VOLUME 1, ISSUE 1 (2024)

PJSET

PHILIPPINE JOURNAL OF SCIENCE, ENGINEERING, AND TECHNOLOGY

RESEARCH ARTICLE

Unveiling Public Sentiment Towards ChatGPT: A Comprehensive Sentiment and Thematic Analysis of X (formerly Twitter) Discourse

Vince Ryan L. Arboleda[1*] | Brian Steven G. Pajarillo[1] | Louther Jan C. Adarle[1] | Dawn Angelika M. Arsenal[1] | Eddie G. de Paula Jr.[1]

[1]Graduate School, University of Saint La Salle, Bacolod City, 6100 Philippines

***Correspondence:** vinceryan.arboleda@gmail.com

Article History:
Received: Mar 25, 2024
Revised: May 18, 2024
Accepted: Jun 28, 2024

Keywords:
Generative AI
Sentiment Analysis
Social Media Analysis
Natural Language Processing
Public Perception
X (formerly Twitter)

Abstract

As generative AI technologies like ChatGPT become increasingly integrated into various aspects of daily life, understanding public perception is crucial for guiding responsible development and ethical deployment. This study conducts a comprehensive sentiment analysis of Twitter discourse, utilizing an innovative approach that integrates Plutchik's Wheel of Emotions, the NRC Word-Emotion Association Lexicon, and the VADER algorithm. By analyzing a dataset of 39,051 tweets, the research aims to identify predominant emotions, sentiment intensity and distribution (positive, negative, and neutral), and underlying themes within the discourse. The findings reveal that trust, anticipation, and joy are the most frequently expressed emotions, reflecting a generally positive reception of ChatGPT. Specifically, 54.4% of the tweets conveyed positive sentiments, 17.02% were negative, and 28.58% were neutral. Thematic analysis, facilitated by Latent Dirichlet Allocation (LDA) and Gibbs Sampling, uncovers key themes related to ChatGPT's potential, functionality, and utility. This study contributes to a deeper understanding of public attitudes towards generative AI technologies, providing valuable insights for developers, policymakers, and researchers in addressing the ethical, practical, and societal implications of AI integration into everyday life.

1 | INTRODUCTION

The integration of artificial intelligence (AI) into everyday life has accelerated rapidly, driving transformative changes across industries, enhancing personal productivity, and reshaping entertainment. Among the many advancements in AI, ChatGPT by OpenAI stands out as a significant milestone in natural language processing (NLP) [1]. Launched in 2020 as part of the Generative Pre-trained Transformer (GPT) series, ChatGPT has undergone multiple iterations to improve its ability to generate human-like text. Its versatile applications range from customer service and content creation to education and programming assistance, making it a widely used tool across professional and personal domains.

As ChatGPT becomes increasingly widespread, critical questions arise about its deployment. Understanding public sentiment towards AI technologies like ChatGPT is crucial for developers, policymakers, and educators [2]. According to Gupta et al. [3], sentiment analysis is the computational study of people's opinions, sentiments, attitudes, and emotions toward entities such as products, services, organizations, and technologies. It leverages NLP, text analysis, and computational linguistics to systematically identify, extract, and study subjective information in text data [4]. For generative AI, sentiment analysis provides valuable insights into public reception, uncovering areas of enthusiasm, concern, fear, and misunderstanding.

Sentiment analysis is important for AI technologies because it provides a detailed understanding of public

Cite as: V.L. Arboleda, B.G. Pajarillo, L.C. Adarle, D.M. Arsenal, and E. de Paula Jr., "Unveiling Public Sentiment Towards ChatGPT: Sentiment and Thematic Analysis of X (formerly Twitter) Discourse," PJSET, vol. 1, no. 1, 2024.

opinion. For developers and companies, it offers direct feedback on consumer reception, helping to shape product development and marketing strategies. Policymakers can also use these insights to draft regulations that balance consumer protection with encouraging innovation [5]. For academics and researchers, sentiment analysis sheds light on societal trends and concerns, fostering interdisciplinary studies on technology and its social impact [1].

