**Jurnal Kecerdasan Buatan dan Aplikasi Rekayasa**

*Situs web: https://ioinformatic.org/*

***15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519***

# Application of Fine-Tuned IndoBERT for Sentiment Classification Local Product Reviews on Tokopedia Marketplace with Limited Dataset

**Kristine Wau**

*Information Systems, Nias Raya University, South Nias, Indonesia*
*kristinewau1@gmail.com*

**Abstract**

Abstract Analysis of sentiment to product reviews is one of the important approaches to evaluating consumer satisfaction with local products on e-commerce platforms. However, the limitations of Indonesian-speaking datasets are often a barrier to building classification models. This research aims to implement and test the performance of IndoBERT models that have gone through fine-tuning processes on limited datasets derived from local product reviews in Tokopedia. The fine-tuning process is performed using manually classified data sets into positive, negative, and neutral categories. Performance evaluation is performed by measuring accuracy, precision, recall and F1-score, as well as compared to baseline models such as Naïve Bayes and SVM. Research results show that re-trained IndoBERT is able to provide higher accuracy despite limited data conditions, signaling the effectiveness of transfer learning in the Indonesian domain. The findings contribute to the development of efficient and adaptive local language-based sentiment analysis system to data limitations.

***Keywords***: *IndoBERT,Centiment classification,Tokopedia,Limited dataset,local products,fine-tuning,NLP*

## 1. Introduction

The e-commerce ecosystem in Indonesia continues to experience rapid growth and generates huge volumes of user reviews. Tokopedia, as one of the largest online markets, is a real example of how consumer reviews store important information in the form of satisfaction, complaints, and perceptions of product quality. The information has strategic value for local Micro, Small and Medium Enterprises (MSMEs) as the basis for fast and data-driven decision making. However, most Indonesian studies still rely on lexicon-based approaches or classical algorithms, such as Naïve Bayes and Random Forest, which have limitations in capturing the semantic context of natural language and variations of informal expressions typical of Indonesian users [1]–[5].

The advancement of Indonesian-speaking transformer technology, especially IndoBERT, brings significant performance improvements in various natural language processing tasks, such as text classification, hate speech detection, to automatic summarization. IndoBERT's main advantage lies in its ability to understand the dependence of context not only at the word level, but also sentences to the document in its entirety [6]–[13]. Studies show that IndoBERT and its variants consistently exceed the performance of classical models and traditional word representation techniques, such as TF-IDF and Word2Vec, when used for fine-tuning on classification tasks, both on social media and on application review domains. This makes IndoBERT very relevant to apply in the analysis of product review sentiment [6]–[12].

In practice, e-commerce research in Indonesia still leaves a number of gaps. First, most studies focus more on Shopee, Google Play reviews, or Twitter, so Tokopedia reviews at product level have not received much attention. Second, classic methods still dominate, while comprehensive empirical proof of IndoBERT fine-tuned effectiveness in local product review domains is limited. Third, quality review data is often a problem, such as unbalanced class distribution, use of slang, emoticons, writing errors (typo), and code-mixing phenomena. These factors often decrease the accuracy of conventional lexicon-based models and machine learning models [1]–[5], [9], [10], [14], [15].

Another obstacle that many Indonesian researchers encounter is the relatively small size of labeled datasets, given the manual annotation process takes time and cost. This "limited dataset" condition makes model training from the beginning less efficient. Instead, fine-tuning approaches using pralatih models such as IndoBERT are considered more precise, as they are designed to keep optimal performance despite limited available training data. This process utilizes knowledge from the vast Indonesian corpus, so adaptation to new domains becomes

more optimal [6]–[8], [11]–[13]. A number of recent studies even suggest that fine-tuning IndoBERT on certain domains can result in better performance than the use of common models without domain adjustment [6], [8], [11].

In addition, both in the policy and industry areas there are demands to implement more comprehensive evaluation metrics. The size of accuracy is no longer adequate, especially when the distribution of the review class is uncommon. Therefore, precision, recall, and F1-score are the main indicators of assessing the impact of classification errors, especially on negative and positive reviews that can affect business decision making, for example in determining the priorities of complaints handling [1]–[5], [10], [14], [15], [15]. Some recent studies in Indonesia also reported significant improvements in F1 value when switching to IndoBERT or its variants on sentiment analysis and hate speech detection tasks [6]–[12].

