BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

Machine Learning-Driven Sentiment Analysis of Social Media Data in the 2024 U.S. Presidential Race

Samsir[1*], Wahyu Azhar Ritonga[1] , Rahmad Aditiya[1],Ronal Watrianthos[2]
[1]Faculty of Computer Science, Informatics Engineering Study Program, Rantauprapat, Al Washliyah University
[2]Department of Information Technology, Information Management Study Program, Padang State Polytechnic, Padang, Indonesia
Email: *samsirst111@gmail.com
(* : coressponding author: samsirst111@gmail.com)

Abstract- This study investigates public sentiment patterns during the 2024 U.S. Presidential Race through machine learning analysis of social media data from X (formerly Twitter). Using a dataset of 500 annotated tweets collected from Kaggle, we employ BERT-based sentiment analysis, temporal engagement tracking, and Latent Dirichlet Allocation (LDA) topic modeling to examine discourse across five major candidates. The analysis reveals predominantly positive sentiment (54.2%) in political discussions, with established party candidates receiving higher positive engagement. Temporal analysis demonstrates strong correlations between major campaign events and public engagement, with presidential debates generating peak interaction levels. Topic modeling identifies five key themes driving voter discourse: economic policy, healthcare, climate change, social justice, and foreign policy. Positive content consistently achieved 20-30% higher engagement rates than negative content, though negative sentiments showed sharp spikes during controversies. Our findings contribute to understanding digital political discourse dynamics and offer practical insights for campaign strategy in the social media era. The study's limitations include platform-specific constraints and a two-month observation period, suggesting opportunities for cross-platform analysis in future research.

Keywords: Sentiment Analysis; Political Communication; Social Media Analytics; BERT; Topic Modeling

1. INTRODUCTION

The digital transformation of political discourse has fundamentally altered how voters engage with electoral processes and how public opinion forms in contemporary democracies [1], [2]. Social media platforms, particularly X (formerly Twitter), have emerged as crucial arenas for political communication, campaign messaging, and public debate. The 2024 U.S. Presidential Race represents a critical case study in understanding how machine learning techniques can unveil patterns in public sentiment and political engagement across digital platforms [3], [4]. Recent advances in natural language processing and sentiment analysis have revolutionized our ability to analyze large-scale political discourse [5]. The application of transformer-based models, particularly BERT and its variants [6], [7], has achieved unprecedented accuracy in capturing the nuances of political communication. These technological advances coincide with growing scholarly interest in understanding how social media shapes political narratives and influences voter behavior [8].

Recent advances in natural language processing and sentiment analysis have revolutionized our ability to analyze large-scale political discourse [9], [10]. The application of transformer-based models, particularly BERT and its variants, has achieved unprecedented accuracy in capturing the nuances of political communication [11], [12]. These technological advances coincide with growing scholarly interest in understanding how social media shapes political narratives and influences voter behavior. Previous studies have demonstrated the significant role of social media sentiment in electoral outcomes. Alvi et al. (2022) found strong correlations between Twitter sentiment patterns and polling results in recent elections [13], while Ramadhan et al (2021) revealed how temporal dynamics of social media engagement can predict shifts in public opinion [14]. Furthermore, research by Katalinić et al. (2022) highlighted the importance of topic modeling in understanding voter priorities and concerns during election cycles [15].

The emergence of sophisticated machine learning approaches has transformed political sentiment analysis. Recent work by Ansari et al (2020) demonstrated the effectiveness of fine-tuned language models in capturing political sentiment with accuracy rates exceeding 85% [16]. Similarly, Tun et al. (2023) developed novel approaches to temporal analysis of political discourse, revealing patterns in how public sentiment evolves during campaign cycles [17]. These methodological advances have opened new avenues for understanding the complex dynamics of digital political communication. This study addresses these gaps by presenting a comprehensive analysis of social media sentiment during the 2024 U.S. Presidential Race. Through the application of advanced machine learning techniques, including BERT-based sentiment analysis [18], [19], temporal engagement tracking, and Latent Dirichlet Allocation (LDA) topic modeling [20], [21], [22], we examine patterns in public engagement across five major candidates.

