a time-consuming process with only few samples available. Using machine learning techniques, we combine data science along with sociology to evaluate vast amounts of data retrieved from social media. We used hashtags (Metoo) and API calls to mine Twitter data and analyses the results.

X. Han, J. Wang, M. Zhang, and X. Wang [21] demonstrates that Corona Virus Disease (COVID-19) epidemic is a serious global public health crisis. Nowadays, general population obtains information and expresses their thoughts and feelings through online. This study investigated public opinion in China at the primary stages of COVID-19 by looking at Sina-Weibo (China's version of Twitter) conversations in terms of location, time, and topic. The spatial distribution for COVID-19 related Weibo messages was analyzed, as well as the temporal variations within one-hour intervals. Using the random forest technique and latent Dirichlet allocation model, a classification model and a topic extraction was constructed to identify seven COVID-19- relevant themes hierarchically and thirteen sub topics from Weibo messages. The findings show that the number of Weibo texts fluctuated throughout time for various topics and sub-topics, correlating to the event's various stages of growth. There is a link between the COVID-19 epidemic's development at the real world and the countless daily conversations on Weibo.

U. Naseem, S. K. Khan, M. Farasat, and F. Ali [25] provides a deep review in their research demonstrates that on utilizing NLP to detect abusive language on Twitter. In this study, a survey was conducted using several methodologies, and research was conducted on the many sorts of abusive language used in social media, as well as why it is significant. How it was recognized on real-time social media platforms, as well as the performance measures employed by researchers to assess the effectiveness of users' detection of abusive language on Twitter. This study organizes and illustrates the current position of this topic by providing an organized evaluation of previous techniques, including methods, essential characteristics, and fundamental algorithms. The study also discusses the complexities of the hate speech concept, which can be found in a variety of stages and circumstances. This study

has a clear societal impact potential, particularly in digital media and online networks. Finally, it can serve as a resource for other academics looking for material related to their field of study in the identification of abusive language on Twitter.

U. Naseem, I. Razzak, S. K. Khan, and M. Prasad [19] through their article mentions that Word representation was been always a significant research area during the history of (NLP) natural language processing. Understanding such complicated text data is critical, considering how rich it is in information and how extensively is used in a variety of applications. In this review, we look at several word representations models and their capacity of expression, from ancient to modern-day. We explain a range of text representation approaches and model designs, including SOTA LMs, which have flourished in the setting of NLP. Here, bulk amount of data/text can be converted into real vector representations that capture the exact semantic information using these approaches. Furthermore, several machine learning algorithms can use such representations for a range of NLP-related tasks. Finally, this survey briefly examines the most often use evaluation metrics, DL, ML based classifiers, and word embedding applications in various NLP tasks. Text data is a rich source of info and allows for greater exploration of key insights not possible with quantitative data. The goal of several NLP techniques is to get a human-like understanding of the text.

U. Naseem, I. Razzak, K. Musial, and M. Imran [22] describes that the fast development of mobile devices, along with the rise of the Internet, there is widespread use of social media, resulting an explosion of short informal writings. Although sentiment analysis of these articles is significant of reasons, it is often seen as a difficult undertaking because they are typically brief, informal, noisy, and packed with ambiguous vocabulary, such as polysemy. Furthermore, most contemporary sentiment analysis methods are based on clean data. We mention DICET, a transformer-based sentiment analysis approach that encrypts represent from a transformer and uses intelligent contextual embedding at a deep level to improve tweet quality in by reducing noise while considering word sentiments, polysemy, syntax, and semantics. To determine the emotion of a tweet, we use a bidirectional long- and short-term memory network. We put the suggested framework through thorough testing on three benchmark datasets to verify its performance. DICET outperforms the present state of art in sentiment categorization, according to the findings.

### III. METHODOLOGY

The main objective of the proposed system is to implement machine learning algorithm and NLP - Natural language processing approaches. To increase overall performance through algorithm classification or categorization. To effectively classify and predict the tweets and to classify and determine whether the tweets are positive or negative.

