Journal of Information Systems and Informatics

Vol. 6, No. 3, September 2024 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

DOI: 10.51519/journalisi.v6i3.814 Published By DRPM-UBD

Sentiment Analysis on Shopee Product Reviews Using IndoBERT

Suhardi Aras[1], Muhammad Yusuf[2], Reinhard Ruimassa[3], Elli Agustinus Billi Wambrauw[4], Elsa Bura Palalangan[5]

[1,2,3,4,5]Informatics Engineering Study Program, Muhammadiyah University of Sorong,
Sorong City, Southwest Papua Indonesia
Email: [1]suhardi.aras@um-sorong.ac.id, [2]yusuf@um-sorong.ac.id,
[3]reinhardruimassa169@gmail.com, [4]billiwambrauw110803@gmail.com,
[5]elsapalalangan03@gmail.com

Abstract

A marketplace is a place in cyberspace where there are commercial activities between buyers and sellers. Products offered from the marketplace have reviews to review. Shopee is the most visited marketplace by people and offers various products. Product reviews can provide benefits for other consumers in assessing the products offered. By utilizing NLP technology in particular, this study can classify positive sentiment and negative sentiment in product review data. The IndoBERT model is a model that can be used in NLP technology by utilizing the relationship between each input and output element as well as the weights to be calculated simultaneously. By utilizing this technology, sentiment analysis on Shopee product reviews provides maximum accuracy until 93% with different training conditions. This provide that IndoBERT model can show that the performance of the indoBERT model in this research is very good.

Keywords: Marketplace, Shopee, Review, NLP, IndoBERT

1. INTRODUCTION

A marketplace is a place in cyberspace where there are commercial activities between buyers and sellers [1]. There are various kinds of marketplaces, for example the most frequently used is Shopee. Based on data released by iPrice, Shopee visits topped e-commerce with a peak number of visitors of 96 million in Q3 and a sharp increase in Q4 2020 to 129 million. Meanwhile in Southeast Asia, Shopee is also ranked first in e-commerce with the most visits, around 281 million [2]. Due to the high public interest in Shopee, this has resulted in many product reviews being carried out by buyers. Product reviews can provide great benefits for consumers because by reading reviews, consumers get information, meaning they can find out the quality of a product from other people's reviews, people share their experiences regarding related products. [3]. Sentiment Analysis is a technique in NLP. Techniques are used to identify text patterns and classify text into positive, negative and neutral sentiment [4]. This technique allows computers to understand, analyze and respond to human language texts in a humane way [5]. With this technology, the data obtained can be processed efficiently in identifying and classifying a text into predetermined sentiments.

A deep learning model known as BERT (Bidirectional Encoder Representations from Transformers) connects each output element to each input element, and the weights between them are calculated dynamically based on these connections. History shows that language models can only read text from left to right or from right to left. However, such models cannot do both at the same time. BERT is different because it can read in both directions at once. BERT has previously been trained for natural linguistic programming tasks, especially implicit language modeling and next sentence prediction, thanks to this bidirectional capability. [6]. In developing BERT, there are models that can be used, for example IndoBERT which was developed for the needs of Indonesian language BERT.

IndoBERT is a BERT-based trained model for Indonesian society using the Indo4B dataset containing more than 23 GB of Indonesian text data, including 4 billion words both formal and colloquial (vernacular) from various sources including social networks, blogs, and news. and website [7]. In previous IndoBERT research, the effectiveness of the IndoBERT model was evaluated with various word embedding models and previous training models, such as Multilingual BERT and XLM-R. The results obtained were that IndoBERT showed the most satisfactory results out of 8 per 12 classification tasks carried out. This shows that IndoBERT is very superior in terms of classification tasks even though the size of IndoBERT is relatively small compared to other models [6].

Therefore, in this research the author implemented the IndoBERT model to carry out sentiment analysis on review data on Shopee product reviews. It is hoped that the results of this research can provide an IndoBERT performance model with maximum results and high accuracy.