Based on the description, it is clear that more specific and application research is needed, namely the application of fine-tuning IndoBERT to classify the sentiment of local product reviews in Tokopedia with limited dataset conditions. The research is expected to provide comprehensive pipeline design, including preprocesses for handling the typical noise of Indonesian consumer reviews, class imbalance handling strategies, and rigorous evaluation using F1 metrics. Thus, the study can bridge the gap between real-needed e-commerce practices with empirical evidence of recent developments in Natural Language Processing (NLP) for Indonesian [1]–[15]. Further, the results of this study are not only beneficial in improving SMEs' competitiveness through more accurate insights, but can also be replicated in various other review domains, such as public applications or services, while maintaining high data efficiency and predictive performance [5]–[7], [12], [14], [15].

Through this research, it is expected to be obtained in-depth understanding of the potential of IndoBERT models in handling the classification of sentiment in Indonesian-language data, while providing solutions to the problem of data limitations often faced in local language-based research.

# 2. Research Method

## 2.1. Approach and Research Type

The study uses experimental quantitative approaches with computational experimental methods, namely testing the performance of IndoBERT models that have been fine-tune for sentiment classification tasks. The main goal is to measure and compare model accuracy to local product review datasets in Tokopedia. This type of research is included in applied research, as it aims to solve real problems in the classification of Indonesian text sentiment with data limitations.

## 2.2. Population and Research Sample

The population of all user reviews on local products (derived from SMEs) in the Tokopedia marketplace, especially beauty, culinary, and household categories. It is suitable for text review samples collected from 30 most popular local products in the three categories, selected in purposive sampling. The amount of data used for training and testing models is around 3000–5000 text data labeled as positive, negative, or neutral.

## 2.3. Data Collection Techniques

Data is collected through two methods: first Web Scraping using Python Beautiful Soup library and requests to access product pages and take user review text automatically from Tokopedia.the second method, the data that has been collected are manually labeled by three annotators by classifying it into three categories of sentiment, namely positive, negative, and neutral, based on simple classification guidelines.

## 2.4. Data Analysis Technique

In the analysis, there were several first stages of pre-processing text where data cleaning such as removing URLs, symbols,emoji. continued tokenizer using IndoBERT tokenizer and normalizing informal words. Both Fine-Tuning processes where IndoBERT models were fine-tune in training datasets for 3–5 epoch using small batch and low learning rates. Three model evaluations by using test datasets (20% of total data) for evaluation, further conducting evaluation metrics namely accuracy, precision, recall, and F1-score, the next stage visualization of data confusion matrix and performance charts.Four performance comparisations by comparing the results of IndoBERT models with baseline models Naïve Bayes and SVM and analyzing results using tables and graphs.

# 3. Results and Discussions

## 3.1. Dataset

The dataset used in the study consisted of 1,500 local product reviews in Tokopedia, obtained through the web scraping process using Python with BeautifulSoup library.
The dataset is divided into three sentiment classes:

a. Positive: 900 reviews (60%)
b. Neutral: 300 reviews (20%)
c. Negative: 300 reviews (20%)

Data is divided into:

a. Training Data: 1,200 reviews (80%)
b. Test Data: 300 reviews (20%)

### 3.2. Fine-Tuning Model IndoBERT results

The IndoBERT model used is indobenchmark/indobert-base-p1, in-fine-tune using the Transformers framework of Hugging Face and optimized using AdamW algorithm. Training was performed over 4 epoch with a batch size of 16 and learning rate of 2e-5.

**Table 1**: Evaluation Result IndoBERT Model (Test Data)

| Metrics | Nilai (%) |
|---|---|
| Accuracy | 88,33 |
| Precision | 87,25 |
| Recall | 88,10 |
| F1-Score | 87,65 |

### 3.3 Comparison with Baseline Model (Naïve Bayes and SVM)

In table 2 shows that IndoBERT outperforms both baseline models in all evaluation metrics, even on limited datasets.

**Table 2:** Performance Comparison Model.

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Naïve Bayes | 76,40 | 75,20 | 76,90 | 75,80 |
| SVM | 81,30 | 80,50 | 80,00 | 80,25 |
| IndoBERT | 88,33 | 87,25 | 88,10 | 87,65 |