Recent studies demonstrate the growing importance of sentiment analysis in understanding public attitudes towards AI. For instance, [6] highlights how social media shapes public opinion on AI, while [7] showcases its potential in predicting technology adoption rates. In particular, [5] explores the emotional dimensions of AI discourse on Twitter, emphasizing the need for sophisticated sentiment analysis tools. Proactive ethical considerations in AI development are essential in addressing risks and fostering public trust in these emerging technologies [8].

Conducting sentiment analysis on generative technologies involves gathering and analyzing large datasets from various sources, such as social media, forums, and product reviews. This task is supported by advanced algorithms and models that can process natural language, recognize patterns, and quantify sentiments. Tools like the Valence-Aware Dictionary & Sentiment Reasoner (VADER) algorithm and machine learning models have been instrumental in advancing sentiment analysis by improving accuracy and scalability [9].

This research paper investigates the public perception of ChatGPT by conducting a sentiment analysis of Twitter discourse using an innovative approach that combines Plutchik's Wheel of Emotion, the NRC Word-Emotion Association Lexicon, and the VADER algorithm. Specifically, the study seeks to answer the following questions: What are the predominant emotions expressed in Twitter discourse about ChatGPT? What is the intensity and proportion of positive, negative, and neutral sentiments in Twitter discourse about ChatGPT? What are the underlying themes discussed in Twitter discourse about ChatGPT?

2 | METHODOLOGY

A. Research Design

Given online discourse's dynamic nature and human emotions' complexities, a mixed-method approach is particularly well-suited for sentiment analysis research. This approach enables both a quantitative assessment of sentiment polarity and intensity across large datasets and a qualitative exploration of the nuanced expressions of sentiment [10].

The careful selection and customization of sentiment analysis algorithms play a crucial role in quantitative analysis. A dataset representing the broader discourse, collected over a specific timeframe, is essential for capturing sentiment trends. This allows for the application of statistical methods to examine the distribution and variance of sentiments across different conversations. On the other hand, the qualitative component is vital for contextualizing these quantitative results, providing deeper insights into the underlying reasons and emotions behind the sentiments expressed. Through thematic analysis, recurring themes or topics within the data can be identified, offering a more comprehensive understanding of public opinion [1].

B. Dataset

The ChatGPT tweets dataset on Kaggle is a curated collection of 39,051 English-language tweets. Published on December 20, 2022, this dataset includes contributions from researchers, practitioners, journalists, and the public, offering valuable insights into global sentiment surrounding ChatGPT. Along with tweet text, the dataset contains rich metadata such as tweet ID, timestamp, and user ID, capturing diverse perspectives on natural language processing, machine learning, and AI. The inclusion of the hashtag #chatgpt provides a focused view of online discussions related to ChatGPT, making the dataset a useful resource for analyzing public discourse on the technology [14].

C. Experiments Procedure

The experiment procedure, central to this investigation, was designed to analyze sentiment from a large corpus of ChatGPT-related tweets. Figure 1 outlines the steps: data collection, preprocessing, analysis, and interpretation.

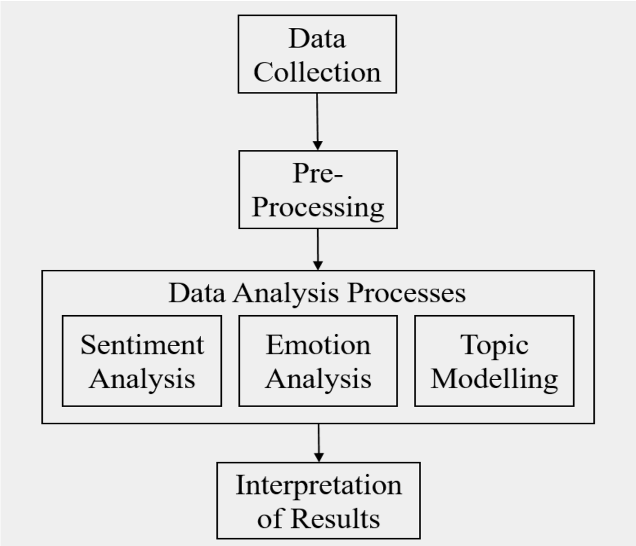

Figure 1. Experiment Procedure.

