

Enhancing Text Classification Performance: A Comparative Study of RNN and GRU Architectures with Attention Mechanisms

Yulita Ayu Wardani¹, Mery Oktaviyanti Puspitaningtyas², Happid Ridwan Ilmi³, Onesinus Saut Parulian⁴

^{1,2,3,4}Nusa Mandiri University, Jakarta, Indonesia
14230021@nusamandiri.ac.id

Article Info

Article history:

Received September 28, 2024

Revised December 23, 2024

Accepted December 23, 2024

Keywords:

Text Classification

RNN

GRU

Attention Mechanisms

Sentiment Analysis

ABSTRACT

Text classification plays a crucial role in natural language processing, and enhancing its performance is an ongoing area of research. This study investigates the impact of integrating attention mechanisms into a recurrent neural network (RNN) based architectures, including RNN, LSTM, GRU, and their bidirectional variants (BiLSTM and BiGRU), for text sentiment analysis. Three attention mechanisms Multihead Attention, Self Attention, and Adaptive Attention are applied to evaluate their effectiveness in improving model accuracy. The results reveal that attention mechanisms significantly enhance performance by enabling models to focus on the most relevant parts of the input text. Among the tested configurations, the LSTM model with Multihead Attention achieved the highest accuracy of 68.34%. The findings underscore the critical role of attention mechanisms in overcoming traditional RNN limitations, such as difficulty in capturing long-term dependencies, and highlight the potential for their application in broader text classification tasks

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Yulita Ayu Wardani

Nusa Mandiri University, Jakarta, Indonesia

Email: 14230021@nusamandiri.ac.id

1. INTRODUCTION

The development of information technology, especially in text mining and TF-IDF can improve efficiency and accuracy. One of the text mining techniques, TF-IDF, is used to assign weights to words in documents [1]. The sentiment here is expressed in posts on social media and is a public criticism of a particular situation. Tools will be used to convert these comments into information and generate text data mining. The purpose of text mining is to extract data and information from several documents. The data source in text mining comes from a sentence or a set of texts in an unstructured or semi-structured format [2]. Sentiment analysis faces two main problems. First, the sentiment generated by the prediction model is different from the actual sentiment. Second, the way a person conveys emotions and sentiments changes over time [3].

Using Recurrent Neural Networks (RNN) models, machines can understand various aspects of human language understanding, such as listening, communicating, recognizing conversations, and understanding accents and grammar. RNN models are very promising for solving various problems related to Natural Language Processing (NLP) [4].

Text mining can transform words and letters, generating numerical data during the creation or processing process. Sentiment analysis is the extraction of information, such as other people's perspectives on

an issue or event, used to express public opinion on an issue with textual data Sentiment analysis is a component of text mining that can categorize text by calculating the value of features contained in the data, text clustering.

Long Short-Term Memory (LSTM) is a component of recurrent neural network (RNN) in deep learning, which is part of the modernization of machine learning to handle big data. Deep Supervised Learning is one of the deep learning approaches that combines deep learning recurrent neural network (RNN) models. This deep supervised learning approach is used to handle big data using the RNN model [5].

Recurrent Neural Network is a type of neural network in which the connections between units form directed cycles, which results in the internal state of the network allowing it to exhibit dynamic behavior. On the other hand, Recurrent Neural Network is naturally suited for processing sequenced and time series data. Recurrent Neural Network-based models classify text as a sequence of words to identify word dependencies. Recurrent Neural Network text structures suffer from the Vanishing Gradient problem or the condition where the long-term component rises exponentially to normal 0, which hinders the model from identifying correlations between temporary conditions. Long Short-Term Memory can mitigate the Vanishing Gradient problem by introducing Memory Cells to remember values over time intervals, and also introducing three gates namely Input Gate, Output Gate, and Forget Gate to regulate the flow of information in and out of the cell to better capture Long-Term Dependencies Gated Recurrent Units are introduced to make each Recurrent Unit adaptively capture dependencies from different time scales. Gated Recurrent Units are a simpler alternative to Long Short-Term Memory and are also quite popular Just like Long Short-Term Memory, Gated Recurrent Units can also mitigate the Vanishing Gradient problem of RNN [6]. One of the most effective RNN development methods to overcome sequential problems is the Gated Recurrent Unit (GRU). To overcome the vanishing gradient problem, GRU has gate update and reset. When compared to LSTM, GRU has a fairly low computational speed [7].

