Implementation of IndoBERT and Gated Recurrent Unit for Sarcasm Detection in Indonesian Text

Andi Ahyaul Wajdi
School of Computing
Telkom University
Bandung, Indonesia
andiyal@student.telkomuniversity.ac.id

Mahendra Dwifebri Purbolaksono
School of Computing
Telkom University
Bandung, Indonesia
mahendradp@telkomuniversity.ac.id

A Adiwijaya
School of Computing
Telkom University
Bandung, Indonesia
adiwijaya@telkomuniversity.ac.id

***Abstract*—The increasing popularity of digital communication has rendered sarcasm a prevalent mode of expression on online platforms, frequently employed to convey implicit criticism or comedy. Identifying sarcasm in Indonesian is notably difficult because of its linguistic intricacies, including several regional dialects, code-mixing, and cultural subtleties. This research introduces a hybrid deep learning model that integrates IndoBERT and Gated Recurrent Unit (GRU) to improve the precision of sarcasm detection in the Indonesian language. IndoBERT is used to create detailed context for the text, while GRU helps the model understand the order of words and long-term context, making it better at recognizing subtle sarcastic comments. The collection comprises 14,000 manually annotated Indonesian texts, representing authentic language usage on social media. The model undergoes extensive preprocessing and data augmentation methods to enhance generalization and resilience. The experimental results show that the IndoBERT-GRU model achieves an F1-score of 76.88%, with a precision of 76.47% and a recall of 77.14%, which is better than traditional machine learning standards. Despite the dataset's relatively modest size, it establishes a significant basis for sarcasm detection in Indonesian and paves the way for additional research. To enhance applicability, cross-domain evaluations across various platforms and situations are advised. This research enhances the reliability of sentiment analysis and content moderation technologies specifically designed for the Indonesian digital landscape, responding to the escalating demand for sophisticated natural language comprehension in varied online settings.**

***Keywords*—Sarcasm detection, IndoBERT, GRU, hybrid model, social media, Indonesian language**

I. Introduction

Digital transformation has changed the way Indonesians communicate, with social media becoming the main platform for sharing information and expressing opinions. Based on a comprehensive study on social media analytics in Indonesia [1], digital communication patterns show increased complexity in the interpretation of meaning, mainly due to the use of informal language and diverse cultural contexts. One of the significant challenges in digital communication is sarcasm detection, which requires a deep understanding of the linguistic and social context [2]. To address this challenge, deep learning-based approaches have evolved as a promising solution, replacing conventional methods in text analysis [3]. The development of Transformer-based language models [4] has shown significant potential in understanding the context and deeper meaning of digital texts.

The evolution of natural language processing technologies has resulted in various approaches to text analysis, especially in the context of sarcasm detection. A comprehensive review of pre-trained models [3] revealed the superior ability of the Transformer architecture in understanding complex linguistic contexts. A recent implementation in language model development [5] showed significant improvements in the accuracy of sentiment analysis and sarcasm detection compared to conventional approaches. Recent research [6] confirms that the integration of deep learning with contextual analysis provides more accurate results in understanding language nuances, including sarcasm.

In the context of Indonesian, several recent studies have shown significant progress in natural language processing. A study of recent developments in Indonesian NLP [6] revealed the importance of adapting neural network architectures to local language characteristics. The optimized IndoBERT implementation [7] has demonstrated improved performance in various language-processing tasks, including sentiment analysis and sarcasm detection. Recent research [8] confirmed the effectiveness of deep learning approaches in handling the complexity of the Indonesians language, with promising accuracy rates in sarcasm detection.

Unlike previous studies that tend to focus on general approaches, this research proposes a hybrid model that integrates the Transformer architecture [4] with the sequential processing capabilities of GRU [9]. This approach is specifically designed to address the challenges in Indonesian sarcasm detection, taking into account the unique characteristics and contextual complexity of the language. The proposed model not only aims to improve the accuracy of sarcasm detection, but also provides a deeper understanding of the linguistic patterns in Indonesian digital communication. The results of this research are expected to make a significant contribution to the development of a more sophisticated text analysis system.