Data Collection. The dataset for this study was sourced from Kaggle. Researchers selected a dataset aligned with the research objectives, downloading it in CSV format for easy access and processing using RStudio. After downloading, the dataset underwent a preliminary integrity check to ensure completeness, consistency, and error-free data.

Preprocessing. Cleaning and preparing the dataset are critical steps in text sentiment analysis to ensure data quality and relevance. The following approaches were employed:

1. **Text normalization.** This process involved techniques to reduce noise, enhance readability, and improve the accuracy of sentiment analysis algorithms [12]. By normalizing text, researchers ensured consistent interpretation across various sources and formats. This step included removing special characters, punctuation, and numbers and converting text to lowercase, ensuring uniformity in representation [13].

2. **Word Tokenization.** Using the document term matrix function in RStudio, text was split into individual words or tokens, enabling sentiment analysis at the word or phrase level. Tokenization aids feature extraction, allowing sentiment algorithms to identify key terms that influence overall sentiment [14].

3. **Stopword Removal.** Frequently occurring words with little semantic meaning were removed to reduce noise. The researchers utilized predefined stopword lists from the `tm` library in RStudio [15]. They created a custom stopword list to exclude irrelevant terms, refining the dataset and improving the accuracy of the sentiment analysis [16], [17].

4. **Word Stemming.** The `PorterStemmer` function in RStudio was used for stemming, a technique that reduces words to their base or root forms. This step consolidated different morphological variants into a single representation [18], reducing the dataset's dimensionality [19], and improving the sentiment algorithm's ability to detect patterns [20], [21].

Data Analysis. The data analysis involved multiple steps to extract insights from the ChatGPT tweets dataset. The researchers applied Term Frequency (TF) as a feature engineering technique to capture syntactic information from the preprocessed data [19]. A sentiment analysis algorithm was then employed to classify sentiments. The document-term matrix, created for text mining, was mapped to a word-emoticon lexicon and analyzed using Plutchik's Wheel of Emotions [9], [17].

D. Data Analysis Tools

Data analysis tools are essential in research, particularly in text sentiment analysis, where they enable researchers to systematically examine large volumes of text and extract insights regarding emotional tone, opinions, and attitudes. Below are the data analysis tools employed in this study:

- **VADER Algorithm.** Hutto and Gilbert [9] designed the VADER algorithm as a lexicon and rule-based sentiment analysis tool for social media text. VADER handles informal language, slang, emoticons, and context-dependent sentiment expressions. It assigns sentiment scores based on a predefined lexicon, with polarity ranging from -1 (negative) to 1 (positive) or neutral, and intensity scores to capture the strength of sentiment. VADER has been widely used to analyze Twitter data, helping researchers gauge public opinion on political events, brand perception, and public health crises [20]. Its efficiency, ease of interpretation, and ability to process social media text make it well-suited for this study's analysis [21], [9].

- **Plutchik's Wheel of Emotions.** Plutchik's Wheel of Emotions is a model that categorizes emotions into eight primary categories: joy, trust, fear, anger, anticipation, surprise, sadness, and disgust, as shown in Figure 2. This framework is frequently applied in psychology, linguistics, and computer science emotion classification tasks [22], [23]. Researchers use the wheel to analyze text data and identify the emotions expressed by users in various settings, such as social media, customer reviews, and online forums. By leveraging Plutchik's framework, this study provides deeper insights into the emotional states and behavioral patterns expressed in the ChatGPT discourse [24].

