GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE

Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

Sentiment Analysis of Pertamax on Social Media and MyPertaminta Data Using the IndoBert Algorithm

Dzulfan Yumna Azis[1], Anthony Dewantoro [2], Kumara Galan Pramana [3], Mahazam Afrad[4*], Hari Widi Utomo[5]
[1,2,3,4] Information Systems, Faculty of Industrial Engineering, Purwokerto, Indonesia

[1]dzulfanyumnaazis@student.telkomuniversity.ac.id, [2] anthonydewantoro@student.telkomuniversity.ac.id, [3] kumaragalanp@student.telkomuniversity.ac.id, [4]mahazama@telkomuniversity.ac.id, [5]hariiu@telkomuniversity.ac.id

Submit: 24-04-2026, Revisi: 30-05-2026, Diterima: 15-06-2026, Publikasi: 26-06-2026,

Abstract— **The rapid growth of digital services in Indonesia has accelerated the adoption of online platforms for fuel distribution through the MyPertamina application developed by PT Pertamina. Public responses toward the application and related fuel distribution policies are widely expressed through social media and application reviews. This study aims to analyze public sentiment toward MyPertamina using multi-platform data collected from Google Play Store, Instagram, and Twitter. The research employed a Natural Language Processing approach using the Transformer-based IndoBERT model. The methodology included data collection, data integration, text preprocessing, sentiment labeling, model fine-tuning, performance evaluation, and result visualization. The collected textual data were classified into positive and negative sentiment categories to represent public opinion. Experimental results showed that IndoBERT achieved an accuracy of 92.04%, with balanced precision, recall, and F1-score values. These findings demonstrate that IndoBERT effectively handles unstructured and informal Indonesian text from multiple digital platforms. Overall, integrating multi-platform data with IndoBERT-based sentiment analysis provides comprehensive insights into public perceptions of MyPertamina and supports strategic decisions. Future studies should expand data sources, increase dataset size, compare additional Transformer models, and evaluate broader sentiment patterns across diverse digital environments effectively.**

Keywords: ***IndoBERT, Multi-Platform Data, MyPertamina, Natural Language Processing, Sentiment Analysis.***

I. INTRODUCTION

Crude oil plays a crucial role in the global economy as it influences various macroeconomic variables, including inflation, gross domestic product (GDP), stock market returns, interest rates, and exchange rates. In addition to being a widely traded commodity, crude oil is also a significant factor in portfolio selection, risk management, and option pricing for investors. This strategic role of oil is also evident at the national level, particularly in Indonesia, where crude oil serves as a primary raw material in the production of fuel oil (BBM) that supports a wide range of social and economic activities. Fuel prices have a direct impact on economic stability and often act as a trigger for inflation. With the continuous growth in the number of motor vehicles, fuel oil has become the most used type of energy by the public. This condition makes Indonesia's national fuel consumption one of the highest in Southeast Asia, placing the country fifth in the Asia–Pacific region [1].

To address the high level of fuel consumption in Indonesia, PT Pertamina (Persero) developed the MyPertamina application as a digital service aimed at facilitating fuel purchases for the public. Through this application, users can purchase various types of fuel, such as Pertamax Turbo, Pertamax, Pertalite, and others, online using a smartphone. Its implementation follows region-specific policies, and users are required to register by providing personal information as well as vehicle details in accordance with the type of fuel used [2]. Along with the rapid development of digital technology, sentiment analysis on application-based services such as MyPertamina has become increasingly relevant for understanding public perception. One widely used approach is Natural Language Processing (NLP), a computational method aimed at analyzing and representing human language text so that it can be interpreted by computer systems. Through NLP techniques, sentiments expressed in textual data can be classified into positive or negative categories, thereby facilitating the identification of opinions, trends, and user

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

satisfaction levels toward a particular service. In the context of MyPertamina as a fuel purchasing application and a digital financial service platform at gas stations, the application of NLP plays a crucial role in objectively evaluating user experiences based on the growing volume of reviews and public discussions. [3].

In addition to application reviews, social media platforms such as Twitter (X) and Instagram have become major channels through which the public expresses opinions regarding fuel quality, pricing policies, and services related to Pertamax and PT Pertamina. Unlike application reviews, discussions on social media are generally more spontaneous and reflect real-time public reactions to current issues, including fuel price adjustments, distribution policies, and user experiences. Consequently, sentiment analysis based solely on application review data may not fully capture public perception. Integrating social media data with MyPertamina reviews enables a more comprehensive understanding of public opinion by combining service-related feedback with broader societal discussions regarding Pertamax.

