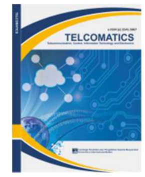
Available online at journal.uib.ac.id/index.php/telcomatics

TELCOMATICS

ISSN (online) 2541-5867



Public Sentiment Analysis Toward the Ministry of Finance 2025 Using Recurrent Neural Network Methods Based on Data from Sosial Media X

Aditya Rahman, Sofyan Fauzi Dzaki Arif, Muhammad Regi Abdi Putra Amanta, M. Syamsuddin Wisnubroto*, Fajri Farid

Department of Data Science, Faculty of Science, Institut Teknologi Sumatera, Jalan Terusan Ryacudu, Lampung Selatan, Indonesia

ARTICLE INFO

Keywords:
Ministry of Finance, X, Sentiment Analysis, Long Short-Term Memory, Gated Recurrent Unit

Received: December 12, 2025
Revised: March 6, 2026
Accepted: April 26, 2026

***Corresponding author:**
E-mail: syamsuddin.wisnubroto@sd.itera.ac.id

DOI: 10.37253/telcomatics.v11i1.11645

ABSTRACT

The Ministry of Finance plays a strategic role in maintaining national economic stability through fiscal policy management, taxation, public debt administration, and state budget control. In today's digital era, social media platforms such as X have become important channels for the public to express opinions about government policies. This study analyzes public perceptions of the Ministry of Finance's performance using machine-learning-based sentiment analysis and identifies the most effective classification model. Data were collected from public posts on X and processed using text mining and Natural Language Processing (NLP). Three Recurrent Neural Network (RNN) models were tested: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and an improved variant, LSTM_G. The findings show that negative sentiment dominates at 43.0%, followed by neutral at 33.9% and positive at 23.1%. Among the models, LSTM_G achieved the highest accuracy of 78.98%, indicating strong capability in capturing sequential patterns in dynamic, unstructured social media text. These results reflect substantial public concerns regarding fiscal policies and demonstrate the usefulness of sentiment analysis as a data-driven tool for decision-making and for strengthening public communication strategies to enhance the Ministry's digital reputation.

I. INTRODUCTION

The Ministry of Finance holds a vital role in maintaining national economic stability through fiscal management, taxation, debt financing, and control of the State Budget (APBN) [1]. Each policy issued by this ministry has broad implications for the public economy, ranging from changes in the prices of basic necessities to subsidy levels and purchasing power [2]. Therefore, understanding public perceptions of the Ministry's performance and policies is a strategic aspect for maintaining public trust and supporting the effectiveness of government actions.

In recent years, social media platforms such as X have become major spaces for people to express opinions, criticism, and support regarding economic policies [3]. Their fast, interactive, and massive nature makes them relevant for studying how public opinion forms and spreads. Every policy issued by the Ministry of Finance tends to trigger new discussions on X, making the platform an important source of data for understanding public responses to fiscal issues. The use of digital conversation analysis has even become a supporting approach in evaluating fiscal policy, especially when real-time public sentiment is needed for more responsive, data-driven decision-making.

In this context, sentiment analysis emerges as a highly relevant method. Through text mining and Natural Language Processing (NLP), it identifies whether opinions expressed in text are positive, negative, or neutral[4]. One commonly used method is Long Short-Term Memory (LSTM), which is designed to retain important information across word sequences while ignoring irrelevant parts [5]. This ability to preserve contextual meaning makes LSTM effective for processing unstructured and emotionally dynamic social media text.

Recurrent Neural Networks (RNN), specifically LSTM and GRU were selected because sentiment analysis on social media text inherently involves sequential and contextual information, where the meaning of a word depends on preceding and following words. Unlike traditional machine learning models that treat text as independent features [6], RNNs process input as ordered sequences and maintain a hidden state that allows the model to capture contextual dependencies across words, which is essential for interpreting nuanced opinions in short and informal posts on X. Furthermore, standard RNN limitations in learning long-term dependencies are effectively addressed by LSTM through memory cells and gating mechanisms, while GRU offers a computationally efficient alternative with fewer parameters and faster convergence, making it well suited for large-scale social media data. Compared to transformer-based models, RNN architecture provides a balanced trade-off between contextual modeling capability, data requirements, and computational efficiency, which aligns with the dataset characteristics and research objectives of this study.

