



## **Comparing Optimizer Strategies For Enhancing Emotion Classification In IndoBERT Models**

**Julius Immanuel Theo Krisna<sup>1\*</sup>, Ardytha Luthfiarta<sup>1</sup>, Leno Dwi Cahya<sup>1</sup>, Sri Winarno<sup>1</sup>, Adhitya Nugraha<sup>1</sup>**

<sup>1</sup>Faculty of Computer Science, Universitas Dian Nuswantoro Semarang, Jl. Imam Bonjol No. 207 Semarang, Central Java, Indonesia

\*[juliusimmanueltheo@gmail.com](mailto:juliusimmanueltheo@gmail.com)

**Abstract.** Emotions are one of the reactions of human when they receive physical or verbal action. Every human action is based on emotion. Every opinion expressed in the comments column also contains the author's emotions. This research aims to classify five emotions, Marah, Takut, Senang, Cinta, and Sedih and evaluate the performance of three commonly used optimizer, Adam, RMSProp, and Nadam. The processed data used IndoBERT model for Indonesian text classification. The research purpose to search the best optimizer for text classification. The result shows classification used Adam Optimizer 90,21%, RMSProp Optimizer 82.11, and Nadam Optimizer 88.61%. The Adam optimizer applied to the IndoBERT model yielded the best results. This shows a significant improvement from previous studies, which had emotion classification.

**Keywords:** NLP, IndoBERT, Optimizer, Adam, RMSProp, Nadam

*(Received 2024-01-20, Accepted 2024-01-31, Available Online by 2024-02-08)*

### **1. Introduction**

Emotions are one of the human reactions when receiving physical or verbal actions. Emotions can be shown through facial expressions or actions.[1] Emotions are experiences that humans experience consciously to characterize a psychological state that is important as part of human nature, such as joy, anger, love, sadness, fear, and so on.[2] Every human action is based on emotions. Every opinion expressed in the comments column also contains the author's emotions. Dissemination of information can influence public opinion. The emotional content in this opinion is interesting to analyze in research, especially the emotional patterns contained in this opinion. Text classification aims to analyze, process, and extract information contained in text. In emotion classification, a comment will be taken to obtain the information contained in it so that we can find out the emotion in the comment. [3]

Many companies currently use comments from various social media in the form of Instagram, Twitter, and Facebook.[4] Through this, they can understand customers' feelings about the company. This opinion processing usually uses sentiment analysis which has developed in various fields of business, government, and organizations.[5] Natural Language Processing (NLP) plays an important role in the field of Artificial Intelligence (AI).

Natural Language Processing enables machines to understand, interpret, and respond to human language, thereby facilitating human-computer interactions. NLP is tightly integrated with machine learning (ML) and AI, enabling the development of voice-controlled systems, language translation, sentiment analysis, and a variety of other applications. The combination of NLP, AI, and ML drives the automation of data analysis processes and changes the way we interact with technology, making advanced tools more accessible and user-friendly. [6] The development of artificial intelligence technology has been enormous in recent years due to the use of the Internet, big data, the Internet of Things, and the use of very massive processing power.[7], [8]

Nowadays pre-trained language models have helped increase the sophistication in many areas of NLP (Natural Language Processing). Currently, many machine learning models are being developed for sentiment analysis. This includes conventional models such as SVM (Support Vector Machine), NBC (Naïve Bayes Classifier), and currently the use of deep learning models based on Neural Networks. One of the developments in the deep learning model is Transformer, including Bidirectional Encoder Representations from Transformers (BERT). [9], [10], [11]