II. Literature Review

Sarcasm detection in Indonesian faces unique challenges due to the complexity of cultural contexts and diverse linguistic variations in everyday language use. Recent studies have introduced benchmark evaluations for Indonesian sarcasm detection [10], identifying that the use of sarcasm in

Indonesian social media has special characteristics in informal language use and implicit context that require a deep understanding of the local culture. Social media platforms are a very significant source of data for the development of sarcasm detection models as they provide real-time interaction and natural expressions of their users in the context of digital communication [11].

The implementation of deep learning for sarcasm detection has shown very promising results in recent developments. In [12], the ensemble learning model achieved an accuracy rate of 71.67% - 73.21% on the sarcasm dataset, significantly surpassing conventional machine learning approaches. Other research [13] implemented a combination of BERT and USE architectures with an accuracy of 75% - 76% to evaluate model performance and solve specific problems in sarcasm detection. Based on a study conducted [14] in sarcasm detection using fine-tuned IndoBERT achieved an accuracy of 95% on an Indonesian language dataset, proving the importance of model optimization for specific language characteristics.

As a new approach in Indonesian sarcasm detection, this research develops a hybrid architecture of IndoBERT and GRU that is expected to overcome the limitations of previous models in understanding local cultural context and linguistic patterns. This integration utilizes IndoBERT strength in contextual understanding and GRU ability to process sequences, specifically aimed at improving the accuracy of sarcasm detection in the Indonesian context. The implementation of this model contributes significantly to the development of sentiment analysis systems that are more accurate and adaptive to local linguistic characteristics, and provides a technological foundation that can be practically implemented in social media platforms and Indonesian text analysis applications.

III. Methodology

This research develops a sarcasm detection system for Indonesian texts by integrating IndoBERT in understanding language context and GRU hybrid model capabilities in recognizing sarcasm patterns. IndoBERT acts as a contextual feature extractor that understands the nuances of Indonesian language, while GRU with model attention mechanism to focus on the part of the text that indicates sarcasm. The overall system design can be seen in Figure 1.

Figure (1) illustrates the workflow of the sarcasm detection system developed in this study. The process starts with processing the dataset through a preprocessing stage to prepare the data. Next, the dataset is divided into two parts, namely training data and testing data. The training data is processed using IndoBERT. Then, the parameter optimization process is performed using GridSearch and RandomSearch methods to determine the best configuration, before the model is trained using Gated Recurrent Unit (GRU). The final stage includes system performance evaluation to measure sarcasm detection capability.

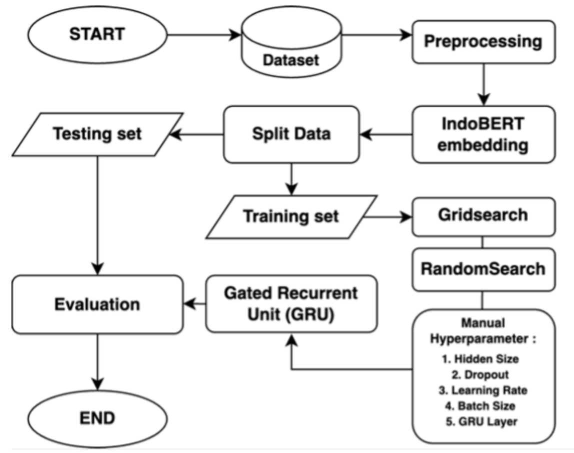

Fig. 1. Sytem Workflow

A. Dataset

This research utilizes a dataset obtained from HuggingFace [11], comprising Indonesian tweets categorized as either sarcasm or non-sarcasm. It contains 14,000 manually annotated texts from the Twitter platform, encompassing various topics and settings, making it suitable for training sarcasm detection models. The dataset has been annotated to identify sarcastic statements in social media dialogues, as shown in Table 1. Nonetheless, despite its size, the dataset may not comprehensively encapsulate the linguistic variety of Indonesian, particularly in terms of dialectal variation and code-mixing. Consequently, future research should involve evaluating the model using external datasets from various platforms and domains to enhance robustness and generalizability.

TABLE I
QUANTITY OF CRAWLED DATA

Dataset	Label	Total
"Mantap pak!! Sudah 8 tahun memimpin, hutang nambah 2.000T, kemiskinan ekstrim nambah 4jt, pengangguran nambah 2jt. PRESTASI YANG SANGAT MEMBANGGAKAN!!"	1 (Sarcasm)	9,800
"Selamat pagi sahabat, semoga hari ini kita diberi kesehatan dan rejeki yang berlimpah ya"	0 (Nonsarcasm)	4,200

The dataset is divided into two, namely 9,800 tweets labeled with sarcasm (Label 1) that contain sarcasm, and 4,200 tweets with non-sarcasm labels (Label 0). This distribution is used as a basis for training IndoBERT and GRU models to detect sarcasm in Indonesian texts.

