JIKO (Jurnal Informatika dan Komputer) Accredited KEMENDIKTI SAINSTEK, No.0173/C3/DT.05.00/2025
Vol. 8, No. 2, Agustus 2025, pp. 121-127 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v8i2.9988 e-ISSN: 2656-1948

EVALUATION OF INDOBERT AND ROBERTA: PERFORMANCE OF INDONESIAN LANGUAGE TRANSFORMER MODELS IN SENTIMENT CLASSIFICATION

M. Adnan Nur[1], Najirah Umar[2], Zhipeng Feng[3], Hamdan Gani[4]

[1,2] Department of Informatics Engineering, Handayani Makassar University, 90231, Makassar, Indonesia
[3]School of Culture Creativity and Media, Hangzhou Normal University, 311121, Hangzhou, Zhejiang, China
[4]Department of Machinery Automation System, ATI Makassar Polytechnic, 90211, Makassar, Indonesia

Email: [1]adnan@handayani.ac.id, [2]najirah@handayani.ac.id*, [3]20200058@hznu.edu.cn, [4]hamdangani@atim.ac.id

(Received: 31 May 2025, Revised: 26 June 2025, Accepted: 11 July 2025)

Abstract

Natural Language Processing (NLP) plays an important role in understanding public opinion, especially on social media. The Indonesian language presents unique challenges due to its morphological complexity, dialectal variations, and dynamic informal vocabulary. This study aims to evaluate and compare two popular transformer models, IndoBERT (Indonesia Bidirectional Encoder Representations from Transformers) and RoBERTa Indonesia (Robustly Optimized BERT Pretraining Approach), in sentiment classification using the Indonesian General Sentiment Analysis Dataset. This comparison is important because the two represent two different pre-trained approaches: IndoBERT is trained specifically for Indonesian, while RoBERTa is an adaptation of a multilingual model with a more aggressive pretraining approach. This study integrates Bayesian Optimization to obtain the best hyperparameter configuration. The evaluation was conducted using precision, recall, F1-score, and accuracy metrics, as well as further analysis using confusion matrix and training loss. The results showed that IndoBERT with Bayesian Optimization achieved an accuracy of 71%, while RoBERTa achieved an accuracy of 68%. IndoBERT also showed more stable and balanced performance across the three sentiment classes, with loss training values and better training efficiency. The application of Bayesian Optimization was proven to improve the performance of both models on all evaluation metrics. The results of this study show better accuracy than previous studies using the same dataset. However, the accuracy is still relatively low, indicating the limitations of the model in handling unbalanced class distributions and limited data. This study contributes to the selection of the optimal NLP model for Indonesian and demonstrates the importance of hyperparameter optimization to improve the effectiveness of sentiment classification.

Keywords: *sentiment analysis, transformert model, IndoBERT, RoBERTa, bayesian optimization*

**Corresponding Author: Najirah Umar*

1. INTRODUCTION

Natural Language Processing (NLP) is a rapidly evolving field of technology that enables computers to understand, interpret, and generate human language automatically [1], [2]. NLP has become an integral part of various modern applications, ranging from automatic translation engines, chatbots, sentiment analysis, to recommendation systems. However, most significant advancements in NLP remain concentrated on English and other major languages that have abundant data and resources available [3]. Indonesian, as the national language, has a wide variety of dialects and unique linguistic characteristics that pose challenges for NLP development [4], [5]. One of the main challenges in Indonesian NLP is the complexity of morphology, such as rich affixation, compound word formation, and regional variations in language usage that affect sentence context comprehension [6]. In addition, the availability of high-quality labeled datasets for tasks such as sentiment analysis is still very limited compared to other languages [7], so the development of accurate and adaptive models is essential.