In recent years Bidirectional Encoder Representations from Transformers (BERT) combines the representation of words and sentences in a Transformer on a large scale. Determining the right model with a dataset is a way to overcome large and heavy computations.[12] In BERT there are various types of models such as IndoBERT, ALBERT, RoBERTa, and many more. Indonesian is the 10th most widely used language. Because of its wide use, NLP practitioners use the IndoBERT model to organize existing language resources.[1] IndoBERT is a transformer model by adapts BERT itself but is drilled purely as a masked language model drilled using huggingface by following the BERT base configuration.[13] BERT models are a development of Transformer by adapting the dataset used by the user. The use of models that match the language in the dataset also reduces the computational complexity.[4] Optimization is one aspect of improving the performance of the model used.[14]

This research aims to implement the Indonesian Transformer model, namely the IndoBERT model, to compare the three optimization methods Adam (Adaptive Moment Estimation), Nadam (Nesterov-accelerated Adaptive Moment Estimation), and RMSProp (Root Mean Square Propagation). Optimization plays an important role in managing the accuracy by adjusting the learning rate of the model during the training process to get optimal prediction results.

**Table 1.** Literature Study

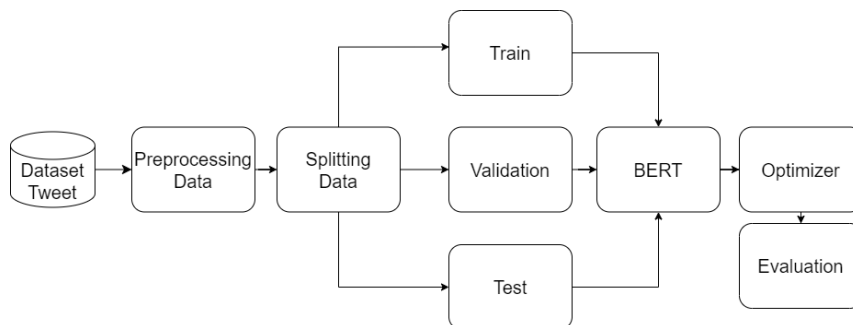
<b>Author</b>	<b>Year</b>	<b>Method</b>	<b>Result or Finding</b>
<b>Bagus [1]</b>	2022	Comparison BERT Uncased and IndoBERT with Adam Optimizer	Classifying emotions using BERT and Adam, resulting in an accuracy value of 90% in BERT Uncased and producing result 81% accuracy in using the IndoBERT.
<b>Hulliyah, Rayya, Bakar [15]</b>	2022	IndoBERT with Adam Optimizer for develop Chatbot	The use of optimizers in Natural Language Processing using IndoBERT with the Adam optimizer to develop chatbots for emotion classification obtained accuracy, F1 score, recall, and precision values of 89%, 89%, 89%, and 90% on train data. To validate the data, the accuracy, F1 score, recall, and precision values were 70%, 71%, 70%, and 72%.
<b>Wijaya [16]</b>	2021	IndoBERT with Adam Optimizer	In implementing IndoBERT with another Adam optimizer Disabled News classification, getting an accuracy value of 0.853, recall 0.853, precision 0.872, and F1 0.853.
<b>Cahya [12]</b>	2024	Handling Imbalanced Dataset using IndoBERT	Experiment with AdamW optimizer the result is Augmentation technique performance by up to 20%, with accuracy 78%, precision 85%, recall 82%, and an

		Model	F1-score of 83%. With SMOTE technique, the evaluation results achieve accuracy to a high 82% with precision at 87%, recall at 85%, and an F1-score of 86%.
<b>Nugroho, Bachtiar</b> [17]	2021	Text-Based Emotion Recognition in Indonesian Tweet using BERT	Experiment use Adam optimizer for BERT-Base got 0,6478 accuracy and IndoBERT-Base got 0,7673 accuracy.

Based on the provided literature study, it is possible to utilize IndoBERT-Base for classifying emotion from Dataset Twitter. The usage of IndoBERT for emotion classification is still relatively uncommon. However, in some studies, the effectiveness of IndoBERT is reported to be suboptimal compared to BERT-Uncased. [1] Several research works have explored the potential of optimizing IndoBERT's performance, with findings suggesting that the use of the Adam Optimizer can yield improved results. [1], [15], [16], [17] IndoBERT-Base remains relevant for classification purposes to date.