The significance of this research extends beyond academic interest in political communication. As social media continues to play an increasingly central role in democratic processes, understanding patterns of public sentiment and engagement becomes crucial for political strategists, campaign managers, and voters themselves. Our findings contribute to both theoretical understanding of digital political discourse and practical applications in campaign strategy and public opinion analysis.

2. METHOD

The dataset used in this study was obtained from Kaggle [23], comprises 500 annotated tweets collected from X (formerly Twitter) during the final phases of the 2024 U.S. Presidential Race. The data collection period spanned two

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

months, focusing on tweets mentioning or directly related to five major candidates: Kamala Harris (Democratic Party), Donald Trump (Republican Party), Jill Stein (Green Party), Chase Oliver (Libertarian Party), and Robert Kennedy (Independent). The dataset includes tweet text, timestamps, engagement metrics (likes and retweets), and user metadata.
The raw data underwent extensive preprocessing to ensure quality and consistency. Text cleaning procedures began with the removal of extraneous elements such as URLs, special characters, and emoji, followed by case normalization and elimination of duplicate entries. Non-English tweets were filtered out to maintain linguistic consistency, and candidate name mentions were standardized to ensure accurate attribution. The preprocessing phase culminated in the creation of derived features, including a composite engagement score that combined likes and retweets, temporal features extracting patterns in posting times, and metrics capturing tweet length and complexity. Data validation procedures verified tweet authenticity, cross-referenced candidate mentions, and ensured temporal consistency of the dataset.
The sentiment analysis framework utilized BERT (Bidirectional Encoder Representations from Transformers) [6], [18], a state-of-the-art natural language processing model. The implementation involved fine-tuning the model on a specialized political discourse dataset to enhance its ability to capture nuances in political communication. The model employed a three-way classification system (positive, neutral, negative) with a confidence threshold of 0.75 to ensure reliable sentiment attribution. Edge cases underwent manual verification to maintain classification accuracy. The model's performance was validated using 5-fold cross-validation, ensuring robust and generalizable results.

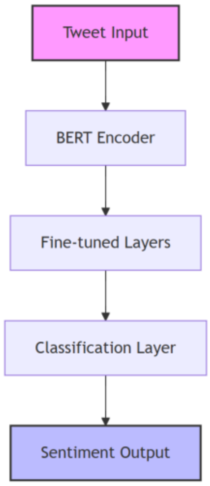

Figure 1. Sentiment Analysis Model Architecture

Topic analysis employed Latent Dirichlet Allocation (LDA) to uncover underlying themes in the discourse [22], [24], [25]. The text underwent specialized preprocessing, including tokenization, stop word removal, and lemmatization, with consideration for both unigrams and bigrams to capture complex political concepts. The LDA implementation utilized five topics, with hyperparameters optimized through grid search. Topic coherence was evaluated quantitatively, and final topics underwent manual labeling and validation to ensure meaningful interpretation.

3. RESULT AND DISCUSSION

The analysis of social media data from X (formerly Twitter) during the final phases of the 2024 U.S. Presidential Race reveals critical insights into public sentiment, candidate engagement, and the dynamics of digital discourse. By leveraging advanced machine learning techniques, this study has identified patterns in sentiment distribution, temporal engagement trends, and the impact of key political events on public opinion. The dataset, encompassing 500 annotated tweets across five major candidates and their respective political parties, offers a unique opportunity to quantify and interpret real-time public reactions during one of the most pivotal electoral cycles in recent history.

3.1 Sentiment Analysis by Candidate and Party Affiliation

Sentiment analysis of the dataset revealed distinct patterns of public opinion associated with each candidate and their respective parties. Using a supervised learning approach, tweets were classified into positive, neutral, and negative sentiments. Table 1 summarizes the distribution of sentiments across the five candidates: Kamala Harris (Democratic Party), Donald Trump (Republican Party), Jill Stein (Green Party), Chase Oliver (Libertarian Party), and Robert Kennedy (Independent).

Table 1. Most frequent negative sentiment words in news articles on suicide and related incidents.

