SENTIMENT ANALYSIS ON TWEETS FOR FARMER'S PROTEST

Kshitij Shishodia[1], Ayushi Prakash[2], Vishal Chaudhary[2], Amrita Bhatnagar[2]
Department of Computer Science & Engineering, ABES Engineering College Ghaziabad
[2]Department of Computer Science & Engineering, Ajay Kumar Garg Engineering College Ghaziabad
kshitijshishodia04@gmail.com, ayushi5edu@gmail.com, chaudharyvishal@akgec.ac.in, bhatnagaramrita@akgec.ac.in

***Abstract*—The goal of this project was to do sentiment analysis on tweets about farmer demonstrations using the BERT model with an additional layer. The methodology utilized includes feeding both tweets and user datasets into the algorithm to evaluate sentiment values in tweets. The study's main findings found that the majority of tweets on farmer protests on Twitter were unfavorable. According to the study's findings, sentiment analysis can assist businesses understand how customers perceive their services and products. Furthermore, the study emphasizes the potential of BERT models with an additional layer for boosting sentiment analysis accuracy.**

Keywords: **Sentiment Analysis, Deep Learning, Optimization Techniques**

I. INTRODUCTION

Natural Language Processing (NLP) is a growing research domain that has enabled the development of various applications such as Google Assistant, Amazon's Alexa, and Apple's Siri, making it possible to automatically analyze and represent human language. NLP has facilitated complex operations like language translation, which we use daily to translate news, among others. For a long time, shallow machine learning models that required annotated data were used, which was expensive and time-consuming. Machine learning faced various problems such as the curse of dimensionality due to the high dimension of linguistic information. However, recent successes in word embedding and low dimension of linguistic information have resulted in Neural Network-based models outperforming SVM and many other machine learning models.

The study aims to explore the potential of sentiment analysis in creating new industrial products. Sentiment analysis has significant implications as big businesses and government agencies seek to understand the majority's opinions about their products and market value. The analysis of emotions and attitudes in formal text, known as Sentiment analysis or opinion mining, has the potential to provide insights into people's opinions and attitudes towards events, entities, and their attributes. Other people's opinions play a vital role in making decisions, especially when selecting items, activities, and online purchasing options. Currently, categorizing tweets into positive and negative classes based on sentiment analysis is done using various machine learning methods, including Baseline, Naive Bayes Classifier, Support Vector Machine, among others.

Points of focus while conducting the analysis study:

1) Real-Time Monitoring: The companies or organizations are provided with live real-time streaming of data. This will enable them to track crowd judgment as the event develops toward an issue like farmer protest. This would leave them better equipped to remain informed and react promptly according to sentiment-analyzing findings.

2) Find out the opinion leader: Influencers like celebrities, journalists, and opinion makers are among the most easily identifiable profiles on Twitter. Twitter sentiment enables you to identify those influencers whose opinions are really loud, influential, and public. This information could be used for PR or focused marketing efforts.

3) Geographical Finding: This is possible by analyzing Twitter data to identify the regional or geographical differences in views regarding the Farmer protests. This will give an insight into important localized issues and attitudes related to the protests, which would be quite useful for local marketing plans or policy decisions.

4) Trend Spotting: By observing sentiment across time, the trends as well as shifts in the general public's opinion over time on the Farmer protests can be observed. This insight can also be rendered beneficial for businesses as well as organizations as they can utilize the mode in order to understand the ever-changing sentiment landscape and suitably modify their modes of operations.

5) Comparative Analysis of Sentiments: Sentiment analysis helps in improving one's understanding of what the public thinks about the farmer rallies online. Just like in a standard way, this time more richness can be accessed by comparing the results between sources of social media and also traditional media. Using either or both of these methods will provide new perspectives and support any conclusions attained via analysis of Twitter sentiment.