This research endeavors to surpass previous studies by employing advanced methods and algorithms. The investigation employs IndoBERT-Base with a focus on emotion classification, utilizing a dataset with five labels yaitu Marah (Angry), Takut (Fear), Senang (Joy), Cinta (Love), and Sedih (Sad) from 9,480 Tweets. The dataset is larger than that used in [17]. Additionally, three different optimizers are employed to identify the most effective one for achieving optimal classification results. The primary goal of this study is to specifically determine the best optimizer for emotion classification using pre-trained IndoBERT-Base.

## 2. Methods



**Figure 1.** Research Methods

The research begins by carrying out a dataset search stage which can be illustrated in Figure 1. The dataset has been adjusted through the Preprocessing Stage for model training. Next, the dataset will be divided into 3 parts: Train Data, Validation Data, and Test Data. In the training process, the model is processed in Data Train, and its performance is assessed using various evaluation metrics, including graphical accuracy, graphical loss, confusion matrix, and classification results. Evaluation of this Matrix can provide new insights into the effectiveness of the model and its ability to classify and predict results correctly.

### 2.1. Dataset

In this research, we use the Emotion Dataset from public opinion from Twitter. This research uses a multi-label emotion dataset in Indonesian. This dataset has 5 labels, namely Marah (Angry), Takut (Fear), Senang (Joy), Cinta (Love), and Sedih (Sad). This dataset has a total of 9,480 tweets along with labels.

**Table 2.** Dataset Emotion

Tweet	Label
bikin gue emosi aja	Marah
takut bgt jujur kalo yangti udh turun tangan	Takut
waa enak tuu, semoga libur mu bahagia yaa!~	Senang

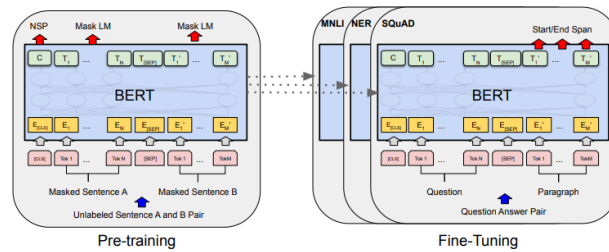
suka nih netijen yang kayak gini :) otak" detektif gw like banget	Cinta
mandi sore, soalnya mo les. sedih malem minggu malah les	Sedih

## 2.2. Text Preprocessing

This research uses several preprocessing, including: case folding such as lower case to change capital text to lower case and delimiters to remove parts that can influence results such as tags, urls, emoji, punctuation and whitespace. Next, the data is given a tokenizer provided in the BERT model. In addition, data that null will be removed.

## 2.3. Modelling IndoBERT

In this research, classification is carried out using the IndoBERT model, which is a development of BERT (BiDirectional Encoder Representations Transformers). This difference is because the architecture of this model focuses on processing and training data in Indonesian based on the Transformer architecture. In the Transformer architecture, there are two parts, namely the decoder and the encoder. Specifically for BERT, it only uses the Encoder architecture.[18]



**Figure 2.** BERT architecture [18]

The BERT model architecture is a multi-layer bidirectional Transformer. There are two stages carried out in BERT. There are two stages in BERT, namely the pre-training stage and the fine-tuning stage. In the pre-training stage, the BERT model architecture is drilled using unlabeled data and trained with different tasks, while for the fine-tuning stage, the BERT model is first initialized with pre-drilled parameters, and all parameters are tuned using labeled data.[13]

In the pretrain stage, the model is drilled with a labeled dataset. For the IndoBERT model architecture, a large dataset was collected, consisting of about four billion words and about 250 million sentences from the Indonesian text collection. This dataset includes news texts from various sources such as local online, social media, Wikipedia, online articles, subtitle texts from video recordings, and a parallel data set known as Indo4B. Indo4B includes formal word data, informal word data and casual words in Indonesian.[12]