The Gated Recurrent Unit (GRU) operates similarly to LSTM but with a simpler and more efficient structure. GRU uses "gating" mechanisms to determine which information to keep or discard, allowing faster training without significant accuracy reduction [5]. This efficiency makes GRU suitable for analyzing short text such as posts on X, particularly when processing speed is essential.

This study examines public perceptions of the Ministry of Finance's policies and performance using conversational data from X. The approach employs deep learning based on Recurrent Neural Networks (RNN), focusing on two main architectures: LSTM and GRU. Through these models, the study aims to portray public opinion on Indonesia's fiscal policies and assess whether discussions regarding a potential Ministry of Finance reshuffle align with sentiments expressed in public discourse. Additionally, the findings are expected to provide valuable input for the government to improve public communication strategies and enhance the effectiveness of national financial management.

II. LITERATURE REVIEW

A. Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is a fundamental architecture designed to process sequential data through what is known as a recurrent connection [8]. However, its main weakness lies in its difficulty in retaining long-term information, which leads to vanishing and exploding gradient problems. To address this issue, Long Short-Term Memory (LSTM) was developed, featuring three main gates—forget gate, input gate, and output gate—to regulate the flow of information. The core computations of LSTM are mathematically defined as in (1) – (6):

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{3}$$

$$\widetilde{C_t} = tanh\ \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{5}$$

$$h_t = o_t \cdot tanh\ \tanh(C_t) \tag{6}$$

Through this mechanism, LSTM can selectively store or forget information as needed, making it effective for data with long-term dependencies [9].

B. Gated Recurrent Unit (GRU)

Another, more simplified model is the Gated Recurrent Unit (GRU), which uses only two gates update gate and reset gate. Its simpler architecture allows GRU to be trained more quickly while producing competitive results compared to LSTM. The formulas are written as in (7) – (11):

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z) \tag{7}$$

$$r_t = \sigma(W_i[h_{t-1}, x_t] + b_r) \tag{8}$$

$$\widetilde{h_t} = tanh\ \tanh(W_h[h_{t-1}, x_t] + b_h) \tag{9}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t} \tag{10}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \widetilde{h}_t \tag{11}$$

With this architecture, GRU can still capture sequential context without the complexity of LSTM [10].

C. Bidirectional LSTM (biLSTM)

A further development is the Bidirectional LSTM (biLSTM), which processes sequential data in two directions forward and backward allowing for a more complete understanding of word context. Its computation is expressed as in (12) – (14):

$$\vec{h}_t = LSTM\left(x_t, , \vec{h}_{t-1}\right) \tag{12}$$

$$\overleftarrow{h}_t = LSTM\left(x_t, \overleftarrow{h}_{t-1}\right) \tag{13}$$

$$h_t = \left[\vec{h}_t; \overleftarrow{h}_t\right] \tag{14}$$

This approach produces word representations enriched by both preceding and following context in a sentence [11].

D. LSTM GloVe (LSTM_G)

Beyond the basic architecture, models may be improved using external embeddings. LSTM_G is an LSTM model that uses GloVe (Global Vectors for Word Representation) embeddings as input. Here, xt is not a one-hot vector but a semantic vector generated by GloVe as in (15):

$$x_t = GloVe(w_t) \tag{15}$$

The remaining computations follow standard LSTM formulas but incorporate global semantic representations [12].

E. Bidirectional LSTM GloVe (biLSTM_G)

Similarly, biLSTM_G combines bidirectional mechanisms with GloVe word representations. The formulas remain identical to biLSTM, but the input xt is a GloVe vector as in (16) – (18):

$$\vec{h}_t = LSTM\left(GloVe(w_{t)}, \vec{h}_{t-1}\right) \tag{16}$$

$$\overleftarrow{h}_t = LSTM\left(GloVe(w_t), \overleftarrow{h}_{t-1}\right) \tag{17}$$

$$h_t = \left[; \overleftarrow{h}_t\right] \tag{18}$$

This enhances the model's ability to understand context with richer semantic meaning. GRU_G applies the same concept to GRU, using GloVe embeddings as input while retaining the efficiency of GRU.

F. LSTM_FASTTEX

Another widely used embedding is FastText, which is based on subword information and can therefore recognize unseen words or words with spelling variations. LSTM_FASTTEXT uses this embedding as input so that standard LSTM operations are performed on (19):

$$x_t = FastText(w_t) \quad (19)$$

Thus, the model captures not only long-range dependencies but also morphological variations [13].