Transformer NLP models, especially BERT (Bidirectional Encoder Representations from Transformers), have become the new standard in NLP due to their ability to capture bidirectional context and complex semantic relationships in sentences compared to traditional machine learning models [8], [9]. This model enables efficient transfer learning so that it can be adapted to various NLP tasks with relatively short fine-tuning [10]. IndoBERT (Indonesia Bidirectional Encoder Representations from Transformers) is a transformer model developed specifically for the Indonesian language, trained on a large and diverse Indonesian language corpus [11]. In addition, there is RoBERTa (Robustly Optimized BERT Pretraining Approach) which supports multilingual languages, including Indonesian. RoBERTa is a variant of BERT that optimizes the training procedure by using a larger dataset and longer training, thereby potentially improving performance [12]. RoBERTa Indonesia is an adaptation of the RoBERTa model tailored to the Indonesian language [13]. The comparison between IndoBERT and RoBERTa Indonesia is relevant because they represent two different pretraining approaches. IndoBERT is specifically trained on Indonesian language data, while RoBERTa adopts a multilingual approach with more intensive training. This study aims to examine the extent to which these pretraining approaches affect sentiment classification performance in the context of the Indonesian language.

Previous research has shown that RoBERTa has high classification accuracy compared to several transformer model variants, especially in handling multiple languages [14], [15], meanwhile, IndoBERT shows high accuracy in handling Indonesian language classification [16], [17], [18]. Sentiment analysis is a major focus due to its wide practical applications in marketing, politics, and social media, where understanding public opinion is crucial. This study compares the performance of IndoBERT and RoBERTa Indonesia in sentiment classification using a labeled Indonesian tweet dataset from a previous study conducted by Ridi Ferdiana et al [7]. This study aims to examine the extent to which the pretraining approach affects sentiment classification performance in the context of Indonesian. To improve the objectivity of the evaluation, this study also applies the Bayesian Optimization method to find the best hyperparameter combination for each model. This approach can improve the performance of the model in performing classification [19], [20]. The application of this combination of models and optimization techniques is expected to optimize performance and contribute more strongly to the selection of Indonesian NLP models.

2. RESEARCH METHOD

This study evaluates the performance of the IndoBERT and RoBERTa Indonesia models in sentiment analysis. Each model applies the Bayesian Optimization method to find the optimal hyperparameter combination. Each stage is designed to ensure the validity of the results, from dataset processing to model evaluation. In the training process, both transformer models are tested using uniform parameters to maintain consistency in comparison. The evaluation was conducted using a confusion matrix in accordance with text classification standards, such as precision, recall, F1-score, and accuracy, to measure each model's ability to understand sentiment context.

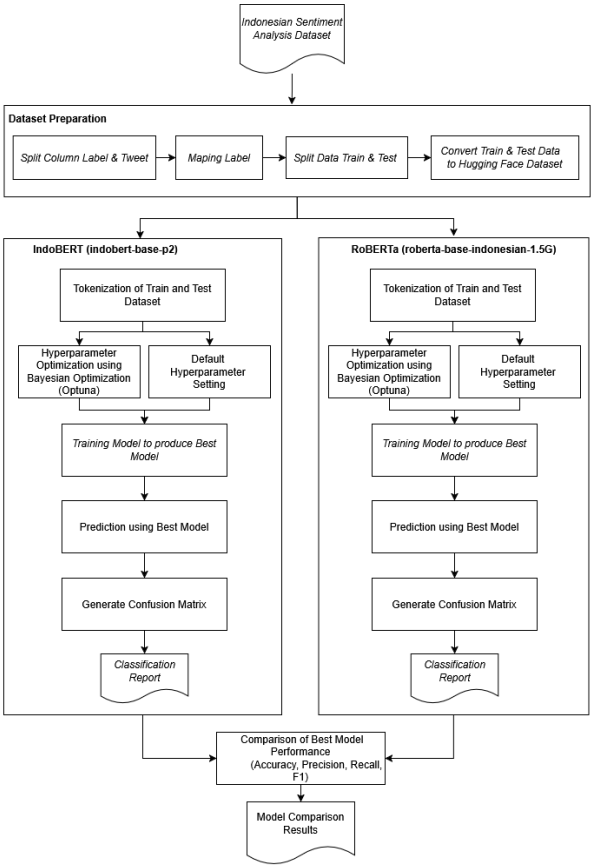

Figure 1. Stages of Model Implementation

2.1 Dataset

The dataset used is the Indonesian General Sentiment Analysis Dataset, which is a dataset of Indonesian-language tweets from research by Ridi Ferdiana et al [7]. The dataset consists of 10,806 tweets that were randomly obtained and labeled with three sentiments: positive (1), neutral (0), and negative (-1). This dataset has a ratio of 2:1:1, with 2 for neutral, 1 for positive, and 1 for negative. To maintain a balanced dataset distribution and the validity of the testing, the dataset is divided into two parts: 80% for fine-tuning/training the model and 20% for testing data using the stratified split technique. This technique ensures a balanced distribution of sentiment classes to avoid bias, allowing fine-tuning and model testing to be conducted representatively across all sentiment classes [21].

