ISSN: 2093-5374 / E-ISSN: 2093-5382

Optimizing BERT Models with Fine-Tuning for Indonesian Twitter Sentiment Analysis

Lasmedi Afuan[1*], Nurul Hidayat[2], Hamdani Hamdani[3], Heru Ismanto[4], Brian Cahya Purnama[5], and Dzakwan Irfan Ramdhani[6]

[1*]Associate Professor, Department of Informatics, Universitas Jenderal Soedirman, Indonesia.
lasmedi.afuan@unsoed.ac.id, https://orcid.org/0000-0003-4493-4684

[2]Associate Professor, Department of Informatics, Universitas Jenderal Soedirman, Indonesia.
nurul@unsoed.ac.id, https://orcid.org/0000-0003-1959-4884

[3]Professor, Department of Informatics, Universitas Mulawarman, Indonesia.
hamdani@unmul.ac.id, https://orcid.org/0000-0002-4255-7662

[4]Associate Professor, Department of Informatics, Universitas Musamus, Indonesia.
heru@unmus.ac.id, https://orcid.org/0000-0002-0379-4008

[5]Undergraduate Student, Department of Informatics, Universitas Jenderal Soedirman, Indonesia.
brian.purnama@mhs.unsoed.ac.id, https://orcid.org/0009-0003-2937-9720

[6]Undergraduate Student, Department of Informatics, Universitas Jenderal Soedirman, Indonesia.
dzakwan.ramdhani@mhs.unsoed.ac.id, https://orcid.org/0009-0003-6068-0490

Received: March 07, 2025; Revised: April 11, 2025; Accepted: May 16, 2025; Published: June 30, 2025

Abstract

Twitter has emerged as a critical platform for capturing public sentiment, offering a valuable source for sentiment analysis. This study presents a comparative evaluation of two BERT (Bidirectional Encoder Representations from Transformers) models—baseline and fine-tuned—targeted at analyzing Indonesian-language tweets. Employing the CRISP-DM framework, the methodology encompasses automated data crawling, comprehensive text pre-processing (including case folding, cleaning, tokenisation, normalisation, and data augmentation), and model development using the IndoBERT-base-p1 architecture. The experimental results reveal that the fine-tuned BERT model achieves significantly improved performance over the non-optimized model, with accuracy, precision, recall, and F1-score values reaching 91%, 0.91, 0.90, and 0.91, respectively. These findings indicate the fine-tuned model's superior ability to capture linguistic subtleties and contextual sentiment features within informal social media text. Furthermore, the model is deployed in a web-based application for real-time sentiment classification, demonstrating its practical applicability. This study underscores the effectiveness of Fine-tuning in enhancing BERT-based sentiment analysis for low-resource languages. It highlights its potential for informing decision-making in digital communication, marketing, and policy research contexts.

Keywords: BERT, Fine-tuning, Sentiment Analysis, Social Media, Text Mining, Indonesian Twitter

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 16, number: 2 (June), pp. 248-267. DOI: *10.58346/JOWUA.2025.I2.016*

*Corresponding author: Associate Professor in the Department of Informatics at Universitas Jenderal Soedirman, Indonesia.

1 Introduction

Twitter, as a widely used social media platform, generates large volumes of text reflecting user sentiment and opinion, making it a valuable resource for sentiment analysis. The vast amount of textual data generated on these platforms presents both opportunities and challenges for sentiment analysis. Accurately identifying sentiment expressed in tweets can offer valuable insights for businesses, governments, and researchers (Ghorbanali & Sohrabi, 2023; Qi & Shabrina, 2023). However, the inherent complexity and nuances of natural language, coupled with the informal and often ambiguous nature of social media text, pose significant hurdles for traditional sentiment analysis techniques (Ahmedshaeva et al., 2025).

To overcome the challenges in sentiment analysis, researchers have explored various methods and algorithms, including deep learning models and ensemble techniques. One promising approach is to utilise sophisticated language models, such as BERT, with Fine-tuning methods (Zhang et al., 2020). Although BERT is a powerful model, it has limitations in capturing nuanced and domain-specific characteristics, particularly in tasks such as sentiment analysis, as well as differences in data distribution between BERT training data and sentiment analysis data (Sun et al., 2021). By training on suitable datasets, BERT can comprehend the language configuration and contextual interpretation of sentiment, emotional expressions, sarcasm, and subjective opinions. BERT's ability to understand sentiment, especially in its finer points, is a direct result of its sensitivity to language and context (Biswas & Tiwari 2024). This is only further enhanced through the process of fine-tuning (Pathak et al., 2021). BERT with Fine-tuning can discover more linguistic patterns by training on specific-sentiment datasets in tiny, domain-specific steps. This makes it an even sharper instrument for sentiment analysis on tasks that are, in effect, socials than the version of BERT that simply exists. (Gupta et al., 2024).

Various real-world applications have been subjected to different categorization techniques, and BERT has surfaced as one of the most preferred algorithms. For example, in (Bilal & Almazroi, 2023), they take online customer reviews and classify them as falling into the "valuable" or "worthless" categories using a refined BERT model. By aiming to improve the generalization of the BERT model across contexts, these authors are targeting current machine learning benchmarks in generalizarion (Yang et al., 2024).

Employing a text mining approach, this study aims to compare the baseline BERT model with the upgraded BERT model to determine which one performs better in classifying sentiment in Twitter data. Sentiment classification offers a valuable approach for analyzing unstructured text data, enabling a deeper understanding of its underlying meaning or structure. Thus, this study also serves as a stepping stone to examine BERT more critically in the context of such data. The main focus of the assessment is on the three main evaluation benchmarks: Precision, Recall, and Accuracy. These are used to find the best way to analyze the sentiment of tweets that contain a certain keyword or phrase that the user provides.

2 Research Methodology

The research steps that use CRISP-DM include Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Implementation. To analyze a problem, we need to know how businesses work. To get data, we need to know how to read it. Data preparation is crucial for the manual removal, pre-processing, and annotation of data. Modelling is crucial for data Modelling, word weighting, and classification using the BERT model. Assessment is crucial for verifying the accuracy

of the data through a confusion matrix. Deployment is crucial to ensure the web application operates smoothly using the Python language and the Streamlit framework. The deployment of the web application enables its publication for access and utilization by other users (Rojas & Mejía 2021).