The twitter dataset was used as input in this current system. Using the API key, the input data was retrieved

from the dataset repository. The data pre-processing phase must then be implemented. To avoid inaccurate predictions, we must manage missing values and encode the label for input data at this step. The sentiment must next be analyzed using natural language processing. We must delete punctuation, stop words, and stemming in this phase. The dataset must next be divided into test and training groups. The ratio is used to separate the data. The majority of the data will be present in the train. A reduced fraction of data will be present in the exam. The training part of this model will evaluate it, while the testing part will be utilized to predict it. The vectorization must then be implemented. To construct feature vectors, the text must be encoded as integers or numeric values. The classification algorithm (i.e., machine learning) must then be implemented. Logistic regression is one of the parts of machine learning algorithm. Finally, this experiment suggest that performance criteria like accuracy, precision, and recall are important.

#### A. Data Selection

The input data was collected from dataset repository. In our process, the twitter dataset is used. The data selection is a kind of process to find the tweets to be either positive or negative. The input dataset was taken from dataset repository such as UCI repository. The dataset contains the 'n' number of tweets and id.

#### B. Data Pre-Processing

Data pre-processing is defined as cleansing / removing all the unnecessary data from the dataset. Pre-processing data transformation operations are employed to convert the dataset into a machine-learning-friendly structure. This process also involves cleaning the dataset by deleting any extraneous or corrupted data that could impair the dataset's correctness, making it more efficient. Removal of Missing

Data: During this process, all the null values like Nan values and missing values are replaced with zero (0). Duplicate and missing values would be removed and data will be cleaned by any of the abnormalities. Encoding of categorical data: This can be stated as finite set of labeled values for a set of variables. Numerical input and output variables are required by the majority of machine learning algorithms.

#### C. NLP Techniques

NLP is one of machine learning (ML) technique, with the ability of a computer to understand, grasp, adapt, and possibly synthesize human language.

Cleaning (or pre-processing) data usually entails number of steps.

Remove punctuation: Punctuation provides grammatical context to help us interpret a sentence.

Tokenization: Tokenization is the act of dividing big chunks of text into smaller chunks, such as phrases or words. It gives shape to the previously unstructured text. e.g.: Blata o Blomo-> 'Blata', 'o', 'Blomo'.

Stemming: Stemming will help to reduce a word to its stem form.

**Padding:** Naturally, there will be sentences of various lengths in any raw text data. All neural networks, also, require inputs of the same size. Padding is used for this reason.

#### D. Data Splitting

During the process of machine learning, the data is needed so that learning can take place. In addition to the data required for training, test data are needed for the evaluation of the performance of the algorithm in order to see how well it works. In our current process, we have identified 60%-70% of the input dataset to be the training data and the remaining 30%-40% to be the testing data. Data splitting is kind of process to divide accessible data by two halves for cross-validation reasons. Here, we would take one part of the data to create a predictive model, while the other is used to assess the model's performance. Separating the data into data sets as training and testing is really an important part of evaluation in data mining models. Once the data is divided into a training set and a testing set, the majority of the data is used for training and a smaller piece is utilized for testing.

#### E. Data Classification

In our process, we have to implement machine learning algorithm such as logistic regression. The most common Machine Learning algorithms is logistic regression - A Supervised Learning approach, where a categorical dependent variable's output is predicted using logistic regression. A set of independent data variables are utilized for the prediction of the categorical variable depended.

#### F. Result Generation

The Final Result will be based on the overall classification and projection. The effectiveness of the suggested technique is assessed using a variety of metrics, including:

**Accuracy** - Accuracy of classifier refers to the ability of classifier. It will help in the successful prediction of the class label, and accuracy of predictor refers to how effectively a predictor can suggest the value of a predicted attribute for the incoming data

$$\text{Accuracy Classifier (AC)} = (TP + TN) / (TP + TN + FP + FN)$$

**Precision** - The no. of true positives added to the no. of the false positives and then be divided with the number of true positives to get Precision.