2.2 Model and Tokenization

The models used are IndoBERT (indobert-base-p2) and RoBERTa Indonesia (roberta-base-indonesian-1.5G), both of which are pre-trained transformer models provided by the Indonesian NLP community and accessible through the Hugging Face Transformers library [22]. In implementing tokenization, IndoBERT uses the WordPiece tokenizer for Indonesian to handle morphology and language structure more effectively [23]. Meanwhile, RoBERTa Indonesia implements Byte Pair Encoding (BPE), a commonly used tokenizer capable of handling text from various languages, making it more widely applicable [24]. These differences in tokenization techniques can potentially affect how text is processed by each model.

2.3 Fine Tuning

The training was implemented on the Google Colab platform using GPU runtime and Python language, as well as the PyTorch framework as the deep learning backend. Fine-tuning was performed to adjust the pre-trained model to the sentiment classification task in the context of Indonesian. In this study, two training approaches were applied separately:

a. Training with Default Parameters

Both models were trained using basic parameter configurations commonly used in text classification tasks. The default parameters used include:

Table 1. Default Parameter

Parameter	Value
Learning rate	0,000003 (3e-5)
Training Bacth Size	8
Evaluation Batch Size	8
Epoch	3
Weight Decay	0,01
Gradient Accumulation Steps	2

The purpose of using this default configuration is to obtain a baseline performance for each model directly without modification.

b. Training with Bayesian Optimization

To improve model performance and address the limitations of default parameters, the best hyperparameters were searched for using Bayesian Optimization through the Optuna library. This optimization process involved searching the following parameter space:

Table 2. Parameter Bayesian Optimization

Parameter	Value
Learning rate	1e-5 to 5e-5
Training Bacth Size	8 or 16
Epoch	2 to 4
Weight Decay	0.0 to 0.3

Optuna performs optimal search through 10 trials (n_trials=10) for each model. The goal of this process is to find the combination of parameters that produces the highest evaluation accuracy (eval_accuracy). This process enhances the value of the research by introducing efficient and adaptive unsupervised hyperparameter tuning to the data.

Once the best hyperparameters have been found for each model, a retraining process is carried out using the optimized parameters to ensure that the model has fully adjusted to the optimal configuration. The entire training and evaluation process is carried out separately between the default version and the optimized version, so that the comparison of results is valid and reliable.

2.4 Evaluation

The evaluation was conducted on 20% of the test data from the dataset. The model was evaluated using accuracy, precision, recall, and F1-score metrics. To support deeper analysis, a confusion matrix was used to determine the distribution of predictions for each class, as well as loss graphs (training and evaluation) and training time per epoch to assess model efficiency. This approach aims not only to compare final performance but also to examine the training behavior of both models under default parameter settings and after optimization.

3. RESULT AND DISCUSSION

3.1 Model Evaluation Results

The evaluation results show that the IndoBERT model with hyperparameter optimization using Bayesian Optimization achieved the highest accuracy of 71%, while RoBERTa Indonesia achieved a maximum accuracy of 68% after optimization. With the default hyperparameter configuration, each model produced an accuracy of 68% for IndoBERT and 66% for RoBERTa Indonesia. Although the accuracy difference between the models appears small, analysis of other metrics such as precision, recall, and F1-score provides a clearer picture of each model's performance across each sentiment class (negative, neutral, positive).