B. Preprocessing

Preprocessing is the initial stage of processing text data to improve the quality of input for sarcasm detection models. The main purpose of preprocessing is to standardize text format, reduce noise, and convert unstructured text into a format suitable for processing by IndoBERT and GRU models. The preprocessing stages are as follows:

1) Case Folding: The process of normalizing text by converting all characters into lowercase letters to standardize the data format and reduce unnecessary variations in the next analysis stage [15].
2) Remove Punctuation: The process of eliminating punctuation and special symbols from the input text to simplify the data structure and reduce noise that can affect analysis accuracy [16].
3) Remove Numbers: A stage of text cleaning by removing all numeric characters, both stand-alone and integrated in the text, to focus the analysis on the linguistic content [16].
4) Stemming: A word processing technique that converts a compound word into a base word by removing prefixes, insertions, and suffixes in accordance with Indonesian morphological rules [15].
5) Slang Word Normalization: The process of converting informal words, slangs, and local dialects into standardized forms using a socially-developed normalization dictionary [17].
6) Tokenization: The stage of segmenting text into smaller linguistic units (tokens), which allows the model to process and understand language structures more effectively [18].

A normalization dictionary was employed to tackle dialectal variances and code-mixing of prevalent slang and mixed-language expressions. Nevertheless, further preprocessing methods like domain adaptation and embedding refinement are necessary to address greater contextual variability.

C. Data Splitting

Dividing the data into training and testing sets is an important step in sarcasm detection. The dataset is divided into two ratios, 70:30 (70% training data and 30% testing data) and 80:20 (80% training data and 20% testing data). This ratio allows the GRU model to learn the linguistic patterns of sarcasm on the training data, while the testing data is used to evaluate the model's generalization ability on new data.

D. IndoBERT embedding

IndoBERT adapts the BERT-Base architecture [4] with a Transformer Encoder that includes 12 encoder layers with a hidden dimension of 768, equipped with Multi-Head Attention and Feed-Forward Network [19]. In Multi-Head Attention, each word (token) is assessed from multiple heads in parallel, while Scaled Dot-Product Attention acts as the core mechanism to determine inter-token relevance weights [4].

The main difference of IndoBERT from the original BERT lies in its pre-training corpus which is fully Indonesian-based [19]. At the input stage, the Input Embedding Layer combines Word Embedding (word meaning), Positional Embedding (word position), and Segment Embedding (segment markers) so that local morphological and syntactic characteristics are better represented. The mathematical form of the attention mechanism can be seen in Formula 1 [3]:

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{1}$$

Formula 1 illustrates the attention mechanism, where Q, K, and $V$are query, key, and value matrices, respectively, which calculate the inter-token relevance. The QK^T operation generates a similarity score normalized by $\sqrt{d_k}$ (the dimension of the key vector) to maintain the stability of the score, which is then processed through the softmax function to generate a weight distribution. These weights are multiplied by a V matrix to form a contextualized representation of the text [3].

E. GRU

Gated Recurrent Unit (GRU) is a recurrent neural network architecture that applies a gate mechanism to process sequential data efficiently [20].This model manages the flow of information with two main gates, the update gate and the reset gate, which adaptively control the "remember" and "forget" processes at each time step [21]. This approach allows the GRU to retain relevant information from the previous hidden state and adjust the processing of new inputs, making it more efficient in dealing with dynamically changing data patterns. A detailed implementation of the GRU unit is shown in Figure 2 [22].

