International Journal of Informatics and Applied Mathematics
e-ISSN: 2667-6990 Vol. 8, No. 1, 39-50

Utilizing Natural Language Processing, Observational Metrics for Predictive Analysis of Generative Artificial Intelligence

Oluwasefunmi Arogundade, Adebayo Abayomi-alli, and Joel Akerele

Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria
akerelejl@gmail.com arogundadeot@funaab.edu.ng abayomiallia@funaab.edu.ng

Abstract. In the era of massive language models, it is essential to understand the evolving interest in tools like ChatGPT, a generative artificial intelligence (GenAI) model. It's doubtful that consumers' present feelings and degree of interest will last over time. This work used observational metrics and natural language processing to predict future sentiments and search trends regarding GenAI. Time-bound web analytics data and Twitter metrics related to GenAI were collected using Google Trend and the Twitter API on Orange Data Mining Toolkit. Google trend data was forecasted using Autoregressive Integrated Moving Average (ARIMA), whereas sentiment polarities and search interest time series were predicted using Naive Bayes. The experiment's results indicated a limited correlation between tweet sentiment polarity scores and engagement metrics. Five subjects in all were returned by the topic modeling: doubts or skepticism about OpenAI and Microsoft, Microsoft and AI Use, French discussions on ChatGPT, ChatGPT arguments and usage, and making something funny in relation to intelligence and analysis. Among 50 predicted sentiment instances, 82% were positive, 8% neutral, and 10% negative—indicating a generally optimistic outlook. These findings underscore the value of analyzing sentiment and interest trends to understand GenAI model evolution.

Keywords: Autoregressive · Sentiment · GenAI · ChatGPT · Predictive Analytics.

Received: 15 February 2025 | Revised: 07 March 2025 | Accepted: 11 June 2025
DOI: 10.53508/ijiam.1640467

1 Introduction

The "artificial intelligence" (AI) academic conference, convened by [1], in 1955 marked the beginning of the discussion regarding the feasibility of a thinking computer system. The concept proven to be a game-changer in the field of computer science and is currently driving unprecedented disruptive innovations across the information technology value chain. In order to surpass humans in decision-making, artificial intelligence (AI) refers to machines and systems that exhibit human cognition, according to [2]. The argument for the possibility of a thinking computer system has been reinforced by the previously unheard-of computing results that have been shown by machine learning (ML), natural language processing (NLP), expert systems, and other AI use cases [3]. Natural language processing (NLP) has recently shown great potential as a disruptive machine intelligence use case that might leverage human language to develop a more sophisticated AI in the language domain. A wide range of NLP problem domains are targeted by the NLP applications, also referred to as conversational AI [4] or language models [5]. These include the Generative AI (GenAI) model [6], chatbots [7], virtual assistants [8], sentiment analysis tools [9], text summarization tools [10], speech recognition tools [11], text-to-speech tools [12], named entity recognition tools [13], topic modeling tools [14], and more. It is becoming increasingly crucial to ascertain the future of the enduring interest in using ChatGPT as an example of GenAI models. However, its perceived usefulness, usability, security and privacy concerns, and ethical considerations may affect its ongoing adoption and use. By studying consumers' existing attitudes and time-aware interest in these products, a market study will be able to determine the variables driving the continued adoption and deployment of GenAI solutions in use case professional areas. The main objective of the study is to develop a framework that can forecast future attitudes and interests of GenAI users.

2 Related Works

[15] investigated the Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. A novel multimodal paradigm for predicting consumer product ratings is put forth in this study. It integrates information from several sources, such as physiological signals and international reviews of certain brands and products. Evaluations from viewers worldwide are retrieved and processed using Natural Language Processing (NLP) technology to calculate a compound score that is utilized to calculate the overall rating. Results from the study showed the Variance problem eliminated by the ensemble modelling. However, there were no test of multicollinearity on predictive attributes. [16] examined the Text Analysis of ChatGPT as a Tool for Academic Progress or Exploitation. The study's goal was to find out how people feel about large language models (LLMs) like ChatGPT because of the intense arguments they cause. Through data mining, it seeks to further understand its likely impact on the education sector by examining end-user viewpoints. The results from

