

# Sentiment and Topic Analysis of Public Opinion on Indonesia's Minister of Finance Using IndoBERTweet, TF-IDF, and Latent Dirichlet Allocation

Surya Sujarwo<sup>1\*</sup>, Jeklin Harefa<sup>2</sup>, Alexander<sup>3</sup>

<sup>1-3</sup>Computer Science Department, School of Computer Science  
Bina Nusantara University,  
Jakarta, Indonesia 11480  
surya.wang@binus.ac.id; jeklin.harefa001@binus.ac.id;  
alexander001@binus.ac.id

\*Correspondence: surya.wang@binus.ac.id

**Abstract** – In today's technology-based society, people share their opinions on online social media platforms, which can be used as data for sentiment analysis. One of the most popular platforms for obtaining publicly accessible data is X. This study analyzes public views of the Ministry of Finance (MoF) by examining 9,543 tweets gathered from February to September 2025. The data collected was preprocessed through cleaning, name entities grouping, and keywords filtering, then evaluated using IndoBERTweet, and keywords were extracted using the Term Frequency-Inverse Document Frequency (TF-IDF). For topic modelling, Latent Dirichlet Allocation (LDA) was used, and sentiment distributions were tracked over time through temporal aggregation. To obtain more specific public opinion sentiment analysis, a neutral classification was added to differentiate from the previous studies that used only positive and negative classifications. To support this approach, a pre-trained model with three sentiment classifications was used. The results show that neutral sentiment dominated the tweets followed by negative sentiment then positive sentiment, especially during the transition to the new Ministry of Finance, showing the relevance of real-world events to online public opinion on X. Based on topic trends, public opinion shows the trend change from fiscal policy and leadership to criticism and leadership change.

**Keywords:** X, Sentiment Analysis; IndoBERTweet; Topic Modelling; Temporal Trend Analysis; Ministry of Finance

## I. INTRODUCTION

The important factors affecting national economic confidence and indication of public trusts in fiscal policy is public opinion of government representatives, especially the finance minister. The most important duty of the finance minister is to oversee the nation's budget, taxes, government expenditure, and economic swings. Thus, it is important for analysts and policymakers to know how public opinion of the Minister's choices and actions.

The significance of social media platforms X, formerly Twitter (Zhang, Harris, & Zheng, 2025), as platforms for political communication and public debate has increased. People express their opinions, sentiments, and evaluations of political figures and policies on these platforms, which can be used as an extensive, real-time data. X's significant impact in Indonesia makes it a valuable tool for knowing popular opinion of political figures. (Mailo & Lazuardi, 2019).

This study investigates public opinion of Indonesia's finance minister from tweets on X between February 19 and September 21, 2025, when there was a intense discussion about the minister by using transformer-based architectures in Natural Language Processing (NLP) to analyze sentiment distribution and semantic themes from the tweets.

This study focuses to: (1) find a sentiment distribution of public opinion toward the Finance Minister; (2) find the main topics of the discussions; and (3) assess the performance of natural language processing techniques in examining social media data. The results show insights about the relationship between digital communication, political accountability, and public opinion in Indonesia.

IndoBERTweet has shown strong success in a variety of industries. Damayanti et al. (Damayanti, Ariningtyas, Icham, & Sari, 2025) correctly

identified positive, negative, and neutral emotions with 95% accuracy in their examination of #BTSComeback tweets. Similarly, IndoBERT achieved 95% accuracy, outperforming BERT, SVM, and Random Forest when Setiawan et al. (Setiawan, Iswavigra, & Anggiratih, 2025) analyzed 23,796 YouTube comments about rulings by constitutional courts. Sayarizki et al. (Sayarizki, Hasmawati, & Nurrahmi, 2024) obtained 80% accuracy in political sentiment analysis on tweets related to the 2024 presidential election; nevertheless, the model had trouble handling complicated tweets that had contradictory opinions on several candidates.

Rizkia et al. (Rizkia, Wufron, & Roji, 2025) evaluated manual, lexicon-based (InSet), and transformer-based labeling methods using 8,035 tweets regarding Coretax. While IndoBERTweet labeling yielded the highest F1-score (0.9802) but ran the risk of neutral class dominance, lexicon-based techniques achieved balanced representation, highlighting complementary features.