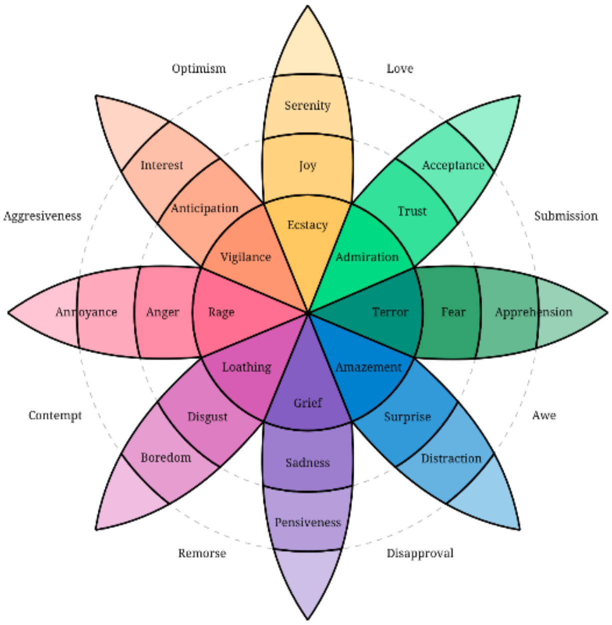

Figure 2. Plutchik's Wheel of Emotion.

- **NRC Word-Emotion Lexicon.** The NRC Word-Emotion Lexicon, developed by the National Research Council, is a comprehensive resource that associates words with emotional categories [25]. The lexicon categorizes thousands of words based on eight core emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. It assigns binary indicators to words based on the presence or absence of a particular emotion. This lexicon is commonly used in NLP applications, such as chatbots and sentiment analysis models, to classify text based on emotional content [26].

- **Latent Dirichlet Allocation (LDA).** LDA is a generative probabilistic model used for topic modeling, enabling the discovery of latent topics within text corpora [27]. The model assumes that each document is a mixture of a few topics, with words assigned to one of these topics based on co-occurrence patterns. LDA has proven versatile in analyzing scientific literature, social media discussions, and online forums, making it an effective tool for identifying themes and trends within the ChatGPT tweets dataset [28].

- **Gibbs Sampling Method.** In the context of LDA, Gibbs sampling is used to estimate the latent topic assignments for document words. Gibbs sampling helps estimate topic distributions and identify thematic patterns in the text by iteratively updating topic assignments based on word frequencies [29]. Researchers commonly use Gibbs sampling for topic modeling in various text corpora, including social media and scientific literature, to uncover key topics and thematic trends [30]. This study utilizes Gibbs sampling to support LDA in analyzing trends and themes within the ChatGPT-related tweets.

E. *Ethical Consideration*

The researchers followed Kaggle's ethical guidelines and data usage policies throughout the study. Proper dataset citation was crucial to ensure credit to the original contributors, uphold transparency in research practices, and comply with relevant data protection regulations and guidelines [28]. The dataset used in this study is publicly available and does not contain personally identifiable information or sensitive data, ensuring that individuals' privacy is protected and maintained.

3 | RESULTS AND DISCUSSIONS

A. *Predominant Emotions*

Based on the NRC Word-Emotion Lexicon and Plutchik's Wheel of Emotions, the emotion analysis revealed insightful patterns in public sentiment towards ChatGPT. The top three emotions are shown in Table 1. These findings highlight a generally positive reception, suggesting distinct perspectives that have significant implications for generative AI development, use, and public perception.

TABLE 1. Emotion Analysis Result.

Emotions	Frequency	Percentage
Trust	23,304	24.31
Anticipation	18,580	19.38
Joy	16,069	16.77
Fear	10,148	10.59
Surprise	7,999	8.34
Sadness	7,804	8.14
Anger	7,375	7.70
Disgust	4,570	4.77
Total Instances	**95,849**	**100.00**

The leading emotion, trust (24.31%), reflects users' confidence in ChatGPT's capabilities, attributed to its effectiveness in tasks like writing assistance, coding, or engaging in meaningful conversations. This strong sense of trust suggests that users view ChatGPT as a reliable and valuable tool, underscoring the importance of accuracy and consistency to maintain public confidence [2], [5].

Anticipation (19.38%) ranks second, reflecting the public's excitement and hopeful expectations for the future of ChatGPT and similar technologies. This emotion may stem from the innovative possibilities ChatGPT offers, such as creativity, productivity, and problem-solving potential, which users find promising in various domains [1], [6]. Joy (16.77%), the third most prevalent emotion, highlights users' pleasure and satisfaction from interacting with ChatGPT. This positive emotional response suggests that users appreciate its capabilities and potential, particularly in enhancing user experiences across fields like education, entertainment, and professional work [7], [24].