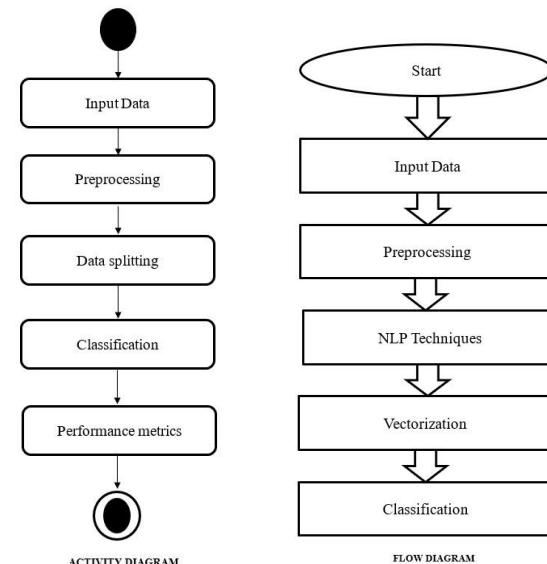
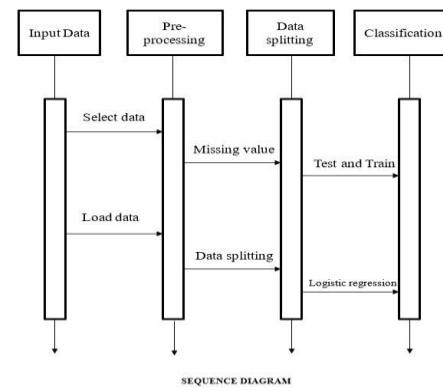
$$\text{Precision} = TP / (TP + FP)$$

**Recall** - The number of right results to divide by the number of results that should have been returned is referred to as Recall. Sensitivity is the term used in binary classification to describe recall. It can be considered as the likelihood of the query returning a relevant document.

$$\text{Recall} = TP / (TP + FN)$$

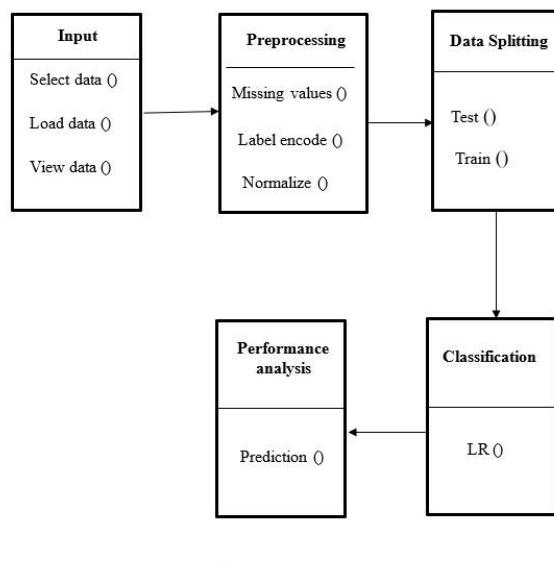
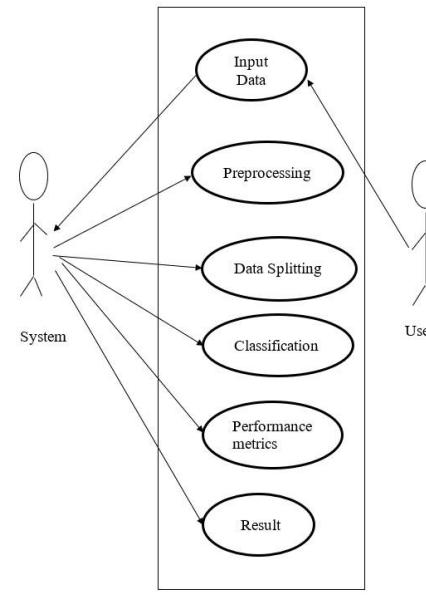
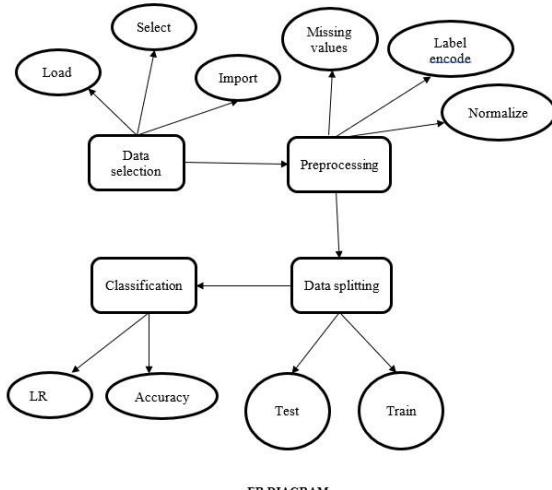
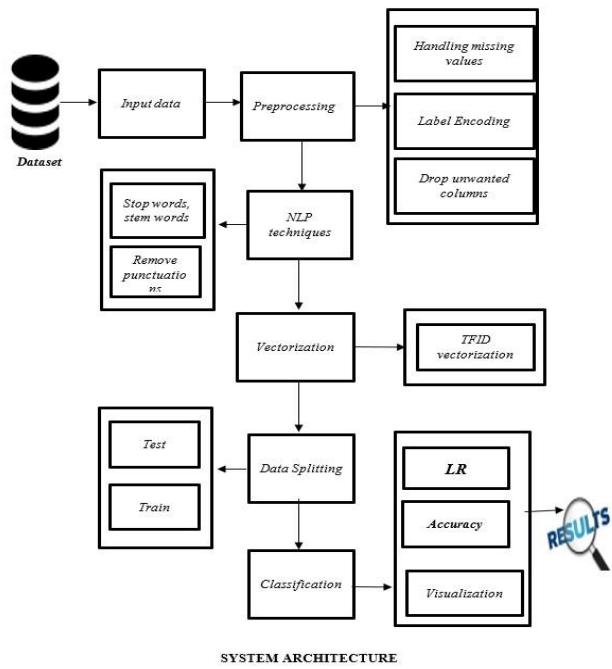
## IV. TWITTER DATA FOR RESEARCH