Several studies have explored efficient alternatives to computationally intensive transformers. Khairul et al. (Khairul, Mutawalli, Bagye, & Tanton, 2025) integrated Chunking and Rule-Based Machine Translation for automated labeling of 225,000 Gojek reviews, achieving 89.9% accuracy with Multilayer Perceptron and Random Forest. Ahmadian et al. (Ahmadian, Abidin, Riza, & Muchtar, 2023) combined IndoBERT with BiLSTM, achieving 0.92 accuracy for sentiment analysis and 0.76 for emotion classification on IndoNLU benchmarks. Lin and Nuha (Lin & Nuha, 2023) proposed a weighted ensemble of BERT and DistilBERT with TCN and Bi-LSTM classifiers, achieving 85.13% accuracy on multi-topic datasets.

Large-Scale Studies and Reviews. Romadhony et al. (Romadhony, Faraby, Rismala, Wisesti, & Arifianto, 2024) demonstrated that BiLSTM provided the highest accuracy on the 700,000-entry FDReview dataset, while SVM and MNB remained competitive on smaller datasets. Setiawan's (Setiawan B., 2024) review of 75 studies from 2023–2024 found that shallow learning models (Naïve Bayes, SVM, KNN) remain prevalent, though interest in deep learning architectures like IndoBERT is increasing, with language-specific tools like Sastrawi stemmer and InSet lexicon being critical for handling Indonesian linguistic complexities.

Domain-Specific Applications. Sejati et al. (Sejati, Alzami, Marjuni, Indrayani, & Puspitarini, 2024) integrated LDA topic modeling with IndoBERT sentiment classification on 10,000+ Ministry of Finance tweets, identifying four main topics (economy, budget, employees, taxes) with predominantly negative sentiment. Firdaus et al.

(Firdaus, Asror, & Herdiani, 2021) applied lexicon-based sentiment analysis with InSet to student feedback, achieving 90.9% document-level accuracy. Qi and Shabrina (Qi & Shabrina, 2023) compared lexical and ML-based approaches on 77,000 UK COVID-19 tweets, finding SVC with Bag of Words or TF-IDF achieved 71% accuracy, demonstrating social media's utility for policy analysis.

These prior studies provide a foundation for the present research on public sentiment toward Indonesia's Minister of Finance. By using pretrained indoBERTweet tailored for Indonesian language tweet, this work aims to provide a temporal sentiment and topic trend for ministry of finance in social media discourse.

## II. METHODS

This study employs a systematic approach to analyze the semantic sentiment of public opinion regarding Indonesia's Minister of Finance on X. The methodology consists of several key stages: data collection, preprocessing, sentiment analysis including sentiment classification, keyword extraction, topic modelling, topic trend analysis, and temporal trend analysis, as can be seen in Figure. 1.

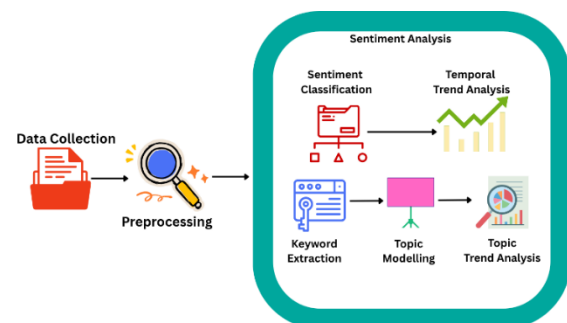


Figure. 1. The Research Flow

### A. Data Collection

Tweets containing keywords relevant to Indonesia's Minister of Finance using Indonesian language keyword “menteri keuangan” were collected over a focused period from February 19 to September 21, 2025. The data was obtained using Twitter scraping tools (Tweet harvest (Satria, 2023)) to capture real-time public discourse during a high-activity period for fiscal policy discussions. After initial crawling, irrelevant tweets such as spam, duplicates, retweets, and non-Indonesian language posts were removed, resulting in a cleaned dataset of 9,543 tweets.

### B. Preprocessing

The raw tweets went through a lot of preparation to get them ready for analysis. Several steps included removal of URLs, mentions (@usernames), hashtags, emoji, retweets, and special characters.

Tokenization, and lowercasing were applied for model readiness.