B. *Intensity and Proportions of Sentiments*

The sentiment analysis using the VADER algorithm reveals a compelling narrative about public sentiment toward ChatGPT. Positive sentiments dominate the dataset at 54.40%, reflecting a favorable view among Twitter users (Table 2). This enthusiasm likely stems from ChatGPT's perceived utility, innovation, and efficiency in various tasks. The widespread positive sentiment underscores the potential market demand and societal acceptance of generative AI technologies [3], [9].

TABLE 2. VADER Sentiment Analysis Result.

Sentiment	Frequency	Percentage
Positive	21, 243	54.40
Negative	6, 648	17.02
Neutral	11,160	28.58
Total	**39,051**	**100.00**

Despite the overall positive sentiment, 17.02% of the tweets expressed negative emotions, highlighting key areas for improvement. These concerns may arise from unmet expectations, ethical issues, fears of job displacement, or the dehumanization of interactions with AI. Addressing these criticisms is crucial for developers to improve user experience and build trust [4], [8]. The neutral sentiment of 28.58% suggests that many users either hold ambivalent views about ChatGPT or use it in contexts that don't evoke strong emotions. This allows developers to convert neutral users into advocates by enhancing functionality and demonstrating practical value in various applications [6], [20].

To validate these findings, a Chi-Square test was performed to determine if the distribution of sentiments was statistically significant. The high proportion of positive sentiment (54.40%) was found to be statistically significant, indicating a meaningful real-world impact of ChatGPT on user satisfaction and adoption [9], [19].

C. *Underlying Themes*

The researchers employed LDA and Gibbs Sampling to uncover the key topics discussed by users in relation to ChatGPT. The study provided a concise summary of the main subjects discussed in the dataset by identifying the top 20 terms associated with each theme, as shown in Table 3.

TABLE 3. Top 20 Terms to Topic Generation

No.	Topic 1	Topic 2	Topic 3
1	ChatGPT	ChatGPT	use
2	OpenAI	ask	like
3	new	write	answer
4	think	code	question
5	now	GPT	make
6	one	Google	work
7	know	create	human
8	good	chatbot	see
9	people	day	help
10	thing	time	learn
11	future	AI	try
12	world	prompt	way
13	Elon Musk	first	tool
14	come	story	model
15	say	technology	look
16	chat	year	language
17	talk	content	generate
18	start	Search	give
19	job	Twitter	conversation
20	play	generate	better

- **The Potential and Impact of ChatGPT**: This theme centers around the transformative potential of ChatGPT, highlighting its societal and technological impact. Keywords like "future," "world," and "job" suggest discussions on how ChatGPT might influence the future of work, innovation, and human-computer interactions [1], [6].
- **The Functionality and Capabilities of ChatGPT**: The second theme explores ChatGPT's technical features, such as writing, coding, and text generation. The inclusion of terms like "GPT," "Google," and "Artificial Intelligence" underscores the comparison of ChatGPT to other technologies, emphasizing its versatility across different applications [3], [19].
- **ChatGPT as a Tool for Humans**: The third theme focuses on the interaction between humans and ChatGPT, particularly its role in learning, communication, and problem-solving. Words like "answer," "question," and "help" suggest ChatGPT's utility in assisting users with information and tasks, further emphasizing its value as an AI tool [2], [23].

4 | CONCLUSION

The sentiment analysis of Twitter conversations about ChatGPT reveals that public perception is largely positive, with trust, anticipation, and joy being the most dominant emotions. These sentiments reflect strong confidence in ChatGPT's reliability, optimism about its future potential, and satisfaction with its current capabilities. This favorable reception underscores the importance of maintaining high ethical standards and consistent development to sustain public trust and enthusiasm. The widespread feelings of anticipation and joy suggest that people see AI, including ChatGPT, as a tool that enhances productivity, sparks creativity, and brings personal satisfaction—highlighting the value placed on its positive impact on daily life. However, the analysis also identifies a smaller but notable presence of negative sentiments, pointing to areas where ChatGPT may fall short or raise concerns. These concerns often relate to issues such as content authenticity, privacy, and the potential impact of AI on jobs and social interactions. Addressing these criticisms is essential to alleviating fears and fostering a more ethical and inclusive approach to AI development.