The following table presents the best results from each model:

Table 3. Comparasion Model Result

Model	Label/Class	Precision	Recall	F1-Score	Accuracy
IndoBERT (Default)	Negative	0.63	0.64	0.63	0.68
	Neutral	0.73	0.70	0.72	
	Positive	0.64	0.68	0.66	
IndoBERT (Bayesian Optimization)	Negative	0.65	0.70	0.67	0.71
	Neutral	0.77	0.71	0.74	
	Positive	0.67	0.71	0.69	
RoBERTa (Default)	Negative	0.61	0.60	0.60	0.66
	Neutral	0.72	0.70	0.71	
	Positive	0.61	0.66	0.64	
RoBERTa (Bayesian Optimization)	Negative	0.63	0.64	0.63	0.68
	Neutral	0.75	0.72	0.73	
	Positive	0.62	0.66	0.64	

Here are the results of the confusion matrix:

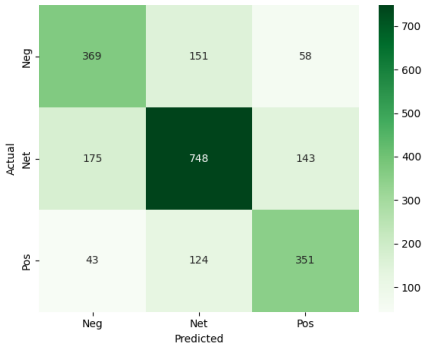

Figure 2. IndoBERT Confusion Matrix with Default Parameters

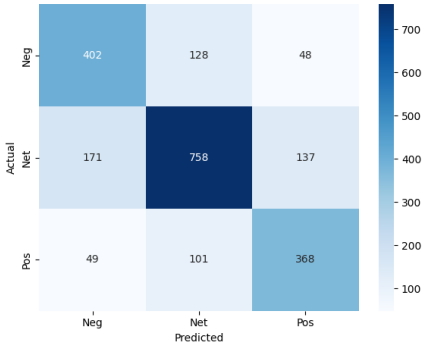

Figure 3. IndoBERT Confusion Matrix with Bayesian Optimization

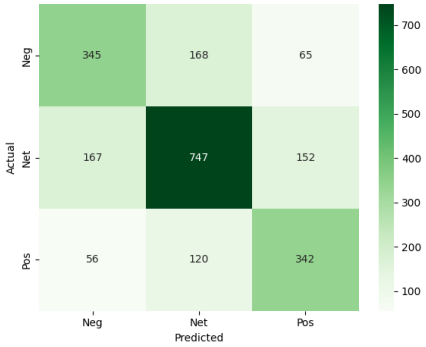

Figure 4. RoBERTa Confusion Matrix with Default Parameters

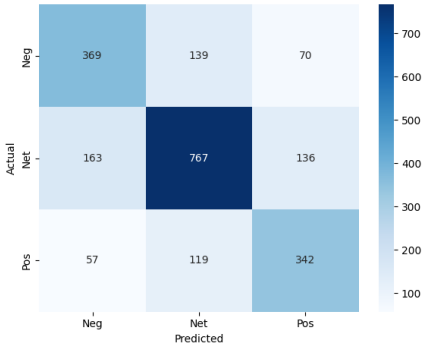

Figure 5. Confusion Matrix of RoBERTa with Bayesian Optimization

Based on model evaluation and confusion matrix, it appears that RoBERTa tends to perform better on the Neutral sentiment class, but often misclassifies the Negative and Positive classes. In contrast, IndoBERT performs more consistently and evenly across all three classes, demonstrating its ability to handle differences in expression and context in Indonesian.

IndoBERT better performance is due to several factors, namely the IndoBERT pretraining corpus, which includes variations of formal and informal Indonesian, and the use of the WordPiece tokenizer, which is more suitable for handling affixation and local word structures. Meanwhile, RoBERTa uses the more common byte-level BPE tokenizer, which is suitable for multilingual texts but is not optimal for capturing the morphological nuances of Indonesian. RoBERTa more aggressive architecture also requires a large dataset to generalize effectively, so in a limited dataset, this model is less than optimal. Both models showed improved performance after Bayesian Optimization using Optuna, particularly in terms of precision and F1-score across all classes. This demonstrates that hyperparameter selection plays a crucial role in enhancing model effectiveness and distinguishing generalization quality.

The best accuracy result obtained from IndoBERT with Bayesian Optimization was 0.71, which was better than previous studies that used the same dataset with the highest accuracy values of 0.629 and 0.6374 [7], [25]. Although this study has better accuracy results than previous studies, it is still relatively low in sentiment analysis. This is due to several factors, namely the dataset used has an uneven class distribution, with the Neutral class dominating at 49.3%, while the Negative and Positive classes only account for 26.7% and 24.0% of the total data, respectively. This imbalance causes the model to be biased toward the Neutral class and makes it difficult to distinguish between the Negative and Positive classes. Additionally, the dataset consists of only 10,806 samples, which is relatively small for training large-scale models such as IndoBERT and RoBERTa. The small amount of data limits the model's ability to learn more complex patterns.