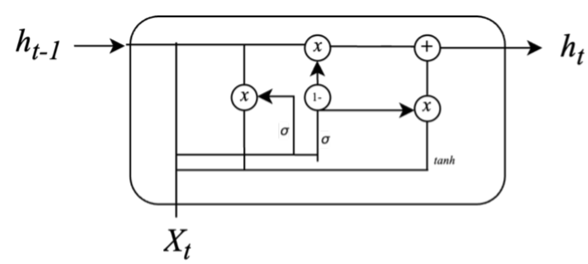

Fig. 2. Architecture GRU

In the architecture above, the symbol σ marks the gate that applies the sigmoid function to control the flow of information, while $\tanh$ denotes the $\tanh$ activation to shape the hidden state candidates. Part $1-$ is the sigmoid gate's output value subtracted from 1, signifying the portion of the old hidden state that is still retained. Symbol $+$ is a summation operation to combine the old memory result with the new memory candidate, and finally produce h_t as the latest hidden state. Input x_t and the previous hidden state h_{t-1} emphasize how each gate and activation function interact with each other to adaptively combine old and new information [20], [22].

This research extracts IndoBERT embeddings from the last encoder layer and inputs them into a two-layer GRU with a hidden size of 256 and a dropout rate of 0.2. The GRU output is subsequently processed by a dense layer utilising a sigmoid activation function to generate the final prediction. The model is trained via the Adam optimiser, with a learning rate of 0.0001 and a batch size of 16.

IV. RESULT

The research employs a sarcasm detection methodology that integrates IndoBERT and GRU models. The assessment reveals a notable enhancement in accuracy following the application of diverse optimisation methods across three experimental conditions. Each scenario is crafted to evaluate performance metrics and ascertain ideal model settings. The assessment employs F1-score, precision, and recall to evaluate the model's efficacy.

1) Scenario 1 tests two dataset split ratios (70:30 and 80:20) to determine the best configuration for maximizing performance.
2) Scenario 2 evaluates the impact of stemming (applied vs. non-applied) on detection accuracy.
3) Scenario 3 compares two hyperparameter tuning methods: GridSearch and RandomSearchCV.
4) Model Result presents the best performing models from each scenario.
5) Analysis Result summarizes and compares findings from all experiments.

A. Scenario 1

In the first scenario, two different dataset sharing ratios of 70:30 and 80:20 were evaluated to determine the optimal proportion for training and testing the sarcasm detection model that combines the GRU architecture with IndoBERT.

TABLE II
QUANTITY OF LABELED DATA

Split Rasio	F1-Score (%)	Recall	Precision
70 : 30	75.04%	76.14%	75.23%
80 : 20	**76.88%**	**77.14%**	**76.47%**

The results showed that the 80:20 ratio was better than 70:30 in all metrics, with F1-score increasing by almost 2%. With more training data, making a significant increase in F1-score from 75.04% to 76.88% shows that the model with the 80:20 ratio has better balance [23].

B. Scenario 2

In the second scenario, the effect of stemming techniques on the performance of the sarcasm detection model was evaluated. This experiment compares the performance of the model with and without the application of stemming techniques to analyze the impact on detection accuracy.

TABLE III
QUANTITY OF LABELED DATA

Parameter	F1-Score (%)	Recall	Precision
NonStemming	**76.88%**	**77.14%**	**76.47%**
Stemming	75.04%	76.14%	75.23%

Based on Table III, the NonStemming approach is more effective for sarcasm detection than Stemming, which achieves an F1-score of 75.04%, indicating that preserving the original form of the word is more important to accurately detect sarcasm [24].

C. Scenario 3

In this scenario, we test the results of the GRU model using the GridSearch and Random-SearchCV methods. To describe the performance of sarcasm detection, the GridSearch and RandomSearchCV methods are applied with several parameters as follows :

TABLE IV
HYPERPARAMETER GRIDSEARCH & RANDOMSEARCHCV PARAMETERS

Hidden Size	Dropout	Learning Rate	Batch Size	GRU Layer	Epoch
128	0.1	0.0001	16	1	5
256	0.2	0.0003	32	2	5
512	0.3	0.0001	–	–	–

Based on the table above, the sarcasm detection model requires proper hyperparameter configuration for optimal performance. Hidden size determines the capacity of the model, dropout prevents overfitting, learning rate regulates the speed of weight updates, batch size controls the amount of data per iteration, GRU layer strengthens context understanding, and sufficient epochs prevent overfitting. Through GridSearch and Random-Search, the best combination of parameters for sarcasm detection was obtained [25].

This research uses the GridSearch method to optimize the hyperparameters of IndoBERT and Gated Recurrent Unit (GRU) based sarcasm detection model, focusing on F1-score, Recall, and Precision metrics to achieve optimal configuration in Indonesian text. As shown in table V.