the study showed an efficient data visualization approach. However, the study encountered non-inclusive/non heterogenous primary data. [17] investigated Sentiment analysis using product review data. The data used in this study came from online product reviews collected from Amazon.com. Experiments on classification at the review level and categorization at the sentence level are carried out with promising outcomes. One fundamental problem in sentiment analysis that is addressed in this paper is the classification of sentiment polarity. Results from the study showed an Automated ground truth labelling as well as effective Feature extraction through Bag of-Words approach. However, topic modelling will better extract themes from product reviews to achieve study aims. [18] examined Sentiment Analysis and Classification for Software as a Service Reviews. This study examines how cloud user reviews represent users' experiences with Software as a Service (SaaS) applications. The polarity of each review—that is, whether a positive, negative, or neutral sentiment was being conveyed— was ascertained by employing sentiment analysis to examine almost 4000 online reviews that were collected from multiple websites. Results from the study showed that Cross-validation approach significantly reduced variance problem as well as outperforming term occurrence approaches. However, the predictive abilities of independent variables not pre-determined. [19] investigated Attention-based sentiment analysis using convolutional and recurrent neural network. This study created a new model that integrates the benefits of attention mechanism architectures based on Recurrent Neural Networks (RNN) and Convolutional Neural Networks into a single model. CNN initially learns the important sentence aspects from the input representation in the suggested architecture. An attention mechanism is employed to direct the model's attention to the characteristics that significantly contribute to the prediction job by calculating the attention score from the features context produced by CNN filters. Results from the study showed that the Computation of attention score optimized the performance of the modelling. However, Out-of-Sample approach will better evaluate the performance of the hybrid predictive model.

3 Methodology

3.1 Conceptual framework of the research methodology

The eight-phase conceptual framework used in this study is shown in Figure 1. Each phase of the framework will be carried out through an experimental process, which includes the following experimental activities: Data Acquisition (Section 3.2), Data Pre-processing (Section 3.3), Topic Modeling (Section 3.4), Sentiment Analysis (Section 3.5), Exploratory Data Analysis (Section 3.6), Prediction of Future Sentiment towards GenAI (Section 3.7), and Evaluation of the Predictive Models (Section 3.8).

3.2 Data Acquisition

The study's primary data, which includes textual expressions concerning GenAIs and the pattern of inquiries into GenAI over time, was obtained from two differ-

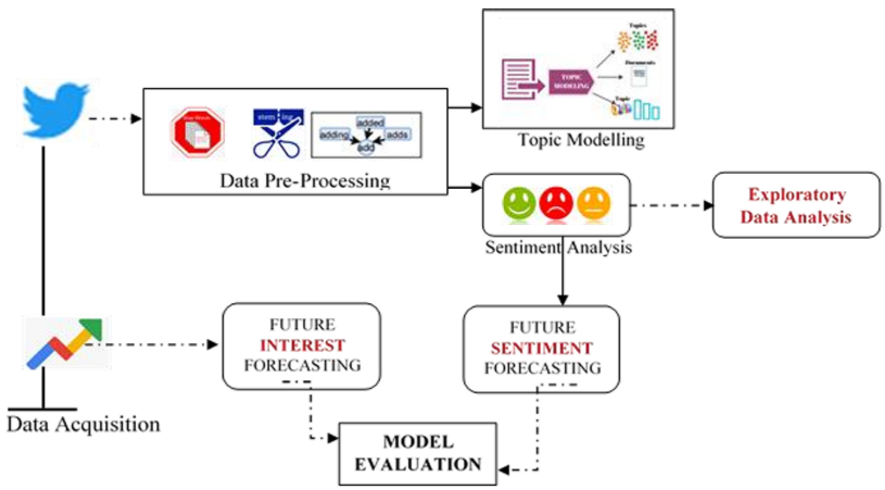

Fig. 1: The conceptual framework of the study methodology.

ent sources. To minimize the likelihood of irrelevant data and to ensure inclusive data representation, the data is gathered using popular hashtags and targeted search terms.

3.3 Data Pre-processing

In order to clean acquired tweets on GenAI, NLP techniques such as lemmatization, tokenization, stop word removal, etc., must be used at this stage. The unstructured corpus of textual expressions known as tweets needs to be appropriate for further NLP use cases, such as topic modeling and sentiment analysis. The textual tweet will be processed by the NLP techniques, which will then turn it into a collection of distinct words or keywords known as tokens.

3.4 Topic Modelling

The penultimate step, topic modeling, groups related tweet posts according to their vectorized representation using the clustering method k-means. Finding a group of subjects or themes that best capture the corpus is the aim.