The main idea of this research is to determine if a user's view on a certain issue is neutral, negative, or favorable based on tweets sent on the Twitter network. Here, we use Sentiment analysis to solve this. Sentiment analysis employs natural language processing (NLP), computational linguistics, text analysis and biometrics to identify, extract, measure, and assess subjective data and emotional states. We employed sentiment analysis with Logistic Regression in system currently being implemented. Here, we are solving the binary classification problem using logistic regression classification. The outcome is often defined as 0 or 1 in models with a twofold situation. We utilized a sample Twitter data set received from the Twitter API for this research. More than thousand tweets are included in the Twitter API data collection where the size/number of the tweets can be increased or decreased based on the system configuration. We used the NLTK library in Python with VADER to forecast the sentiment of various tweets from twitter. VADER is a one among the tools for sentiment analysis that uses a lexicon and rules to analyze emotions expressed on social media. It analyzes the emotion of material which has both the positive and negative polarity.



## V. SYSTEM DESIGN / ARCHITECTURE

Below is system architecture, flow diagrams designed for the system is as follows:



## VI. APPLICATION

### A. Import Packages

We are importing all the python packages required for implementation

### B. Data Selection

We are collecting the size of tweets by giving the <No. of Tweets> for a Specific user by providing the <Username> from Twitter API. The data is then copied to the CSV file

```
=====
Data Selection
=====
0 narendramodi ... Addressing a programme in Prayagraj. #NariShaktiDeshKiShakti https://t.co/2njX6mz9zB
1 narendramodi ... Greetings to AP CM Shri @ysjaganGuru on his b...
2 narendramodi ... Interacted with leading CEO's from different s...
3 narendramodi ... बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...
4 narendramodi ... Money would be transferred to over 1 lakh bene...
5 narendramodi ... At the programme in Prayagraj tomorrow, Rs. 10...
6 narendramodi ... In the afternoon tomorrow, 21st December, I wo...
7 narendramodi ... Spoke with my friend, President Putin, to foll...
8 narendramodi ... Met the Foreign Ministers of Kazakhstan, Kyrgy...
9 narendramodi ... Congratulations to @srikidambi for winning a n...
9 narendramodi ... Birthday greeting to Union Minister Shri RK Si...
[10 rows x 4 columns]
```

### C. Reading the Input Data

Here, we are using pandas to read the data from CSV file for the Specific user and then writing the encoding field used in the read\_csv () command to convert the data into ASCII based alphabet characters. By using “error\_bad\_lines=False”, we are skipping the invalid rows from the data sheet.

### D. Data Pre-processing

During Data Pre-Processing we are transforming the raw data as a formatted data set and verifying for any missing values.

```
=====
Data Selection
=====
0 narendramodi ... Addressing a programme in Prayagraj. #NariShaktiDeshKiShakti https://t.co/2njX6mz9zB
1 narendramodi ... Greetings to AP CM Shri @ysjaganGuru on his b...
2 narendramodi ... Interacted with leading CEO's from different s...
3 narendramodi ... बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...
4 narendramodi ... Money would be transferred to over 1 lakh bene...
5 narendramodi ... At the programme in Prayagraj tomorrow, Rs. 10...
6 narendramodi ... In the afternoon tomorrow, 21st December, I wo...
7 narendramodi ... Spoke with my friend, President Putin, to foll...
8 narendramodi ... Met the Foreign Ministers of Kazakhstan, Kyrgy...
9 narendramodi ... Congratulations to @srikidambi for winning a n...
9 narendramodi ... Birthday greeting to Union Minister Shri RK Si...
[10 rows x 4 columns]
```