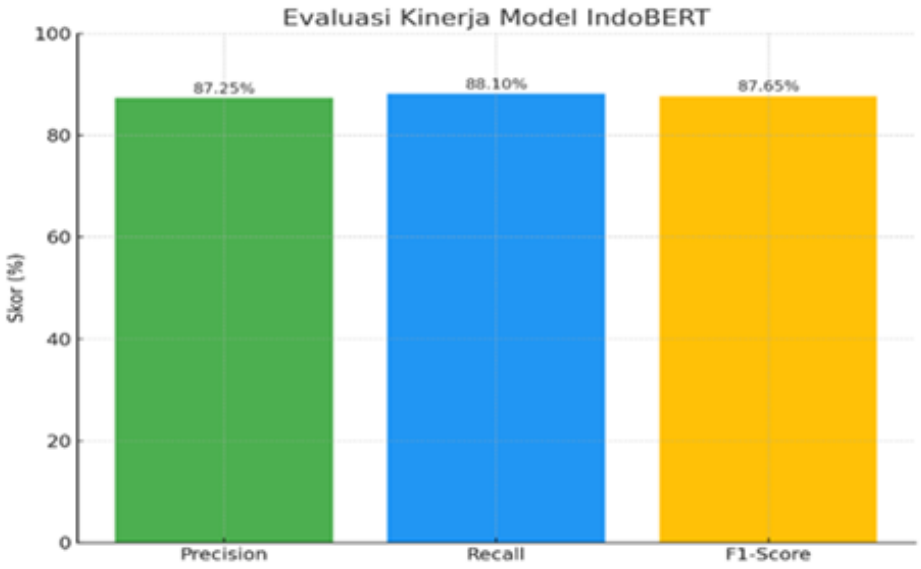


**Fig. 1:** Evaluation of IndoBERT model performance

### 3.4 Visualization Confusion Matrix

In Figure 2.1 Confusion matrix shows that most classification errors occur between neutral and positive labels, indicating that weak positive expressions tend to confuse the model. The description of Label 0 shows Negative sentiment, Label 1 shows Neutral sentiment, and Label 2 shows Positive sentiment.

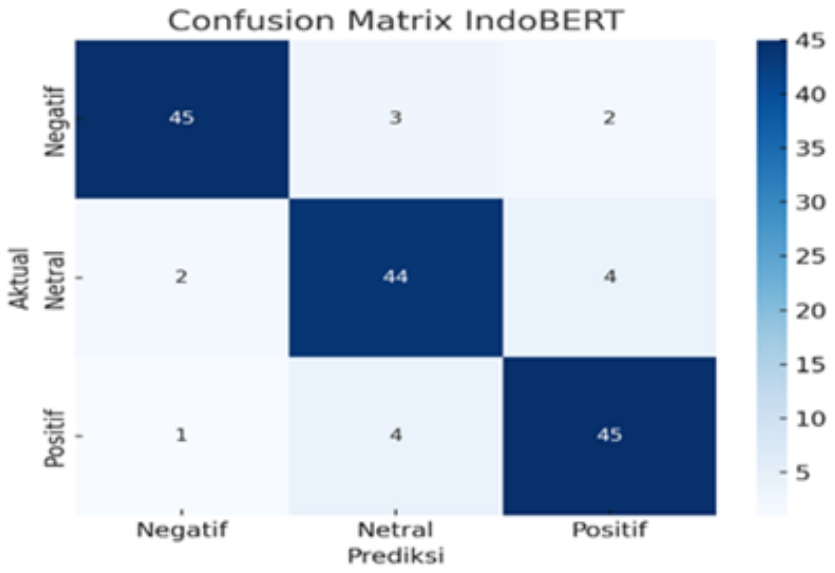


**Fig. 2:** Confusion matrix IndoBERT

The results of the study showed that the Fine-Tuned IndoBERT model delivers superior performance in the classification of Indonesian-speaking sentiment despite the dataset used relatively small. This is in line with the research conducted by Wilie et al. (2020) which demonstrates IndoBERT's excellence in the Indonesian language classification task over the traditional approach.

Achievement of 88.33% accuracy shows that models are able to understand the context of sentences in product reviews, including informal language structures that are often used by Tokopedia consumers. It proves the strength of a Transformer-based pretrained model in understanding the two-way context (bidirectional). When compared to Naïve Bayes and SVM, IndoBERT is not only superior to the accuracy side but also more stable in precision and recall. Baseline models tend to be sensitive to word variations and are unable to capture semantic meaning, especially in languages that have non-raw structures such as consumer reviews.

## 4. Conclusion

The research aims to implement IndoBERT models that have gone through fine-tuning processes in the classification of sentiment against local product reviews on Tokopedia's e-commerce platform using limited datasets. Based on the results of training and evaluation of models, it was concluded that the first IndoBERT model in fine-tune was significantly able to handle the problem of classification of Indonesian sentiment, even with the limited amount of training data. This is evidenced by the value of accuracy, precision, recall, and F1-score relatively high and consistent during the training process. Both preprocessing text such as punctuation cleaning, stopwords removal, and tokenizer with IndoBERT contribute to improving model performance against informal and unstructured data, such as customer reviews in marketplace.Third using limited datasets (1,500 data), models are able to generalize quite well to test data, showing that IndoBERT has an edge in data efficiency over other deep learning models that require large amounts of data.The five visualization training loss and evaluation metrics during training process indicates that the model is subject to stable convergence and does not overfitting, thanks to fairly convergence and proportionate data division.