3.5 Sentiment Analysis

The VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment methodology is a rule-based approach to sentiment analysis. It employs a vocabulary of words and their accompanying sentiment ratings to determine the sentiment of a text. Among other things, the method offers rules for handling negation, capitalization, and punctuation. For sentiment analysis, the following

Algorithm 1 Topic Modelling Algorithm

Input: Textual tweet corpus T, number of topics K
Output: Topic-word distribution matrix π, document-topic distribution matrix Θ
Initialize π and Θ randomly
Initialize topic assignments for all words in T
repeat
  for each document d in T **do**
    Compute topic distribution for document d using current π and Θ
    for each word w in document d **do**
      Compute topic distribution for word w using current π and Θ
      Sample a new topic for word w based on the distribution
      Update topic counts for the new assignment
    end for
  end for
  Update π and Θ based on topic counts and hyperparameters
until convergence
return final π and Θ

procedures are used. Positive sentiment score: the total of the word's ratings across all positive sentiment categories. Score for negative sentiment: sum of all negative sentiment categories' scores that is negative The term is a part of the total of the scores for each category to which the word belongs is known as the neutral sentiment score. The compound sentiment score, which ranges from -1 (most negative) to +1 (most positive), is a normalized, weighted composite value of the positive, negative, and neutral scores. The word "w" in each GenAI tweet is given a numerical value of 1, 0, or -1 to indicate a positive, neutral, or negative sentiment in its calculations. Consequently, a tweet with the polarity "T" is:

$$T = \{w_1, w_2, w_3, \ldots, w_n\} \tag{1}$$

while Equation 1 is computed on the frequency of words w in T which occurs in z. The $\mathrm{pos}(T, z)$ and $\mathrm{neg}(T, z)$ are positive and negative words from T that occur in z:

$$\mathrm{sum}(T, z) = \mathrm{pos}(T, z) - \mathrm{neg}(T, z) \tag{2}$$

hence, sentiment $s_1(z)$ of a feature z under polarized lexicon T is derived by:

$$s_1(z) = \begin{cases} T & \text{if } \mathrm{sum}(T, z) > 0 \\ 0 & \text{if } \mathrm{sum}(T, z) = 0 \\ -T & \text{if } \mathrm{sum}(T, z) < 0 \end{cases} \tag{3}$$

Compound score C is given as:

$$C = \frac{p - n}{\sqrt{p^2 + n^2 + v^2}} \tag{4}$$

where p is the positive polarity score and n is the negative polarity score.

Algorithm 2 VADER-based Sentiment Analysis Algorithm

Input: A GenAI-related tweet

Output: A sentiment score (between -1 and 1) and a sentiment category (positive, negative, neutral, and compound score)

```
Load the VADER lexicon, which contains a list of words and their associated
  sentiment scores
Tokenize the text into individual words
Initialize the positive, negative, and neutral sentiment scores to 0
for each word in the text do
    if the word is not in the VADER lexicon then
        continue
    else
        Get the sentiment score of the word from the lexicon
        if the sentiment score is positive then
            Add it to the positive sentiment score
        else if the sentiment score is negative then
            Add it to the negative sentiment score
        else
            Add it to the neutral sentiment score
        end if
    end if
end for
```

3.6 Exploratory Data Analysis (EDA)

EDA is a method for examining and distilling datasets to gain knowledge and identify trends, correlations, and abnormalities. As the first step in the data analysis process, exploratory data analysis (EDA) usually involves statistical and graphical approaches. To further provide actionable insights into the topic of GenAI as it relates to the future feelings of its users, this study will make use of certain EDA methodologies.

3.7 GenAI Future Sentiment Forecasting with Naïve Bayes

In the literature, Naïve Bayes is frequently suggested for sentiment polarity predictive analytics, particularly when it comes to social media metrics [20]. A popular probabilistic machine learning approach for text classification tasks, such as sentiment analysis, is Naïve Bayes. The Naïve Bayes model in this study was trained using predictive features, which include the sentiment polarity scores of positive, negative, neutral, and other social media metrics linked to each GenAI tweet. In order to approximate the ground truth of either positive, negative, or neutral sentiment—the set of dependent variables—the predictive independent variables are the collection of social media indicators. A sentiment score—that is, a value of 1 for positive, a value of -1 for negative, and a value of 0 for neutral—was assigned to each tweet. The Naïve Bayes algorithm then learnt to predict the mood of new, unlabeled GenAI tweets based on the word patterns and linkages in the training data.