TABLE V
GRIDSEARCH PERFORMANCE

F1-Score	Recall	Precision
76.88%	77.14%	76.47%

Table V shows the best performance of the model from GridSearch testing, with optimal results in F1-score 76.88%, Recall 77.14%, and Precision 76.4%.

TABLE VI
GRIDSEARCH PARAMETER

GridSearch					
Hidden Size	**Dropout**	**Learning Rate**	**Batch Size**	**GRU Layer**	**Epoch**
256	0.2	0.0001	16	2	5

Table VI displays the best parameter combinations of the GridSearch method. Settings such as Hidden Size 256, Dropout 0.2, dan Learning Rate 0.0001 help the model understand sarcastic texts better, while Batch Size 16, GRU Layer of 2, and Epoch 5 ensure an effective training process for optimal results.

RandomSearch is applied in this study to adjust the hyperparameters of the sarcasm detection model integrating IndoBERT

TABLE VII
RANDOMSEARCHCV PERFORMANCE

F1-Score	Recall	Precision
75.04%	76.14%	75.23%

and Gated Recurrent Unit (GRU), with special attention to the F1-score, Recall, and Precision metrics.

The table above is the model's best result from the RandomSearch trial, with F1-score 75.04%, Recall 76.14%, and Precision 75.23%. These results show the efficiency of RandomSearch in parameter exploration, although its performance is below that of Gridsearch.

TABLE VIII
RANDOMSEARCHCV PARAMETER

RandomSearchCV					
Hidden Size	**Dropout**	**Learning Rate**	**Batch Size**	**GRU Layer**	**Epoch**
512	0.3	0.0001	16	2	5

Table VIII shows the optimal parameters selected through RandomSearchCV. With Hidden Size 512, Dropout 0.3, Learning Rate 0.0001, Batch Size 16, and GRU Layer 2, followed by epoch 5, which makes the training process effective enough for adequate performance.

D. Model Result

Based on the implementation of the sarcasm detection model using IndoBERT and GRU, testing was conducted with two different scenarios to find the optimal configuration. The test scenarios include GridSearch and RandomSearchCV approaches.

Figure 3 shows that the F1-Score curve reaches its highest point at the 5th epoch. Precision increased from 73.10% to 76.47%, indicating the model is getting more accurate in predicting sarcasm cases. There was a fluctuation in the 4th epoch where all metrics decreased, but the model managed to improve and reached the best performance in the last epoch.

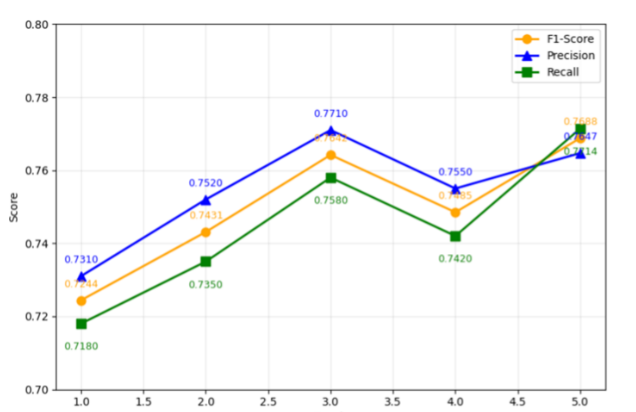

Fig. 3. Model Result GridSearch

The model showed a good balance between precision and recall at the end of training, as seen from the minimal gap between the two metrics (76.47% and 77.14%). The convergence of the three metrics at the 5th epoch indicates that the model has reached the optimal point in learning without signs of overfitting.

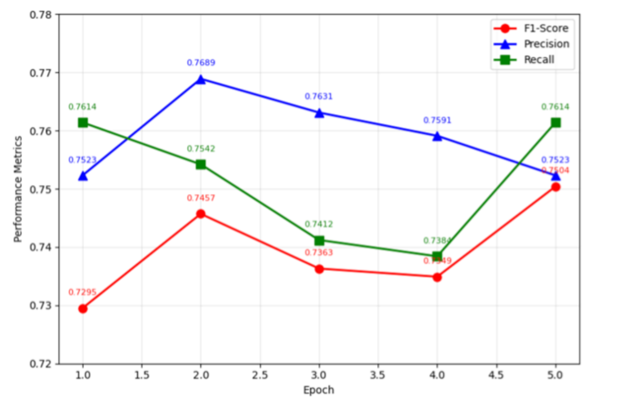

Fig. 4. Model Result RandomSearchCV

Figure 4 below shows the performance during training. F1-Score increased from 72.95% to 75.04% at the 5th epoch. Precision peaked at 76.89% on the 2nd epoch, then dropped slowly to 75.23%. Recall forms a U pattern, starting from 76.14%, decreasing to 73.84% in the 4th epoch, then returning to 76.14% in the last epoch.