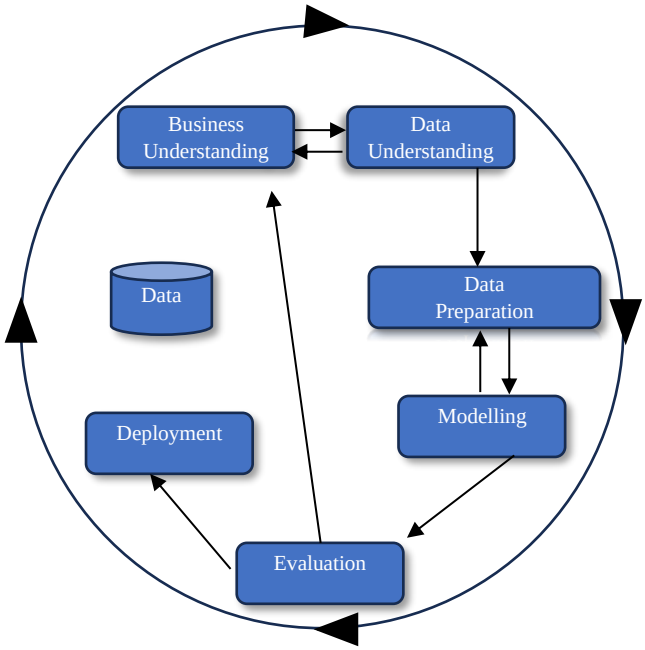

Figure 1: CRISP-DM Model

Figure 1 shows the stages of CRISP-DM (Schröer et al., 2021). This study aims to perform sentiment classification of tweets into positive, negative, and neutral categories by employing two approaches: the baseline BERT model without optimization and the BERT model enhanced through Fine-tuning techniques.

Business Understanding

The initial stage of this research involved identifying and analysing key challenges associated with sentiment analysis on Twitter using text mining approaches. One of the main obstacles lies in the efficiency of data acquisition; previous manual methods were not only time-intensive but also susceptible to human error. To overcome this limitation, the study introduces the implementation of the Tweet-Harvest package from Node Package Manager (NPM), executed via Python. The automatic and real-time crawling method employed in this study enables the speed and scalability of the data collection process, facilitating more accurate and efficient processing of large volumes of tweets. Another major challenge lies in the accuracy of sentiment classification. Traditional methods sometimes miss the small details and contextual meaning that are in social media messages. To address these issues, this study employed IndoBERT, a language model trained on Indonesian text data (Richardson & Wicaksana, 2022). People believe that IndoBERT is better at handling the complexity of the Indonesian language, as it can grasp context more effectively. This makes it easier to do sentiment analysis. In this study, we compare the performance of the standard BERT model with a version of BERT that has undergone fine-tuning. The primary focus is to determine the extent to which fine-tuning can influence the results of sentiment classification, particularly in the context of analyzing Indonesian-language tweets on the Twitter platform.

Data Understanding

The second stage of this study, namely data collection, plays a crucial role. At this stage, Twitter was chosen as the data source because it is a widely used and easily accessible platform. Data collection is carried out automatically using the crawling method, based on keywords entered by users. In this study, the keyword used is "biznet", which refers to one of the leading telecommunications companies in Indonesia.

The focus on Indonesian-language tweets was carried out to ensure that the collected data aligned with the characteristics of the IndoBERT model, which was specifically designed for sentiment analysis in Indonesian. To maintain objectivity, the distribution of sentiment in the dataset was analyzed thoroughly. The results showed that 21.6% of tweets had positive sentiment, 31.1% were neutral, and 47.3% contained negative sentiment. Interestingly, this distribution is quite balanced, although it tends to be more in the neutral and negative categories. This pattern aligns with the findings of previous studies on public perceptions of internet service providers in Indonesia on social media (Srinivasareddy et al., 2025). This relatively balanced distribution supports the use of the keyword "biznet" as a neutral representation in sentiment analysis, thereby avoiding the dominance of overly positive or negative opinions.

The crawling process itself is a common technique in text mining, where content is automatically collected from various online platforms—such as social media, websites, or blogs—based on pre-determined keywords.. In this case, it helped us quickly find a suitable dataset in the right language for further analysis. (Khder, 2021). This technique typically utilizes a script or software tool to locate, extract, and retrieve text-based data that is useful.

For the purpose of this study, tweet data was crawled from Twitter using the NPM twitter-harvest package. This method enables the collection of data in an automatic and efficient manner, saving considerable time and effort compared to manually accomplishing the same task.

Data Preparation

The Twitter data that has been gathered will now undergo several thorough pre-processing steps before it is examined more closely. Once this step is completed, the data will be divided into two groups: one group will consist of the standard BERT model, which will not be optimized, and the other group will consist of the BERT model that has been improved via a fine-tuning procedure. Pre-processing is a crucial step that significantly influences the accuracy of the BERT model's classifications. In this study, the pre-processing procedure is structured around several key steps (Snousi et al., 2022). These are: case-folding, data-cleaning, tokenization, normalization, augmentation, and dataset-splitting. All these steps follow the criteria set forth by BERT with an eye to obtaining classification results that are as good and as accurate as possible. (Bello et al., 2023).

Case Folding

The objective of this step is to convert every letter in the text to lowercase, thereby addressing the inconsistent use of uppercase and lowercase letters that frequently plagues unstructured data. After this procedure, words that should be equivalent (like "Happy" and "happy") are no longer represented as if they were different entities. This is a straightforward way to reduce the unnecessary redundancy that plagues text as it is represented in our computational system (Akbar et al., 2020). In this study, we used Python's lower() method to do our case-folding step. This is a nice, simple, and effective way to reduce

the detrimental effect of inconsistent use of uppercase and lowercase letters on the accuracy of our sentiment categorization task.

Data Cleaning

This is a very important stage in the preparation of social media text for sentiment analysis. The aim of this process is to remove irrelevant or distracting elements, such as mentions, hashtags, and links (URLs). Filtering out parts that do not provide essential information prior to the conduct of the kind of analysis we shall shortly describe significantly enhances the accuracy and reliability of the social media sentiment analysis model. In this study, we removed user mentions (e.g., @username), hashtags (like #topics), and unnecessary links. Without cleaning the data adequately, there was too much "noise" that could have interfered with the model's ability to predict sentiment and to be efficient in capturing true patterns (Chauhan et al., 2024).

Tokenization

Tokenization is the process of dividing a sentence into words or phrases. It is of utmost significance to recognize and detach significant parts known as "tokens." These tokens are employed to figure out the sentiment contained in a text. (Cho et al., 2021).

Normalization

Sentiment analysis has an important step, which is normalization, that directly affects its result. Normalization helps to ensure that text data is uniform and consistent (Mehmood et al., 2020). This process involves several key subtasks or steps, such as: 1. Replacing nonstandard words or slang with standard (e.g., changing "gak" to "enggak"); 2. Correcting misspellings (e.g., "sech" to "sih"); 3. Removing unnecessary symbols, emojis, or inappropriate punctuation; 4. Converting all text to lowercase so that there is no difference in meaning between uppercase and lowercase letters (e.g., "INDONESIA" versus "Indonesia").