### E. NLP Techniques

Here, in this step, we are using the NLP – Natural Language Processing techniques to cleanse the data including updating the text to lower case for base lining and then applying stop words for filtering the unnecessary data. The regular expression (re) library should be introduced before the stemming procedure. Then with the split( ) function, the words of the sentences are then separated into parts and with the command sub( ), then the regular patterns we called in the regular expression to search for the number of iterations given. Then we apply tokenization to separate and classify parts of a string of input characters. It is to separate every word in the sentence.

```
=====
Before applying NLP techniques
=====
0 Greetings to AP CM Shri @ysjaganGuru on his b...
1 Interacted with leading CEO's from different s...
2 बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...
3 Money would be transferred to over 1 lakh bene...
4 At the programme in Prayagraj tomorrow, Rs. 10...
5 In the afternoon tomorrow, 21st December, I wo...
6 Spoke with my friend, President Putin, to foll...
7 Met the Foreign Ministers of Kazakhstan, Kyrgy...
8 Congratulations to @srikidambi for winning a n...
9 Birthday greeting to Union Minister Shri RK Si...
Name: Tweets, dtype: object
=====
```

```
=====
After applying NLP techniques
=====
0 greet AP CM shri ysjagan garu birthday. may al...
1 interact lead ceo' differ sectors. We discuss ...
2 बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...
3 money would transfer 1 lakh beneficiari mukhya...
4 At programme prayagraj tomorrow, rs. 1000 crore...
5 In afternoon tomorrow, 21st december, I would ...
6 spoke friend, presid putin, follow discuss rec...
7 met foreign minist kazakhstan, kyrgyz republic...
8 congratul srikidambi win histor silver medal. ...
9 birthday greet union minist shri RK singh ji. ...
Name: even, dtype: object
=====
Vocabulary size: 1613
```

Vocabulary size: 1613

Example:

Sentence:  
0 greet AP CM shri ysjagan garu birthday. may al...  
1 interact lead ceo' differ sectors. We discuss ...  
2 बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...  
3 money would transfer 1 lakh beneficiari mukhya...  
4 At programme prayagraj tomorrow, rs. 1000 crore...  
...  
194 grate peopl kongthong kind gesture. the govern...  
195 tune 11 AM tomorrow. #mannkibaat  
196 thi wonder know. I known mlkhattar Ji year som...  
197 जल जीवन मिशन का असरी उद्देश्य भी तो यही है. जि...  
198 If get opportunity, tast organ appl ladakh. th...  
Name: even, Length: 199, dtype: object

After tokenizing :  
[134, 455, 456, 135, 457, 42, 73, 221, 222, 223, 458, 16, 459]

### F. Sentiment Analysis

Here, in this steps we are using SentimentIntensityAnalyzer() from VADER, where VADER works in concurrence with NLTK for sentiment analysis on longer texts...i.e., decomposing paragraphs, articles/reports/publications, or novels into sentence-level analysis

### G. Data Splitting

Here, the data in the dataset is fragmented/splitted according to the random\_state ratio as training part and testing part as per the specified criteria

```
=====
Vocabulary size: 1613
=====
Example:
=====
Sentence:
0 greet AP CM shri ysjagan garu birthday. may al...
1 interact lead ceo' differ sectors. We discuss ...
2 बड़ी के लिए यह न केवल दिलचस्प कार्यक्रम होगा...
3 money would transfer 1 lakh beneficiari mukhya...
4 At programme prayagraj tomorrow, rs. 1000 crore...
...  

194 grate peopl kongthong kind gesture. the govern...  

195 tune 11 AM tomorrow. #mannkibaat  

196 thi wonder know. I known mlkhattar Ji year som...  

197 जल जीवन मिशन का असरी उद्देश्य भी तो यही है. जि...  

198 If get opportunity, tast organ appl ladakh. th...  

Name: even, Length: 199, dtype: object
=====
After tokenizing :
[134, 455, 456, 135, 457, 42, 73, 221, 222, 223, 458, 16, 459]
```