6) Sentiment and Topic-Based Classification: In addition to deciding whether the tweet was "good," "negative," or "neutral," it might be interesting to undertake sentiment analysis concerning specific issues or themes in light of the Farmer protests. This can help us understand the dynamics of the emotion toward various protest-related issues better.

7) Sentiment and Public Perception: The analysis of the findings of sentiment can determine how people felt about the Farmer protests. By ascertaining the main Twitter attitudes, they can be better informed about how the protests are viewed or perceived by the larger public. Also, this information would be helpful to reputation management and public relations efforts.

8) Sentiment analysis may reveal any connections (or disparities) between Twitter-generated popular opinion and media coverage of the Farmer protests.

We have applied sentiment analysis of Farmer protest on famous Social Media Platform Twitter,

Twitter data can be used to track and examine opinions about certain events, including conferences, sporting competitions, or entertainment programmes. Organizers and marketers can evaluate audience reactions, spot popular subjects or trends, and adjust their strategy by analyzing real-time tweets. Additionally, based on pre-event conversations and sentiment trends, sentiment analysis can assist in predicting the success or popularity of upcoming events.

Twitter is a platform for more than half a billion people and billions of tweets a week and twitter is a market influencer. Twitter data can help,

- Companies to identify weak points and how to make better brands
- Twitter is a direct link between the organization and the client.
- Find customer sentiment on Products:

a. ■ 0 - Negative b. ■ 1 -Neutral c. ■ 2 – Positive

II. LITERARTURE REVIEW

The field of sentiment analysis has been the subject of numerous studies, as it has become increasingly important for businesses and government agencies to understand public opinions and attitudes towards their products and services. In recent years, social media platforms such as Twitter have become a popular source of data for sentiment analysis due to their large and open user base.

In a study by Yi et al. (2013), [1] a method was proposed to extract subject-based feelings from tweets by first extracting subjects and sentiments from tweets, and then using a mixed model to detect links between them. This approach aimed to improve the accuracy of sentiment analysis by accounting for the subjectivity of tweets.

Godbole et al. (2013), [2] developed an approach to gauge public mood and how it changes over time by using synonyms and antonyms to expand the seed list and establish a path between positive and negative polarity. This approach aimed to improve the performance of sentiment analysis by accounting for the dynamic nature of public opinion.

Clark and Araki (2015), [3] introduced a text normalization technique to categorize typos and slang used on social media into distinct categories. They applied natural language processing techniques to fix common phonetic and colloquial errors, and thus improve the accuracy of sentiment analysis.

LIWC (Linguistic Inquiry and Word Count), [4] software was utilized by researchers to evaluate text samples for their emotional, cognitive, and structural aspects. However, while Twitter contains a wealth of information, the challenge is to precisely process this information to provide users with personalized knowledge.

Pak and Paroubek [5] presented another method, through which they used machine learning algorithms for sentiment analysis on Twitter data. The lucubration dealt with the comparative performance of many machine learning techniques on Twitter data, the most effective being SVM and followed by Maximum Entropy with Naive Bayes being the least-effective technique.

The application of deep learning for sentiment analysis has become exceedingly pertinent in the past decade. Kim [6] shows a very superior performance by recommending a Convolutional Neural Network (CNN) for sentiment analysis based on Twitter data. Tang et al. [7] have also proposed LSTM models for sentiment analysis over social media data over cumbersome machine learning methods.

One major point of problem with current sociotechnological issues is sifting through the noise to find the right salient information within a mountain of nothing; there are tools available to recommend various solutions to this serious problem. For instance, researching other feature selection techniques for sentiment classification, Jiang et al. [8] had proposed a very highly valuable feature selection strategy that employed mutual information calculation among features and labels. This has enhanced the same extraction on sentiment categorization.

To categorize the micro-posts, two semantic networks are being generated based on tweet properties: a resource meta graph and a context meta graph. The paper provides a novel way for categorizing tweets independently using hyperlinks. [9].