In the process of recurrent neural networks, also known as back propagation neural networks, vanishing gradient is an issue that arises because Recurrent Neural Network (RNN) models cannot relate very long information [8]. Wijanarko et al. [9], the model used was a Recurrent Neural Network which was used to build translation machines for several regional languages in Indonesia but has never been used for Indonesian-Banyumasan language, especially the Gated Recurrent Unit as a result The parallel corpus used contained 1,302 sentences with an average sentence length of 20 words per sentence. The required training time is 72 hours. The resulting metric score is 34.1 which means that the model and the parallel corpus produce translation quality that can still be improved.

Wahyuzi [10], aims to be able to extract electricity consumption data, as well as data on internal and external conditions of office buildings. The model evaluation shows that BiGRU has superior performance compared to GRU. This can be seen from the lower MSE, RMSE, and MAE values, as well as a higher R2 for BiGRU. In the testing phase, the BiGRU model produces more accurate predictions, with an MAE of 0.0188 compared to GRU which has an MAE of 0.0231. In addition, the difference in total daily electricity consumption between the prediction and actual data was 11.76 kWh for BiGRU, while GRU had a difference of 23.68 kWh. Despite the longer training time of BiGRU, this model remains the best choice thanks to its higher prediction accuracy.

Astriningsih and Fudholi [11], looking at various hotel reviews aims to identify more than one aspect in a review sentence and sentiment by using deep learning methods, namely LSTM, it is hoped that the model will be able to effectively handle long strings of words. The test results obtained are a 2-aspect combination accuracy of 79% in two-aspect and three-aspect combination sentences, F1_score of 85.7% in one-aspect sentences, and F1_score of 83% in sentiment sentences. The results show that the model built has been able to perform multi-aspect sentiment on hotel reviews.

As technology develops and the volume of textual data generated every day increases, the need for more efficient and accurate text classification methods becomes increasingly important. RNN and GRU are two types of neural network architectures that are often used for NLP tasks due to their ability to handle sequential data and retain context information from previous data sequences. However, they have different characteristics and performance, which prompts the need for comparative studies to determine which method is more effective in various situations.

Attention mechanisms are innovations in the field of neural networks that aim to improve the model's ability to focus attention on important parts of the input data during the training process. By integrating attention mechanisms, the model can give greater weight to relevant information and ignore less important information, thereby improving classification accuracy. This article aims to explore the extent to which attention mechanisms can improve the performance of RNN and GRU in text classification tasks.

This research involves experiments conducted on various text datasets with different characteristics. The results of these experiments are expected to provide greater insight into the strengths and weaknesses of

each architecture as well as the impact of applying attention mechanisms. The results of this research can contribute to the development of more accurate text classification methods that can be applied in various industrial and research contexts.

Focuses on the development and performance comparison of text classification models utilizing Recurrent Neural Networks (RNN) and Gated Recurrent Units (GRU) architectures with attention mechanisms. Text classification is one of the important tasks in natural language processing (NLP) used in various applications, such as sentiment analysis and document categorization.

The purpose of this research is to examine the effect of attention mechanisms in improving the performance of RNN and GRU on text classification, determine how much these mechanisms contribute to the results obtained by implementing these models on relevant datasets, and evaluate performance based on predetermined metrics, such as accuracy, precision, recall, and F1-score.

2. METHOD

This research uses an experimental approach to evaluate the performance of various Recurrent Neural Network (RNN) models in Twitter text sentiment analysis, with and without the use of an attention mechanism. The methodological steps taken in this research are described as follows:

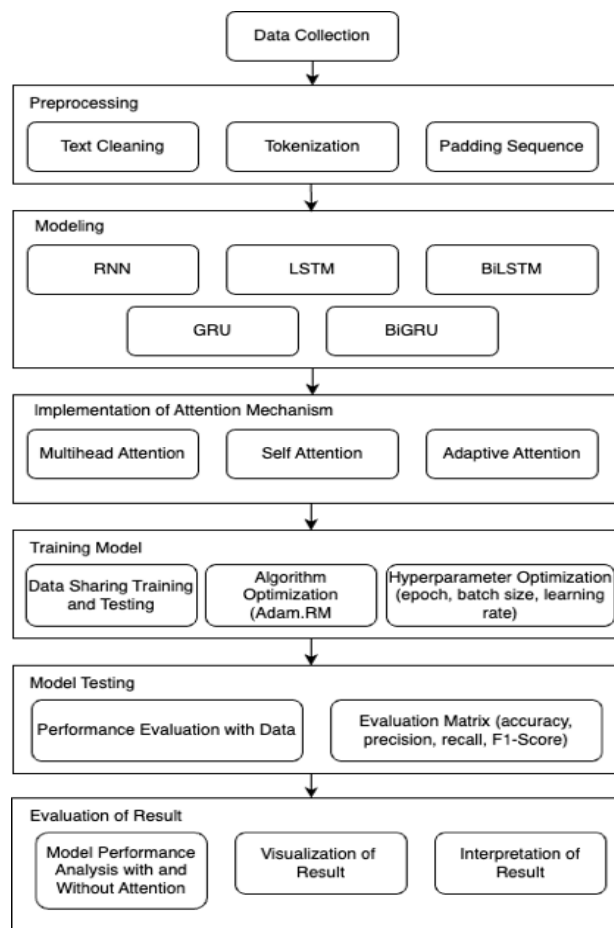


Figure 1: Research Stages

2.1 Dataset

The dataset used in this research is a tweet sentiment analysis dataset downloaded from Kaggle. This dataset consists of two main files, train.csv and test.csv. The train.csv file is used to train the model, while test.csv is used to test the performance of the model that has been trained. Each file contains tweets in text form and corresponding sentiment labels, indicating whether the tweet has a positive, negative, or neutral sentiment. This dataset was chosen because it includes a variety of language expressions commonly found on social media platforms.

Enhancing Text Classification Performance: A Comparative Study of RNN and GRU Architectures with Attention Mechanisms ... (Yulita Ayu Wardani)

2.2 Preprocessing

Preprocessing is an important stage to prepare the data before it is fed into the model. The first step in preprocessing is text cleaning, which involves removing punctuation marks, numbers, URLs, and other special characters irrelevant to sentiment analysis. After that, the text is converted to lowercase to ensure consistency. Tokenization is performed to break the text into smaller word units. The next step is sequence padding, which adds a zero value (padding) to each token so that all sequences have the same length and can be processed by the model in batches. This process is important to maintain the uniformity of input data that will be fed into the neural network.

2.3 Modeling

The models used in this study include several RNN variants: RNN, LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM), GRU (Gated Recurrent Unit), and BiGRU (Bidirectional GRU). Each model was trained using the processed data. The RNN was used as a baseline to understand the basic performance without a long-term memory mechanism. LSTM and GRU were chosen for their ability to handle long-term dependencies in sequential data. BiLSTM and BiGRU are used to exploit information from both forward and backward directions in the sequence data, which is expected to improve prediction accuracy.

2.4 Implementation of Attention Mechanism

To improve model performance, this research also implements various attention mechanisms in each type of RNN. The attention mechanisms used include Multihead Attention, Self Attention, and Adaptive Attention. Multihead Attention allows the model to focus on different parts of the input in parallel, which helps in capturing various patterns in the data. Self Attention allows the model to assess the importance of each part of the input independently, improving contextual understanding in the sequence. Adaptive Attention is used to adapt attention based on different contexts in the data, providing flexibility in information processing.

2.5 Training and Testing

The dataset is divided into training data 60% and testing data 40%. Each model is trained using the training data with optimization algorithms such as Adam or RMSprop to accelerate convergence. Hyper-parameters such as number of epochs=10, batch size=32, and learning rate=0.001 are optimized through experiments to get the best performance. Once the training is complete, the model is tested using test data that has not been seen by the model during training. The performance of the model is evaluated based on accuracy, precision, recall, and F1-score metrics to provide a comprehensive picture of the model's reliability in sentiment prediction.