3.2 Loss and Overfitting Analysis

Another important aspect in evaluating model performance is Training Loss analysis. The loss per-epoch graph is shown in the image below:

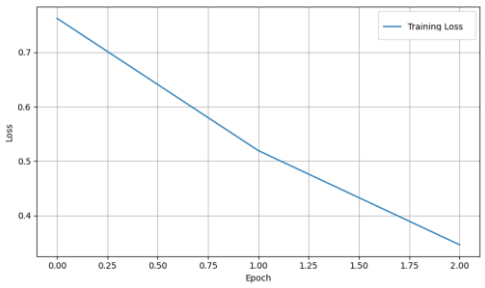

Figure 6. Loss per epoch of IndoBERT with default parameters

The graph above shows a steady and significant decrease from 0.77 to 0.33 over two epochs. This substantial decrease in loss indicates that the default parameters of IndoBERT are already sufficiently

effective in learning the patterns of the Indonesian language dataset, with a stable and efficient training process without excessive fluctuations. The continuously decreasing loss value up to the second epoch suggests that the model has not yet reached overfitting, allowing for extended training if needed to achieve optimal results.

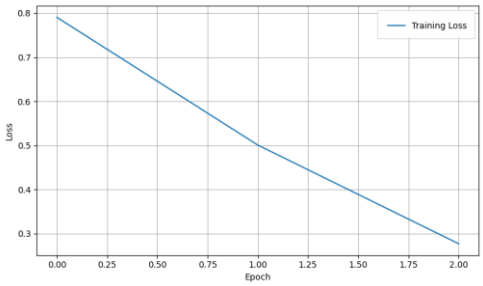

Figure 7. Loss per epoch of IndoBERT with Bayesian Optimization

IndoBERT Training Loss graph with Bayesian Optimization shows a decrease in loss from 0.79 to 0.275 over two epochs, which means a better decrease than the default parameters. This indicates that hyperparameter tuning using Bayesian Optimization successfully improves the learning efficiency of IndoBERT, accelerating the decrease in loss within the same number of epochs. This reduction also demonstrates that the tuning successfully identified a more suitable combination of learning rate, batch size, and weight decay for the dataset and model, enabling the model to learn faster and more stably.

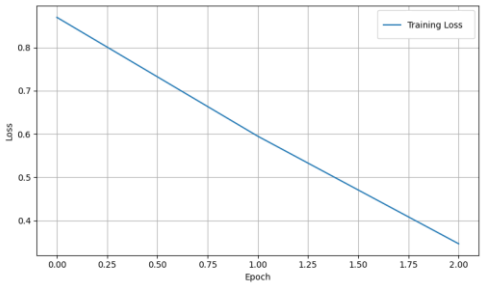

Figure 8. Loss per epoch of RoBERTa with default parameters

RoBERTa graph with default parameters shows a steady decrease in loss from 0.87 to 0.34 over two epochs. This decrease indicates that RoBERTa is quite effective in learning the Indonesian language dataset with the default configuration, and the learning process proceeds without any signs of divergence. This fairly low final loss value also demonstrates RoBERTa ability to understand data patterns well.

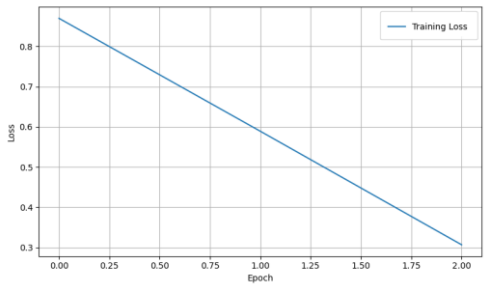

Figure 9. Loss per epoch of RoBERTa with Bayesian Optimization

RoBERTa graph with Bayesian Optimization shows a better loss reduction compared to the default parameters, from 0.87 to 0.31 in two epochs. This reduction indicates that hyperparameter tuning via Bayesian Optimization successfully enhances the learning efficiency of the RoBERTa model, enabling it to learn from the dataset more quickly and effectively. With a lower final loss, this tuning demonstrates improved model performance, making RoBERTa increasingly optimal for Indonesian language processing tasks that require higher efficiency and accuracy.