In line with the importance of sentiment analysis in understanding public opinions toward the MyPertamina service, this study asserts that such techniques can effectively assist in grouping textual data into specific sentiment categories. Through this mapping process, MyPertamina stakeholders can more easily evaluate which features or services receive positive appreciation, while simultaneously identifying various complaints or negative feedback from users. The sentiment labeling procedure employed in this study is presented in Table 1.

Table 1. Sentiment Label Determination

Sentiment	Provision
Positive	Includes statements containing praise, recommendations, or users' tendencies to choose and use the MyPertamina application.
Negative	Reflects complaints, negative experiences, usage difficulties, as well as doubtful or skeptical attitudes toward MyPertamina.

Various models can be employed in sentiment analysis, and the selection of an optimal model is influenced by several factors, including data size and quality, the type of sentiment being analyzed, problem complexity, and computational requirements. Based on previous studies, sentiment analysis methods can be broadly classified into two main categories: sentiment dictionary–based approaches and machine learning–based approaches [4]. The sentiment dictionary–based approach relies on lexicon resources containing words with predefined sentiment labels or scores to determine the sentiment orientation of a text. In contrast, the machine learning–based approach employs classification algorithms that learn sentiment patterns from labeled training data to predict the sentiment of unseen text [5].

Machine learning based sentiment analysis methods are implemented by applying various algorithms that are trained on historical data to generate predictions on new, unseen data. This approach generally achieves higher accuracy than sentiment dictionary–based methods; however, its performance strongly depends on the quality of the corpus labeled with sentiment polarity. A corpus refers to a collection of selected textual data that is organized and stored within a database [4]. In sentiment analysis and text classification, two primary approaches are commonly employed. The first is a classifier-based approach that utilizes machine learning techniques to learn sentiment patterns from labeled training data. The second is a sentiment dictionary–based approach, which relies on lexicon resources containing lists of words with predefined sentiment labels or scores to determine the sentiment orientation of the analyzed text [5].

By leveraging the contextual representation capabilities of the IndoBERT model, this study analyzes public sentiment toward Pertamax by integrating textual data collected from Google Play Store, Instagram, and Twitter. This multi-platform approach enables the model to capture public opinions originating from both application reviews and social media discussions, thereby providing a more comprehensive representation of user perceptions regarding Pertamax products and MyPertamina services. Furthermore, the sentiment classification results are analyzed to identify dominant issues discussed by the public, supporting PT Pertamina in evaluating service quality and understanding community responses to fuel-related policies.

II. LITERATURE REVIEW

Sentiment analysis research in Indonesia has shown significant progress through the utilization of Transformer-based models capable of deeply understanding local language contexts. Previous studies have demonstrated the effectiveness of IndoBERT in analyzing user sentiment toward QRIS adoption on TikTok, particularly when handling unstructured social media data [6]. The application of IndoBERT has also been extended to the digital healthcare domain through the analysis of application reviews on Google Play Store,

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

achieving an accuracy of up to 96% [7]. In addition, previous studies have applied IndoBERT using an Aspect-Based Sentiment Analysis (ABSA) approach on student academic portal reviews and achieved an accuracy of 97.3% through hyperparameter optimization. The similarity among these three studies and the present research lies in the use of IndoBERT as the primary model, owing to its strong capability in handling contextual nuances, informal language, and the linguistic complexity of the Indonesian language [8].

Although employing a similar algorithmic approach, this study offers novelty through the integration of diverse data sources and a distinct analysis object. Unlike previous studies that focused on a single platform or specific domain, this research combines data from social media and MyPertamina application reviews to represent public opinion more comprehensively. This approach enables analysis that covers both technical service issues and public perceptions of national energy distribution policies, thereby providing strategic contributions toward improving the quality of digital services in the energy sector with more heterogeneous user characteristics.

IndoBERT offers several advantages over conventional machine learning methods and multilingual language models for Indonesian text processing. Since it is pre-trained on a large Indonesian corpus, IndoBERT can capture linguistic characteristics, contextual meanings, and informal language variations commonly found in social media and application reviews. Previous studies have shown that IndoBERT consistently achieves superior performance in sentiment classification tasks, demonstrating higher accuracy and F1-score compared with several baseline models [9]. These advantages make IndoBERT a reliable and effective model for analyzing public opinion in the Indonesian language[10].