2.6 Evaluation of Results

The results of each model with and without the attention mechanism were compared to assess the impact of using attention on prediction accuracy. Further analysis was conducted to understand the strengths and weaknesses of each approach, as well as the conditions under which one model outperforms another. Graphs and tables were used to visualize the results and facilitate data interpretation. This research also explored the interpretability of the model with attention to understanding which parts of the input the model considers important during the prediction process.

3. RESULTS AND DISCUSSION

This research evaluates the performance of five types of RNN models: RNN, LSTM, BiLSTM, GRU, and BiGRU in a text sentiment analysis task. Each model was tested with and without the use of attention mechanisms (Multihead Attention, Self Attention, and Adaptive Attention). The results of these tests are outlined in the following table:

Table 1. Model Performance Results Table

Model	Without Attention	Multihead Attention	Self-Attention	Adaptive Attention
RNN	58.29%	63.58%	60.45%	61.78%
LSTM	64.75%	68.34%	66.89%	67.45%
BiLSTM	65.59%	69.02%	67.53%	68.12%
GRU	66.49%	70.12%	68.23%	69.44%
BiGRU	59.59%	64.78%	62.87%	63.95%

The results show that using the attention mechanism significantly improves the model's performance in the text sentiment analysis task. The RNN model without the attention mechanism has the lowest validation accuracy of 58.29%. However, with the addition of multihead attention, the accuracy of the model increased to 63.58%. This shows that the attention mechanism helps the model to focus more on the important parts of the input text, thus improving the prediction accuracy.

The LSTM model shows a more significant improvement when using the attention mechanism. Without the attention mechanism, LSTM has an accuracy of 64.75%, but with multihead attention, the accuracy increases to 68.34%. This improvement shows that LSTM, with its ability to handle long-term dependencies, is more effective when combined with an attention mechanism that allows the model to focus on relevant information in the text.

BiLSTM and BiGRU also show significant performance improvement with the use of the attention mechanism. BiLSTM without an attention mechanism has an accuracy of 65.59%, which increases to 69.02% with multihead attention. Similarly, BiGRU without attention has an accuracy of 59.59%, which increases to 64.78% with the addition of the attention mechanism. This shows that bidirectional models (BiLSTM and BiGRU) that process information from two directions (forward and backward) can better understand the context with the help of attention, thus improving prediction accuracy.

Other attention mechanisms, such as self-attention and adaptive attention, also provide performance improvements, although not as great as multihead attention. Self-attention improved the accuracy of the standard RNN from 58.29% to 60.45%, while adaptive attention improved the accuracy to 61.78%. In the LSTM model, self-attention improved the accuracy to 66.89% and adaptive attention to 67.45%. Similar results are also seen in BiLSTM and GRU, where self-attention and adaptive attention provide a significant increase in accuracy.

Overall, the results of this study show that attention mechanisms are highly effective in improving the performance of RNN models in text sentiment analysis tasks. The LSTM model with multihead attention showed the best performance with the highest validation accuracy of 68.34%. This shows that the combination of LSTM and multihead attention is the most optimal for this task. The attention mechanism allows the model to assess and focus on important information in the text, which helps in providing more accurate predictions.

These results show that the addition of attention mechanisms to RNN models provides significant performance improvements, with LSTM and GRU showing the most significant improvements. This research makes an important contribution to the field of text sentiment analysis by demonstrating that attention mechanisms can effectively improve the performance of machine learning models.

4. CONCLUSION

This research underscores the effectiveness of combining attention mechanisms with RNN-based architectures for text sentiment analysis. The methodical integration of multihead attention, self-attention, and adaptive attention into RNN, LSTM, GRU, and their bidirectional variants revealed significant improvements in accuracy across all models. The preprocessing steps ensured clean and standardized input for the models, enabling a fair comparison of their performance under varying configurations. These methodological decisions laid a strong foundation for the insightful results obtained.

The results showed that attention mechanisms, especially multihead attention, significantly enhance the ability of RNN-based models to capture relevant features in the text. The LSTM model with multihead attention achieved the highest accuracy of 68.34%, demonstrating its superior capacity to handle sequential dependencies and focus on critical information. Bidirectional models, such as BiLSTM and BiGRU, also benefited greatly from the attention mechanisms, with improved contextual understanding and enhanced performance compared to their unidirectional counterparts.