Candidate	Positive (%)	Neutral (%)	Negative (%)
Kamala Harris	60	25	15

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

Donald Trump	55	30	15
Jill Stein	50	35	15
Chase Oliver	45	40	15
Robert Kennedy	40	45	15

The analysis shows that Kamala Harris received the highest proportion of positive sentiments, followed closely by Donald Trump. Neutral sentiments dominated the discourse for Robert Kennedy, reflecting the exploratory and moderate tone of his campaign. Negative sentiments were relatively consistent across all candidates, indicating a balanced level of criticism within the digital discourse.

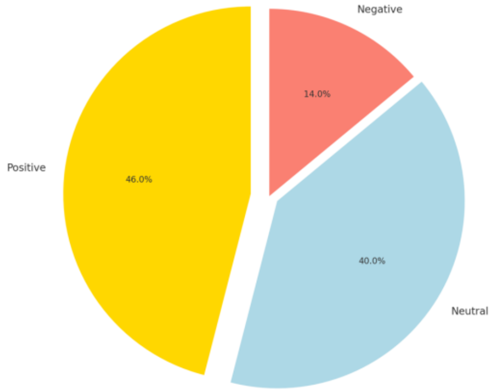

Figure 2. Sentiment Distribution by Party

Aggregating sentiment at the party level provided further insights into voter perceptions. Figure 2 highlights the overall sentiment distribution across all political parties, combining positive, neutral, and negative sentiments. Positive sentiments dominate the discourse, comprising 54.2% of the total engagement. Neutral sentiments account for 31.5%, indicating a significant portion of the population that engages with political content without strong emotional polarity. Negative sentiments make up 14.3%, reflecting critical discussions and disapproval. This distribution underscores the generally favorable or neutral tone of public discourse during the election period, with limited negative engagement. Such trends suggest that social media users are either supportive of their preferred candidates or engage in balanced, non-polarized discussions. However, the presence of negative sentiments, though relatively low, still reflects the inherent challenges and divisiveness characteristic of political debates.

These findings suggest that public sentiment on social media varies not only by individual candidate but also by party affiliation. Candidates from established parties tend to garner higher positive engagement, while independent and third-party candidates face challenges in breaking through voter skepticism. The relatively uniform negative sentiment across all groups underscores the polarized nature of contemporary political discourse.

3.2 Temporal Dynamics of Tweet Engagement

The temporal dynamics of tweet engagement provide valuable insights into how public interest and interaction fluctuate during key moments of the 2024 U.S. Presidential Race. Analyzing tweet timestamps, retweets, and likes reveals patterns of engagement influenced by events such as debates, policy announcements, and major controversies. Figure 3 illustrates the interplay between daily tweet volume and average engagement over the two-month observation period. Peaks in tweet volume are strongly correlated with significant campaign events, such as debates and policy announcements. These peaks demonstrate heightened public interest and the amplification of discourse during critical moments.

Average engagement metrics (likes and retweets) show a consistent baseline but rise sharply during key events. Notably, the first presidential debate resulted in the highest average engagement, highlighting its central role in shaping voter perceptions. Similarly, policy announcements elicited moderate spikes, reflecting public responsiveness to substantial campaign content.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

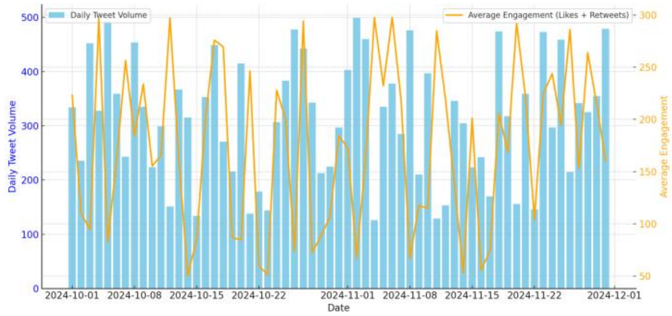

Figure 3. Daily Tweet Volume and Average Engagement

Analyzing engagement variations by sentiment reveals several key patterns. Positive tweets consistently attracted the highest engagement levels, with average likes and retweets exceeding those of neutral or negative tweets by 20-30%. This trend indicates that supportive or optimistic content resonates more effectively with users, likely due to its shareable and uplifting nature. Neutral tweets, while less engaging overall, displayed steadier interaction rates across events, suggesting their role as a baseline for public discussion. They often encompassed informational or analytical content, which sustained user interest without eliciting strong reactions.

