iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: https://doi.org/10.22214/ijraset.2023.56738

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 11 Issue XI Nov 2023- Available at www.ijraset.com

Twitter Sentiment Analysis on Chatgpt Tweets

Anurag Jadhav[1], Vaishanavi Bolli[2], Sanket Kewale[3], Atharva Kharat[4]
[1, 2, 3, 4]*Department of Computer Engineering, All India Shree Shivaji Memorial Society College of Engineering, Pune*

Abstract: ***This study investigates multilingual sentiment analysis within tweets on ChatGPT, an AI conversational model, employing Support Vector Machines (SVM) and BERT, an advanced language model. It aims to detect and classify emotions, including emoji identification, embedded within diverse messages across multiple languages on Twitter.***
By leveraging SVM's text classification and BERT's contextual understanding in various languages, the research delves into preprocessing techniques and feature engineering for sentiment analysis, encompassing multilingual and emoji detection.
Furthermore, it explores the fusion of traditional SVM methods with BERT's state-of-the-art model for multilingual sentiment analysis, emphasizing emotion and emoji detection in AI-generated content on multilingual social media platforms like Twitter.
This research yields insights into the successful detection of multilingual sentiment nuances and emotions, including emoji identification. It offers implications for advancing multilingual sentiment analysis in natural language processing across diverse linguistic contexts.
Keywords: ChatGPT, Text classification, Twitter, Sentiment analysis, Emotion detection, multilingual, SVM, BERT, natural language processing.

I. INTRODUCTION

In this research, we aim to analyze the sentiments expressed in tweets related to OpenAI's ChatGPT utilizing text processing methodologies and machine learning algorithms. Twitter, being a platform that offers real-time and concise messages, provides a substantial dataset for this examination. Our study will detail the methodology, encompassing data collection, processing techniques, analytical methods, and the resultant findings from this sentiment analysis. Conventional sentiment analysis often falls short in comprehending intricate emotions and contextual nuances. To address these limitations, our approach focuses on three fundamental aspects: identifying emotions, scrutinizing specific contextual elements, and accommodating multilingual capabilities. By amalgamating these features, our advanced sentiment analysis system strives to offer a more nuanced, precise, and language-independent comprehension of emotions embedded within textual content. This approach aims to overcome the constraints observed in traditional sentiment analysis techniques.

II. LITERATURE REVIEW

1) *Sentiment Analysis of Using ChatGPT in Education [1]:* A research on the application of Chat Generative Pretrained Transformer (ChatGPT) in education is presented in this article. A sentiment analysis model of tweets pertaining to the employment of ChatGPT in education has been presented in this study. Furthermore, the research examines the opinions shared in tweets on ChatGPT's use in education by employing four distinct classifiers: Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes (NB). The SVM classifier has the maximum accuracy of 81.4 percent, based on the findings.
2) *A Study of Sentiment Analysis Task and It' s Challenges [2]:* This study helps to answer the queries like what is sentiment analysis, how to perform it, and what challenges one has to face while developing a sentiment analysis system. In this paper they have covered different levels of sentiment analysis and a detail discussion over aspect-based sentiment analysis is given. The important challenges to this research area like named entity recognition, sentiment polarity detection, subjectivity detection etc. have been described with suitable example.
3) *Stock Market Prediction based on Social Sentiments using Machine Learning [3]:* This study focuses on the sentiment analysis system diagram in addition to the model that will be used. With the help of the Twitter API and the closing prices of different stocks, we are able to gather tweets and work toward developing a system that can predict changes in the stock prices of different firms. The SVM model provides superior accuracy.
4) *Comparative study of Twitter Sentiment On COVID- 19 Tweets [4]*: Three distinct algorithms—Vader Sentiment Analysis, BERT, and Logistic Regression—were used in this research. The three algorithms' scores have been adjusted to fall between -1 and 1. This is done to make the comparison transparent and equitable. Upon comparison, BERT outperforms VADER and Logistic Regression in terms of accuracy (92%). Because BERT looks for the aspect of the sentences, it is more accurate than other algorithms.