### C. Sentiment Analysis

- 1) Sentiment classification using pre-trained IndoBERTweet model (Ardiyanto, 2024), the tweets were classified into positive, negative, or neutral categories.
- 2) Keyword extraction was conducted using Term Frequency-Inverse Document Frequency (TF-IDF) by removing stop words, including common conjunctions in the Indonesian language, and applying a list of named entities to group names, in order to identify key terms related to Ministry of Finance topics.
- 3) Topic modelling and trend analysis over time using Latent Dirichlet Allocation (LDA) by using the result from TF-IDF to find main topics people talked about and see how these topics changed over time.
- 4) Temporal trend analysis by aggregating monthly sentiment results to observe how public opinion changes in public opinion over time.

## III. RESULTS AND DISCUSSION

For sentiment distribution of public opinion regarding Indonesia's Minister of Finance reveals diverse range of opinions regarding Ministry of Finance in social media: Out of 9,543 tweets, the sentiment classification show that the majority of tweets express a neutral sentiment, constituting approximately 64.0% (6111 tweets) of the total dataset. This reflects that many users primarily shared information or commented without strong emotional polarity. negative sentiment made up 28.1% (2682 tweets) while positive sentiment accounted for 7.9% (750 tweets). According to this distribution, the majority of the public discussion about the Ministry of Finance is factual or instructive, with a minority expresses dissatisfaction follow by few support. It provides useful insights into how public opinion works and identifies areas where policy communication needs to be improved.

The sentiment classification was performed using IndoBERTweet, a state-of-the-art transformer model for Indonesian social media text, which ensures a high level of reliability in the sentiment categorization. The model achieved strong performance during evaluation: accuracy = 0.9068, F1-score = 0.9071, precision = 0.9078, and recall = 0.9068 (Ardiyanto, 2024). Figure 2 shows the sentiment distribution of tweets.

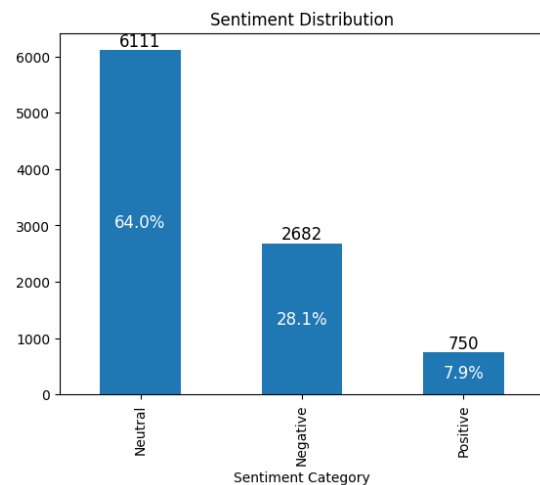


Figure 2. Sentiment Distribution

The TF-IDF keyword extraction produced a weighted ranking of terms that frequently appeared in the dataset while accounting for their relative distinctiveness, the top 20 keywords can be seen on Figure 3. The results indicate that discourse about the Ministry of Finance is strongly centered around the figure of the finance minister (Sri Mulyani Indrawati), institutional roles (menteri, menkeu), and national issues such as “negara”, “pajak”, “ekonomi”, “rakyat”. The recurrent mention of “2025”, “baru”, “Purbaya Yudhi Sadewa” suggests that discussions were also linked to fiscal year agendas and new policy directions for new ministry of finance. The words cloud for the text used for the keyword extraction can be seen on Figure. 4.

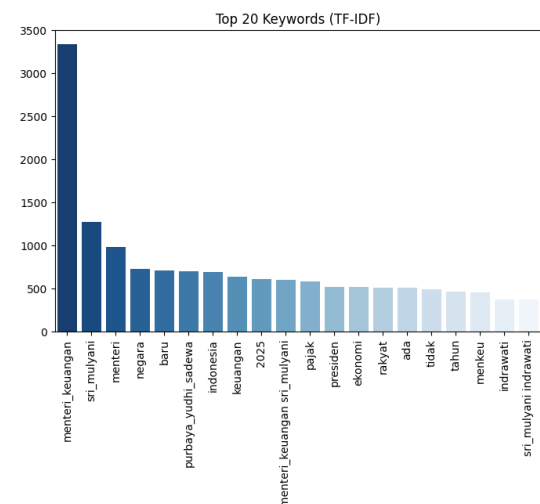


Figure 3. Top 20 Keywords for TF-IDF

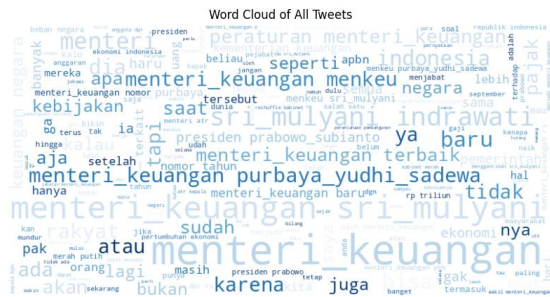


Figure 4. Word Cloud of text for keyword extension

The LDA model with four topics revealed meaningful clusters of discourse surrounding the Ministry of Finance:

- Topic 1: Economic and Fiscal Policy

The top keywords for this topic are displayed in Figure 5, which highlights structural and macroeconomic topics. The inclusion of pajak (taxes) and ekonomi (economy) indicates that talks on national economic performance and fiscal planning are the main topics covered by this cluster, about 1,854 tweets (19.4%).

- Topic 2: Leadership and Ministerial Role

This topic focuses on Sri Mulyani's function as the finance minister and her place in the presidential and governmental structure, as the list of keywords displayed in Figure 6. This cluster had 2,295 tweets (24.0%).

- Topic 3: Criticism and Accountability

This cluster covers key discussions, including skepticism and discontent, which are significant assertions in Figure 7. Lexical terms like tidak (not), ada (there is/are), and rakyat (the people) suggest that this topic represents views on public welfare, tax duties, and governmental responsibility. 18.0% of the 1,722 tweets were composed of this.

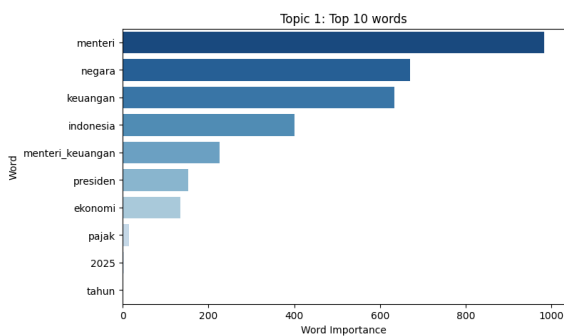


Figure 5. Top 10 words for topic 1

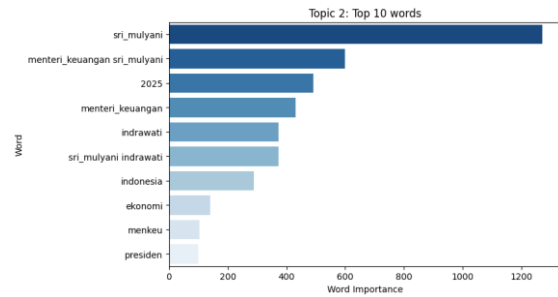


Figure 6. Top 10 words for topic 2

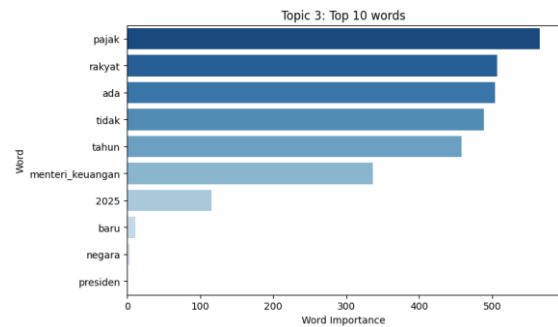


Figure 7. Top 10 words for topic 3

- Topic 4: New Ministry Leadership – Purbaya Yudhi Sadewa

The sentences in Figure 8 demonstrate how this problem emphasizes the shift in ministerial leadership. This cluster shows talks about the appointment of Purbaya Yudhi Sadewa as the new Minister of Finance and the implications of the leadership transition for the direction of economic policy in 2025, given the significant association between baru (new) and purbaya\_yudhi\_sadewa. With 3,672 tweets (38.5%), this was the most popular topic in the collection.

Together, these four topics show the range of topics discussed on X regarding the Ministry of Finance, including technical matters pertaining to tax and fiscal policy, ministerial performance and leadership, public accountability and criticism, and the importance of leadership changes. Figure 9 displays the topics' distribution over the tweets.