2.1. My Pertamina

MyPertamina is a digital payment application integrated with LinkAja and provided by a state-owned enterprise to facilitate fuel purchase transactions, as well as to offer loyalty point features and monthly fuel expenditure monitoring [5]. The application was developed to support the targeted distribution of subsidized fuel through mandatory registration of personal and vehicle data for subsidy recipients, enabling more accurate monitoring of fuel distribution. In addition, MyPertamina provides an integrated digital payment system, transaction recording, and a nearby gas station search service to enhance user convenience [11].

2.2. Sentiment Analysis

Sentiment analysis encompasses a wide range of tasks, including sentiment extraction, sentiment classification, opinion summarization, review analysis, sarcasm detection, and emotion identification. Since the early 2000s, sentiment analysis has evolved into one of the prominent research areas in Natural Language Processing (NLP). Numerous studies have generally focused on specific aspects of the sentiment analysis process, ranging from task types, technologies and methods employed, levels of analysis granularity, to application domains [12].

2.3. Natural Language Processing

Sentiment analysis covers a broad spectrum of tasks, including sentiment extraction, sentiment classification, opinion summarization, review analysis, sarcasm detection, and emotion recognition. Since the early 2000s, sentiment analysis has emerged as a prominent research field within Natural Language Processing (NLP). Existing studies generally emphasize components of the sentiment analysis pipeline, such as task formulation, applied technologies and methods, levels of analytical granularity, and diverse application domains [13].

2.4. Multi-platform Data (Play Store, Instagram, Twitter)

The utilization of data from multiple digital platforms (multi-platform data) has become an important approach to obtaining a broader and more representative overview of public opinion. Each platform, such as application stores and social media, possesses distinct user characteristics and communication contexts; therefore, integrating multiple data sources can reduce single-platform bias and provide a more comprehensive perspective on the analyzed topic. This approach enables a more dynamic understanding of public opinion compared to analyses that rely on a single data source [14].

Multi-platform sentiment analysis emphasizes the importance of combining communication features from various digital media. Previous studies have demonstrated that integrating data from Google Play Store and Twitter can enrich the context of public opinion through the application of web scraping techniques and machine learning algorithms for sentiment classification. Furthermore, comprehensive reviews across social media platforms such as Instagram and Twitter highlight differences in language style, content structure, and platform-specific features—such as emojis and hashtags—must be accommodated through adaptive Natural Language Processing (NLP) approaches. By combining application reviews with social media data, sentiment analysis can produce a more holistic and relevant representation of public opinion toward a particular entity or service [15].

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

2.5 IndoBERT

IndoBERT is a pre-trained language model based on the Bidirectional Encoder Representations from Transformers (BERT) architecture, specifically developed for the Indonesian language. The model is trained on a large and diverse Indonesian text corpus, including Wikipedia, online news, and web data, enabling it to generate bidirectional contextual representations. This capability allows IndoBERT to understand word meanings more deeply based on sentence context, including syntactic and semantic structures as well as informal language variations commonly found in social media data and application user reviews [16]. The superiority of IndoBERT in sentiment analysis tasks has been demonstrated through benchmarking studies that compare IndoBERT with various NLP models, such as mBERT, XLM-R, CNN, and BiLSTM, using Indonesian digital government service review data. The evaluation results indicate that IndoBERT achieves the best performance with the highest F1-score, highlighting its superior ability to handle unstructured and noisy Indonesian text. These findings confirm that Transformer models trained specifically on the Indonesian language are more effective than multilingual or classical models for sentiment classification, making IndoBERT a relevant and reliable method for public opinion analysis in the Indonesian context [17]. BERT (Bidirectional Encoder Representations from Transformers) is a Transformer-based language model that employs a bidirectional encoder architecture to learn contextual word representations by considering both preceding and succeeding words within a sentence. Unlike traditional sequential language models, BERT utilizes a self-attention mechanism to capture semantic relationships among words, enabling a deeper understanding of sentence context. This architecture has significantly improved the performance of various Natural Language Processing (NLP) tasks, including text classification, question answering, and sentiment analysis, by producing rich contextual embeddings that better represent textual information [18].