The model shows steady performance improvement with an optimal F1-Score of 75.04% at the 5th epoch. There is an interaction pattern between Precision and Recall, where an increase in one metric is often followed by a decrease in the other. This suggests that the model continues to adjust the balance between the accuracy of prediction and the ability to recognize all relevant cases.

V. ANALYSIS RESULT

Scenario 1 illustrated that an 80:20 training-test split facilitates more equitable learning and enhances performance, resulting in a 1.84% increase in F1-score. This verifies that an increase in training data improves the model's generalisation in sarcasm detection.

Scenario 2 shown that stemming lowered performance. The NonStemming configuration received superior ratings, signifying that preserving original word forms aids the model in retaining crucial semantic indicators necessary for identifying sarcasm, particularly in informal or expressive circumstances.

Scenario 3 contrasted two hyperparameter optimisation techniques. GridSearch yielded superior results through more meticulous tuning, whereas RandomSearchCV attained somewhat lower accuracy but with less computing duration.

Despite yielding robust results, the proposed method occasionally misclassifies sarcasm that relies on cultural allusions or external events, particularly in politically nuanced expressions and ironic remarks grounded in shared community knowledge, indicating the need for enhanced semantic understanding and contextual modeling in future work.

The model was trained on an NVIDIA RTX 3080 GPU. The setup chosen by GridSearch required around 2.5 hours for training. The inference latency for each sentence is roughly 0.03 seconds, demonstrating that the model is computationally economical and appropriate for real-time sarcasm detection applications.

VI. Conclusion

This research introduces a hybrid IndoBERT GRU model for detecting sarcasm in Indonesian texts, attaining a maximum F1 score of 76.88% via hyperparameter optimisation using GridSearch. The amalgamation of IndoBERT's contextual comprehension with GRU's sequential processing exhibits robust performance, particularly with an 80 to 20 data division and in the absence of stemming, which more effectively retains linguistic subtleties. Despite being trained on a sample of 14,000 tweets, the model occasionally misclassifies culturally veiled or politically nuanced sarcasm, revealing shortcomings in its ability to capture deeper semantic context. Notwithstanding these obstacles, the model exhibits computational efficiency, only roughly 0.03 seconds per sentence, rendering it appropriate for real-time applications such as content moderation and sentiment analysis. This research enhances sarcasm identification in under-represented languages and establishes a basis for future investigations into cross-domain robustness and cultural context modelling in low-resource natural language processing. In conclusion, the proposed model provides a practical and scalable solution for improving the accuracy of sarcasm detection systems in the Indonesian digital landscape, particularly for social media analysis and automated moderation tools.

References

[1] R. H. Pandjaitan, "The social media marketing mix trends in Indonesia for 2024: Communication perspective," *Jurnal Komunikasi Ikatan Sarjana Komunikasi Indonesia*, vol. 9, no. 1, pp. 251–269, Jun. 2024, doi: 10.25008/jkiski.v9i1.1005.

[2] R. Batchalakuri and S. Badugu, "A comprehensive review on sarcasm detection," 2024, [Online]. Available: https://www.researchgate.net/publication/382336805. [Accessed: 5-May-2025].

[3] G. Verma, S. Raskar, M. Emani, and B. Chapman, "Cross-feature transfer learning for efficient tensor program generation," *Applied Sciences (Switzerland)*, vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020513.

[4] Y. Sun, "The evolution of transformer models from unidirectional to bidirectional in natural language processing," *Applied and Computational Engineering*, vol. 42, no. 1, pp. 281–289, Feb. 2024, doi: 10.54254/2755-2721/42/20230794.