Data Augmentation

The back-translation technique is used to enrich the data. It consists of translating materials in the Indonesian language into English and then translating them back into the Indonesian language. (Sujana & Kaos, 2023). Back-translation is executed in two stages: the first is to translate the original text into English, and the second is to translate the English text back into Indonesian. The aim of using back-translation is to create a semi-paraphrased data set. Semi-paraphrased data sets contain variations of the same content, allowing a model that is trained on such data to better generalize.

Split Dataset

The term "biznet" served as the basis for data collection in this study. We then parsed this data into three distinct groups, enabling a more effective analysis and evaluation of the data. To accomplish this, we utilized a sizable 70% chunk of the total collected data to coalesce an ensemble of two BERT models. These incipient models, both trained on our curated dataset, consist of the standard BERT model and a fine-tuned variant. The choice of word in our initial data parsing had both a reason and a rationale. (Alzubaidi et al., 2021).

Moreover, as much as 20% of the data was reserved for validation, which serves to monitor the training process and prevent overfitting (Vabalas et al., 2019). With this generous amount of data, the validation

process also allows for some dynamic model adjustments during training. In our situation, we reserved 10% of the data for testing, and let me tell you, this was done with the utmost regard for an even playing field. We rigorously kept the testing data segregated from the training and validation data so that the model's primary performance could be evaluated in as fair and objective a manner as possible. We used the common go-tos for model evaluation, such as accuracy, precision, recall, and F1-score.

Modeling

The modeling phase is the fourth part of our study, since it is a crucial part of the data processing pipeline. We first explored a method to classify the data using a regular, pre-trained BERT model without any fine-tuning. Then, we looked at a better version of BERT that had been fine-tuned to work better.

BERT without Optimization

As stated by Abdel-Salam and Rafea (2022), BERT is a language transformer model that has been trained a single time and shown to be able to do many different things in the field of Natural Language Processing (NLP). Unlike other multistage, multitask procedures, which for some researchers may seem natural, fine-tuning, the most common way to get the best outcomes for jobs like sentiment analysis, is basically Greene's (2019) and Reimers and Gurevych's (2019) push-button way of achieving prepotency for BERT. Using BERT (either the base or large version) without fine-tuning is like using a well-preserved, long-dead body for some purpose without any animating spirit. It is not really using BERT the way the makers of BERT intended (or expected) us to use it. And yet, this is what Greene (2019) does and what Wang et al. (2021) too seem to do.

BERT with Fine-tuning

Generally, two stages of training are conducted to fine-tune the BERT model for specific tasks. Pre-training is the first stage, in which BERT is trained on a vast amount of text data. This stage is necessary because BERT must first learn the basic structure and common patterns of the aforementioned data, which serves as a foundation for the subsequent and more specialized training stages. Next, the model is adjusted to solve a specific task, such as sentiment analysis, the BERT model is trained to recognize relevant language patterns in the context of the task, applying the concept of tuning (Meng et al., 2023).

Actually, this fine-tuning technique is part of Transfer Learning, where the general knowledge gained in pre-training can be used for a specific task that is smaller and has finite data (Devlin et al., 2019; 2021). However, there are many researchers who argue that BERT has limitations in reasoning and understanding knowledge, the knowledge needed to solve complex problems. That is why the purpose of fine-tuning is not to develop a comprehensive understanding, but to limit overfitting and increase accuracy on specific data. Through this approach, BERT is more reliable in understanding the context of input data and its performance on many advanced tasks is significantly improved (Ko & Choi, 2020; Lamsiyah et al., 2023).

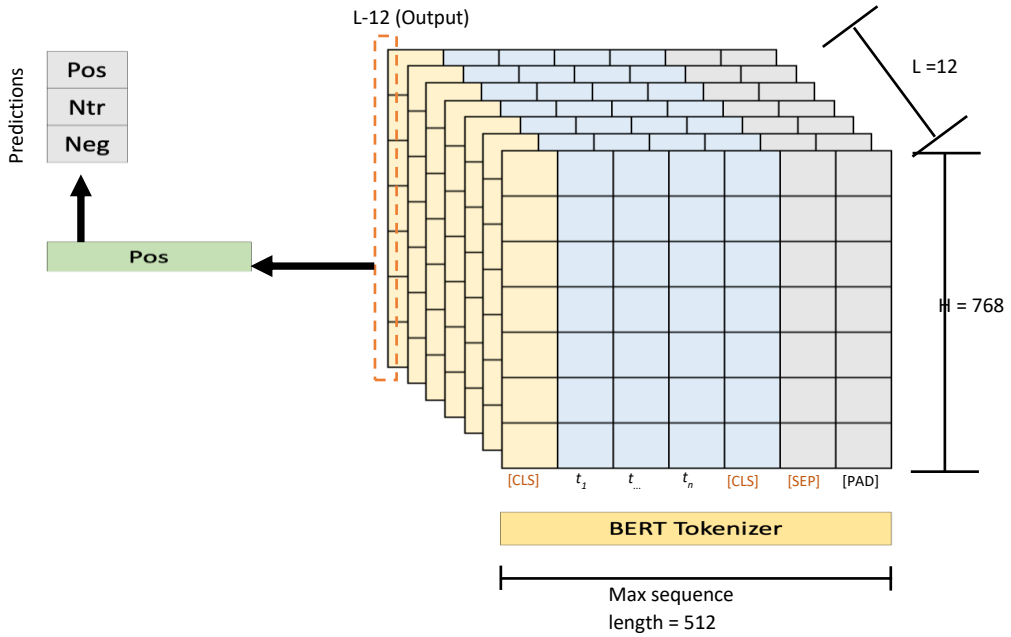

Figure 2: BERT Tokenizer

This study focuses on the IndoBERT-base-p1 model which is a variant of BERT-base. Like any BERT model, the dataset requires pre-processing in the form of tokenization through BertTokenizer, which maps sentences to tokens, as shown in Figure 2. This is essential as the BERT model expects the data in a predetermined format. Each text is tokenized into 512 token fragments. During the fine-tuning process, these tokens are converted into embedding vectors through several layers, namely token embedding, segment embedding, and position embedding. These layers function to embed contextual and positional information of each token within the sentence.

The tokens that have been transformed into embedding vectors are then further processed by BERT's encoder stack, which consists of 12 structurally identical layers. Each layer applies a self-attention mechanism to capture the contextual relationships between tokens within the sentence. As a result, BERT produces vector representations for each token. However, the primary focus is placed on the [CLS] token, which is specifically designed to represent the overall meaning of the entire sentence. This [CLS] token vector is then passed into a classifier, which outputs a logit—a preliminary prediction indicating the likelihood that the sentence falls into one of the sentiment categories: positive, negative, or neutral (Liu et al., 2023). The Softmax function is used to transform the logits into probabilities ranging from 0 to 1. To enhance BERT's performance on specific tasks, several hyperparameters can be adjusted, such as batch size, number of epochs, and learning rate. In this study, the applied configurations include a maximum encoded sequence length of 512, a batch size of 32, eight workers, a learning rate of 5e-5 using the AdamW optimizer, and training conducted over five epochs. By applying fine-tuning techniques, BERT can adapt its previously acquired general knowledge to suit specific tasks, such as sentiment classification, thereby enabling more accurate and optimal performance in these applications.