G. Bidirectional LSTM FASTTEXT (biLSTM_FASTTEXT)

For bidirectional processing, biLSTM_FASTTEXT integrates BiLSTM with FastText embeddings, computed as in (20) – (22):

$$,\vec{h}_t = LSTM\left(FastText(w_{t)}, \vec{h}_{t-1}\right) \quad (20)$$

$$\overleftarrow{h}_t = LSTM\left(FastText(w_t), \overleftarrow{h}_{t-1}\right) \quad (21)$$

$$h_t = \left[;\overleftarrow{h}_t\right] \quad (22)$$

This model excels in handling complex and highly varied text.

2.1 GRU_FASTTEXT

Finally, GRU_FASTTEXT is a GRU model with FastText embeddings as input. Its formulas follow the standard GRU framework but use subword-based representations as in (23) – (26):

$$z_t = \sigma(W_z[h_{t-1}, FastText(w_t)] + b_z) \quad (23)$$

$$r_t = \sigma(W_r[h_{t-1}, FastText(w_t)] + b_r) \quad (24)$$

$$\widetilde{h_t} = tanh \tanh(W_h[r_t \cdot h_{t-1}, FastText(w_t)] + b_h) \quad (25)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \widetilde{h_t} \quad (26)$$

With its simple architecture and flexible embeddings, GRU_FASTTEXT serves as a fast yet effective alternative.

III. Research Methods

A. Type of Research

This study employs an applied quantitative research design with a computational approach, aiming to classify public sentiment on social media regarding the Ministry of Finance of Indonesia. The research relies on natural language processing (NLP) and deep learning methods to analyze textual data originating from public conversations on the social media platform X (formerly Twitter). The primary goal of this study is to measure and categorize sentiment polarity based on public expressions about government activities and financial policies. The use of quantitative methods enables a systematic and measurable evaluation of sentiment patterns, while deep learning provides a high level of adaptability in processing informal language typically found in social media content. The study also integrates supervised learning techniques, in which labeled text data is required to train various sentiment classification models.

B. Data and Data Sources

The dataset used in this study consists of public posts (tweets) collected from X (Twitter) during the observation period of January 1, 2025, to September 27, 2025, using the keyword “Kemenkeu”, representing discussions related to the Ministry of Finance. Data scraping was conducted using authenticated access through browser cookies to comply with platform policies and obtain public tweets relevant to the topic. Only Indonesian-language tweets were retained by applying language filtering, and the data was saved in CSV format for further processing. Following collection, an initial cleaning process was performed to remove advertisements, spam, duplicate content, and irrelevant tweets, ensuring that the dataset contains only valid and contextually relevant discussions. The processed text was then normalized through several preprocessing steps, including tokenization, slang normalization, abbreviation expansion, stopword removal, and stemming using an Indonesian stemming algorithm. After normalization, a manual labeling process was performed to categorize each tweet into three sentiment classes (positive, neutral, and negative), supervised by multiple annotators to maintain consistency and reliability. The labeled dataset was then utilized for model training and evaluation.

C. Research Flowchart

Based on **Figure 1** illustrates the systematic workflow proposed in this study, beginning with the data collection phase followed by data preprocessing and tokenization to clean and structure the text. Once the class attributes are defined, the process moves to the classification stage, which conducts a comparative analysis of multiple deep learning architectures. This includes testing LSTM, GRU, and BiLSTM models using different embedding techniques such as standard, GloVe (denoted as _G), and FastText variants. To assess performance, each model undergoes a rigorous evaluation using standard metrics including accuracy, precision, recall, and F1-score. The pipeline concludes by identifying and selecting the model that demonstrates the best overall accuracy.

D. Model Evaluation

To determine the most effective sentiment classification model, each deep learning architecture tested in this research was evaluated using the same labeled dataset. The evaluation focused on four commonly used performance metrics in supervised text classification: **accuracy, precision, recall, and F1-score** as in (27) – (30):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (27)$$

$$Precision = \frac{TP}{TP + FP} \quad (28)$$

$$Recall = \frac{TP}{TP + FN} \quad (29)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (30)$$

Accuracy measures the overall correctness of model predictions, while precision reflects the model's ability to identify sentiments accurately without misclassification. Recall evaluates the completeness of predictions by measuring the proportion of correctly recognized sentiment classes, and the F1-score provides a balanced assessment between precision and recall, particularly useful when class distributions are imbalanced. By comparing these metrics across all models—including the baseline RNN and its variants such as LSTM, GRU, LSTM_G, biLSTM_G, GRU_G, LSTM_FASTTEXT, biLSTM_FASTTEXT, and GRU_FASTTEXT—this study identifies the optimal architecture for sentiment classification on Indonesian social media text related to the Ministry of Finance. The evaluation approach ensures objective assessment and facilitates model selection based on measurable predictive performance.