2.5.2 Fine-Tuning

Fine-tuning is the process of adapting a pre-trained language model to a specific downstream task by training it on a labeled dataset. During this stage, the pre-trained knowledge acquired by the model is refined to recognize patterns relevant to the target classification problem while preserving its contextual language understanding. In sentiment analysis, fine-tuning enables IndoBERT to effectively classify positive and negative sentiments by optimizing model parameters using task-specific data, resulting in improved classification performance and generalization [19].

2.7. Model Evaluation

Model evaluation is an important stage in machine learning to assess the performance of a classification model. The evaluation is commonly based on a confusion matrix, which compares predicted labels with actual labels. In this study, model performance is evaluated using four metrics: accuracy, precision, recall, and F1-score.

Accuracy measures the proportion of correctly classified instances among all observations and provides an overall indication of model performance. It is calculated using Equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision measures the proportion of correctly predicted positive instances among all predicted positive instances, indicating the reliability of positive predictions. It is calculated using Equation (2).

$$Precision = \frac{TP}{TP + FP}$$

Recall measures the proportion of actual positive instances that are correctly identified by the model, reflecting its ability to detect relevant positive samples. It is calculated using Equation (3).

$$Recall = \frac{TP}{TP + FN}$$

F1-score is the harmonic mean of precision and recall, providing a balanced measure of model performance, especially for classification tasks. It is calculated using Equation (4).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where TP denotes True Positive, TN denotes True Negative, FP denotes False Positive, and FN denotes False Negative [20].

III. RESEARCH METHODS

This study adopts a quantitative approach using Natural Language Processing (NLP) and a Transformer-based IndoBERT model to perform sentiment analysis on public opinions regarding the MyPertamina application. The research workflow consists of several sequential stages, including data collection from multiple digital platforms, data integration, text preprocessing, sentiment labeling, IndoBERT model fine-tuning, model evaluation, and results analysis with visualization. Each stage is designed to ensure high-quality data processing and reliable sentiment classification performance. The overall research methodology is illustrated in Figure 1.

*[4]mahazama@telkomuniversity.ac.id

JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

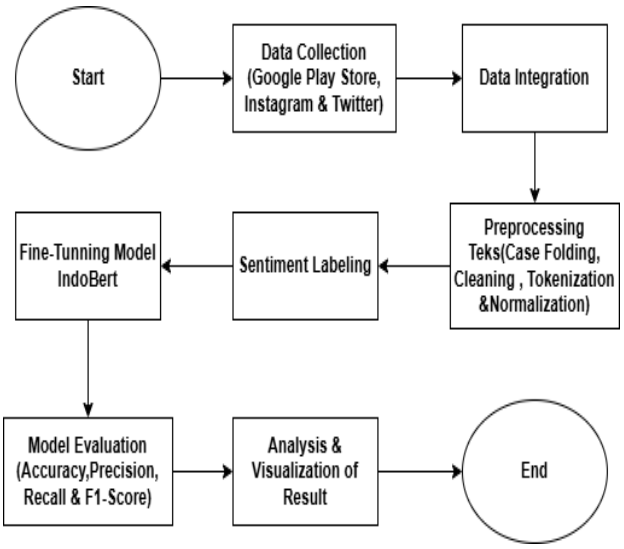

Figure 1. Research Flow

3.1. Data Collection

The initial stage of the study involved collecting textual data related to Pertamina and the MyPertamina application from various digital platforms, namely Google Play Store, Instagram, and Twitter. The collected data consisted of application user reviews, as well as public comments and tweets that represent public opinions regarding services and fuel distribution policies.

3.2. Data Integration

Data obtained from multiple platforms were merged into a unified dataset. This integration process aimed to standardize data formats, remove duplicate entries, and ensure data structure consistency so that the dataset could be processed uniformly in subsequent stages.

3.3. Text Preprocessing

The preprocessing stage was conducted to improve the quality of textual data prior to analysis. This process included case folding (converting all characters to lowercase), data cleaning (removing URLs, mentions, hashtags, symbols, and non-relevant characters), tokenization, and word normalization. These steps are essential for reducing noise and improving the accuracy of the sentiment classification model.

3.4. Sentiment Labeling

The preprocessed textual data were then labeled into two sentiment classes: positive and negative. These labeled data were used as training and testing datasets in the supervised learning process applied to the IndoBERT model.