The consistent improvements across different architectures validate the methodological framework used in this study. By leveraging attention mechanisms, the models not only achieved higher accuracy but also demonstrated the potential for broader application in natural language processing tasks. The findings confirm the hypothesis that attention mechanisms are crucial in overcoming the limitations of traditional RNN architectures, such as difficulty in handling long-term dependencies.

In conclusion, the combination of a well-structured methodological approach and impactful results highlights the potential of attention mechanisms in advancing text sentiment analysis. The integration of multihead attention into LSTM and bidirectional models emerged as the most optimal strategy for achieving high accuracy. Future work can expand on this study by exploring additional datasets, testing more advanced attention mechanisms, or integrating hybrid architectures to further enhance the performance and applicability of text classification models.

REFERENCES

- [1] Y. A. Putri Gabriella, "Optimasi Penerimaan Siswa Baru Dengan Penerapan Algoritma Text Mining Dan Tf-Idf," *J. Comput. Informatics Res.*, vol. 2, no. 3, pp. 110–117, 2023, doi: 10.47065/comforch.v2i3.941.
- [2] T. E. Hidayat and A. Rosid, "Analysis of Community Sentiments Regarding Plans to Relocate National Capital Using the Naïve Bayes Method Analisa Sentimen Masyarakat Tentang Rencana Pemindahan Ibukota Negara Dengan Metode Naïve Bayes," *Network, Comput. Sci.*, vol. 3, no. 2, pp. 43–49, 2020.
- [3] R. Siringoringo and J. Jamaludin, "Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online," *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 1, pp. 41–48, 2019, doi: 10.34012/jutikomp.v2i1.456.
- [4] T. A. Zuraiyah, D. K. Utami, and D. Herlambang, "Implementasi Chatbot Pada Pendaftaran Mahasiswa Baru Menggunakan Recurrent Neural Network," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 2, pp. 91–101, 2019, doi: 10.35760/tr.2019.v24i2.2388.
- [5] D. Sintia Amelia, N. Cahyana Aminuallah, and S. Informasi, "Teks Dan Analisis Sentimen Pada Chat Grup Whatsapp Menggunakan Long Short Term Memory (Lstm)," *J. Teknol. Terkini*, vol. 3, no. 2, p. 1, 2023, [Online]. Available: <http://teknologiterkini.org/index.php/terkini/article/view/354>
- [6] I. N. Husada and H. Toba, "Pengaruh Metode Penyeimbangan Kelas Terhadap Tingkat Akurasi Analisis Sentimen pada Tweets Berbahasa Indonesia," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 400–413, 2020, doi: 10.28932/jutisi.v6i2.2743.
- [7] F. A. Verilia, R. Firdaus, and H. D. Septama, "Pengembangan Pengenalan Aktivitas Manusia Secara Real Time Menggunakan Metode Convolutional Neural Network dan Deep Gated Recurrent Unit," *ULIL ALBAB J. Ilm. Multidisiplin*, vol. 2, no. 2, pp. 899–909, 2023, [Online]. Available: <https://journal-nusantara.com/index.php/JIM/article/view/1351>
- [8] S. J. Pipin, R. Purba, and H. Kurniawan, "Prediksi Saham Menggunakan Recurrent Neural Network (RNN-LSTM) dengan Optimasi Adaptive Moment Estimation," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 806–815, 2023, doi: 10.47065/josyc.v4i4.4014.
- [9] A. Wijanarko, A. N. Al Haura, I. Puspitaningrum, D. Intan, and S. Saputra, "Model Recurrent Neural Network-Gated Recurrent Unit untuk Membangun Mesin Penerjemah Bahasa Indonesia-Banyumasan," *J. Eksplora Inform.*, pp. 218–226, 2023, doi: 10.30864/eksplora.v13i2.977.
- [10] Z. Wahyuzi, "Analisis dan Prediksi Konsumsi Listrik Smart Office Berbasis IoT Terhadap Faktor Internal dan Eksternal Menggunakan Deep Learning," 2024.
- [11] W. Astriningsih and D. H. Fudholi, "Multi Aspek Sentimen Analisis pada Review Hotel Menggunakan Deep learning," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 10, no. 3, p. 433, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/5321>