IV. Results and Discussion

The data collected from social media platform X using keywords related to the Ministry of Finance was processed through a series of text preprocessing stages to ensure its suitability for sentiment modeling. Three primary techniques were applied: stopword removal, slang normalization, and abbreviation normalization. Stopword removal was conducted to eliminate high-frequency words such as "yang," "di," and "dan" that do not convey meaningful information for classification. Slang normalization was performed using a predefined dictionary to convert informal or non-standard words commonly found on social media into their formal equivalents, for example, "gmn" to "bagaimana" and "yg" to "yang." The final step utilized an abbreviation dictionary to expand commonly used shortened terms, such as changing "pajak" to "pajak negara" and "apbn" to "anggaran pendapatan dan belanja negara." These preprocessing procedures resulted in a cleaner and more standardized text corpus ready for further analysis.

Following preprocessing, tokenization was applied to break each sentence into smaller units (tokens) that can be processed computationally. The resulting tokens facilitated the conversion of text into numerical vector representations suitable for machine learning and deep learning models. Subsequently, sentiment categories were defined as the target class attributes consisting of three labels: positive, negative, and neutral. The labeling process was conducted manually and semi-automatically with the aid of a sentiment lexicon to maintain consistency across annotations. Finally, a sentiment distribution check was performed to observe the proportion of positive, negative, and neutral sentiments in the dataset, providing insights into the overall polarity of public opinion.

Based on the results of the attribute-level analysis at **Figure 2**, the sentiment classification shows that 23% of the data are categorized as positive, 43% as negative, and 33.9% as neutral. Subsequently, the dataset will be classified using eight RNN models, where the accuracy of each model will be evaluated and obtained as follows:

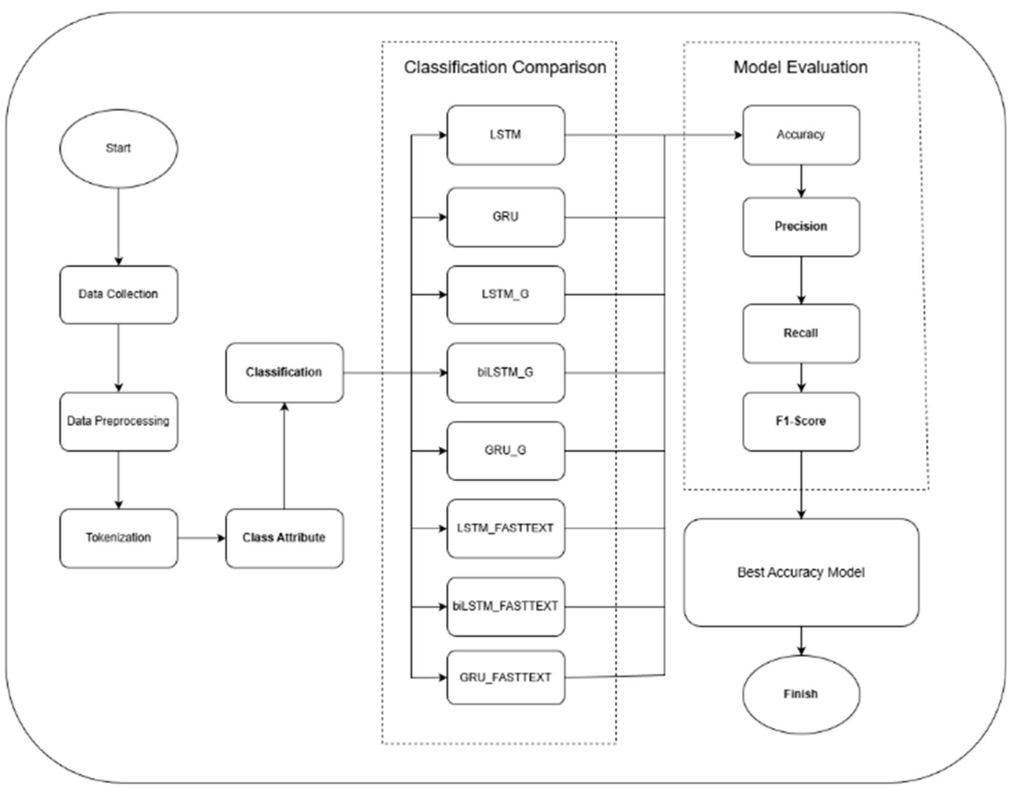


Figure 1. Research Flowchart

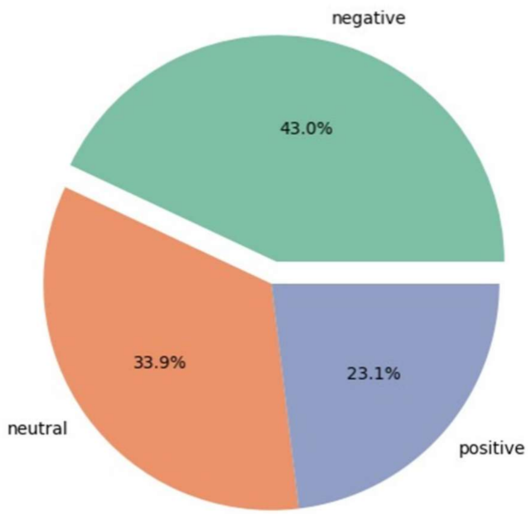


Figure 2. Sentiment Distribution

Table 1. LSTM Result.

LSTM	Precision	Recall	F1-Score
Negative	0.76	0.83	0.79
Neutral	0.79	0.82	0.81
Positive	0.72	0.55	0.63
Accuracy			0.76
Macro avg	0.76	0.74	0.74
Weighted avg	0.76	0.76	0.76

Table 2. GRU Result.

GRU	Precision	Recall	F1-Score
Negative	0.78	0.83	0.81
Neutral	0.84	0.81	0.83
Positive	0.72	0.65	0.68
Accuracy			0.78
Macro avg	0.78	0.77	0.77
Weighted avg	0.78	0.78	0.78

Table 3. LSTM_G Result.

LSTM_G	Precision	Recall	F1-Score
Negative	0.78	0.83	0.81
Neutral	0.84	0.81	0.83
Positive	0.72	0.65	0.68
Accuracy			0.78
Macro avg	0.78	0.77	0.77
Weighted avg	0.78	0.78	0.78

Table 4. BiLSTM_G Result.

bLSTM_G	Precision	Recall	F1-Score
Negative	0.82	0.76	0.79
Neutral	0.77	0.90	0.83
Positive	0.69	0.61	0.65
Accuracy			0.77
Macro avg	0.7699	0.76	0.75
Weighted avg	0.77	0.77	0.77

Table 5. GRU_FastText Result.

GRU_FastText	Precision	Recall	F1-Score
Negative	0.70	0.86	0.77
Neutral	0.82	0.82	0.82
Positive	0.75	0.43	0.55
Accuracy			0.75
Macro Avg	0.76	0.70	0.71
Weighted Avg	0.75	0.75	0.73

Table 6. LSTM_FastText Result.

LSTM_FastText	Precision	Recall	F1-Score
Negative	0.67	0.83	0.74
Neutral	0.87	0.68	0.76
Positive	0.64	0.57	0.61
Accuracy			0.72
Macro Avg	0.73	0.69	0.70
Weighted Avg	0.73	0.72	0.72

Based on **Table 1**, the LSTM model achieves a moderate overall performance with an accuracy of 0.76, showing relatively strong results for the negative and neutral classes, while performance on the positive class remains weaker, as indicated by a lower F1-score. **Table 2** demonstrates that the GRU model outperforms the standard LSTM, achieving a higher accuracy of 0.78 along with improved macro and weighted average F1-scores, suggesting better generalization across sentiment classes. As shown in **Table 3**, the LSTM_G model produces identical results to GRU, indicating that the inclusion of gating mechanisms enhances model stability and classification consistency. **Table 4** reveals that the BiLSTM_G model excels in capturing neutral sentiment with a high recall value, although this improvement is accompanied by reduced performance in the positive class, resulting in a slightly lower overall accuracy. For models incorporating FastText embeddings, **Table 5** shows that GRU_FastText attains an accuracy of 0.75 with strong performance on negative and neutral sentiments but suffers from a substantial decline in positive sentiment recall. Similarly, **Table 6** indicates that LSTM_FastText yields the lowest overall accuracy among all models, highlighting that the FastText–LSTM combination is less effective for identifying positive sentiment in this dataset.