3.8 Model Evaluation

The Confusion Matrix's True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) components will be used to compute the metrics. The confusion matrix, which is separated into two dimensions, such as the Predicted Values and True Values, along with the total number of predictions in each category, is used to determine how well a classification model performs on a particular test set of data.

4 Implementation and Results

4.1 Data Acquisition

The API produced 801 tweets and data in total, including the date, tweet instances, likes, retweets, recounts, quotation counts, and more. The timestamp and the quantity of search entries per minute are included in the Google trend statistics. 53 instances in total are returned by the Google trend. This is shown in Figure 2.

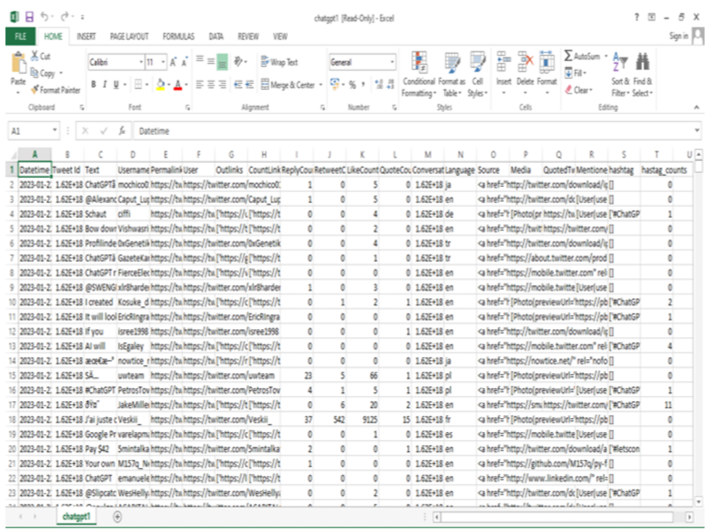

Fig. 2: Screenshot of acquired tweets and other engagement metrics.

4.2 Topic modelling of acquired tweets

The obtained 801 tweets are put through topic modeling in accordance with the conceptual framework shown in Figure 1 in order to identify consensus topics within the corpus. Five (5) topics and the weights of the words that make up each topic are returned by the LDA-based modeling. Table 1 displays the topic modeling result, which includes the subjects and their corresponding weights.

Table 1: Topics returned by VADER modelling.

Output Topics	**Topical Words**	**Weights of Words**
Topic 0	de, en, chatgpt, use, su	0.030, 0.015, 0.012, 0.012, 0.012
Topic 1	chatgpt, ai, 're, way, using	0.018, 0.018, 0.015, 0.008, 0.008
Topic 2	le, chatgpt, de, 's, pa	0.015, 0.015, 0.009, 0.009, 0.009
Topic 3	openai, microsoft, ai, n't, never	0.016, 0.013, 0.013, 0.010, 0.010
Topic 4	make, fun, exam, n't, intelligence	0.013, 0.009, 0.009, 0.009, 0.009

4.3 Summary Statistics of Tweet Engagements

The sentiment polarity that the initial sentiment analysis returned are concatenated with the engagement metrics that correlate to each of the obtained tweets. Table 2 displays the statistical summary of the data (twitter engagements) on generative AI tweets. Important information about the distribution and properties of different observational attributes is provided by the data. Retweet counts, like counts, quote counts, and hashtag counts are all included in the dataset in Table 2. As indicated by the comparatively high standard deviations, one noteworthy finding is the existence of a significant variance in the quantity of replies, retweets, and likes. The presence of tweets with a much greater level of involvement is demonstrated by the fact that, for example, the highest value reaches 3098, while the mean reply count is 7.07. Similar to this, the dataset's tweets' varying levels of popularity are highlighted by the mean like count, which is 97.80 with a maximum value of 56073.