### H. Vectorization

Here, in this step we process the text input, and convert them into vectors. We used sklearn's CountVectorizer() object to extract all the word features converting the data to lower case and removing the stop words. Then fit and transform is performed for the input data every single time and converts the data points

### I. Classification

Now we are ready to apply our logistic regression model using sklearn for initializing, fitting and then predicting and further retrieving the accuracy score based on

the prediction We are going to display the classification review based compound score calculated from the tweets based on the analysis

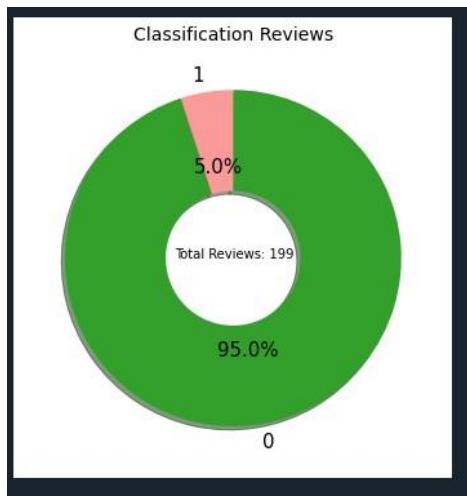
Logistic Regression					
Accuracy : 94.0 %					
	precision	recall	f1-score	support	
0	1.00	0.94	0.97	50	
1	0.00	0.00	0.00	0	
accuracy			0.94	50	
macro avg	0.50	0.47	0.48	50	
weighted avg	1.00	0.94	0.97	50	

#### J. Prediction

Here, we are going to identify whether the tweet is Positive/Negative/Neutral based on the Index number provided

Logistic Regression					
Accuracy : 94.0 %					
	precision	recall	f1-score	support	
0	1.00	0.94	0.97	50	
1	0.00	0.00	0.00	0	
accuracy			0.94	50	
macro avg	0.50	0.47	0.48	50	
weighted avg	1.00	0.94	0.97	50	

#### K. Visualization



#### CONCLUSION

Twitter become a common starting point for research initiatives requiring significant amounts of user-generated or user-driven data. Due to the tremendous value of big data in today's market, Twitter's free API only exposes a small portion of the entire data set that a paying user of the API would have the access. Although cost-effective, using

Twitter's API has only yielded about 1% of the full data set on occasion. We demonstrated our in-house created scalable Twitter platform.

Our platform has demonstrated that it can improve the amount/size of data available through Twitter's API as a dataset is considered adequate for our scientific research. Thus we conclude that, the twitter dataset from Twitter's API was taken as input. We have implemented the NLP techniques and classification algorithms (i.e.) such as machine learning algorithm which is logistic regression. Finally, the result shows that the accuracy for above mentioned algorithm and visualize the output in the form of graph. Then, analyze the tweets is positive or negative.