Based on the evaluation results, the model with the highest performance is LSTM_G, achieving an accuracy of 78.97%. This result indicates that the application of GloVe embedding in the LSTM architecture significantly improves performance compared with the standard model and other variants. This improvement can be attributed to GloVe's ability to capture global semantic representations of words, enabling the model to better map the context of public opinions expressed on social media.

Overall, LSTM_G proves to be effective in analyzing negative and neutral sentiments, while still facing challenges in identifying positive sentiments. With an accuracy approaching 80%, this model is considered the most optimal for analyzing public sentiment toward the Ministry of Finance. Future research may address the limitations related to sarcasm and mixed opinions by integrating multi-modal approaches or more complex transformer-based architectures.

The confusion matrix analysis at **Figure 3** shows that the model performs best in predicting the negative class (562 correctly classified instances), followed by the neutral class (427 correctly classified instances). However, misclassifications are still present, particularly in the positive class, where 87 positive instances were incorrectly predicted as negative and 40 were misclassified as neutral. These errors are commonly influenced by the presence of sarcasm and ambiguous opinions, which remain challenging for text-based models to interpret accurately.

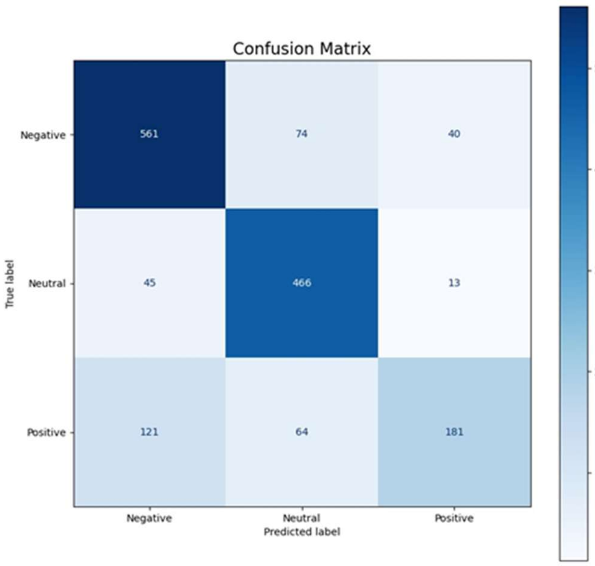


Figure 3. Confusion Matrix

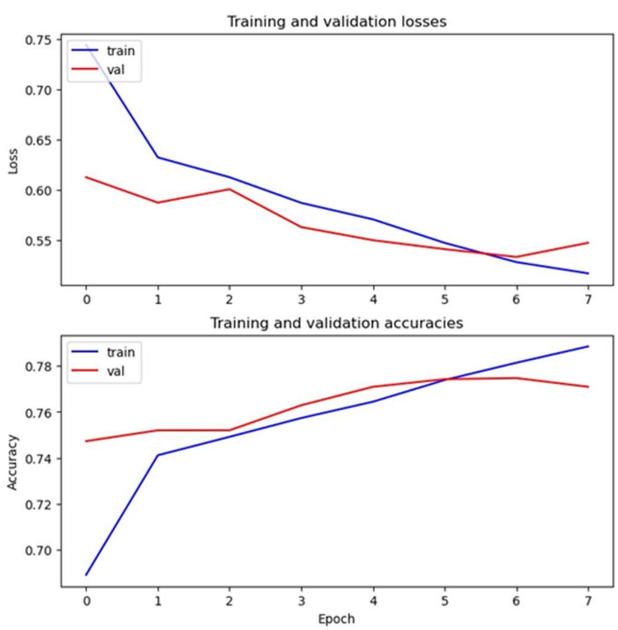


Figure 4. Training and validation losses

The training results of the LSTM_G model are presented in **Figure 4**. The loss curves show a clear downward trend for both the training and validation data, decreasing from 0.72 to approximately 0.52 by the seventh epoch. The validation loss remains stable with only a small gap from the training loss, indicating the absence of significant overfitting.

The accuracy plot also demonstrates consistent improvement: training accuracy increases from 0.69 to 0.79, while validation accuracy rises from 0.75 to 0.77 with minor fluctuations after the sixth epoch. This pattern indicates that the model learns progressively and maintains strong generalization performance.