4.4 Sentiment Analysis by Naïve Bayes

The priors were calculated by the Naïve Bayes. The class priors for the Negative, Neutral, and Positive classes are calculated as [0.3625, 0.040625, and 0.596875] in this context. These values signify the probabilities of encountering each sentiment class in the overall training data. Breaking down these class priors further, it suggests that approximately 36.25% of the tweets are classified as Negative sentiment, 4.06% as Neutral sentiment, and a predominant 59.69% as Positive sentiment. An intriguing finding regarding the sentiment polarity of tweets about GenAI is highlighted by this distribution. The high likelihood of Positive class priors suggests that there is a strong positive sentiment in this dataset, indicating that a significant percentage of tweets are optimistic or have positive views about GenAI. Conversely, a lower frequency of negative or neutral attitudes in the examined tweets is shown by the lower probabilities for the Negative and Neutral classes. This knowledge is essential for analyzing, using the given dataset, the sentiment landscape in the field of generative AI as a whole.

Table 2: Summary statistics of engagement metrics of acquired tweets

	Reply_Count	Retweet_Count	Like_Count	Quote_Count	Hashtag_Counts
Mean	7.07	11.38	97.80	2.62	0.69
Std	120.86	244.74	2019.06	68.85	1.55
Min	0.00	0.00	0.00	0.00	0.00
25%	0.00	0.00	0.00	0.00	0.00
50%	0.00	0.00	1.00	0.00	0.00
75%	1.00	0.00	2.00	0.00	1.00
Max	3098.00	6815.00	56073.00	1947.00	15.00

4.5 Naïve Bayes for Future Sentiment Forecast

Given that there are five instances of predicted negative sentiment (10%), four instances of neutral sentiment (8%), and the remaining eighty-two percent of positive sentiments, the ARIMA-based future sentiment forecast is used to show the expected trend of sentiments on the topic of GenAI. The Results showed that for the forecasted sentiment polarity scores for each time out of the 50 forecasted times, 41 instances are predicted to have a positive sentiment, four (4) are predicted to be neutral, and five (5) are predicted to have a negative sentiment. The performance evaluation plot of the Naïve Bayes and the Comparative search trend plot between acquired and forecasted time series data are shown in Figure 3 and 4 respectively. Figure 3 showed that the Mean Absolute Error (MAE) had the highest value of 15 while the Root Mean Squared Error (RMSE) had the lowest value of 4. Figure 4 showed that precision had the highest value of 0.9 while the accuracy and recall both had the least value of 0.5.

5 Conclusion

The aim of this study was to estimate future interest trends and obtain insights into the attitudes surrounding GenAI models by utilizing web analytics data and natural language processing techniques. The study aims to give a thorough grasp of the dynamics surrounding GenAI by gathering time-bound site analytics data and Twitter metrics and using techniques including sentiment analysis, topic modeling, and time series forecasting. The results of the experiment indicate different levels of future attitude and interest across time. Over the designated time periods, the projected sentiment analysis shows a largely favorable sentiment trend with some noteworthy oscillations but an overall upbeat prognosis. The majority of the time, the Naive Bayes model predicts positive sentiments,

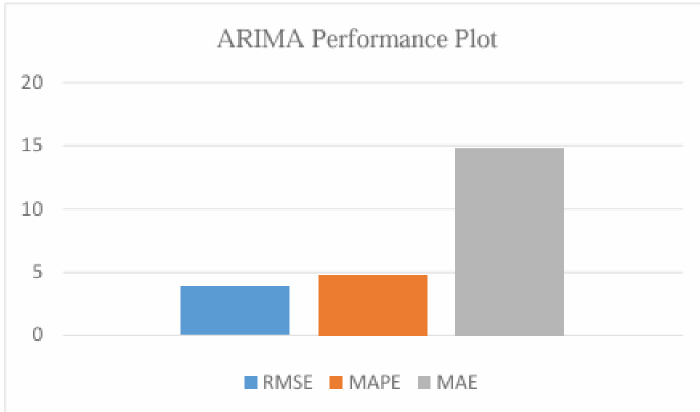

Fig. 3: The performance evaluation plot of the Naïve Bayes

which is consistent with the public's favorable opinions about generative artificial intelligence (GenAI). However, the ARIMA prediction for search interest in GenAI shows a steady drop in interest over time, pointing to a possible drop in public interest or involvement. These divergent patterns might indicate that interest in GenAI is plateauing or declining, even while sentiment is still positive. The study's conclusions offer insightful information for stakeholders and decision-makers in the AI sector by highlighting the significance of taking into account both sentiment and interest patterns in order to obtain a thorough grasp of the changing environment around GenAI models.