©IJRASET: All Rights are Reserved

5) *RoBERTa: A Robustly Optimized BERT Pretraining Approach [5]:* This research is a replication study of BERT pretraining (Devlin et al., 2019) in which the effects of several important hyperparameters and training data size are meticulously measured. This model discover that BERT was remarkably undertrained, and it can perform as well as or better than any model released after it. State-of-the-art results on GLUE, RACE, and SQuAD are achieved by our top model.
6) *Tracking public attitudes toward ChatGPT on Twitter using sentiment analysis and topic modeling [6]:* In this research, they employ sentiment analysis and topic modeling techniques to Twitter data in order to examine public opinions on ChatGPT using natural language processing tools. This has led to the conclusion that the best accurate algorithm for the analysis is BERT.
7) *Hybrid Deep Learning Models for Sentiment Analysis [7]:* The purpose of this research is to evaluate the validity of many hybrid strategies using a range of datasets from diverse fields. The purpose of our research questions is to find out if hybrid models with diverse domains and types of datasets may perform better than single models. On all kinds of datasets, the hybrid models outperformed single models in terms of sentiment analysis accuracy, particularly when combining SVM and deep learning models.

III. METHODOLOGY

1) *Data Collection:* Gathering data stands as a crucial initial step in analyzing sentiments within ChatGPT-related tweets on Twitter. It starts by identifying the specific sentiments to study, such as positive, negative, or neutral expressions. Accessing real-time and historical tweet data mandates creating a Twitter Developer account, generating an application, and procuring API keys. When handling extensive datasets, considering a sample subset of tweets for an initial overview proves beneficial. Safeguarding collected data is paramount; employing databases or file systems aids in organized storage for further analysis. Strive for strict adherence to data protection regulations and pertinent laws throughout the entire process.
2) *Data Preprocessing:* Exclude extraneous information from the data, such as mentions, URLs, and special characters. Put the text in words or phrases to tokenize it. To minimize dimensionality, eliminate stop words and use stemming or lemmatization. Python is widely used for natural language processing and sentiment analysis tasks. Developers can utilize libraries like NLTK (Natural Language Toolkit) and Text Blob for text processing tasks. Alternatively, you can use programming languages like JavaScript or Java, depending on your team's expertise and project requirements. Proper functioning of these tools is necessary for accurate sentiment analysis. Updates or changes in these tools might affect the analysis.
3) *Feature Extraction:* Feature extraction encompasses the conversion of text data into numerical formats tailored for machine learning algorithms. Widely used methods include TF-IDF (Term Frequency-Inverse Document Frequency), which evaluates term significance within documents. Alongside this, word embeddings like Word2Vec or GloVe are utilized to create numerical representations of words, capturing semantic meanings and interrelations between them. These techniques enable machine learning models to understand and process textual data by transforming words into numerical features essential for analysis and classification purposes.
4) *Model Selection:* The initial stage in a Twitter sentiment analysis project centered on ChatGPT tweets is to gather a labeled dataset of tweets that have been tagged with sentiments (positive, negative, or neutral). The next step is data preparation, which includes cleaning, tokenization, and feature extraction with the use of methods like word embeddings or TF-IDF. Because of their superiority in text categorization, Support Vector Machines (SVM) and Naive Bayes are taken into consideration while choosing a model. Furthermore, the potential of deep learning models—more precisely, an LSTM neural network built using TensorFlow and Keras—to extract semantic associations from the tweets is investigated. Evaluation metrics, such accuracy and classification reports, are used to evaluate the performance of the model after it has been trained. Hyperparameter modifications are used to fine-tune the models, and optional deployment for real-time sentiment analysis is taken into consideration.
5) *Model Training:* Divide the dataset into sets for testing and training. Utilizing the training data, train the chosen model. Optimize performance by adjusting the hyperparameters. Analyze the model's accuracy and generalizability using the testing set.
6) *Model Evaluation:* Both Support Vector Machines (SVM) and BERT models go through a methodical evaluation procedure when it comes to Twitter sentiment analysis models for ChatGPT tweets. The labeled dataset is split into training and testing sets. BERT uses its tokenization and padding techniques, whereas SVM uses TF-IDF for feature extraction. The training set is used to train and fine-tune both models, while the test set is used to assess their performance using measures like accuracy, precision, recall, and F1 score. This analysis aids in determining how effectively each model categorizes the sentiment present in tweets produced by ChatGPT. Computational resources and the capacity to grasp contextual subtleties are two aspects that comparative analysis takes into account. The results direct prospective efforts at fine-tuning both models with the goal of optimizing their performance in the unique ChatGPT sentiment analysis setting.