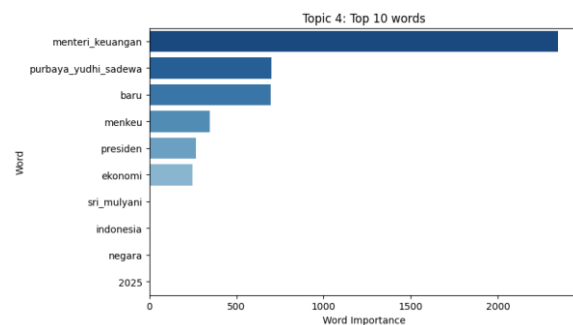


Figure 8. Top 10 words for topic 4

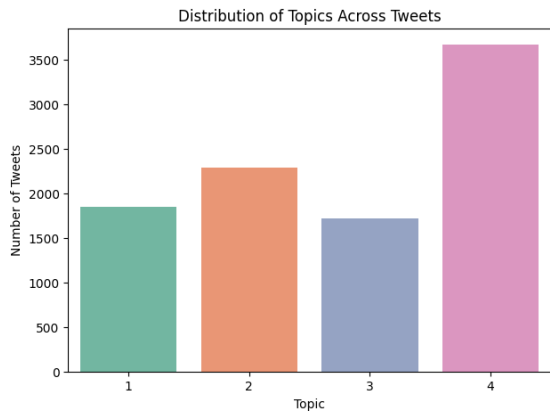


Figure 9. Distribution of Topic Across Tweets

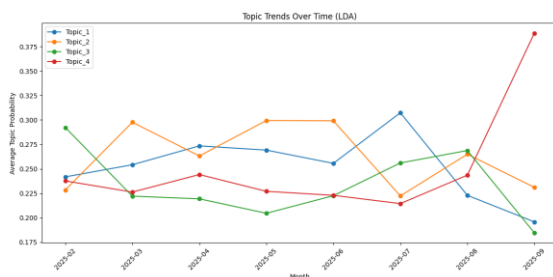


Figure 10. Topic Trends Over Time (LDA)

The monthly subject trend in Figure 10 shows how debate has changed over time. Topic 1 (economic policy) peaked in February as a consequence of early discussions and then gradually declined until August, when the ministerial change took place. With the exception of February and July, Topic 2 (leadership) was discussed often during most of the months. After first emerging in February, Topic 3 (criticism) made a significant resurgence in July and August, perhaps in response to the Ministry's reorganization. The nomination of the new Minister of Finance caused Topic 4 (leadership change) to dominate, peaking in August and September.

Figure 11 illustrates how the sentiments count was aggregated each month to represent the changes of popular opinion. The result shows three separate stages of sentiment growth from February to September 2025 are revealed by the results:

- **Stable Early Period (Feb–Jun 2025):**  
Neutral tweets continuously predominated over this time, and mood remained mostly unchanged. Positive sentiment stayed low (less than 30 tweets per month), whereas negative sentiment was moderate (28 to 142 tweets per month). This suggests that early conversations were more descriptive or instructive and less divisive.
- **Rising Polarization (Jul–Aug 2025):**  
In comparison to previous months, there were more negative (236) and good (61) tweets in July. By August, the number of unfavorable tweets increased

dramatically (923) while the number of positive tweets increased less (97). The increase points to increased public expectations for political reforms and examination of ministerial performance.

- **Leadership Transition (Sep 2025):**  
The biggest change was in September, when the number of tweets expressing negative emotion peaked at 1037, while the number of tweets expressing good sentiment peaked at 482. Even while neutral tweets were still prevalent at 2888, they demonstrated that the level of debate intensity had peaked. Strong polarization brought on by the nomination of a new Finance Minister is indicated by the concurrent increase in both positive and negative mood.

Overall, the temporal sentiment analysis shows that although neutral tweets predominated throughout the year, political developments and changes in leadership had a major impact on the rise of divisive viewpoints.