3.5. IndoBERT Model Fine-Tuning

The pre-trained IndoBERT model was further adapted through a fine-tuning process using the research dataset. This step aimed at enabling the model to recognize sentiment patterns that are specific to the context of the MyPertamina service and national energy policies

3.6. Model Evaluation

The model's performance was evaluated using standard classification metrics, including accuracy, precision, recall, F1-score and Confusion Matrix. This evaluation was conducted to measure the effectiveness of the IndoBERT model in accurately classifying sentiment on the test data.

3.7. Results Analysis and Visualization

The evaluation results and sentiment distribution were further analyzed and presented using visualizations such as graphs or charts. This stage aimed to facilitate result interpretation and identify trends in public opinion toward the MyPertamina service.

IV. RESULTS AND DISCUSSION

4. Data Preprocessing and Dataset After Processing

Before conducting sentiment classification, the collected textual data from Google Play Store, Instagram, and Twitter underwent a preprocessing stage to improve data quality and reduce irrelevant information. Since social media data generally contains noise such as URLs, mentions, hashtags, emojis, punctuation marks, repeated characters, and inconsistent writing styles, preprocessing is essential to produce clean and structured text suitable for model training. In this study, the preprocessing process consisted of case folding, text cleaning, tokenization, stopword removal, and word normalization.

Table 2. Before and after Text Preprocessing

Original Text	Text After Preprocessing
Biasanya orang yang banyak uang itu juga glowing...	[biasa, orang, banyak, uang, glowing, tidak, ...]
Mau download QRCode di app Pertamina masa gagal...	[download, qrcode, aplikasi, pertamina, masa, gagal, ...]
Aplikasi pemerintah selama ini gak ada yang bener...	[aplikasi, perintah, lama, tidak, benar, ...]
Gimana mesin GK cepat rusak, tanki berkarat...	[bagaimana, mesin, tidak, cepat, rusak, tangki, ...]

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

Original Text	Text After Preprocessing
Barcode di scan tidak bisa, petugas Pertamina...	[barcode, scan, tidak, bisa, petugas, pertamina, ...]

Through these preprocessing stages, each text was transformed into a standardized representation while preserving its semantic meaning. This process enables the IndoBERT model to focus on meaningful contextual information instead of irrelevant textual components, thereby improving sentiment classification performance. Table 2 presents several examples of textual data before and after preprocessing.

4. Distribution Labeling

After completing the preprocessing stage, each text instance was assigned a sentiment label to prepare the dataset for supervised learning. The labeling process categorized the textual data into two sentiment classes, namely positive and negative, based on the overall opinion expressed toward Pertamina and the MyPertamina application. Positive labels represent comments expressing satisfaction, appreciation, or support, whereas negative labels indicate complaints, criticism, dissatisfaction, or problems experienced by users.

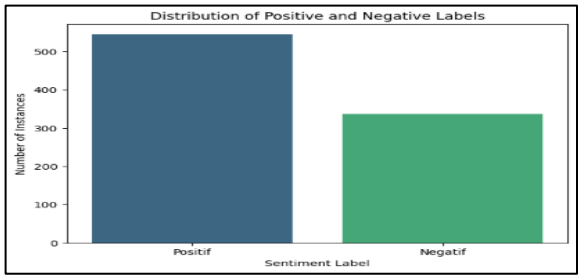

Figure 2. Distribution Of Positive and Negatif Labels

Figure 2 illustrates the distribution of sentiment labels in the dataset. The dataset consists of 547 positive instances and 337 negative instances, indicating that positive sentiment is more dominant than negative sentiment. This distribution suggests that, although users generally express favorable opinions regarding Pertamina and its digital services, a considerable number of negative comments still exist, primarily related to application performance, transaction issues, and service quality.

The presence of both sentiment classes provides sufficient variation for training the IndoBERT model. Although the dataset is moderately imbalanced, the difference between the two classes is not excessive, allowing the model to effectively learn sentiment patterns while maintaining stable classification performance during the fine-tuning process.

To illustrate the labeling process, several examples of the processed textual data are presented in Table 3. Each text was manually assigned a sentiment label according to the expressed opinion. Positive labels were assigned to comments containing appreciation, satisfaction, or support for Pertamina and the MyPertamina application. In contrast, negative labels were assigned to comments expressing dissatisfaction, complaints, criticism, or technical problems experienced during the use of the application or related services.

