

Social Media Sentiment Analysis Using Twitter Dataset

Thanuj R

Department of Computer Science &
Engineering
Amrita School of Computing Bengaluru
Amrita Vishwa Vidyapeetham, India
Bl.en.u4aie22140@bl.students.amrita.edu

Mani Sankar D

Department of Computer Science &
Engineering
Amrita School of Computing Bengaluru
Amrita Vishwa Vidyapeetham, India
Bl.en.u4aie22112@bl.students.amrita.edu

Prem Reddy YS

Department of Computer Science &
Engineering
Amrita School of Computing Bengaluru
Amrita Vishwa Vidyapeetham, India
Bl.en.u4aie22166@bl.students.amrita.edu

Abstract — this project is developing an automated sentiment analysis of Twitter data for a chosen theme. It is composed of the following: collecting various Twitter datasets, preprocessing texts, initial data analysis, and feature engineering. The specific sentiment analysis model, which varies from traditional to deep learning techniques, is learned and assessed for precision and efficiency. The goal is to offer a valuable data analytics tool with plug-and-play capabilities and an interactive interface for the purpose of real-time public sentiment analysis.

Keywords- Sentiment analysis, Social Media, Hyperparameters, feature fusion, lifelong learning, Noisy Data, Supervised learning, Artificial Intelligence.

I. INTRODUCTION

The prevalence of various social media platforms, especially Twitter, has revolutionized the means through which people express their opinions, share stories, and partake in global discussions from the daily life point of view. The topic of this document is a comprehensive platform for sentiment analysis in terms of its dynamics, which depend on the temporal characteristic of user-generated content on Twitter. The skill of distinguishing and decoding sentiments enclosed in posts is an essential competence for companies, scientists and individuals, who have the task of processing the finer details of public opinion covering various topics, products and events.

While social media and Twitter, in particular, are invaluable archives of the most natural and pristine expressions, they are really amazing sources of getting an understanding of people's moods. Analyzing this endless and continuously changing data source needs advanced analytical methods, including Natural Language Processing and Machine Learning. This paper recommends a three level system, which starts with the data collection, then goes to the preprocessing, exploratory data analysis, feature engineering, model selection, training and evaluation. The objective is to supply a systematic facility for understanding emotions hence bringing out information on the transient environment of public talk in social media.

The following paragraphs will detail the contents of sections, which expounds on the methodology of the sentiment analysis framework, covering data acquisition processes, in-depth preprocessing steps, the array of modeling techniques, and the possibility of real-time implementation. With this extensive examination, we strive to cast some light on the hitherto neglected topics in the sentiment analysis methods for the Twitter, giving the exact resource needed for the persons analyzing public sentiment in the digital era.

II. LITERATURE SURVEY

The paper explains the use of a lexicon-based approach for analyzing Turkish shares in social media with a similarity around 80% in the classification of the sentiment aspect. [1] The offered paper covers the topic on the efficacy of Bidirectional Encoder Representations From Transformers (BERT) in regard to the sentiment analysis on bulk social media datasets. Leveraging a state of the art language model, is able to handle the varying dimensions of sentiment expressions and also context under the fast dynamic and diverse nature of the social media contents. [2] The paper study looks at the application of sentiment analysis to examine the communication of Indian companies on the social media platforms like Twitter. The instrument does not explicitly refer to assessing social responsible stances and attitudes with the help of sentiment analysis. [3] The paper proposes a multimodal sentiment analysis approach based on a top-layer fusion strategy in social networks that outperforms the baselines on the MVSA-Single dataset. This approach allows the use of multiple models in one, combining different types of data, showing its advantage in the context of sentiment analysis accuracy improvement for social media content. [4] The paper analyzes the impact of sentiments expressed in tweets on the stock market using the VADER model for sentiment analysis. [5] The paper proposes a multilingual sentiment classifier

to analyze how Malaysians react on social media during disasters, achieving 0.862 accuracy and 0.864 F1-score. [6] A paper is presented that suggests a distantly supervised lifelong learning scheme for the social media sentiment analysis on the large scale. This approach, therefore, seeks to reconfigure the ways of dealing with a large amount of data from the social media platforms. [7] This paper is about exploring the significance of sentiment analysis and examines existing research on evaluating and comparing relevant data mining techniques. [8] The paper contributes to a growing body of research on sentiment analysis within the context of social media using some different classifiers like support vector machines and random Forest etc. [9] This paper is about Identifying Fake News in Social Media Using Sentiment Analysis using the relations negative sentiment and the likelihood of news being fake.[10]

III. REFERENCES.

- [1] H. Karamollaoglu, İ. A. Doğru, M. Dörterler, A. Utku and O. Yıldız, "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach," 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, Bosnia and Herzegovina, 2018, pp. 45-49, doi: 10.1109/UBMK.2018.8566481.
- [2] T. Bikku, J. Jarugula, L. Kongala, N. D. Tummala and N. Vardhani Donthiboina, "Exploring the Effectiveness of BERT for Sentiment Analysis on Large-Scale Social Media Data," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-4, doi: 10.1109/CONIT59222.2023.10205600.
- [3] M. V. S. P. Mastan Rao and S. Babu, "Evaluating Social Responsible Attitudes and Opinions using Sentiment Analysis – An Indian Sentiment," 2022 3rd International Conference on Computing, Analytics and Networks (ICAN), Rajpura, Punjab, India, 2022, pp. 1-7, doi: 10.1109/ICAN56228.2022.10007315.
- [4] C. Li and Z. Hu, "Multimodal Sentiment Analysis of Social Media Based on Top-Layer Fusion," 2022 IEEE 8th International Conference on Computer and Communications (ICCC), Chengdu, China, 2022, pp. 1-6, doi: 10.1109/ICCC56324.2022.10065604.
- [5] A. Singh, J. Singh and A. Ghosal, "Impact Of Social Media On Stock Market- A Case Of Sentiment Analysis Using Vader," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 300-304, doi: 10.1109/COM-IT-CON54601.2022.9850668.
- [6] M. J. Fuadvy and R. Ibrahim, "Multilingual Sentiment Analysis on Social Media Disaster Data," 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Denpasar, Indonesia, 2019, pp. 269-272, doi: 10.1109/ICEEIE47180.2019.8981479.
- [7] R. Xia, J. Jiang and H. He, "Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis," in *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 480-491, 1 Oct.-Dec. 2017, doi: 10.1109/TAFFC.2017.2771234.
- [8] Farogh, Sakhawia. (2023). A Comprehensive Evaluation and Comparative Analysis of Data Mining Techniques for Sentiment Analysis in Social Media. *International Journal of Advanced Research in Science, Communication and Technology*. 40-45. 10.48175/IJARSCT-11609.

- [9] (2024). Sentiment Analysis using Support Vector Machine and Random Forest. *Journal of Informatics and Web Engineering*. 3. 67-75. 10.33093/jiwe.2024.3.1.5.
- J. Wilson, S. E. Hosseini and S. Pervez, "Identification of Fake News in Social Media Using Sentimental Analysis," 2023 IEEE Industrial Electronics and Applications Conference (IEACon), Penang, Malaysia, 2023, pp. 220-224, doi: 10.1109/IEACon57683.2023.10370300.