## 2.4. Optimizer

### 2.4.1. Adam (Adaptive Moment Estimation)

Adam is one of the optimizations that is widely used in training models today. This optimization can be used in machine learning or deep learning. Adam is a gradual optimization algorithm that uses an adaptive learning rate. The Adam optimizer combines the first and second moments of the gradient to update the parameters.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (1)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (2)$$

$$\widehat{m}_1 = \frac{m_1}{1 - \beta_1^t} \quad (3)$$

$$\widehat{m}_2 = \frac{m_2}{1 - \beta_2^t} \quad (4)$$

$$p = \frac{lr \times \widehat{m}_1}{\sqrt{\widehat{m}_2 + \epsilon}} \quad (5)$$

In this formula,  $m$  is the momentum,  $\beta$  is the repetition parameter,  $g$  is the gradient of the parameter loss function,  $p$  is the parameter,  $t$  is the iteration,  $lr$  is the learning rate, and  $\epsilon$  is a small number to avoid divide it by zero.

#### 2.4.2. RMSProp (Root Mean Square Propagation)

RMSProp is the default optimization algorithm used in some neural network models. These algorithms have adjustments for the problem of decreasing the learning rate too quickly.

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \cdot g_t \quad (6)$$

In this formula  $\theta_t$  is the parameter at iteration time (t),  $g_t$  is the gradient parameter at iteration time (t),  $E[g^2]_t$  is the storage of the average square of the gradient at iteration time (t),  $\eta$  is the learning rate, and  $\epsilon$  is the epsilon is a small number to avoid divide it by zero.

#### 2.4.3. Nadam (Nesterov-accelerated Adaptive Moment Estimation)

Nadam is a development of Adam. Nadam is a combination of the Nesterov Accelerated Gradient (NAG) optimizer with Adam (Adaptive Moment Estimation).

$$\theta_{t+1} = \theta_t - \alpha \left( \beta_1 m_t \frac{(1-\beta_1)\nabla f(\theta_t)}{1-\beta_1^t} \right) \cdot \left( \frac{\beta_2 \sqrt{v_t}}{\sqrt{1-\beta_2^t}} + \epsilon \right)^{-1} \quad (7)$$

In this formula  $\theta_t$  is the parameter at iteration time (t),  $\alpha$  is the learning rate,  $\beta_1$  and  $\beta_2$  are the decay factors for gradient and gradient square respectively.  $m_t$  is the average change in gradient,  $v_t$  is the average change in the squared gradient, and  $\epsilon$  is a small value that prevents division by zero.

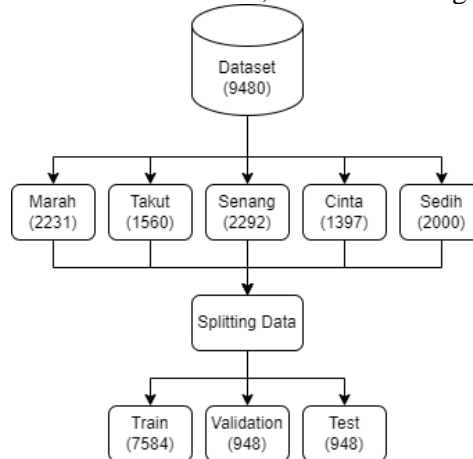
#### 2.5. Evaluation

In this research, evaluation was carried out with three models, namely Adam, RMSProp, and Nadam. Evaluation has been carried out by comparing several optimizers with a confusion matrix. The results of the confusion matrix will be compared and explained to determine the performance of each optimizer. The performance will be compared in terms of precision, recall, f1-score, and accuracy.

### 3. Results and Discussion

#### 3.1. Dataset

In this research, we use the Emotion Dataset from public opinion from Twitter. This dataset contains two columns containing tweets and labels. This research uses a multi-label emotion dataset in Indonesian. This dataset has five labels, namely Marah (Angry), Takut (Fear), Senang (Joy), Cinta (Love), and Sedih (Sad). This dataset has a total of 9,480 tweets along with labels.