Evaluation

The fifth stage focuses on evaluating the model that has been implemented. Evaluation is done by conducting tests. In this research, the evaluation is carried out by dividing the dataset into three subsets, namely the training subset, validation subset, and test subset, then, the performance of BERT before optimization with Fine-tuning, and BERT after optimization with Fine-tuning on the dataset will be

evaluated using confusion matrix, such as accuracy, precision, recall, and F1-score (Geetha & Karthika Renuka, 2021).

Table 1: Confusion matrix

True sentiment	Predicted sentiment		
	Positive	Neutral	Negative
Positive	TP	FNt	FN
Neutral	FP	TNt	FN
Negative	FP	FNt	TN

Using the equation (1), precision means figuring out how many positive cases there are.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (1)$$

As shown in equation (2), recall is the calculation of the estimated percentage of positive cases that were correctly identified.

$$Recall = \frac{TP}{TP+FNt+FN} \times 100\% \quad (2)$$

The F1-score is a metric that measures both precision and recall simultaneously. It provides a summary of how well the model can identify positive cases while minimising false positives. It is written down in equation (3).

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \times 100\% \quad (3)$$

Equation (4) shows that accuracy is the ratio of the total number of correct predictions.

$$Accuracy = \frac{TP+TNt+TN}{TP+FP+TNt+FNt+TN+FN} \times 100\% \quad (4)$$

Where: TP : True Positive

TNt : True Neutral

TN : True Negative

FP : False Positive

FNt : False Neutral

FN : False Negative

Deployment

The sixth stage is a key part of this study's implementation. At this point, the Streamlit web app will use the crawling, pre-processing, and best model from the training and evaluation process. Users of this app can type in certain keywords or phrases and get predictions about how people feel about them based on the processed Twitter data. The predictions will show if the sentiment is positive, negative, or neutral. Individuals or groups aiming to perform sentiment analysis on Twitter data can efficiently utilize the results of this research through a web application developed with Streamlit. This application was built using an optimized IndoBERT model and features a well-structured architecture, making it suitable for real-world implementation and replication. The application was developed using the Python programming language, with a user interface (frontend) created using Streamlit and a backend powered by FastAPI. The backend is responsible for running the sentiment analysis model and processing user requests in real-time. Meanwhile, the front end offers a straightforward interface for users to enter

keywords. We use an automated data pipeline to crawl recent tweets with these keywords. Then we pre-process the data and classify the sentiment. The processed results are then displayed clearly and intuitively on the same interface. Users can access the system via the URL https://soeara-sentweet.streamlit.app, enter any Indonesian-language keyword, and view the sentiment distribution of the corresponding tweets. This deployment demonstrates how the proposed model can be applied in real-world applications and serves as a reference for future implementations.

3 Result

This section is a discussion of the research that has been done. Starting from data understanding, data preparation, modelling, evaluation, and deployment.

Crawling Result

The crawling process is done to collect tweet data from Twitter using the keyword “biznet”. The crawling technique, combined with the NPM tweet-harvest package, enables automatic and real-time data collection. From the crawling results in Table 2, a dataset of 3,051 tweet data was obtained which will be used for the training and evaluation process of the sentiment analysis mode.

Table 2: Crawling Result

No	Tweet	Sentiment
1	Everyday Biznet is like this, how can I stream, man? :(	Negative
2	Actually, Biznet is good... if there are no interruptions.	Neutral
3	Thank you to those who came & sorry for the stuttering at the beginning. Biznet is being stupid lately. 2 months ago 12k bitrate was fine, now even 4k sometimes turns red/yellow.	Negative
4	@ibnux It's not just IndiHome, apparently XL Home is exactly the same. Biznet is even worse.	Negative
5	Must be a Biznet user because it's the same.	Neutral
...	...	...
3050	Biznet, represented by Senior Manager of Marketing Biznet Adrianto Sulistyo, received the Indonesia Most Innovative Business Award 2019.	Positive

From the results of crawling 3,051 data using the keyword “biznet”, the distribution of sentiment in the dataset is obtained as follows: 660 positive sentiments (21.6%), 949 neutral sentiments (31.1%), and 1,442 negative sentiments (47.3%).

Case Folding Result

After obtaining the dataset from the crawling results, the first pre-processing stage performed is Case folding. By performing Case folding, the entire text is converted into lowercase letters, as shown in Table 3, thereby eliminating the distinction between uppercase and lowercase letters. This helps improve data consistency and reduce redundancy in the tokenization and sentiment modeling process.

Table 3: Case Folding Result

No	Tweet	Sentiment
1	everyday biznet is like this, how can i stream, man? :(	Negative
2	actually, biznet is good... if there are no interruptions."	Neutral
3	"thank you to those who came & sorry for the stuttering at the beginning. biznet is being stupid lately. 2 months ago, 12k bitrate was fine, now even 4k sometimes turns red/yellow.	Negative
4	@ibnux it's not just indihome, apparently xl home is exactly the same. biznet is even worse.	Negative
5	must be a biznet user because it's the same.	Neutral
...	...	...
3050	"biznet, represented by senior manager of marketing biznet adrianto sulistyo, received the indonesia most innovative business award 2019."	Positive

Data Cleaning Result

The data cleaning stage is the second pre-processing stage performed after the case folding stage. The data cleaning process, as shown in Table 4, aims to remove noise, thereby improving the model's accuracy in sentiment analysis. Additionally, removing mentions, hashtags, and URLs can also help optimise the next tokenisation process and text normalisation.

Table 4: Data Cleaning Result

No	Tweet	Sentiment
1	biznet is like this every day how can we stream bro	Negative
2	actually biznet is good if there are no interruptions	Neutral
3	thank you for coming & sorry for the initial lag. biznet is acting weird a month ago k bitrate was fine now even k sometimes turns red yellow	Negative
4	not just indihome apparently xl home is exactly the same biznet is even worse	Negative
5	it must be a biznet user because its the same	Neutral
...	...	...
3050	biznet represented by senior marketing manager adrianto sulistyo received the indonesia most innovative business award	Positive

Tokenization Result

The next pre-processing stage is tokenization. In Table 5 of the tokenisation results, it is evident that the text has been separated into individual tokens. This separation of text into tokens allows the model to more easily recognize language patterns.

