

Covid19 Sentiment Analysis for a Large Scale Benchmark Twitter Data Set

Mamidi Prathima Deva Sena¹, Dr. G. BabuRao²

¹Student, Department of Computer Science, Visakhapatnam, Andhra Pradesh, India

²Assitant Professor, Department of Computer Science, Visakhapatnam, Andhra Pradesh, India

Abstract— Sentiment analysis over Twitter offer formation a fast and effective way to monitor the publics’ feelings towards their brand, business, directors, etc. A wide range of characteristic and methods for training sentiment classifiers for Twitter datasets have been researched in recent years with varying results. Twitter offers formations a fast and productive way to analyse customers’ perspectives toward the critical to success in the market place. Expanding a program for sentiment analysis is an approach to be used to reckoning measure customers’ perceptions. Sentiment analysis is also known as “opinion mining” or “emotion Artificial Intelligence” and suggests to the utilization of natural language processing (NLP), text mining, computational linguistics, and bio measurements to methodically recognize, calculate, evaluate, and examine emotional states and subjective information. The system is developed, the dataset was collected from twitter. The employees work from home tweets dataset as input was collected from twitter by using API key. Then, we have to analyse the sentiment by using the NLP techniques, text classification and deep learning algorithm. The experimental results show that, the performance metrics such as accuracy and analyse the sentiment based on sentiment analyser into positive, negative and neutral.

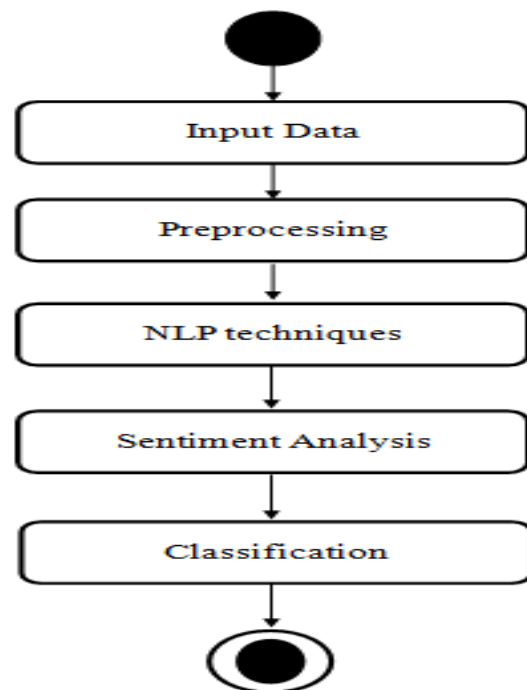
I. INTRODUCTION

Sentiment analysis is the process of instinctive perceive whether a text segment contains emotional or assertive content, and it can further more classify the text’s contradiction. Twitter sentiment classification aims to classify the sentiment contradiction of a tweet as positive, negative or unbiased.

Initially it is performed extensively in existing approaches, especially in machine learning-based approaches. However, few studies focus on the consequence of pre-processing method on the execution of Twitter sentiment analysis. This paper

focus on inspect various pre-processing methods for elevating the execution of Twitter sentiment analysis. The experimental results show that the precision of sentiment organization may be notably improved using relevant features and representation after pre-processing.

II. FRAMEWORK



ACTIVITY CASE DIAGRAM

III. METHODOLOGY

- Data selection
- Data preprocessing
- NLP techniques
- Classification
- Result Generation

1 Data Selection:

- The input data was collected from twitter by using the API key.
- In our process, work from home employees tweet dataset is used.
- The data selection is the process of analyzing the sentiment into positive, neutral and positive.

2 Data Preprocessing:

- Data pre-processing is the process of removing the unwanted data from the dataset.
- Pre-processing data modification operations are used to transform the dataset into a structure suitable for machine learning.
- This step also includes cleaning the dataset by removing immaterial or corrupted data that can affect the accuracy of the dataset, which makes it more well organized.

3 Nlp Techniques:

- NLP is a field in machine learning with the ability of a computer to understand, analyze, manipulate, and potentially generate human language.
- Cleaning the data typically consists of a number of steps:

4 Data Splitting:

- During the machine learning process, data are needed so that learning can take place.
- In addition to the data required for training, test data are needed to evaluate the performance of the algorithm in order to see how well it works.
- In our process, we considered 70% of the twitter dataset to be the training data and the remaining 30% to be the testing data.

5 Classification:

- In our process, we have to implement the deep learning algorithms such as LSTM.
- LSTM Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning.

6 Result Generation:

The Final Result will get based on the overall classification and prediction. The performance of this proposed slant is evaluated using some measures like,

Accuracy

Accuracy of classifier refers to the ability of morpheme. It forecasts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted quality for a new data.

$$AC = (TP+TN) / (TP+TN+FP+FN)$$

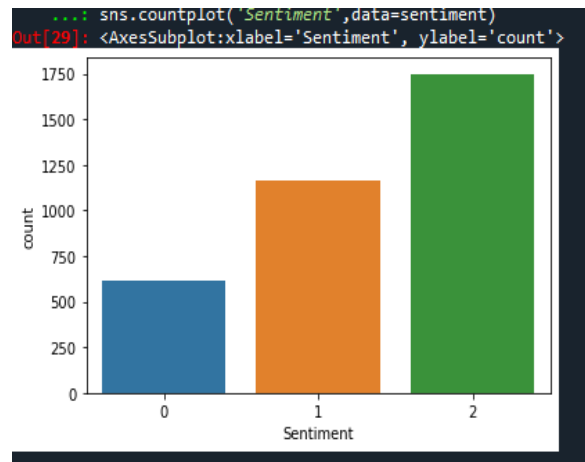
Then, the result shows that, analyse and visualize the sentiment into positive, negative and neutral in the form of graph.

V. RESULTS

```

=====Checking Missing Values=====
Date          3527
Reviews       3527
Reviews title  3527
City          3527
Sentiment     3527
dtype: int64

=====After Handling Missing Values=====
Date          0
Reviews       0
Reviews title  0
City          0
Sentiment     0
dtype: int64
    
```



VI. CONCLUSION

We conclude that, the employees twitter dataset was taken as input. The input dataset was collected from twitter by using API key. We are implemented the NLP techniques and deep learning algorithm such as Long short term memory (LSTM). Then envision the

sentiment into positive, unbiased and negative in the form graph.

REFERENCE

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.
- [2] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, "TwitPersonality: calculate personality traits from tweets using word immerse and supervised learning," *Information*, vol. 9, no. 5, p. 127, May 2018.
- [3] X. Carreras and L. Màrquez, "Boosting trees for anti-spam email filtering," 2001, arXiv:cs/0109015. [Online]. Available: <https://arxiv.org/abs/cs/0109015>.
- [4] J. P. Carvalho, H. Rosa, G. Brogueira, and F. Batista, "MISNIS: An intelligent objective for Twitter topic mining," *Expert System Appl.*, vol. 89, pp. 374–388, Dec. 2017.