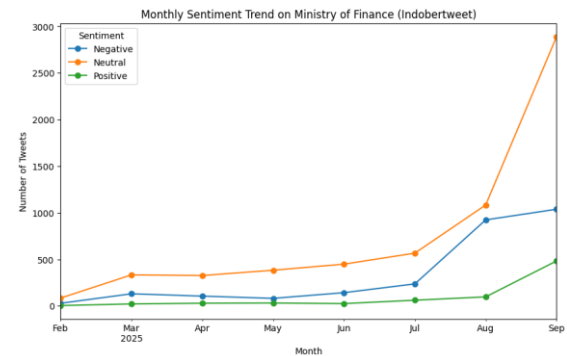


Figure 11. Monthly Sentiment Trend

## IV. CONCLUSION

This study used sentiment analysis, topic modeling, and temporal trend analysis to examine 9,543 tweets about the Ministry of Finance from February to September 2025. Most tweets were neutral, according to the results, but negative emotion predominated, particularly in the second half of the year. Three phases were identified by temporal trends: early stability, polarization in the middle of the year, and increased reactivity in September associated with the formation of a new Ministry of Finance. Topic trends followed a similar pattern, moving from debates about leadership and economic policies to criticism and, ultimately, leadership turnover, which emerged as the main issue. The result shows that the pre-training IndoBERTweet can be used for the sentiment analysis of Ministry of Finance topic. Future research could expand by applying semantic embeddings, event alignment, cross-platform comparisons, and multilingual analysis like translating to English.

## AUTHOR'S CONTRIBUTION

S.S: writing, datasets curation, modeling, experimenting, analysis, validation, and editing; J.H: writing, modeling, experimenting, analysis; A.A: writing, modeling, validation, visualization.

## AVAILABILITY DATA AND MATERIALS

The dataset used can be accessed via the link <https://github.com/suryawang/indonesia-sentiment-analysis-dataset>.

## REFERENCES

- Ahmadian, H., Abidin, T. F., Riza, H., & Muchtar, K. (2023). Transformer-Based Indonesian Language Model for Emotion Classification and Sentiment Analysis. *2023 International Conference on Information Technology and Computing (ICITCOM)*, (pp. 209-214). Yogyakarta, Indonesia.
- Ardiyanto, M. (2024). *Aardiiiiy/indobertweet-base-Indonesian-sentiment-analysis*. (Hugging Face) Retrieved September 2025, from <https://huggingface.co/Aardiiiiy/indobertweet-base-Indonesian-sentiment-analysis>
- Damayanti, N. M., Ariningtyas, I. D., Icham, M. I., & Sari, A. P. (2025). ANALISIS SENTIMEN PUBLIK PADA TAGAR #BTSCOMEBACK DI PLATFORM X MENGGUNAKAN INDOBERTWEET . *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, 13(3), 1142-1153.
- Firdaus, R., Asror, I., & Herdiani, A. (2021). Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation. *Ind. Journal on Computing*, 6(1), 1-12.
- Khairul, I., Mutawalli, L., Bagye, W., & Tantoni, A. (2025). Automated Label Extraction for Sentiment Analysis in Indonesian Text. *International Journal on Advanced Science, Engineering and Information Technology*, 718-728.
- Lin, C.-H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*.
- Mailo, F. F., & Lazuardi, L. (2019). Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia. *Journal of Information Systems for Public Health*, 4(1), 28-36.
- Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*.
- Rizkia, A. S., Wufron, W., & Roji, F. F. (2025). Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, Transformers-Based, dan Lexicon-Based pada Performa IndoBERT . *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(3), 1037-1048.
- Romadhony, A., Faraby, S. A., Rismala, R., Wisesti, U. N., & Arifianto, A. (2024). Sentiment Analysis on a Large Indonesian Product Review Dataset. *Journal of Information Systems Engineering and Business Intelligence*.
- Satria, H. (2023, July). *Crawl Data Twitter Menggunakan Tweet Harvest - Juli 2023*. Retrieved September 2025, from <https://helimisatria.com/blog/crawl-data-twitter-menggunakan-tweet-harvest/>
- Sayarizki, P., Hasmawati, & Nurrahmi, H. (2024). Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates. *Indonesian Journal on Computing (Indo-JC)*, 61-72.
- Sejati, P. T., Alzami, F., Marjuni, A., Indrayani, H., & Puspitarini, I. D. (2024). Aspect-Based Sentiment Analysis for Enhanced Understanding of 'Kemenkeu' Tweets. *Journal of Applied Informatics and Computing*, 487-498.
- Setiawan, B. (2024). A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024. *Journal Information Engineering and Educational Technology*.
- Setiawan, V. D., Iswavigra, D. U., & Anggiratih, E. (2025). Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates. *Scientific Journal of Informatics*, 397-406.
- Zhang, P., Harris, R. D., & Zheng, J. (2025). GNN-based social media sentiment analysis for stock market forecasting. *Expert Systems With Applications*, 291.