**Figure 3.** Splitting Dataset

This research uses a dataset of 9,480 text data from the Emotion Dataset in Indonesian which is divided into 5 labels love 1397, anger 2231, sadness 2000, happy 2292, fear 1560. After being combined the dataset is divided into three parts for train, validation, and test of 8:1:1. The amount of data divided into each is 7580 for train data, 948 for validation data, and 948 for test data.

#### 3.2. Percentage Training Result

Emotion classification experiments were carried out using the IndoBERT model. The dataset comes from Tweets that have been labeled. The existing dataset is classified into five labels. The labels consist

of 'Love', 'Angry', 'Sad', 'Happy', and 'Fear'. The control variables in training used in this model consist of the dataset, IndoBERT model, learning rate 1e-6, and batch size 8. In the experimental process, 25 epochs were used by applying an early stop and patient of 5 epochs.

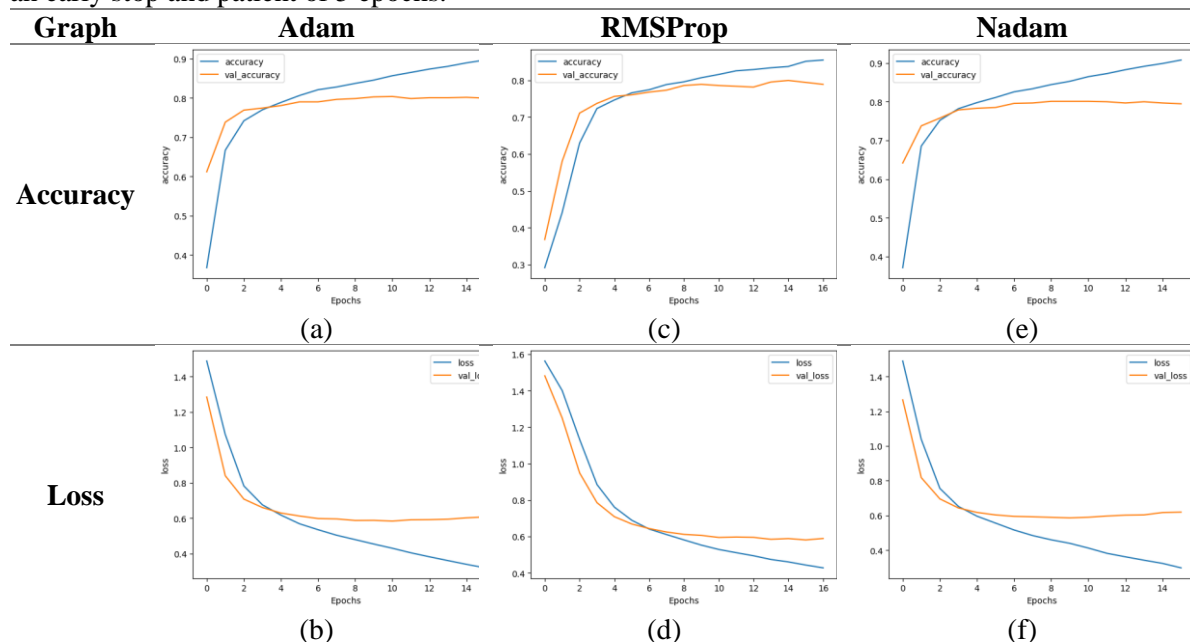
**Table 3.** Comparison Accuracy and Confusion Matrix

Optimizer	Accuracy		Average Confusion Matrix		
	Train	Test	Precision	Recall	F1-Score
<b>Adam</b>	90,21%	80.59%	81%	81%	81%
<b>RMSProp</b>	82.11%	78.37%	79%	79%	79%
<b>Nadam</b>	88.61%	80.90%	80%	80%	80%

In this experiment, the highest accuracy results were obtained on the Adam optimizer. On the other hand, the results of the RMSProp optimizer are more stable compared to other optimizers. In the Average Confusion Matrix results, each optimizer does not have a striking difference because the distance between each optimizer is not far.