Table 5: Tokenization Result

No	Tweet	Sentiment
1	day, day, biznet, like, this, how, can, you, stream, bro	Negative
2	actually, good, biznet, ., if, it, doesn't, have, issues	Neutral
3	thanks, to, those, who, came, &, sorry, for, the, choppy, start, ., biznet, is, acting, up, ., last, month, k, bitrate, was, fine, now, even, k, sometimes, goes, red, yellow	Negative
4	not, just, indihome, apparently, xl, home, is, the, same, biznet, is, worse	Negative
5	definitely, biznet, users, because, it's, the, same	Neutral
...	...	...
3050	biznet, represented, by, senior, marketing, manager, biznet, Adrianto, Sulistyo, received, the, Indonesian, Most, Innovative, Business, Award	Positive

Normalization Result

In Table 6, we present the results of the text normalisation process. By performing normalization, the text becomes cleaner, structured, and uniform. This helps machine learning models recognise patterns more easily and extract relevant features from the text, thereby improving accuracy in tasks such as sentiment analysis.

Table 6: Normalization Result

No	Tweet	Sentiment
1	every day biznet like this how want to stream bro	Negative
2	actually good biznet. if not disruption	Neutral
3	thank you to those who have come sorry at the beginning a bit broken. biznet is having its issues. last month the bitrate was safe now just sometimes red yellow	Negative
4	not just indihome it turns out xl home same exactly biznet more cruel	Negative
5	must be biznet user because same	Neutral
...	...	...
3050	biznet represented by senior manager marketing biznet adrianto sulistyo received award indonesia most innovative business award.	Positive

Data Augmentation Result

From the initial dataset of 3,051 data points, the amount of data increased to 6,016 data points after augmentation with this technique, as shown in Table 7—the addition of 2,965 data or about 49.3% of the total data. The addition of this data aims to increase data diversity and help the model learn new patterns, thereby improving sentiment classification performance.

Table 7: Augmentation Result

No	Tweet	Sentiment
1	every day biznet like this how want to stream bro	Negative
2	actually good biznet. if not disruption	Neutral
3	thank you to those who have come sorry at the beginning a bit broken. biznet is having its issues. last month the bitrate was safe now just sometimes red yellow	Negative
4	not just indihome it turns out xl home same exactly biznet more cruel	Negative
5	must be biznet user because same	Neutral
...	...	...
6015	why is my biznet. why is my indosat.	Negative

Split Dataset Result

From the dataset of 6,016 data, it was divided into three subsets: the training subset of 4,211 data (70%), the validation subset of 1,209 data (20%), and the test subset of 596 data (10%). The details of the split result are shown in Table 8.

Table 8: Split Dataset Result

Train	Validation	Test
70%	20%	10%
4.211	1.209	596

In the training subset consisting of 4,211 data points, the proportions of positive, negative, and neutral sentiment are 911 (21.6%), 1,991 (47.3%), and 1,309 (31.1%), respectively. In the validation subset of 1,209 data, the proportions of positive, negative, and neutral sentiments are 262 (21.7%), 571 (47.2%), and 376 (31.1%), respectively. Finally, the test subset consisting of 596 data has a distribution of 129 data (21.6%) with positive sentiment, 282 data (47.3%) with negative sentiment, and 185 data (31%) with neutral sentiment. In general, the pattern of sentiment distribution in the three subsets is quite balanced.

Result Analyze

To evaluate the contribution of each preprocessing step, we trained models incrementally and measure d their impact on accuracy using the validation set. The results are summarized in Table 9.

Table 9: the contribution of each pre-processing step

Pre-processing Step	Accuracy (%)
Raw Tweets (no pre-processing)	58.2
+ Case Folding	62.4
+ Cleaning	68.7
+ Tokenization	73.1
+ Normalization	77.5
+ Data Augmentation	80.4

As shown, each step of preprocessing incrementally improves model accuracy, especially normaliz ation and augmentation. After pre-processing the data, configuring the model is essential to achieve optimal performance. This research investigates two configurations: BERT without optimization and BERT with Fine-tuning.

The results of each configuration will be analyzed using metrics such as accuracy, precision, recall and F1 score. By comparing performance metrics across configurations, this research aims to identify the model that achieves the most robust and accurate results in twitter sentiment analysis.

BERT without Optimization

Tests were conducted on the BERT model without optimization using example tweets outside the dataset, as follows:

Text: "wifi ‘biznet’ lambat sekali", Label: neutral (41.313%)

Text: "wifi ‘biznet’ stabil atau tidak?", Label: neutral (40.774%)

Text: "wifi ‘biznet’ cepat dan lancar", Label: neutral (40.166%)

From the three example sentences, it can be observed that the BERT model without optimisation tends to predict all sentences as neutral sentiment, despite the first sentence being negative and the third sentence being positive. The BERT without Optimisation Confusion Matrix can be seen in Table 10.

Table 10: BERT without Optimization Confusion Matrix

True sentiment	Predicted sentiment		
	Positive	Neutral	Negative
Positive	36	534	1
Neutral	25	347	4
Negative	25	236	1

Furthermore, in Table 10. BERT without Optimisation Confusion Matrix: We can see the model's performance on the validation data subset. From the confusion matrix, the results are presented in Table 4—Classification Report on BERT without Optimisation, with an accuracy of only 0.32, or 32%. The precision, recall, and F1-Score values for each sentiment class are also relatively low, as shown in Table 11.

Table 9: Classification Report on BERT without Optimization

	Precision	Recall	F1-Score	Support
Negative	0.17	0.00	0.01	262
Neutral	0.31	0.92	0.46	376
Positive	0.42	0.06	0.11	571
Accuracy			0.32	1209
Macro Avg	0.30	0.33	0.19	1209
Weighted Avg	0.33	0.32	0.20	1209

These results indicate that the BERT model without optimisation is less effective in recognising sentiment patterns in Twitter data. This is because the BERT model used is still general and has not been adapted to the characteristics of Twitter sentiment data. Therefore, it is necessary to optimise the model using fine-tuning techniques to improve its performance in sentiment analysis on Twitter data. Fine-tuning will help the BERT model adjust its weights to the characteristics of Twitter sentiment data, allowing the model to recognise better language and context patterns associated with sentiment expressions.

BERT with Fine-tuning

At this stage, the BERT model, pre-trained on a large dataset, is fine-tuned using a more specific dataset for Twitter sentiment. This adjustment is achieved through several additional training epochs, as shown in Table 12, by further learning the language and context patterns associated with sentiment analysis on Twitter data.