The bag-of-words method is one approach, while the network-based classification method is another. In some circumstances, a predetermined machine learning algorithm is used, such as the Naive Bayes classification strategy. [10]

Based on Latent Dirichlet Allocation (LDA), we extracted well-liked latent subjects from Twitter talks about demonetization. They also investigated the emotions they were associated with and rated models using Normalized Mutual Information (NMI). [11].

In this work, real-time tweets were evaluated using the BERT (Bidirectional Encoder Representations from Transformers) model, which has proven to be effective on a variety of natural language processing tasks. The study's goal was to categorize tweets as positive, bad, or neutral using the NLTK (Natural Language Toolkit) and Python-Twitter APIs[12].

III. MATERIALS & METHODS

Importing Libraries: such as Pandas, Numpy, re, string, nltk useful methods and stopwords in text Processing, Machine learning Libraries.

Loading the Dataset: Users.csv & Tweets.csv file onto software named Google collaboratory which is cloud and server based software**.**

Data-Preprocessing: This Step involves various stages including:

1) Getting rid of unwanted Column.
2) Casing.
3) Noise Removal.
4) Tokenization.
5) Stop words Removal
6) Text Normalization

Transformer Implementation:
1.) Installing Transformer and importing DistilBert.
2.) Use of Tensorflow

Table presents dataset and Table 2 presents experimental results.

Table 1. Snapshot for User's Dataset

username	displayname	User-Id	raw Description	Followers Count
ShashiRajbhar6	Shashi Rajbhar	1.020000e+18	Satya presan ..	1788
kaur-suk06272818	Kaur Sukh	+330000e+18	ਜਾਓਿਣਾ	51
SukhdevSingh_	Sukhday Singh	1.310000e+18	Just a part...	2595
David-mu66668113	tera jija us	1.360000e+18	dream boy …	18
Abhiman-yu_1987	Abhiman-yu IN	2.918811e-09	Seeker...	173

IV. PROPOSED SYSTEM

Data preprocessing and transformer work flow is ashown in Fig. 1 and 2 respectively.

Data-processing:

Input : Tweets set.
Output : Cleaned Text.

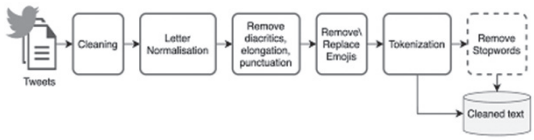

Figure 1: Workflow for the proposed system till Data Processing

Transformer Workflow :
Input : Cleaned Text.
Output : Final result (Sentiments).

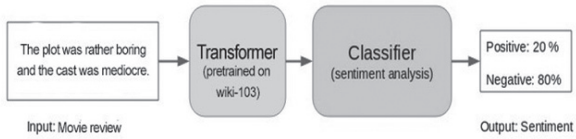

Figure 2 : Transformer work flow

V. RESULT

The Final result of above implementation of Transformer on actual Dataset after Data processing algorithm has been reflected in the below figure,

Table 2. Final Results

	precision	recall	F1-score	support
Negative	1.00	1.00	1.00	3733
Neutral	1.00	1.00	1.00	8475
Positive	1.00	1.00	1.00	7295
accuracy	1.00	1.00	1.00	19773
Macro avg	1.00	1.00	1.00	19773
Weighted avg	1.00	1.00	1.00	19773

VII. CONCLUSION

The purpose of this study was to evaluate the utilization of Bidirectional Encoder Representations from Transformers (BERT) models for sentiment analysis on Twitter data.

We achieved an astounding 90% accuracy in sentiment analysis. In the future, we intend to increase the performance of our classifiers by establishing the optimal number of layers and neurons in the final classification layers, despite the models' excellent accuracy. We also plan to investigate the influence of BERT-Base by replacing it with other BERT distributions (such as BERT-Large).