#### REFERENCES

- [1] R. Bhat, V. K. Singh, N. Naik, C. R. Kamath, P. Mulimani, and N. Kulkarni, "COVID-2019-outbreak:-The-disappointment-in-Indian-teachers," Asian J. Psychiatry, vol. 50, Apr. 2020, Art. no. 102047.
- [2] Bandi and A. Fellah, "Socio-analyzer:-A-sentiment-analysis-using-social-media-data," in Proc. 28th Int. Conf. Softw. Eng. Data Eng., in EPiC Series in Computing, vol. 64, F. Harris, S. Dascalu, S. Sharma, and R. Wu, Eds. Amsterdam, The Netherlands: EasyChair, 2019, pp.
- [3] E. Karafillakis, R. Preet, Depoux, S. Martin, A. Wilder-Smith, and H. Larson, "The-pandemic-of-social-media-panic-travels-faster-than-the- COVID-19-outbreak," J. Travel Med., Apr. 2020, Art. no. taaa031.
- [4] C. Aggarwal and C. K. Reddy, Data-Clustering:-Algorithms-and- Applications. Boca Raton, FL, USA: CRC Press, 2013.
- [5] F. Barbieri and H. Saggion, "Automatic-detection-of-irony-and-humour-in-Twitter," in Proc. ICCC, 2014, pp. 155–162.
- [6] M. Blei, A. Y. Ng, and M. I. Jordan, "Latent-Dirichlet-allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, Jan. 2003.
- [7] P. Boldog, T. Tekeli, Z. Vizi, A. Dénes, F. A. Bartha, and G. Röst, "Risk-assessment-of-novel-coronavirus-COVID-19-outbreaks- outside-China," J. Clin. Med., Feb. 2020.
- [8] M. De Choudhury, E. Horvitz and S. Counts, "Predicting-postpartum- changes-in-emotion-and-behavior-via-social-media," in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., Apr. 2013,
- [9] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, "TwitPersonality:-Computing-personality-traits-from-tweets-using-word-embeddings-and-supervised-learning," Information, May 2018.
- [10] M. E. El Zowalaty and J. D. Järhult, "From-SARS-to-COVID-19:-A- previously-unknown-SARS-related-coronavirus-(SARS-CoV-2)-of- pandemic-potential-infecting-humans—Call for a one health approach," One Health, vol. 9, Jun. 2020, Art. no. 100124.
- [11] N. Bandi and J. Siddique, "Personality-assessment-using-Twitter- tweets," Procedia Comput. Sci., Sep. 2017.
- [12] T. Ahmad, H. Ahmed and A. Ramsay, "Detecting-emotions-in-English-and-Arabic-tweets," Information, Mar. 2019.
- [13] L. Márquez, X. Carreras, "Boosting-trees-for-anti-spam-email-filtering," 2001, arXiv:cs/0109015. J. P. Carvalho, H. Rosa, G. Brogueira, and F. Batista, "MISNIS: An intelligent platform for Twitter topic mining," Expert Syst. Appl., vol. 89, pp. 374–388, Dec. 2017.
- [14] B. K. Chae, "Insights-from-hashtag-#supplychain-and-Twitter- analytics:-Considering-Twitter-and-Twitter-data-for-supply-chain- practice-and-research," Int. J. Prod. Econ. Jul. 2015.
- [15] Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting- elections-with-twitter:-What-140-characters-reveal-about-political- sentiment," Icwsm, 2010.
- [16] K. Toutanova, M.-W. Chang, J. Devlin, K. Lee, "BERT--- Pretraining--- of-deep-idirectional-transformers-for-language-understanding," in Proc. Conference of North America un. 2019
- [17] M. S. Kamel, K. M. Hammouda, "Efficient-phrase-based-document- indexing-for-Web-document-clustering," IEEE Trans. Knowl. Data Eng., vol. 16, no. 10, pp. 1279–1296, Oct. 2004.
- [18] Fung et al., "Pedagogical--demonstration-of--Twitter-data--- analysis:-A-case---study-of-world-AIDS-day,-2014," Jun. 2019.

[19] Naseem, U., Razzak, I., Khan, S. K., & Prasad, M. (2020). A- Comprehensive---Survey---on-Word-Representation-Models:- From- Classical-to-State-Of-The-Art-Word-Representation- Language- Models. arXiv preprint arXiv:2010.15036.

[20] S. Lehal, V. Gupta, G. "A-survey-of-text-mining-techniques-and-applications," J. Emerg. Technol. ug. 2009.

[21] X. Han, J. Wang, M. Zhang, and X. Wang, "Using-social-media-to- mine-and-analyze-public-opinion-related-to-COVID-19-in- China," Int. J. Environ. Res. Public Health, vol. 17, no. 8, p. 2788, Apr. 2020.

[22] Naseem U., Razzak I., Musial K., Imran M – "Transformer-based - Deep-Intelligent-Contextual-Embedding(DICE) for Twitter- Sentiment Analysis; Future Generation-Computer-System.

[23] R. Xia, S. Li, C. Zong – "Ensemble-of-feature--sets- and-- classification-algorithms-for--sentiment-classification,"- Information- Sciences,-vol.-181,-no.-6,-pp.-1138-1152,- 2011/03/15/-2011.-

[24] R.-Sharma,-S.-Nigam,-and-R.-Jain,-"Opinion-mining-of-movie- reviews-at-document-level,"-arXiv-preprint-arXiv:-1408.3829,- 2014.-

[25] Naseem-U.,-Khan-S.K.,-Farasat-M.,-Ali-F.: Abusive-language- detection:-A-Comprehensive-Review:-Indian-Journal-of-Science- and- Technology.,-IJST,2019,-DOI:- 10.17485/ijst/2019/v12i45/146538.