4. CONCLUSION

This study evaluated the performance of two Indonesian language transformer models, IndoBERT and RoBERTa Indonesia, with hyperparameter optimization using Bayesian Optimization, in a sentiment classification task using the Indonesian General Sentiment Analysis Dataset. The results show that IndoBERT consistently outperforms RoBERTa Indonesia in various evaluation aspects, including accuracy, F1-score, and training efficiency. IndoBERT with Bayesian Optimization achieved the highest accuracy of 71% and an average F1-score of 0.70, while RoBERTa with similar optimization achieved an accuracy of 68% and an average F1-score of 0.67. In addition, IndoBERT demonstrated more efficient training time with lower training and evaluation loss values, indicating a stable learning process and better generalization ability without signs of overfitting during the epochs run. IndoBERT's advantages are influenced by its architecture and pretraining corpus tailored for Indonesian, as well as the use of the WordPiece tokenizer, which is effective in handling the morphological complexity and structure of the Indonesian language. Conversely, RoBERTa Indonesia has an advantage in the neutral sentiment class, but its performance is less stable in the negative and positive classes, especially on datasets with imbalanced class distributions. The application of Bayesian Optimization has been proven to improve the accuracy, precision, recall, and F1-score values of both models, indicating that hyperparameter selection plays an important role in improving the effectiveness and generalization quality of the model. Although the accuracy results are better than previous studies, this study confirms that the use of larger and more

balanced datasets is still necessary to reduce bias in dominant classes and improve overall model performance in Indonesian sentiment analysis.

5. REFERENCE

[1] R. C. Rivaldi, T. D. Wismarini, J. T. Lomba, and J. Semarang, "Analisis Sentimen Pada Ulasan Produk Dengan Metode Natural Language Processing (NLP) : (Studi Kasus Zalika Store 88 Shopee)," *Elkom: Jurnal Elektronika dan Komputer*, vol. 17, no. 1, pp. 120–128, Jul. 2024, doi: 10.51903/ELKOM.V17I1.1680.

[2] M. F. Naufal and S. F. Kusuma, "Natural Language Processing untuk Otomatisasi Pengenalan Pronomina dalam Kalimat Bahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 5, pp. 1011–1018, Oct. 2022, doi: 10.25126/JTIIK.2022946394.

[3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, Oct. 2018, doi: https://doi.org/10.48550/arXiv.1810.04805.

[4] A. F. Aji *et al.*, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 7226–7249, Mar. 2022, doi: 10.18653/v1/2022.acl-long.500.

[5] D. Sebastian, H. D. Purnomo, and I. Sembiring, "BERT for Natural Language Processing in Bahasa Indonesia," *2022 2nd International Conference on Intelligent Cybernetics Technology and Applications, ICICyTA 2022*, pp. 204–209, 2022, doi: 10.1109/ICICYTA57421.2022.10038230.

[6] I. M. Krisna Dwitama, M. S. Al Farisi, I. Alfina, and A. Dinakaramani, "Building Morphological Analyzer for Informal Text in Indonesian," *Proceedings - ICACSIS 2022: 14th International Conference on Advanced Computer Science and Information Systems*, pp. 199–204, 2022, doi: 10.1109/ICACSIS56558.2022.9923494.

[7] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. Sekar, T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 8, no. 4, pp. 334–339, Nov. 2019.

[8] A. Hussein Mohammed *et al.*, "Survey of BERT (Bidirectional Encoder Representation Transformer) types," *J Phys Conf Ser*, vol. 1963, no. 1, p. 012173, Jul. 2021, doi: 10.1088/1742-6596/1963/1/012173.

[9] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, "Comparing BERT against traditional machine learning text classification," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 4, pp. 352–356, May 2020, doi: 10.47852/bonviewjcce3202838.

[10] A. Pardamean, H. F. Pardede, and A. Pardamean STMIK Nusa Mandiri, "Tuned bidirectional encoder representations from transformers for fake news detection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1667–1671, Jun. 2021, doi: 10.11591/IJEECS.V22.I3.PP1667-1671.