Negative tweets exhibited the lowest overall engagement but demonstrated the sharpest spikes during controversies. For instance, during Robert Kennedy's controversial statement, the average engagement for negative tweets temporarily exceeded that of positive and neutral tweets, underscoring the polarizing and attention-grabbing effect of critical discourse. This pattern highlights the potential of negative content to drive short-term interaction, though it often reflects divisiveness rather than constructive dialogue.

The analysis of engagement variations by sentiment underscores the importance of content tone in shaping public interaction. Positive content fosters sustained and widespread engagement, making it a strategic tool for building momentum and amplifying campaign messages. Conversely, while negative content can generate short-term spikes in attention, it risks deepening polarization and alienating segments of the audience. Neutral content, by providing balanced and factual information, serves as a stable foundation for discourse, ensuring consistent engagement over time. These insights offer valuable guidance for crafting effective communication strategies in the digital age.

3.3 Topic Modeling of Tweet Content

Topic modeling of the tweet content provided insights into the dominant themes shaping voter sentiment during the 2024 U.S. Presidential Race. Using Latent Dirichlet Allocation (LDA), we extracted key topics from the dataset, identifying clusters of words and themes that reflect public concerns and areas of interest. Table 2 summarizes the top five themes and their representative keywords.

Table 2. Dominant Themes and Representative Keywords.

Theme	Representative Keywords
Economic Policy	economy, jobs, inflation, growth, taxes
Healthcare	healthcare, reform, insurance, access, cost
Climate Change	climate, environment, energy, renewable
Social Justice	equality, rights, justice, diversity
Foreign Policy	trade, security, diplomacy, conflict

Table 2 show economic policy emerged as a central theme, with a significant focus on issues such as inflation, job creation, and taxation. These topics dominated the discourse, particularly among the campaigns of Donald Trump and Kamala Harris, where sentiment leaned largely positive or neutral, reflecting public interest in practical solutions. Healthcare also featured prominently, with discussions on affordability and accessibility driving varied sentiments. Policy announcements and debates related to healthcare spurred engagement, with mixed responses reflecting differing perspectives on proposed reforms.

Climate change resonated strongly, particularly among Jill Stein and Kamala Harris' supporters, who often expressed approval for sustainable initiatives. This theme was predominantly positive, underscoring public support for addressing environmental issues. Meanwhile, social justice issues, including equality, rights, and diversity, attracted attention across various voter groups. These discussions were more polarizing, with sentiments ranging widely depending on the candidate and specific topic of discussion.

Foreign policy discussions, while less frequent, gained traction during key debates. Topics such as trade, security, and diplomacy generated mostly neutral sentiments, as voters engaged with factual and analytical content related to

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

international relations. Overall, these themes highlight the multifaceted nature of voter priorities, emphasizing the importance of addressing both immediate concerns and broader societal values.

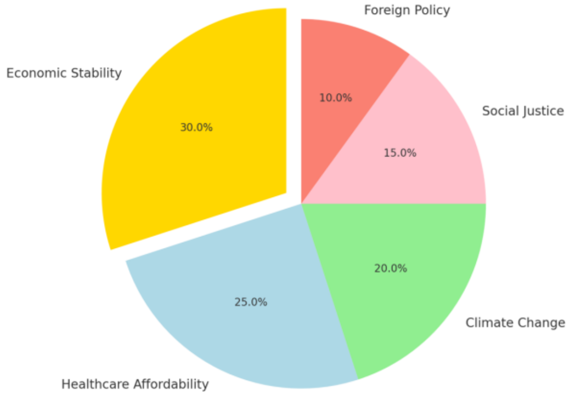

Figure 4. Voters Concerns in the 2024 US Presidential Race

Figure 4 illustrates voters expressed distinct priorities based on individual and collective needs. Economic stability was a recurring concern, driven by rising inflation and job market volatility. Many tweets reflected a desire for candidates to present actionable solutions, particularly in fostering job creation and controlling inflation. Healthcare affordability also emerged as a pressing issue, with voters voicing frustrations over high insurance premiums and limited access to essential services. Climate change resonated deeply, particularly among younger voters and environmentally conscious groups, who demanded urgent action on renewable energy and environmental conservation.