With a final accuracy of 78.97 percent, the results confirm that the combination of LSTM and GloVe embeddings is an effective and optimal approach for analyzing public sentiment related to the Ministry of Finance.

V. CONCLUSION

The sentiment analysis of public conversations regarding the Ministry of Finance on platform X shows that the majority of responses lean negative at 43.0 percent, followed by neutral sentiment at 33.9 percent and positive sentiment at 23.1 percent. These proportions indicate public concern and dissatisfaction toward several ongoing fiscal policies, including issues related to economic stability, perceived uncertainty in policy changes, and public reactions to reshuffle discourse that is seen as potentially affecting trust and regulatory certainty.

From the modeling perspective, this study identifies the LSTM_G architecture as the best-performing model, achieving an accuracy of 0.7898. This performance suggests that an LSTM-based Recurrent Neural Network combined with GloVe word embeddings is effective in capturing temporal patterns and linguistic dynamics present in social media text. Beyond providing a measurable picture of public perceptions of fiscal policy, these findings also offer important insights for policymakers to rethink their public communication strategies to better address societal concerns.

Future research is encouraged to expand the dataset by incorporating additional social media platforms such as Instagram, Facebook, and YouTube to obtain a more comprehensive representation of public sentiment. Exploring more advanced deep learning models, such as BERT or RoBERTa, may also provide deeper semantic understanding. Furthermore, applying temporal analysis may yield additional insights into how public perceptions shift over time, allowing the results to be used strategically in government communication planning and policy evaluation.

VI. REFERENCES

[1] Kementerian Keuangan Republik Indonesia, “Kemenkeu RI,” 2024.

[2] K. K. R. I. Badan Kebijakan Fiskal, “Laporan Kebijakan Fiskal 2023: Menjaga Stabilitas dan Mendorong Pertumbuhan,” 2023.

[3] A. N. dan D. Rusdi, “Jurnal Komunikasi Indonesia,” vol. 10, n, pp. 101–104, 2022, [Online]. Available: https://books.google.co.id/books?hl=en&lr=&id=st52EQAAQBAJ&oi=fnd&pg=PA78&dq=Media+Sosial+dan+Transformasi+Digital+Komunikasi+Publik+di+Indonesia&ots=9BikB_BRA0&sig=rwHLQTWPKmB4S-5aT6DQG8-tjDM&redir_esc=y#v=onepage&q=Media Sosial dan Transformasi Digital

[4] A. M. Zuhdi, E. Utami, and S. Raharjo, “Analisis Sentiment Twitter terhadap Capres Indonesia 2019 dengan Metode K-NN,” *Jurnal Informa : Jurnal Penelitian dan Pengabdian Masyarakat*, vol. 5, no. 2, pp. 1–7, Aug. 2019, doi: 10.46808/INFORMA.V5I2.73.

[5] W. A. Wily, S. Anggai, and T. Tukiyat, “Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial X Di Play Store Menggunakan Algoritma Long Short-Term Memory (Lstm) Dan Gated Recurrent Unit

(Gru)," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 9, no. 1, pp. 63–72, 2025, doi: 10.47970/siskom-kb.v9i1.875.

[6] A. Damar, A. Zahrawati, M. A. Jibran, S. Nurulita, N. Cantika, and A. Dewana, "Analisis Sentimen Review Aplikasi Chat GPT dengan Memanfaatkan Algoritma Support Vector Machine", doi: 10.37253/telcomatics.v10i2.11759.

[7] W. Arisandi and S. Anggai, "Analisis Sentimen Ulasan Pengguna Aplikasi Media Sosial X Di Play Store Menggunakan Algoritma Long Short- Term Memory (LSTM) Dan Gated Recurrent Unit (GRU)," Jurnal Sistem Komputer & Kecerdasan Buatan, vol. IX, no1. September, pp. 63–72, 2025, doi: 10.47970/siskom-kb.v9i1.875.

[8] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1532–1543, 2014, doi: 10.3115/v1/d14-1162.

[9] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, 1994, doi: 10.1109/72.279181.

[10] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1724–1734, 2014, doi: 10.3115/v1/d14-1179.

[11] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[12] M. A. M. S. Saputri, R. Mahendra, "Indonesian-Twitter-Emotion-Dataset," GitHub Repository. [Online]. Available: https://github.com/meisaputri21/Indonesian-Twitter-Emotion-Dataset

[13] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, 2017, doi: 10.1162/tacl_a_00051.