Table 12: Epoch Process on BERT with Fine-tuning

		Accuracy	Precision	Recall	F1-Score
Epoch 1	Train 1	0.76	0.75	0.73	0.74
	Validation 1	0.80	0.80	0.79	0.79
Epoch 2	Train 2	0.91	0.90	0.90	0.90
	Validation 2	0.86	0.87	0.83	0.85
Epoch 3	Train 3	0.96	0.96	0.96	0.96
	Validation 3	0.89	0.89	0.88	0.88
Epoch 4	Train 4	0.98	0.98	0.98	0.98
	Validation 4	0.89	0.88	0.88	0.88
Epoch 5	Train 5	0.98	0.98	0.98	0.98
	Validation 5	0.89	0.89	0.87	0.88

Figure 3 shows the accuracy graph of the BERT model against training and validation data during the Fine-tuning process for five epochs. At epoch 1, the accuracy is still low. However, the accuracy continues to increase significantly in subsequent epochs, both for training and validation data. This indicates that the Fine-tuning process successfully adjusted the model weights to learn sentiment analysis patterns from Twitter data better.

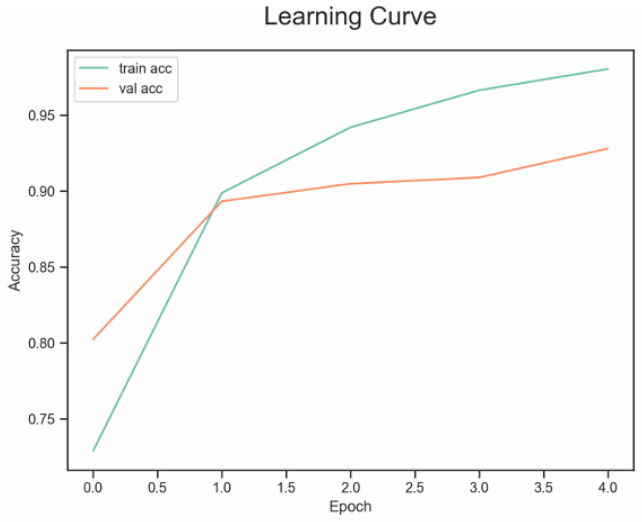

Figure 3: Learning Curve

Tests were also conducted on the BERT model with Fine-tuning using example tweets outside the dataset, as follows:

Text: "wifi ‘biznet’ lambat sekali", Label: negative (99.872%)

Text: "wifi ‘biznet’ stabil atau tidak?", Label: neutral (99.852%)

Text: "wifi ‘biznet’ cepat dan lancar", Label: positive (98.274%)

The test results show that the BERT model, with Fine-tuning, can classify sentiment effectively, taking into account the sentence context. The sentence "WiFi 'biznet' lambat sekali" can be classified by the model as a negative sentiment (poor service). On "wifi ‘biznet’ cepat dan lancar", the model classifies it as positive (good service). This success is attributed to the Fine-tuning process, which enables the model to learn sentiment-related language and context patterns efficiently, resulting in more accurate sentiment classification performance on Twitter data. BERT with Fine-tuning Confusion Matrix Validation Set can be seen on Table 13.

Table 13: BERT with Fine-tuning Confusion Matrix Validation Set

True sentiment	Predicted sentiment		
	Positive	Neutral	Negative
Positive	535	30	9
Neutral	38	329	10
Negative	13	36	209

Table 14: Classification Report on BERT with Fine-tuning Validation Set

	Precision	Recall	F1-Score	Support
Negative	0.91	0.93	0.92	574
Neutral	0.83	0.87	0.85	377
Positive	0.92	0.81	0.86	258
Accuracy			0.89	1209
Macro Avg	0.89	0.87	0.88	1209
Weighted Avg	0.89	0.89	0.89	1209

Results on Validation Data in Table 13. BERT with Fine-tuning Confusion Matrix Validation Set allows us to evaluate the model's performance on the validation data subset after undergoing the Fine-tuning process. Based on Table 14. Classification Report on BERT with Fine-tuning Validation Set, the model accuracy reaches 0.89 or 89%. The precision, recall, and F1-Score values for each sentiment class are also perfect, ranging from 0.81 to 0.93.

Table 15: BERT with Fine-tuning Confusion Matrix Test Set

True sentiment	Predicted sentiment		
	Positive	Neutral	Negative
Positive	272	8	3
Neutral	21	157	8
Negative	9	5	113

Table 16: Classification Report on BERT with Fine-tuning Test Set

	Precision	Recall	F1-Score	Support
Negative	0.90	0.96	0.93	283
Neutral	0.92	0.84	0.88	186
Positive	0.91	0.89	0.90	127
Accuracy			0.91	596
Macro Avg	0.91	0.90	0.90	596
Weighted Avg	0.91	0.91	0.91	596

Results on the subsequent Test Data, in Table 15, BERT with Fine-Tuning Confusion Matrix Test Set, and Table 16, Classification Report on BERT with Fine-tuning Test Set, show the model's performance on a separate subset of test data. There was an increase in accuracy compared to the validation data, from 0.89 to 0.91, or 91%, in the validation data. The precision, recall, and F1-Score values for each class are also still perfect.

The above results show that fine-tuning the BERT model significantly improves its performance in sentiment analysis on Twitter data, surpassing that of BERT without optimisation. Fine-tuning helps the model adjust its weights to the characteristics of Twitter sentiment data, so that it can better recognise language and context patterns associated with sentiment expressions.

The big rise in accuracy, precision, recall, and F1-Score after Fine-tuning shows this. An error analysis was conducted to identify common patterns in misclassification, aiming to understand the model's performance better. The study found that the model often struggled with tweets that were sarcastic, conveyed hidden feelings, or expressed a mix of emotions. The model considered a tweet like "Biznet finally fixed my connection, only took three weeks" neutral, even though it was marked as negative because it was sarcastic. Vague tweets, such as "Biznet is okay, but sometimes annoying," were also often misclassified. These mistakes suggest that the model may not be susceptible to small changes in context and rhetorical nuances, even after Fine-tuning. In future work, these problems could be fixed by adding more advanced sentiment features, sarcasm detection modules, or better integrating outside linguistic resources to help understand complicated phrases in informal social media text

Deployment Result

At this stage, the Streamlit Community Cloud platform, which is a free hosting service from Streamlit, is used to deploy the developed system. The system can be accessed and used by users at the following URL: https://soeara-sentweet.streamlit.app. The deployment process includes putting together the main

parts of the research, such as the data crawling module, the pre-processing methods, and the Fine-tuning optimized BERT model.

4 Discussions

In this study, we compared the BERT model without optimization to the BERT model with fine-tuning optimization for sentiment analysis on Twitter data using text mining techniques. The findings indicate that the BERT model with fine-tuning optimization outperforms the unoptimized BERT model by a substantial margin. The unoptimized BERT model attains a mere 0.32, or 32%, accuracy on the validation dataset. The other classification performance metrics such as precision, recall, and F1-Score for each sentiment category are also relatively low. Thus, the unoptimized BERT model is poorly discriminating sentiment-derived patterns within tweet data. This occurs because the BERT model in question is still a general purpose model and has not been tailored to the specificities of twitter sentiment data.