Acknowledgments

The authors acknowledge the efforts of the reviewers of this paper. We also appreciate their meaningful contribution, valuable suggestions, and comments to this paper which helped us in improving the quality of the manuscript.

References

1. M. Haenlein and A. Kaplan. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4):5–14, 2019.
2. A. Tlili, B. Shehata, M. A. Adarkwah, A. Bozkurt, D. T. Hickey, R. Huang, and B. Agyemang. What if the devil is my guardian angel: Chatgpt as a case study of using chatbots in education. *Smart Learning Environments*, 10(15), 2023.
3. K. Ayoub and K. Payne. Strategy in the age of artificial intelligence. *Journal of Strategic Studies*, 39(5):793–819, 2016.
4. J. Gao, M. Galley, and L. Li. Neural approaches to conversational ai. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1371–1374. ACM, 2018.

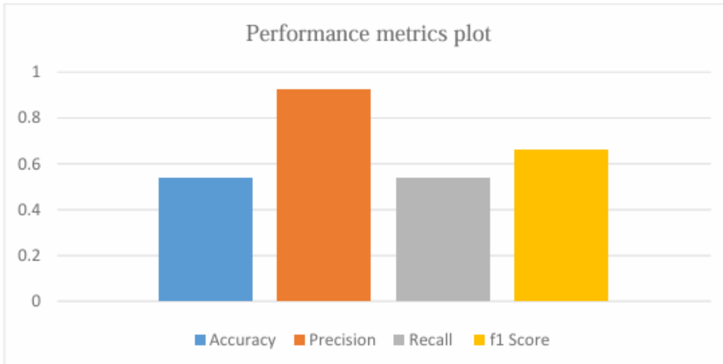

Fig. 4: Comparative search trend plot between acquired and forecasted time series data

5. M. Zaib, Q. Sheng, and W. E. Zhang. A short survey of pre-trained language models for conversational ai: A new age in nlp. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–4. ACM, 2020.
6. J. Qadir. Engineering education in the era of chatgpt: Promise and pitfalls of generative ai for education, 2022. Preprint or unpublished manuscript.
7. T. Lalwani, S. Bhalotia, A. Pal, V. Rathod, and S. Bisen. Implementation of a chatbot system using ai and nlp. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, 6, 2018.
8. A. Vasilateanu and R. Ene. Call-center virtual assistant using natural language processing and speech recognition. *Journal of ICT, Design, Engineering and Technological Science*, pages 40–46, 2018.
9. D. M. Hussein. A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4):330–338, 2018.
10. R. Boorugu and G. Ramesh. A survey on nlp based text summarization for summarizing product reviews. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 352–356. IEEE, 2020.
11. U. Kamath, J. Liu, and J. Whitaker. *Deep learning for NLP and speech recognition*. Cham, 2019.
12. E. E. Adam. Deep learning based nlp techniques in text to speech synthesis for communication recognition. *Journal of Soft Computing Paradigm (JSCP)*, 2(4):209–215, 2020.
13. H. Shelar, G. Kaur, N. Heda, and P. Agrawal. Named entity recognition approaches and their comparison for custom ner model. *Science & Technology Libraries*, 39(3):324–337, 2020.
14. R. Alghamdi and K. Alfalqi. A survey of topic modeling in text mining. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 6(1), 2015.
15. S. Kumar, M. Yadava, and P. P. Roy. Fusion of eeg response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*, 52:41–52, 2019.

16. U. Bukar, M. S. Sayeed, S. F. Razak, S. Yogarayan, and O. A. Amodu. Text analysis of chatgpt as a tool for academic progress or exploitation. *SSRN*, 2023. SSRN 4381394.
17. X. Fang and J. Zhan. Sentiment analysis using product review data. *Journal of Big Data*, 2(1):1–14, 2018.
18. A. M. Alkalbani, A. M. Ghamry, F. K. Hussain, and O. K. Hussain. Sentiment analysis and classification for software as a service reviews. In *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, pages 1–10. IEEE, 2016.
19. M. Usama, B. Ahmad, E. Song, M. S. Hossain, M. Alrashoud, and G. Muhammad. Attention-based sentiment analysis using convolutional and recurrent neural network. *Future Generation Computer Systems*, 113:571–578, 2020.
20. A. M. Rahat, A. Kahir, and A. K. Masum. Comparison of naive bayes and svm algorithm based on sentiment analysis using review dataset. In *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 266–270. IEEE, 2019.