Table 3. Examples of Preprocessed Text with Sentiment Labels

Processed Text	Text
[biasa, orang, banyak, uang, glowing, tidak, ...	Positive
[download, qrcode, aplikasi, pertamina, masa, gagal, ..	Negative
[barcode, salahgunakan, sama, orang, tidak, ...	Positive
[aplikasi, sempurna, tidak, susah, guna, ...	Positive
[bagaimana, mesin, tidak, cepat, rusak, tangki, ...	Negative

Based on Table 3, the processed dataset consists of text that has undergone a series of preprocessing steps, including case normalization, cleaning, tokenization, stop-word removal, and stemming. As a result, each document is represented as a sequence of meaningful root words, while irrelevant elements such as punctuation, numbers, URLs, emojis, and common stopwords have been removed. This preprocessing reduces textual noise and allows the model to focus on semantically important terms during the sentiment classification task. The examples presented in the table show that each processed text is paired with the corresponding sentiment label, either positive or negative. Positive examples generally contain words indicating user satisfaction or supportive opinions, while negative examples include terms related to application issues, transaction failures, or service complaints. The processed dataset serves as input for the IndoBERT model during the fine-tuning phase, enabling the model to learn contextual representations of

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

Indonesian text more effectively. Therefore, the preprocessing stage plays a crucial role in improving the quality of the input data and contributes to the overall performance of the sentiment classification model.

4. Hyperparameter

To optimize the performance of the IndoBERT model, several hyperparameters were configured during the fine-tuning process. The model was trained using a learning rate of 2×10^{-5}, a batch size of 16 for both training and evaluation, a weight decay of 0.01, and 5 training epochs. Model evaluation was performed at the end of each epoch using the validation dataset, while the model with the best performance was automatically selected based on the highest F1 score. The model's performance during the training process is presented in Table 4.

Table 4.Training Results of the IndoBERT Model Across Five Epochs

Epoch	Training Loss	Validation loss	Accuracy	F1
1	No log	0.336484	0.840909	0.875000
2	No log	0.316885	0.897727	0.921739
3	0.342086	0.368309	0.920455	0.938596
4	0.342086	0.358080	0.920455	0.938053
5	0.038718	0.364172	0.920455	0.938053

Based on Table 4 the IndoBERT model showed consistent performance improvement during the early stages of training. The validation accuracy increased from 84.09% in the first epoch to 89.77% in the second epoch and reached 92.04% in the third epoch. Similarly, the F1 score increased from 0.8750 to 0.9385, indicating that the model became increasingly effective at classifying both positive and negative sentiments. Although the validation loss value increased slightly after the third epoch, accuracy and the F1 score remained stable at 92.04% and around 0.938, respectively, during the fourth and fifth epoch. This behavior suggests that the model had largely converged by the third epoch, and additional training epochs yielded only marginal improvements. Furthermore, the stable evaluation metrics indicate that the selected hyperparameter configuration allowed the model to learn robust semantic representations without significant overfitting. Overall, these results demonstrate that the chosen hyperparameter settings enable the IndoBERT model to achieve stable performance on the sentiment analysis task. The highest F1-score of 0.9385 was achieved in the third epoch, indicating that this checkpoint provides the best balance between classification accuracy and generalization performance on the validation dataset.

4.. Model Evaluation Results

The performance of the IndoBERT model was evaluated using a test dataset consisting of 177 samples, classified into two sentiment classes: negative and positive. The evaluation was conducted after the model was fine-tuned using several performance metrics, including accuracy, precision, recall, and F1-score. Based on the experimental results, the IndoBERT model achieved an overall accuracy of 92.04%, indicating that it is capable of accurately classifying public sentiment regarding MyPertamina. These results demonstrate that IndoBERT possesses strong generalization capabilities in recognizing linguistic patterns from unstructured Indonesian text collected from social media and app reviews. Detailed classification performance is presented in the classification report. For the negative sentiment class, the model achieved a precision of 0.96, a recall of 0.82, and an F1-score of 0.89, based on 67 samples. The relatively lower recall for the negative sentiment class indicates that a small number of negative sentiments were misclassified as positive. This may be due to the presence of ambiguous expressions, veiled criticism, or contextual language that is more difficult for the model to distinguish. In contrast, the positive sentiment class achieved a precision of 0.90, a recall of 0.98, and an F1-score of 0.94, based on 109 samples.