Conversely, the BERT model's performance was accelerated after it was subjected to optimisation through Fine-tuning techniques. For validation data, the accuracy was 0.89 (89%). Precision, recall, and F1-Score metrics for each sentiment class ranged between 0.81 and 0.93. For the split test data, accuracy further improved to 0.91 (91%), with other evaluation metrics also yielding exceptional results. Such marked improvement suggests that Fine-tuning had successfully optimised the BERT model weights with respect to the Twitter sentiment data. The model's performance was enhanced in recognizing the language and context patterns associated with the sentiment expression.

These results corroborate with those from other studies which demonstrated that optimised BERT models are more effective for sentiment analysis with fine-tuning. As an example, Muhammad Bilal employed BERT with Fine-tuning for classifying customer reviews.

This method was more effective at predicting the usefulness of reviews than traditional machine learning methods. This study does have some good points, however. For example, it analyzes Indonesian Twitter data and utilizes the IndoBERT model, which was trained specifically for Indonesians. In this study, several pre-processing methods were applied, including data augmentation through back-translation, to improve model performance. The findings of this study are particularly relevant for the enhancement of sentiment analysis techniques for Twitter data in the Indonesian language. The application of BERT in sentiment analysis through fine-tuning offers valuable insights to the vast number of users who leverage Twitter data.

5 Conclusion

This research utilized text mining techniques to evaluate the sentiment analysis performance of a BERT model version with fine-tuning against one without. Analysis outcomes demonstrate that the fine-tuned BERT model substantially outperforms the unoptimized version. The BERT model without optimisation exhibits low accuracy, precision, recall, and F1 Score for each sentiment class when evaluated on the validation data. However, after Fine-tuning, these metrics improve significantly. When using the fine-tuned BERT model, the accuracy and other evaluation metrics are improved in the test data. This indicates that the Fine-tuning process was successful, and the BERT model effectively learned to recognize language and context patterns related to sentiment expressions. Fine-tuning helps the BERT model learn how to do Twitter sentiment analysis by moving what it has already learned about general knowledge to the specific task. The model can learn more about the subtleties and context of sentiment expressions through additional training stages, leading to more accurate sentiment classification.

Overall, this study demonstrates that the fine-tuned BERT model is the most effective for analyzing Twitter data, particularly in Indonesian, and that it can provide Twitter users with useful information.

Acknowledgement

The authors would like to thank everyone and every organization that helped with this research. We would like to thank the Institute for Research and Community Service (LPPM) at Universitas Jenderal Soedirman for its very helpful funding and support.

References

[1] Abdel-Salam, S., & Rafea, A. (2022). Performance Study on Extractive Text Summarization Using BERT Models. *Information (Switzerland)*, *13*(2), 1–10. https://doi.org/10.3390/info13020067

[2] Ahmedshaeva, M., Pulatova, N., Khayitov, S., Sufiyeva, D., & Nizomova, M. (2025). The Role of Natural Language Processing (NLP) in Translating Legal Documents with High Accuracy. *Indian Journal of Information Sources and Services*, *15*(2), 83–90. https://doi.org/10.51983/ijiss-2025.IJISS.15.2.12

[3] Akbar, M. R., Slamet, I., & Handajani, S. S. (2020). Sentiment analysis using tweets data from twitter of indonesian's capital city changes using classification method support vector machine. *AIP Conference Proceedings*, *2296*(1), 1–8. https://doi.org/10.1063/5.0030357

[4] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

[5] Bello, A., Ng, S. C., & Leung, M. F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors*, *23*(1), 1–14. https://doi.org/10.3390/s23010506

[6] Bilal, M., & Almazroi, A. A. (2023). Effectiveness of Fine-tuned BERT Model in Classification of Helpful and Unhelpful Online Customer Reviews. *Electronic Commerce Research*, *23*(4), 2737–2757. https://doi.org/10.1007/s10660-022-09560-w

[7] Biswas, D., & Tiwari, A. (2024). Utilizing Computer Vision and Deep Learning to Detect and Monitor Insects in Real Time by Analyzing Camera Trap Images. *Natural and Engineering Sciences*, *9*(2), 280-292. https://doi.org/10.28978/nesciences.1575480

[8] Chauhan, P., Sharma, N., & Sikka, G. (2024). On the importance of pre-processing in small-scale analyses of twitter: a case study of the 2019 Indian general election. *Multimedia Tools and Applications*, *83*(7), 19219–19258. https://doi.org/10.1007/s11042-023-16158-3

[9] Cho, D., Lee, H., & Kang, S. (2021). An empirical study of korean sentence representation with various tokenizations. *Electronics (Switzerland)*, *10*(7), 1–12. https://doi.org/10.3390/electronics10070845

[10] Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NaacL-HLT*, *1*(1), 4171–4186.

[11] Geetha, M. P., & Karthika Renuka, D. (2021). Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. *International Journal of Intelligent Networks*, *2*(1), 64–69. https://doi.org/10.1016/j.ijin.2021.06.005

[12] Ghorbanali, A., & Sohrabi, M. K. (2023). A comprehensive survey on deep learning-based approaches for multimodal sentiment analysis. *Artificial Intelligence Review*, *56*(1), 1479–1512. https://doi.org/10.1007/s10462-023-10555-8

[13] Gupta, V., Patel, J., Shubbar, S., & Ghazinour, K. (2024). COVBERT: Enhancing Sentiment Analysis Accuracy in COVID-19 X Data through Customized BERT. *CS & IT Conference Proceedings*, *14*(2), 171–182. https://doi.org/10.5121/csit.2024.140212

[14] Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications, 13*(3), 144–168. https://doi.org/10.15849/ijasca.211128.11

[15] Ko, B., & Choi, H. J. (2020). Twice fine-tuning deep neural networks for paraphrase identification. *Electronics Letters, 56*(9), 449–450. https://doi.org/10.1049/el.2019.4183

[16] Lamsiyah, S., Mahdaouy, A. El, Ouatik, S. E. A., & Espinasse, B. (2023). Unsupervised extractive multi-document summarization method based on transfer learning from BERT multi-task fine-tuning. *Journal of Information Science, 49*(1), 164–182. https://doi.org/10.1177/0165551521990616

[17] Li, D., Xiong, Y., Hu, B., Tang, B., Peng, W., & Chen, Q. (2021). Drug knowledge discovery via multi-task learning and pre-trained models. *BMC Medical Informatics and Decision Making, 21*(1), 1–8. https://doi.org/10.1186/s12911-021-01614-7

[18] Liu, C., Zhu, W., Zhang, X., & Zhai, Q. (2023). Sentence part-enhanced BERT with respect to downstream tasks. *Complex and Intelligent Systems, 9*(1), 463–474. https://doi.org/10.1007/s40747-022-00819-1