Using methods like Latent Dirichlet Allocation (LDA) and Gibbs sampling, the thematic analysis further uncovers key discussions about ChatGPT's potential, capabilities, and usefulness in supporting human activities. These themes reflect broader societal questions about AI's role in the future of work, technological advancement, and its integration into various aspects of life. There is a clear need for ongoing research focused on improving transparency, addressing ethical challenges, and ensuring that AI development aligns with societal values. Monitoring public sentiment and engaging with diverse stakeholders will provide valuable feedback to guide the responsible evolution of ChatGPT and similar technologies. Additionally, further exploration of AI's long-term effects on employment, privacy, and social dynamics will ensure that these innovations benefit society in meaningful and equitable ways.

REFERENCES

[1] N. Bakalos, N. Papadakis, and A. Litke, "Public Perception of Autonomous Mobility Using ML-Based Sentiment Analysis over Social Media Data," *Logistics*, vol. 4, no. 2, p. 12, 2020, doi: 10.3390/logistics4020012.

[2] S. Kang, Y. Choi, and B. Kim, "Impact of motivation factors for using generative AI services on continuous use intention: Mediating trust and acceptance attitude," Social Sciences, vol. 13, no. 9, p. 475, 2024, doi: 10.3390/socsci13090475.

[3] P. Gupta, B. Ding, C. Guan, and D. Ding, "Generative AI: A systematic review using topic modeling techniques," Data and Information Management, vol. 8, no. 2, pp. 100066, 2024, doi: 10.1016/j.dim.2024.100066.

[4] H. Taherdoost and M. Madanchian, "Artificial Intelligence and Sentiment Analysis: A Review in Competitive Research," Computers, vol. 12, no. 2, p. 37, 2023, doi: 10.3390/computers12020037.

[5] J. O. Krugmann and J. Hartmann, "Sentiment Analysis in the Age of Generative AI," Customer Need and Solution, vol. 11, no. 3, 2024, doi: 10.1007/s40547-024-00143-4.

[6] T. Davenport, A. Guha, D. Grewal, and T. Bressgott, "How artificial intelligence will change the future of marketing," Journal of the Academy of Marketing Science, vol. 48, pp. 24-42, 2019, doi: 10.1007/s11747-019-00696-0.

[7] H. Saragih and J. Manurung, "Leveraging the BERT Model for Enhanced Sentiment Analysis in Multicontextual Social Media Content," Computational Intelligence and Technology, vol. 16, pp. 82-89, May 2024, doi: 10.35335/cit.Vol16.2024.766.pp82-89.

[8] T. Anderson, S. Sarkar, and R. Kelley, "Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques," Natural Language Processing Journal, vol. 8, pp. 100097, 2024, doi: 10.1016/j.nlp.2024.100097.

[9] C. J. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," Eighth International AAAI Conference on Weblogs and Social Media, 2014, doi: 10.1609/icwsm.v8i1.14550.

[10] G. Bello, H. Menéndez, S. Okazaki, and D. Camacho, "Extracting Collective Trends from Twitter Using Social-Based Data Mining," in Lecture Notes in Computer Science, vol. 8083, pp. 622-630, Sep. 2013. doi: 10.1007/978-3-642-40495-5_62.

[11] S. Prata, "ChatGPT Tweets Dataset," accessed Mar. 2, 2024. [Online]. Available: https://bit.ly/4bzMtrS.

[12] C. Sindhu, B. Sasmal, R. Gupta, and J. Prathipa, "Subjectivity Detection for Sentiment Analysis on Twitter Data," in *Artificial Intelligence Techniques for Advanced Computing Applications*, D. Hemanth, G. Vadivu, M. Sangeetha, and V. Balas, Eds. Lecture Notes in Networks and Systems, vol. 130, Singapore: Springer, 2021, pp. 555-563, doi: 10.1007/978-981-15-5329-5_43.