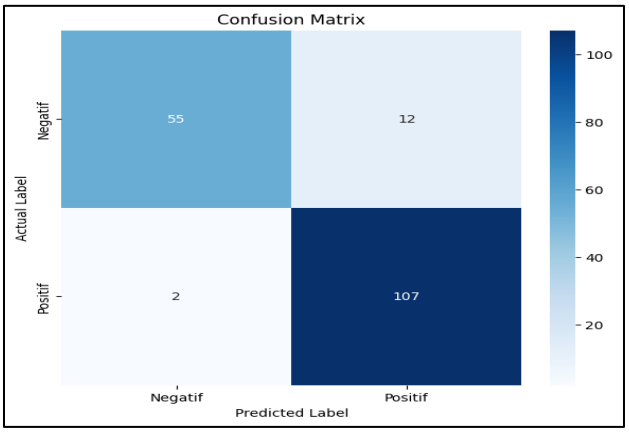

Figure 3. Confusion Matrix

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE
Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

Figure 3 illustrates the confusion matrix of the proposed IndoBERT model on the test dataset. The model successfully classified 55 out of 67 negative sentiment samples and 107 out of 109 positive sentiment samples correctly, resulting in a total of 162 correctly classified examples. Meanwhile, 12 negative samples were incorrectly predicted as positive, while only 2 positive samples were misclassified as negative. The confusion matrix shows that this model performs better at identifying positive sentiment than negative sentiment. The relatively larger number of false positives (12 instances) indicates that some negative opinions contain linguistic expressions like those of positive sentiment, making them more difficult for the model to distinguish. Such cases are common in social media text, where users often employ implicit criticism, sarcasm, or mixed opinions. On the other hand, the very small number of false negatives (2 cases) indicates that IndoBERT is highly effective at recognizing positive sentiment expressed by users. These findings are consistent with the high precision and recall values obtained for the positive class in the classification report. Overall, the confusion matrix confirms that the proposed model produces a low classification error rate and is capable of accurately distinguishing public sentiment toward Pertamax and the MyPertamina app.

Overall, the macro average F1-score of 0.88 indicates that this model demonstrates balanced performance across both sentiment classes without favoring one over the other. Additionally, the weighted average F1-score of 0.92 shows that this model maintains stable performance despite a slight imbalance in the class distribution. Overall, the accuracy, classification report, and confusion matrix show that the proposed IndoBERT model is capable of effectively classifying sentiment in Indonesian while maintaining consistent performance across various evaluation metrics.

4. Word Cloud of Processed Text

After all the data went through the preprocessing stage which included case folding, data cleaning, tokenization, stopword removal, and stemming it was visualized using a word cloud. This visualization aims to provide an overview of the words that appear most frequently in the preprocessed dataset. In a word cloud, the size of a word represents its frequency of occurrence; thus, the larger a word is, the more frequently it appears in the data. This visualization helps identify the dominant topics discussed by users before sentiment classification is performed using the IndoBERT model.

Figure 4. Wrold Cloude

Based on Figure 4, the word “Pertamina” is the largest, indicating that it is the most frequently occurring word in the dataset. This aligns with the focus of the study, which analyzes public sentiment toward Pertamax and the MyPertamina app. In addition, words such as "pakai", "aplikasi", "daftar", "baik", "guna", "masuk", "BBM", and "SPBU" also have high occurrence frequencies. These words indicate that most discussions relate to the use of the MyPertamina app, the registration and login processes, and fuel purchase transactions at gas stations. In addition to these words, there are also terms like "error", "login", "verifikasi", "lama", and "sulit", suggesting that many users discussed technical issues encountered while using the app. On the other hand, the appearance of words such as "baik", "bagus", "cepat", and "mudah" indicates positive experiences shared by some users. However, the word cloud visualization only depicts word frequency and does not directly show sentiment polarity. Therefore, determining whether an opinion is positive or negative is done during the classification stage using the IndoBERT model.