## References

[1] U. T. Ama, D. N. Mulya, Y. P. D. Astuti, dan I. B. G. Prasadhya, "Analisis Sentimen Customer Feedback Tokopedia Menggunakan Algoritma Naïve Bayes," Jurnal Sistem Komputer dan Informatika (JSON), vol. 4, no. 1, pp. 50–55, 2022. doi: 10.30865/json.v4i1.4783.

[2] T. Ernayanti, M. Mustafid, A. Rusgiyono, dan A. R. Hakim, "Penggunaan Seleksi Fitur Chi-Square dan Algoritma Multinomial Naïve Bayes untuk Analisis Sentimen Pelanggan Tokopedia," Jurnal Gaussian, vol. 11, no. 4, pp. 562–571, 2023. doi: 10.14710/j.gauss.11.4.562-571.

[3] M. W. Ramadhani, M. D. A. Pratama, dan F. Rachman, "Implementasi Metode Naive Bayes untuk Klasifikasi Ulasan Aplikasi E-Commerce Tokopedia di Google Playstore," INTECOMS, vol. 6, no. 1, 2022. doi: 10.31539/intecoms.v6i1.5515.

[4] R. Rahmadona dan R. Handayani, "Analisis Sentimen Aplikasi Shopee, Tokopedia, Lazada dan Blibli di Google Play Store Menggunakan Pendekatan Lexicon-Based dan Random Forest," Jurnal Informatika dan Teknologi Elektro Terapan (JITeT), vol. 12, no. 3 (Supl.), 2024. doi: 10.23960/jitet.v12i3s1.5155.

[5] I. N. A. Kusuma dkk., "Analisis Sentimen Ulasan Pengguna Aplikasi Pelayanan Masyarakat dengan Menggunakan Algoritma Random Forest," Jurnal Nasional Teknologi Informasi dan Aplikasi (JNATIA), vol. 1, no. 1, 2022. doi: 10.24843/JNATIA.2022.v01.i01.p43.

[6] R. Merdiansah, S. Siska, dan A. A. Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI), vol. 7, no. 1, pp. 221–228, 2024. doi: 10.55338/jikomsi.v7i1.2895.

[7] N. A. R. Putri dan Ardiansyah, "Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa," Jurnal Sains dan Informatika (JSI), vol. 9, no. 2, pp. 136–145, 2023. doi: 10.34128/jsi.v9i2.649.

[8] H. Jayadianti dkk., "Sentiment Analysis of Indonesian Reviews using Fine-Tuning IndoBERT and R-CNN," ILKOM: Jurnal Ilmiah, vol. 14, no. 3, pp. 348–354, 2022. doi: 10.33096/ilkom.v14i3.1505.348-354.

[9] S. Aras dkk., "Sentiment Analysis on Shopee Product Reviews Using IndoBERT," Journal of Information Systems and Informatics, vol. 6, no. 3, pp. 1616–1627, 2024. doi: 10.51519/journalisi.v6i3.814.

[10] P. F. Supriyadi dan Y. Sibaroni, "Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT," Jurnal Riset Komputer (JURIKOM), vol. 10, no. 1, 2023. doi: 10.30865/jurikom.v10i1.5540.

[11] M. R. H. Prayogo dkk., "Fine-Tuned IndoBERT Based Model and Data Augmentation for Paraphrase Identification in Bahasa Indonesia," Revue

d'Intelligence Artificielle, vol. 37, no. 3, 2023. doi: 10.18280/ria.370322.
[12] Y. Astuti dkk., "IndoBERT for Classifying Hate Speech in Twitter," AIP Conference Proceedings, 2024. doi: 10.1063/5.0199750.
[13] M. Ahsani dkk., "Automatic Summarization of Indonesian Court Decisions using IndoBERT and Pointer-Generator," JOIV: International Journal on Informatics Visualization, vol. 7, no. 2, 2023. doi: 10.30630/joiv.7.2.1811.
[14] R. Nursyamsu dan N. Hidayat, "Analisis Sentimen Publik Terhadap Program Makan Siang Gratis Menggunakan Model BERT," Jurnal Ekonomi Manajemen Sistem Informasi (JEMSI), vol. 6, no. 2, 2025. doi: 10.38035/jemsi.v6i2.3376.
[15] K. C. Pradhisa dan R. Fajriyah, "Analisis Sentimen Ulasan Pengguna E-commerce di Google Play Store Menggunakan Metode IndoBERT," Technol. Sci., vol. 6, no. 1, pp. 92–104, 2024. doi: 10.47065/bits.v6i1.5247.