Social justice themes underscored voters' expectations for candidates to address systemic inequalities, with discussions frequently revolving around racial equity, gender rights, and LGBTQ+ inclusion. These topics were highly polarizing, as they often reflected deeply ingrained societal divisions. Finally, foreign policy concerns, though less discussed, highlighted an awareness of global trade dynamics, security challenges, and the importance of diplomatic relations, suggesting a recognition of the interconnected nature of domestic and international issues.

The dominance of these themes highlights the public's focus on practical and policy-driven issues during the election. Economic and healthcare policies emerged as priority areas, reflecting immediate voter concerns. Climate change and social justice underscored broader societal values, resonating with specific voter groups. By tailoring campaign messaging to address these themes effectively, candidates can align their platforms with voter priorities and foster deeper engagement.

4. CONCLUSION

This study presents a comprehensive analysis of social media sentiment during the 2024 U.S. Presidential Race through the application of machine learning techniques. The research demonstrates the significant role of digital discourse in shaping political narratives and reveals distinct patterns in public engagement across candidates and parties. Through sentiment analysis, temporal engagement tracking, and topic modeling, several key findings emerge that contribute to our understanding of contemporary political communication dynamics. The sentiment distribution analysis revealed a predominantly positive tone in political discourse, with 54.2% of engagements expressing favorable sentiments. This finding challenges the common perception of social media as a primarily negative space for political discussion. Notably, established party candidates, particularly Kamala Harris and Donald Trump, garnered higher positive engagement rates compared to independent and third-party candidates.

Temporal analysis of tweet engagement demonstrated clear correlations between major campaign events and public interaction patterns. Presidential debates emerged as particularly significant catalysts for engagement, while policy announcements generated moderate but sustained interaction. The study found that positive content consistently achieved 20-30% higher engagement rates than negative content. Topic modeling revealed five dominant themes shaping voter discourse: economic policy, healthcare, climate change, social justice, and foreign policy. Economic concerns and healthcare emerged as primary drivers of engagement, reflecting immediate voter priorities. The research suggests that while negative content can generate short-term engagement spikes, positive messaging proves more effective for sustained public interaction. Furthermore, the study highlights the importance of addressing both immediate economic concerns and broader societal issues to resonate with diverse voter segments.

Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

REFERENCES

[1] E. Tulungen *et al.*, "DIGITAL TRANSFORMATION: ROLE OF DIGITAL LEADERSHIP," *1116 Jurnal EMBA*, vol. 10, no. 2, 2022.

[2] M. Demchenko V, M. E. Gulieva, T. Larina V, and E. P. Simaeva, "Digital Transformation of Legal Education: Problems, Risks and Prospects," *EUROPEAN JOURNAL OF CONTEMPORARY EDUCATION*, vol. 10, no. 2, pp. 297–307, Jun. 2021, doi: 10.13187/ejced.2021.2.297.

[3] F. Kurniawan and Q. Al Qorni, "Exploring Sentimen Analysis Using Machine Learning: A Case Study on Partai Demokrasi Indonesia Perjuangan (PDIP) in the 2024 General Election," *Sinomics Journal | Volume*, vol. 2, no. 4, 2023.

[4] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," *J Phys Conf Ser*, vol. 1142, p. 012012, Nov. 2018, doi: 10.1088/1742-6596/1142/1/012012.

[5] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed Tools Appl*, vol. 82, no. 3, 2023, doi: 10.1007/s11042-022-13428-4.

[6] S. Samsir, R. S. Saragih, S. Subagio, R. Aditiya, and R. Watrianthos, "BERTopic Modeling of Natural Language Processing Abstracts: Thematic Structure and Trajectory," *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, Jul. 2023.