[19] Mehmood, K., Essam, D., Shafi, K., & Malik, M. K. (2020). An unsupervised lexical normalization for Roman Hindi and Urdu sentiment analysis. *Information Processing & Management, 57*(6), 102368. https://doi.org/https://doi.org/10.1016/j.ipm.2020.102368

[20] Meng, Q., Song, Y., Mu, J., Lv, Y., Yang, J., Xu, L., Zhao, J., Ma, J., Yao, W., Wang, R., Xiao, M., & Meng, Q. (2023). Electric Power Audit Text Classification With Multi-Grained Pre-Trained Language Model. *IEEE Access, 11*(1), 13510–13518. https://doi.org/10.1109/ACCESS.2023.3240162

[21] Oliveira, F. B., Haque, A., Mougouei, D., Evans, S., Sichman, J. S., & Singh, M. P. (2022). Investigating the Emotional Response to COVID-19 News on Twitter: A Topic Modelling and Emotion Classification Approach. *IEEE Access, 10*(1), 16883–16897. https://doi.org/10.1109/ACCESS.2022.3150329

[22] Pathak, A., Kumar, S., Roy, P. P., & Kim, B. G. (2021). Aspect-based sentiment analysis in hindi language by ensembling pre-trained mbert models. *Electronics (Switzerland), 10*(21), 1–15. https://doi.org/10.3390/electronics10212641

[23] Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining, 13*(1), 1–14. https://doi.org/10.1007/s13278-023-01030-x

[24] Richardson, B., & Wicaksana, A. (2022). Comparison of Indobert-Lite and Roberta in Text Mining for Indonesian Language Question Answering Application. International Journal of Innovative Computing, *Information and Control, 18*(6), 1719–1734. https://doi.org/10.24507/ijicic.18.06.1719

[25] Rojas, C., & Mejía, L. (2021). Enhancing User Engagement on Web Platforms through Emotion-Aware Interaction Design. *International Academic Journal of Innovative Research, 8*(2), 16–20. https://doi.org/10.71086/IAJIR/V8I2/IAJIR0811

[26] Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science, 181*(1), 526–534. https://doi.org/10.1016/j.procs.2021.01.199

[27] Snousi, H. M., Aleej, F. A., Bara, M. F., & Alkilany, A. (2022). ADC: Novel Methodology for Code Converter Application for Data Processing. *Journal of VLSI Circuits and Systems, 4*(2), 46–56. https://doi.org/10.31838/jvcs/04.02.07

[28] Srinivasareddy, S., Narayana, Y. V., & Krishna, D. (2021). Sector beam synthesis in linear antenna arrays using social group optimization algorithm. *National Journal of Antennas and Propagation, 3*(2), 6–9.

[29] Sujana, Y., & Kao, H. Y. (2023). LiDA: Language-Independent Data Augmentation for Text Classification. *IEEE Access,* 10894–10901. https://doi.org/10.1109/ACCESS.2023.3234019

[30] Sun, C., Yang, Z., Wang, L., Zhang, Y., Lin, H., & Wang, J. (2021). Deep learning with language models improves named entity recognition for PharmaCoNER. *BMC Bioinformatics, 22*(1), 1–16. https://doi.org/10.1186/s12859-021-04260-y

[31] Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE, 14*(11), 1–20. https://doi.org/10.1371/journal.pone.0224365

[32] Wang, X., Chao, F., & Yu, G. (2021). Evaluating Rumor Debunking Effectiveness During the COVID-19 Pandemic Crisis: Utilizing User Stance in Comments on Sina Weibo. *Frontiers in Public Health, 9*, 1–18. https://doi.org/10.3389/fpubh.2021.770111

[33] Yang, C. S., Lu, H., & Qian, S. F. (2024). Fine tuning SSP algorithms for MIMO antenna systems for higher throughputs and lesser interferences. *International Journal of Communication and Computer Technologies, 12*(2), 1-10.

[34] Zhang, L., Fan, H., Peng, C., Rao, G., & Cong, Q. (2020). Sentiment analysis methods for hpv vaccines related tweets based on transfer learning. *Healthcare (Switzerland), 8*(3), 1–18. https://doi.org/10.3390/healthcare8030307

Authors Biography

Lasmedi Afuan is an associate professor in the Department of Informatics at Universitas Jenderal Soedirman in Indonesia. He has a PhD in Computer Science from Gadjah Mada University in Indonesia, where he focused on information retrieval. He is also a researcher, and from 2018 to now, he has been getting money from the Indonesian Ministry of Education, Culture, Research, and Technology. His academic interest includes artificial intelligence, more specifically in machine learning, data mining, information retrieval, and text mining. Currently, he is the head of the Department of Informatics at Universitas Jenderal Soedirman.

Nurul Hidayat is an associate professor in the Department of Informatics at Indonesia's Universitas Jenderal Soedirman. From Gadjah Mada University in Indonesia, he received a PhD in Computer Science with a concentration on information retrieval. He additionally holds the position of a researcher and has been funded by the Indonesian Ministry of Education, Culture, Research, and Technology since the year 2019. His research interests include e-learning. He is the Vice Dean of the Engineering Faculty at Universitas Jenderal Soedirman right now.

Hamdani Hamdani is a Professor at the Informatics Study Program at Universitas Mulawarman in Indonesia. His date of birth is June 6, 1979, and he was born in Muara Bengkal, East Kutai Regency. He received a Diploma in Informatics from Universitas Amikom Yogyakarta in 2000, and subsequently earned a Bachelor's degree in Informatics Engineering from Universitas Ahmad Dahlan, Yogyakarta, in 2002. He completed his Master's degree in Computer Science in 2009 and was conferred a Doctorate in the same field in 2018, both from Universitas Gadjah Mada. He also got the professional engineering title (Ir. and IPM) from Universitas Mulawarman in 2020. Professor Hamdani is interested in Group Decision Support Systems (GDSS), Medical Informatics, and Intelligent Systems. In 2023, he was officially promoted to the highest academic rank, which is Professor in Intelligent Systems. He is involved in academic and research activities that help Indonesia's health informatics and intelligent decision-making systems grow.

Heru Ismanto is an associate professor in the Department of Informatics at Universitas Musamus in Indonesia. He has a PhD in Computer Science from Gadjah Mada University in Indonesia, where he focused on ecision Support Systems, Geographic Information Systems (GIS), and Electronic Government Systems (e-Government). Currently, he is lecturer at the Department of Informatics at Universitas Musamus.

Brian Cahya Purnama is an undergraduate student in the Department of Informatics, Universitas Jenderal Soedirman, Purwokerto, Central Java, Indonesia. He is interesting in text mining and Artificial intelligence.

Dzakwan Irfan Ramdhani is an undergraduate student in the Department of Informatics, Universitas Jenderal Soedirman, Purwokerto, Central Java, Indonesia. He is interesting in text mining and Artificial intelligence.