### 3.3. Graph Training Result

In the research, training was carried out using the IndoBERT model. The existing dataset is classified into five labels. The control variables in the training used in this model consist of the dataset, IndoBERT model, learning rate 1e-6, and batch size 8. In the process used, 25 epoch training is applied by applying an early stop and patient of 5 epochs.



**Figure 4.** Graph Result (a) accuracy (b) loss

In Figure 4 (a) and (b), the graph of train and validation data using the Adam optimizer (a) and (b) shows signs of over-fitting. The training results show that with the IndoBERT optimizer model Adam obtained quite high accuracy results compared to the others, but the results obtained showed quite a large difference between train and validation accuracy. In this research, the Adam optimizer can show the potential for overfitting, where the model learns too many details from the training data that may not be common and cannot be applied to new data.

Figure 4 (c) and (d), shows that in the graph using the RMSProp optimizer, the results obtained are quite good because the test results obtained can look stable. Although the obtained RMSProp Optimizer training accuracy is lower than Adam's, the difference between training and validation accuracy is smaller. This graph shows the model tends to be more general and perhaps better at handling new data.

This optimizer may be more resistant to overfitting. The RMSProp optimizer as the default optimizer in the IndoBERT model has a striking difference between train and validation accuracy which is smaller. This shows that the RMSProp optimizer tends to be more stable in processing train, validation, and test data.

Figure 4 (e) and (f), shows that the graph using the Nadam Optimizer in this model has the highest accuracy on train, validation, and test data, and the difference between training and validation accuracy is very large. In this re-search, the Adam optimizer can show the potential for overfitting, where the model is too large. learning de-tails from training data that may be unfamiliar and not applicable to new data.

Data overfitting arises because the distance between the training and validation graphs is quite large. As in the Tabel 3, Adam and Nadam Optimizer graph the train graph is growing but the validation graph is stagnant. The results from RMSProp get a fairly close distance difference between the training and validation graphs. If a model is trained too specifically on the training data, it will likely perform poorly on new data because the model has "memorized" patterns that are not generally patterns that can be applied to new data.

#### 3.4. Anomaly Result

In this experiment, several reasons resulted in the data being difficult for the model to classify. Data in the form of sentences can produce different results when processed. Sentences that contain two or more emotions in them can produce different emotions in each process. On the other hand, in the preprocessing process of removing punctuation marks, several sentences have different outputs because semantically the sentences have different meanings.

## 4. Conclusion

Based on the results obtained, determining the right optimizer is done by conducting experiments according to the desired aspects. This research aims to classify five emotions, Marah, Takut, Senang, Cinta, and Sedih and evaluate the performance of three commonly used optimizer, Adam, RMSProp, and Nadam. The processed data used IndoBERT model for Indonesian text classification. The research purpose to search the best optimizer for text classification. This research uses control variables in the training used in this model consisting of dataset, IndoBERT model, learning rate 1e-6, and batch size 8. In the training process, 25 epochs are used by applying an early stop and patient of 5 epochs.

The result shows classification used Adam Optimizer 90,21%, RMSProp Optimizer 82.11, and Nadam Optimizer 88.61%. The Adam optimizer applied to the IndoBERT model yielded the best results. This shows a significant improvement from previous studies, which had emotion classification. There are still many ways to improve the performance of Indonesian Text Processing but Indonesian text data processing resources are still lacking and still need to be developed further. Emotion classification can be further developed with multilabels by processing each emotion in one data.