[13] I. Istifci and A.D. Ucar, "A Review of Research on the Use of Social Media in Language Teaching and Learning," Journal of Educational Technology and Online Learning, vol. 4, Sep. 2021, doi: 10.31681/jetol.922968.

[14] H. Wickham, "Tidyverse: Easily Install and Load the 'Tidyverse'," R package version 1.3.0, 2019. [Online]. Available: https://CRAN.R-project.org/package=tidyverse.

[15] I. Feinerer, K. Hornik, and D. Meyer, "Text Mining Infrastructure in R," Journal of Statistical Software, vol. 25, no. 5, pp. 1-54, 2008, doi: 10.18637/JSS.V025.I05.

[16] M. Javed and S. Kamal, "Normalization of Unstructured and Informal Text in Sentiment Analysis," *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2018, doi: 10.14569/IJACSA.2018.091011.

[17] P. Dandannavar, S. Mangalwede, and S. Deshpande, "Emoticons and Their Effects on Sentiment Analysis of Twitter Data," in Proceedings, pp. 191-201, Jan. 2020, doi: 10.1007/978-3-030-19562-5_19.

[18] S. Pradha, M. Halgamuge, and Q. Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data," in Proc. 11th Int. Conf. Knowledge and Syst. Eng., Oct. 2019, doi: 10.1109/KSE.2019.8919368.

[19] A. Nandi and P. Sharma, "Comparative Study of Sentiment Analysis Techniques," in *Sentiment Analysis Techniques," vol. 1, p. 188, Jun. 2021, doi:10.1201/9781003202240-72.

[20] K. Jaidka, S. Giorgi, H.A. Schwartz, M.L. Kern, L.H. Ungar, and J.C. Eichstaedt, "Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods," Proc. Natl. Acad. Sci. U.S.A., vol. 117, no. 19, pp. 10165–10171, 2020, doi: 10.1073/pnas.1906364117.

[21] M. HarishRao and S. D.R, "Automatic product review sentiment analysis using Vader and feature Visualization," Int. J. Comput. Sci. Eng. Inf. Technol. Res., vol. 7, no. 4, pp. 53–66, 2017, doi: 10.24247/ijcseitraug20178.

[22] K. K. Aldous, J. An, and B. J. Jansen, "Measuring 9 emotions of news posts from 8 news organizations across 4 social media platforms for 8 months," ACM Trans. Social Comput., vol. 4, no. 4, pp. 1–31, 2021, doi: 10.1145/3516491.

[23] D. Griol, J. M. Molina and J. García-Herrero, "Fusion of sentiment analysis and emotion recognition to model the user's emotional state," 2015 18th International Conference on Information Fusion (Fusion), Washington, DC, USA, 2015, pp. 814-822. Available: https://ieeexplore.ieee.org/document/7266644.

[24] R. Plutchik, "Emotions and psychotherapy: A psychoevolutionary perspective," Emotion, Psychopathology, and Psychotherapy, pp. 3–41, 1990, doi: 10.1016/b978-0-12-558705-1.50007-5.

[25] F. S. Tabak and V. Evrim, "Comparison of emotion lexicons," in 2016 HONET-ICT, 2016, pp. 154–158, doi: 10.1109/honet.2016.7753440.

[26] D. Andrzejewski and X. Zhu, "Latent dirichlet allocation with topic-in-set knowledge," Proc. NAACL HLT Workshop Semi-Supervised Learn. Nat. Lang. Process., pp. 43–48, 2009, doi: 10.3115/1621829.1621835.

[27] D. M. Blei, A. Ng, and M.I. Jordan, "Latent Dirichlet Allocation." Available: https://acsweb.ucsd.edu/~yuw176/report/lda.pdf

[28] J. Arrivillaga, D. Greenleaf, M. Hawthorn, and R. Alvarado, "Revealing the landscape: Detecting trends in a scientific corpus," in 2016 IEEE Syst. Inf. Eng. Design Symp. (SIEDS), pp. 292–297, 2016, doi: 10.1109/sieds.2016.7489317.

[29] T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proc. Natl. Acad. Sci. U.S.A., vol. 101, suppl. 1, pp. 5228–5235, 2004, doi: 10.1073/pnas.0307752101.

[30] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.