[5] A. Taneja, "Sentiment analysis using machine learning: A comprehensive review," 2024. [Online]. Available: https://iaeme.com/Home/issue/JARET?Volume=3&Issue=1

[6] Y. Wang, Z. Li, Z. Deng, H. Song, and J. Yang, "Deep learning for natural language processing," 2023. [Online]. Available: www.intechopen.com

[7] A. F. Aji, G. I. Winata, F. Koto, S. Cahyawijaya, A. Romadhony, R. Mahendra, K. Kurniawan, D. Moeljadi, R. E. Prasojo, T. Baldwin, J. H. Lau, and S. Ruder, "One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia," *Long Papers*, 2022.

[8] R. A. Bagate and R. Suguna, "Sarcasm detection with and without #sarcasm: Data science approach," 2022. [Online]. Available: https://orcid.org/0000-0002-7538-5414

[9] J. Wu, X. Zheng, J. Wang, J. Wu, and J. Wang, "AB-GRU: An attention-based bidirectional GRU model for multimodal sentiment fusion and analysis," *Mathematical Biosciences and Engineering*, vol. 20, no. 10, pp. 18523–18544, 2023, doi: 10.3934/mbe.2023822.

[10] M. A. Rosid, D. Siahaan, and A. Saikhu, "Sarcasm detection in Indonesian-English code-mixed text using multihead attention-based convolutional and bi-directional GRU," no. 1, 2017, doi: 10.1109/ACCESS.2022.Doi.

[11] D. Suhartono, W. Wongso, and A. T. Handoyo, "Idsarcasm: Benchmarking and evaluating language models for Indonesian sarcasm detection," *IEEE Access*, vol. 12, pp. 87323–87332, 2024, doi: 10.1109/ACCESS.2024.3416955.

[12] P. Goel, R. Jain, A. Nayyar, S. Singhal, and M. Srivastava, "Sarcasm detection using deep learning and ensemble learning," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 43229–43252, Dec. 2022, doi: 10.1007/s11042-022-12930-z.

[13] H. K. Bagga, J. Bernard, S. Shaheen, and S. Arora, "Was that sarcasm?: A literature survey on sarcasm detection," Nov. 2024, [Online]. Available: http://arxiv.org/abs/2412.00425

[14] R. A. Fitrianto, A. S. Editya, M. M. H. Alamin, A. L. Pramana, and A. K. Alhaq, "Classification of Indonesian sarcasm tweets on X platform using deep learning," in *Proc. Int. Conf. on Informatics and Computational Sciences*, IEEE, 2024, pp. 388–393, doi: 10.1109/ICICoS62600.2024.10636904.

[15] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method: A case of non-formal Indonesian conversation," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00413-1.

[16] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on Twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.

[17] E. Yulianti, A. Kurnia, M. Adriani, and Y. S. Duto, "Normalisation of Indonesian-English code-mixed text and its effect on emotion classification," 2021. [Online]. Available: https://www.inlingua-edinburgh.co.uk/how-multilingual-is-your-country/

[18] C. Toraman, E. H. Yilmaz, F. Şahinuç, and O. Ozcelik, "Impact of tokenization on language models: An analysis for Turkish," Apr. 2022, doi: 10.1145/3578707.

[19] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A pre-trained language model for Indonesian Twitter with effective domain-specific vocabulary initialization," 2021. [Online]. Available: https://huggingface.co/huseinzol05/

[20] K. Han, W. Wang, and J. Guo, "Research on a bearing fault diagnosis method based on a CNN-LSTM-GRU model," *Machines*, vol. 12, no. 12, p. 927, Dec. 2024, doi: 10.3390/machines12120927.

[21] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble LSTM-GRU model," *IEEE Access*, vol. 10, pp. 39313–39324, 2022, doi: 10.1109/ACCESS.2022.3165621.

[22] A. Flores, H. Tito, and D. Centty, "Improving gated recurrent unit predictions with univariate time series imputation techniques," 2019. [Online]. Available: www.ijacsa.thesai.org

[23] H. Bichri, A. Chergui, and M. Hain, "Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets," 2024. [Online]. Available: www.ijacsa.thesai.org

[24] D. Mustikasari, I. Widaningrum, R. Arifin, W. Henggal, and E. Putri, "Comparison of effectiveness of stemming algorithms in Indonesian documents," 2021. [Online]. Available: http://tiny.cc/rootwords

[25] R. Hossain and D. Timmer, "Machine learning model optimization with hyperparameter tuning approach," 2021.