## References

- [1] A. Tusa Bagus, "KLASIFIKASI EMOSI PADA TEKS MENGGUNAKAN METODE DEEP LEARNING," 2022.
- [2] R. D. Handayani, K. Kusrini, and H. Al Fatta, "Perbandingan Fitur Ekstraksi Untuk Klasifikasi Emosi Pada Sosial Media," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 21, Jul. 2020, doi: 10.30646/sinus.v18i2.457.
- [3] A. S. Aribowo and S. Khomsah, "Implementation Of Text Mining For Emotion Detection Using The Lexicon Method (Case Study: Tweets About Covid-19) Implementasi Text Mining Untuk Deteksi Emosi Menggunakan Metode Leksikon (Studi Kasus: Twit Tentang Covid-19)," *Jurnal Informatika dan Teknologi Informasi*, vol. 18, no. 1, pp. 49–60, 2021, doi: 10.31515/telematika.v18i1.4341.
- [4] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, "Sentiment analysis and opinion mining on educational data: A survey," *Natural Language Processing Journal*, vol. 2, p. 100003, Mar. 2023, doi: 10.1016/j.nlp.2022.100003.

- [5] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," *IEEE Trans Knowl Data Eng*, vol. 35, no. 11, pp. 11019–11038, Nov. 2023, doi: 10.1109/TKDE.2022.3230975.
- [6] M. N. Hoda, Bharati Vidyapeeth's Institute of Computers Applications and Management Delhi, and Institute of Electrical and Electronics Engineers Delhi Section, *NLP & AI Speech Recognition: An Analytical Review*. 2023.
- [7] Z. Li, Y. Wang, Y. Ji, and W. Yang, "A Survey of the Development of Artificial Intelligence Technology," in *2020 3rd International Conference on Unmanned Systems (ICUS)*, IEEE, Nov. 2020, pp. 1126–1129. doi: 10.1109/ICUS50048.2020.9274952.
- [8] W. Hongyuan and D. Mingxing, "OVERVIEW OF THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE TECHNOLOGY," *Int J Res Eng Technol*, vol. 07, no. 08, pp. 92–95, Aug. 2018, doi: 10.15623/ijret.2018.0708011.
- [9] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [10] Y. Yanfi, Y. Heryadi, L. Lukas, W. Suparta, and Y. Arifin, "Sentiment Analysis of User Review on Indonesian Food and Beverage Group using Machine Learning Techniques," in *2022 IEEE Creative Communication and Innovative Technology (ICCIT)*, IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ICCIT55355.2022.10118707.
- [11] S. Saadah, Kaenova Mahendra Auditama, Ananda Affan Fattahila, Fendi Irfan Amorokhman, Annisa Aditsania, and Aniq Atiqi Rohmawati, "Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, Aug. 2022, doi: 10.29207/resti.v6i4.4215.
- [12] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 3, pp. 290–298, Jan. 2024, doi: 10.25077/TEKNOSI.v9i3.2023.290-298.
- [13] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.00677>
- [14] P. Langgeng, W. E. Putra, M. Naufal, and E. Y. Hidayat, "A Comparative Study of MobileNet Architecture Optimizer for Crowd Prediction," *Semarang 123 Jl. Imam Bonjol No*, vol. 8, no. 3, p. 50131, 2023.
- [15] K. Hulliyah, F. Rayyan, and N. S. A. A. Bakar, "Development Of A Chatbot For The Online Application Telegram Chat With An Approach To The Emotion Classification Text Using The Indobert-Lite Method," in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, IEEE, Oct. 2022, pp. 1–4. doi: 10.1109/ICORIS56080.2022.10031483.
- [16] A. Wijaya, "Implementasi Model Neural Network IndoBERT untuk Klasifikasi Berita Ddifabel," Universitas Multimedia Nusantara, Jakarta, 2021.
- [17] K. S. Nugroho and F. A. Bachtiar, "Text-Based Emotion Recognition in Indonesian Tweet using BERT," in *2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, IEEE, Dec. 2021, pp. 570–574. doi: 10.1109/ISRITI54043.2021.9702838.
- [18] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: <https://github.com/tensorflow/tensor2tensor>