[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 757–770, Nov. 2020, doi: 10.18653/v1/2020.coling-main.66.

[12] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, doi: https://doi.org/10.48550/arXiv.1907.11692.

[13] W. Suwarningsih, R. A. Pratama, F. Y. Rahadika, and M. H. A. Purnomo, "RoBERTa: language modelling in building Indonesian question-answering systems," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, no. 6, pp. 1248–1255, Dec. 2022, doi: 10.12928/TELKOMNIKA.V20I6.24248.

[14] J. Carreras Timoneda and S. Vallejo Vera, "BERT, RoBERTa, or DeBERTa? Comparing Performance Across Transformers Models in Political Science Text," *https://doi.org/10.1086/730737*, Jan. 2025, doi: 10.1086/730737.

[15] A. F. Adoma, N. M. Henry, and W. Chen, "Comparative Analyses of Bert, Roberta, Distilbert, and Xlnet for Text-Based Emotion Recognition," *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2020*, pp. 117–121, Dec. 2020, doi: 10.1109/ICCWAMTIP51612.2020.9317379.

[16] M. R. Mahardika, I. P. J. Wijaya, A. R. Prayoga, H. Lucky, and I. A. Iswanto, "Exploring the Performance of BERT Models for Multi-Label Hate Speech Detection on Indonesian Twitter," *2023 4th International Conference on Artificial Intelligence and Data Sciences: Discovering Technological Advancement in Artificial Intelligence and Data Science, AiDAS 2023 - Proceedings*, pp. 256–261, 2023, doi: 10.1109/AIDAS60501.2023.10284596.

[17] C. Jocelynne, L. Tobing, I. G. N. Lanang Wijayakusuma, L. Putu, and I. Harini,

"Perbandingan Kinerja IndoBERT dan MBERT untuk Deteksi Berita Hoaks Politik dalam Bahasa Indonesia," *JST (Jurnal Sains dan Teknologi)*, vol. 14, no. 1, pp. 114–123, May 2025, doi: 10.23887/JSTUNDIKSHA.V14I1.92126.

[18] R. I. Yulfa, B. H. Setiawan, G. G. Lourensius, and K. Purwandari, "Enhancing Hate Speech Detection in Social Media Using IndoBERT Model: A Study of Sentiment Analysis during the 2024 Indonesia Presidential Election," *ICCA 2023 - 2023 5th International Conference on Computer and Applications, Proceedings*, 2023, doi: 10.1109/ICCA59364.2023.10401700.

[19] F. P. Simamora, R. Purba, and M. F. Pasha, "Optimisasi Hyperparameter BiLSTM Menggunakan Bayesian Optimization untuk Prediksi Harga Saham," *Jambura Journal of Mathematics*, vol. 7, no. 1, pp. 8–13, Feb. 2025, doi: 10.37905/JJOM.V7I1.27166.

[20] D. Utami, F. Dwiatmoko, and N. A. Sivi, "Analisis Pengaruh Bayesian Optimization Terhadap Kinerja SVM Dalam Prediksi Penyakit Diabetes," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 140–150, Jan. 2025, doi: 10.29408/JIT.V8I1.28468.

[21] M. Kuhn and K. Johnson, "Applied predictive modeling," *Applied Predictive Modeling*, pp. 1–600, Jan. 2013, doi: 10.1007/978-1-4614-6849-3.

[22] Hugging Face, "Transformers." [Online]. Available: https://huggingface.co/docs/transformers/index

[23] W. Wongso, D. S. Setiawan, S. Limcorn, and A. Joyoadikusumo, "NusaBERT: Teaching IndoBERT to be Multilingual and Multicultural," Mar. 2024, doi: https://doi.org/10.48550/arXiv.2403.01817.

[24] R. Sennrich, B. Haddow, and A. Birch, "Neural Machine Translation of Rare Words with Subword Units," *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, vol. 3, pp. 1715–1725, Aug. 2015, doi: 10.18653/v1/p16-1162.

[25] N. Umar, M. A. Nur, T. Informatika, and H. Makassar, "Application of Naïve Bayes Algorithm Variations On Indonesian General Analysis Dataset for Sentiment Analysis," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 585–590, Aug. 2022, doi: 10.29207/RESTI.V6I4.4179.