V. CONCLUSION

This study conducted a sentiment analysis of public opinion regarding the MyPertamina app and the Pertamina

*[4]mahazama@telkomuniversity.ac.id

GOVITA
JOURNAL GOVERNANCE IT ADOPTION AND TECHNOLOGY ADVANCE

Volume 1, (2) 2026, p. 61-69
E-ISSN: 3124-7180
DOI: 10.25124/govita.v1i2.11679

entity using multi-platform digital data. The Transformer-based IndoBERT model proved effective in handling unstructured and informal Indonesian text, which is commonly found on social media and in app reviews. Experimental results show that IndoBERT achieved an accuracy of 92.04% with a balanced F1 score between positive and negative sentiment classes, indicating strong classification performance. These findings demonstrate that the proposed approach successfully achieved the research objective of accurately classifying public sentiment while providing a more comprehensive understanding of user opinions through multi-platform data integration. Thus, these research results can support strategic decision-making to improve digital services and national energy distribution policies. Despite these promising results, this study has several limitations. The dataset is limited to specific digital platforms and binary sentiment classification, which may not fully capture the diversity and complexity of public opinion. Therefore, future research is recommended to incorporate additional data sources, explore aspect-based or multi-class sentiment analysis, and compare IndoBERT with newer transformer-based language models. These improvements are expected to provide deeper insights into public sentiment and further enhance the performance and implementation of sentiment analysis systems.

REFERENCES

[1] M. V. Santos, F. Morgado-Dias, and T. C. Silva, "Oil Sector and Sentiment Analysis—A Review," Jun. 01, 2023, *MDPI*.

[2] D. Pratmanto, A. Ardiansyah, and S. Aji, "Sentiment Analysis on Fuel Purchase Policy Through MyPertamina Application Using NB and SVM Methods Optimized by PSO as Weight Optimation," 2023. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[3] F. Y. A'la, "Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 5, no. 1, pp. 80–91, Nov. 2022, doi: 10.20895/inista.v5i1.838.

[4] A. W. V. Hutabarat, N. L. S. S. Adnyani, and K. Suryadi, "Analisis Sentimen Data Ulasan Pengguna MyPertamina di Twitter dengan Metode Text Mining," *Jurnal Rekayasa Sistem Industri*, vol. 13, no. 1, pp. 145–154, Apr. 2024, doi: 10.26593/jrsi.v13i1.6958.145-154.

[5] *International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2021.

[6] E. Supriyadi and P. N. Makatita, "Sentiment Analysis of TikTok User Comments on QRIS Adoption in Indonesia Using IndoBERT," *Procedia Comput. Sci.*, vol. 269, pp. 121–130, 2025,

[7] H. Imaduddin, F. Yusfida A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach." [Online]. Available: www.ijacsa.thesai.org

[8] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Applied Sciences (Switzerland)*, vol. 15, no. 1, Jan. 2025.

[9] A. A. Sinurat *et al.*, "MyPertamina Application To Increase Consumer Engagement," 2022.

[10] J. Cui, Z. Wang, S. B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 8469–8510, Aug. 202.

[11] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023.

[12] M. P. Ningrum, R. Mutia, H. Azmi, and H. D. Khalifah, "Sentiment Analysis of Twitter Reviews on Google Play Store Using a Combination of Convolutional Neural Network and Long Short-Term Memory Algorithms," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 2, pp. 107–115, Jan. 2025.

[13] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos, and P. M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," Aug. 01, 2023, *Elsevier Ltd*.

[14] Dhendra and V. Gayuh Utomo, "Benchmarking IndoBERT and Transformer Models for Sentiment Classification on Indonesian E-Government Service Reviews," *Jurnal Transformatika*, vol. 23, no. 1, pp. 86–95, Jul. 2025, doi: 10.26623/transformatika.v23i1.12095.

[15] 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

[16] 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

[17] R. Budianoor, S. W. Saputro, F. Abadi, R. A. Nugroho, and A. Farmadi, "Quantifying the Impact of Text Preprocessing on IndoBERT Fine-Tuning for Indonesian Informal Culinary Sentiment Analysis," *Journal of Computing Theories and Applications*, vol. 3, no. 4, pp. 564–581, May 2026.

[18] M. A. Nur, N. Umar, Z. Feng, and H. Gani, "Evaluation Of Indobert And Roberta: Performance Of Indonesian Language Transformer Models In Sentiment Classification," *JIKO (Jurnal Informatika dan Komputer)*, vol. 8, no. 2, pp. 121–127, Jul. 2025,

*[4]mahazama@telkomuniversity.ac.id