©IJRASET: All Rights are Reserved

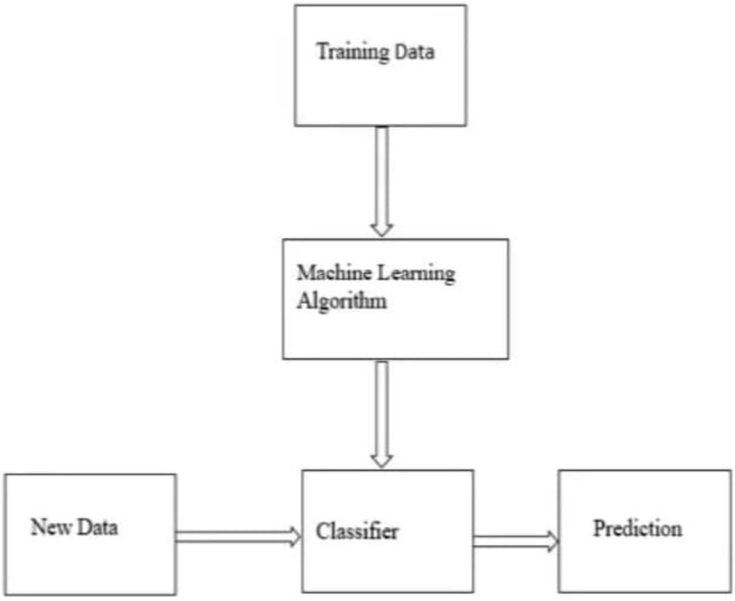


Fig1: System flow Diagram

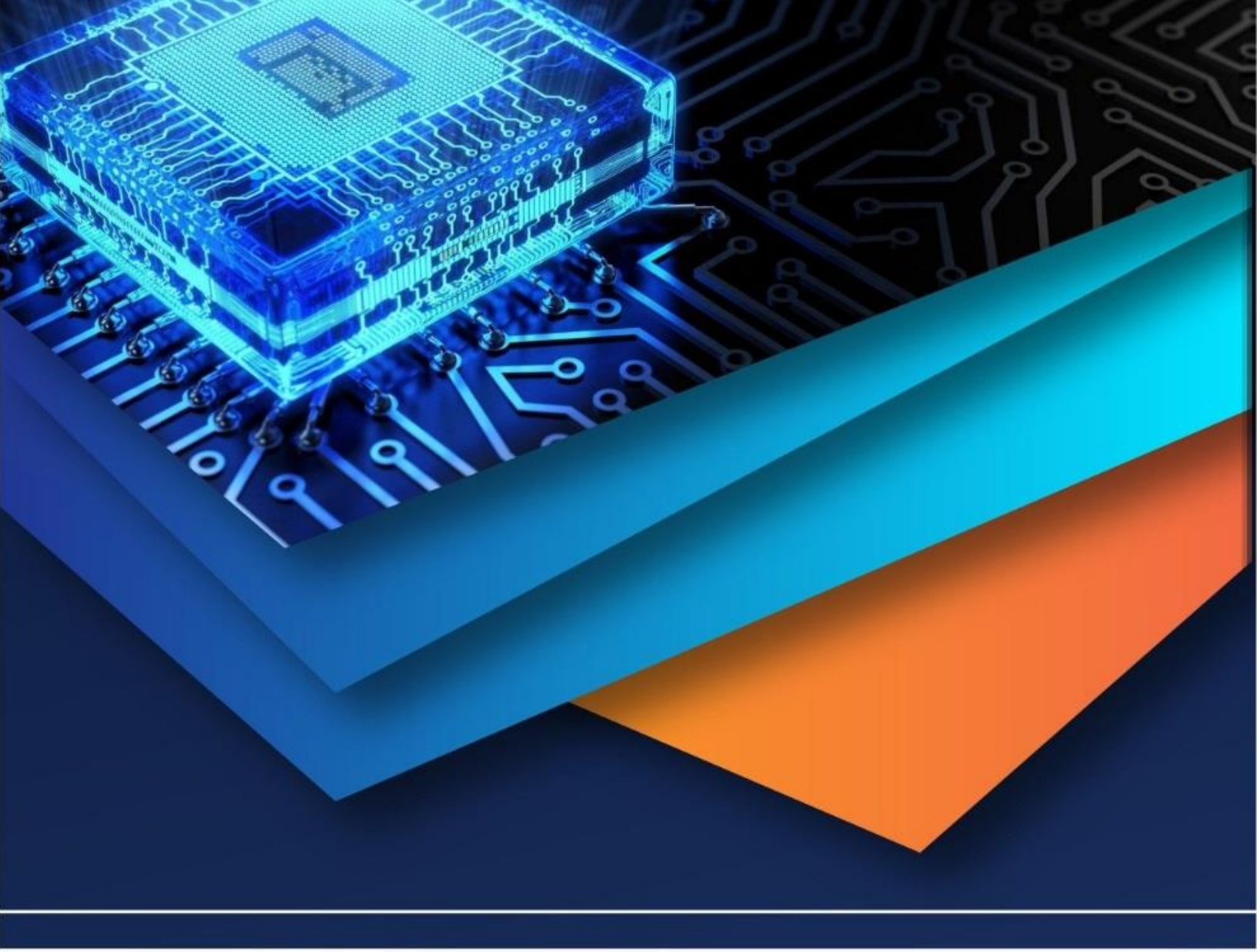
IV. CONCLUSION

In conclusion, the SVM and BERT models used in the Twitter sentiment analysis of ChatGPT tweets have yielded important insights about the tone of emotion in the content that has been created. The SVM model performed admirably in capturing the general sentiment patterns because of its capacity to categorize tweets according to a set of attributes. Conversely, the BERT model demonstrated a sophisticated understanding of the feelings expressed in the tweets by utilizing its contextual knowledge of language. The power of fusing cutting-edge deep learning techniques like BERT with conventional machine learning methods like SVM was made evident by the incorporation of these models into the sentiment analysis pipeline. With this hybrid technique, sentiment analysis across a wide range of lively Twitter conversations may be done in-depth.

REFERENCES

[1] M. Tubishat, F. Al-Obeidat and A. Shuhaiber, "Sentiment Analysis of Using ChatGPT in Education," 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), Istanbul, Turkiye, 2023, pp. 1-7, doi: 10.1109/SmartNets58706.2023.10215977.
[2] S. V. Pandey and A. V. Deorankar, "A Study of Sentiment Analysis Task and It's Challenges," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, India, 2019, pp. 1-5, doi: 10.1109/ICECCT.2019.8869160.
[3] T. Mankar, T. Hotchandani, M. Madhwani, A. Chidrawar and C. S. Lifna, "Stock Market Prediction based on Social Sentiments using Machine Learning," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, 2018, pp. 1-3, doi: 10.1109/ICSCET.2018.8537242. .
[4] A.J. Nair, V. G and A. Vinayak, "Comparative study of Twitter Sentiment On COVID- 19 Tweets," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1773-1778, doi: 10.1109/ICCMC51019.2021.9418320.
[5] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
[6] Koonchanok, R., Pan, Y. and Jang, H., 2023. Tracking public attitudes toward ChatGPT on Twitter using sentiment analysis and topic modeling. arXiv preprint arXiv:2306.12951.
[7] https://www.hindawi.com/journals/complexity/2021/9986920/.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)