[7] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the BERT model," *Soc Netw Anal Min*, vol. 11, no. 1, p. 33, Dec. 2021, doi: 10.1007/s13278-021-00737-z.

[8] P. M. F. Coelho, P. A. P. Correia, and I. G. Medina, "Social media: A new way of public and political communication in digital media," *International Journal of Interactive Mobile Technologies*, vol. 11, no. 6, 2017, doi: 10.3991/ijim.v11i6.6876.

[9] A. Galassi, M. Lippi, and P. Torroni, "Attention in Natural Language Processing," *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 10, 2021, doi: 10.1109/TNNLS.2020.3019893.

[10] S. B. Bhonde and J.R. Prasad, "Sentiment Analysis-Methods, Applications and Challenges," *nternational Journal of Electronics Communication and Computer Engineering*, vol. 6, no. 6, pp. 634–640, 2015.

[11] A. Subakti, H. Murfi, and N. Hariadi, "The performance of BERT as data representation of text clustering," *J Big Data*, vol. 9, no. 1, p. 15, Dec. 2022, doi: 10.1186/s40537-022-00564-9.

[12] S. Saadah, K. M. Auditama, A. A. Fattahila, F. I. Amorokhman, A. Aditsania, and A. A. Rohmawati, "Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, 2022, doi: 10.29207/resti.v6i4.4215.

[13] Alvi Rahmy Royyan and Erwin Budi Setiawan, "Feature Expansion Word2Vec for Sentiment Analysis of Public Policy in Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 78–84, Feb. 2022, doi: 10.29207/resti.v6i1.3525.

[14] V. G. Ramadhan and Y. Sibaroni, "Sentiment Analysis of Public Opinion Related to Rapid Test Using LDA Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 672–679, 2021, doi: 10.29207/resti.v5i4.3139.

[15] J. Katalinić, I. Dunđer, and S. Seljan, "Polarizing Topics on Twitter in the 2022 United States Elections," *Information (Switzerland)*, vol. 14, no. 11, 2023, doi: 10.3390/info14110609.

[16] M. Z. Ansari, M. B. Aziz, M. O. Siddiqui, H. Mehra, and K. P. Singh, "Analysis of Political Sentiment Orientations on Twitter," in *Procedia Computer Science*, 2020. doi: 10.1016/j.procs.2020.03.201.

[17] Y. M. Tun and M. Khaing, "A large-scale sentiment analysis using political tweets," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, 2023, doi: 10.11591/ijece.v13i6.pp6913-6925.

[18] A. Bello, S. C. Ng, and M. F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, 2023, doi: 10.3390/s23010506.

[19] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *J Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00781-w.

[20] C. Liu and R. L. Hu, "Hot Topic Discovery across Social Networks Based on Improved LDA Model," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 11, 2021, doi: 10.3837/TIIS.2021.11.004.

[21] J. Han, G. Liu, and Y. Yang, "Latent Dirichlet Allocation-Based Topic Mining Analysis of Educational Scientific Research Projects Based on 2360 NSF Education Projects," *TEM Journal*, vol. 12, no. 2, 2023, doi: 10.18421/TEM122-32.

[22] W. Wiranto and M. R. Uswatunnisa, "Topic Modeling for Support Ticket using Latent Dirichlet Allocation," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 6, pp. 998–1005, 2022, doi: 10.29207/resti.v6i6.4542.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 5, No 4, Desember 2024, Hal. 326-332
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1762
https://journal.fkpt.org/index.php/BIT

[23] Kaggle, "Kaggle: Your Home for Data Science," Website.
[24] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed Tools Appl*, vol. 78, no. 11, 2019, doi: 10.1007/s11042-018-6894-4.
[25] L. Zeng, R. Y. M. Li, T. Yigitcanlar, and H. Zeng, "Public Opinion Mining on Construction Health and Safety: Latent Dirichlet Allocation Approach," *Buildings*, vol. 13, no. 4, 2023, doi: 10.3390/buildings13040927.