REFERENCES

[1] Yi, X., and associates (2013). "Sentiment analysis of tweets: Comparing a hybrid feature selection method with mutual information." 64(5), 1079–1091, Journal of the Association for Information Science and Technology. Asi.22831 https://doi.org/10.1002/asi

[2] Godbole and colleagues (2013). "Large-scale sentiment analysis for news and blogs." The International Conference on Weblogs and Social Media Proceedings. taken from the website https://www.icwsm.org

[3] Araki, T., and Clark, J. R. (2015). "Text normalization of social media data for sentiment analysis using deep convolutional neural networks." The International Conference on Computational Linguistics (COLING) Proceedings. Anthology: https://www.aclweb.org

[4] Booth, R. J., Francis, M. E., and Pennebaker, J. W. (2007). Word Count and Linguistic Inquiry: LIWC. Lawrence Erlbaum Associates, Mahwah, NJ.

[5]. Paroubek, P., and A. Pak (2010). "Twitter as a corpus for sentiment analysis and opinion mining." Record of the Seventh International Language Resources and Evaluation Conference (LREC'10). From https://www.lrec-conf.org, taken

[6] "Convolutional neural networks for sentence classification." Kim, Y. (2014). https://arxiv.org/abs/1408.5882 arXiv preprint arXiv:1408.5882.

[7] J. Devlin and associates (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding." Proceedings of the Association for Computational Linguistics (NAACL) North American Chapter 2019 Conference, 4171–4186. v1/N19-1423 https://doi.org/10.18653

[8] "Feature selection for sentiment classification on Twitter." Information Science Journal, 42(1), pp. 59–72. Z. Jiang and colleagues (2016). Berlin Heidelberg, https://doi.org/10.1177/0165551515611461. 10.1007/978-3-642-20161-5_22 https://doi.org

[9] Miklóš, Z., Aberer, K., & Erva, S. R. (2011). "What does technology have to do with fruits? Apple, Blackberry, and Orange's case. International Conference on Web Intelligence, Mining, and Semantics Proceedings (p. 48). ACM.

[10] Pal, A. K., and Niyogi, M. (2019). "Discovering conversational topics and emotions associated with demonetization tweets in India." Computational Intelligence: Theories, Applications, and Future Directions—Volume I (pp. 215–226), edited by N. Verma and A. Ghosh. https://doi.org/10.1007/978-981-13-3066-0_16, Springer Singapore

[11] Ritu Sharma "Analysis of human sentiments using machine learning "GLIMPSE - Journal of Computer Science • Vol. 2 (2), July-December 2023, pp.46-51.

[12] Yogendra Prajapati & Arwind Gautam "Advances in Speech Emotion Recognition: Techniques, Challenges, And Applications "Glimpse - Journal of Computer Science •Vol. 3, No. 2, July-December 2024, pp. 44-48

ABOUT THE AUTHORS

Kshitij Sisodia received a BTech degree from CSE Dept, ABESEC, Ghaziabad. His research interest is Machine learning and deep learning. He is currently working as React Native Developer in Infoicon Technology.

Dr. Ayushi Prakash is a **Professor** in the Department of Computer Science and Engineering at Corporate Relation Manager in Ajay Kumar Garg Engineering College, Ghaziabad, She received a B.Tech. , M.Tech, Ph.D. degree in Computer Science Engineering domain. She has **16 years of experience** as an academician and she is also very active in the Research and Training & Placement domain. Prof. Prakash, professional member of IEEE and IAEPT.

Vishal Choudhary is working as an Assistant Professor in the Department of Computer Science & Engineering, AKGEC Ghaziabad. He has completed his M.Tech from Centre for Advanced Studies, AKTU. He has several publications in various journals of repute.

Amrita Bhatnagar is an Assistant Professor in the Department of Computer Science & Engineering, AKGEC, Ghaziabad. She is BE from SRMSCET Bareilly and MTech from BIT Mesra Rachi. Her research area includes Digital Image Processing, Cyber Security, Information Security, and Python. She has published many research papers in the area of cyber security in reputed journals.