2. METHOD

This research includes several stages, including data collection, text preprocessing, labeling, stopwords, tokenize, stemming, data visualization, data splitting, modeling, model evaluation. To better understand, the following is a picture of the research scheme on the Figure 1.

2.1 Data Collection

The review data used in this research was obtained from the scrapping technique on the shopee.co.id website using scrapping tools. This scrapping review data will be saved in CSV form in Indonesian. The review data taken was in the form of positive and negative reviews totaling 1926 review data and data was taken in the period from January last year to June this year.

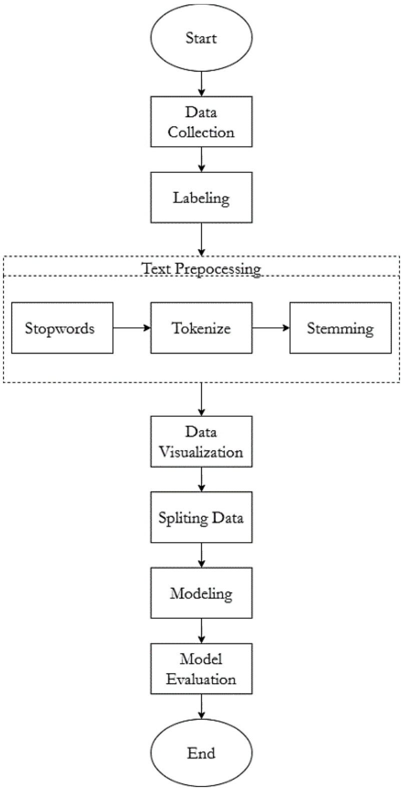


Figure 1. Research Scheme

2.2 Labeling

Labeling is a process to label a text into a specified label [8]. At this stage the text will be labeled based on the review rating. If the review rating shows one and two stars, the review will be categorized as a negative review or 0. Meanwhile, if the review has above two stars, namely three stars, four stars, five stars, then the review will be categorized as a positive review or 1.

2.3 Text Prepocessing

Text Prepocessing is a process to preprocessing text before entering stage modeling [9].The purpose of text preprocessing is to make text data that has been scrapped in CSV form so that it looks neat, such as removing emojis, reducing all letters to lowercase, and removing punctuation. This aims to ensure that in the next stage the text is easy to process at the next stage. There are several stages such as stopwords, tokenize, and stemming.

1) Stopwords
   Stopwords help eliminate unnecessary data. The purpose of this stop word is to remove words that are repeated but do not have much meaning so that the remaining words in the data are only accurate and meaningful data. [10].
2) Tokenize
   Tokenize functions to break sentences into words, phrases, symbols, or other data elements. This is intended to distinguish which data components are useful [11].
3) Stemming
   Stemming is done on words with affixes to change them into the base form of the word. This is intended so that when modeling the model does not read affixed words repeatedly which can cause noise [12].

2.4 Data Visualization

At this stage, data that has passed the above stages will be visualized to see the cleanliness of the data. This is intended to see whether there is still noise such as emojis, affixes, repeated words, etc. At this stage, the data that appears most frequently in the data that has been cleaned will also be displayed.

2.5 Splitting Data

After the data has been cleaned, data is divided into a data train dan data test [13]. based on the total amount of data, which is 1926 review data. The researcher decided to use ratio 8:2 for the split data which mean, data train will divide as much 1541 data and data test will divide as much 385 data.

2.6 Modeling

At this stage the researcher models the split data to be entered into the IndoBERT. As seen in the Figure 2 IndoBert is a leading text analysis model in Indonesian. This architecture is mostly built using the transformer model on BERT, which is more often in English. Like BERT, this technique also uses twelve hidden layers, with each hidden layer limited to 786 dimensions. In addition, this method also uses twelve attention heads [14] as seen in the figure . For this research, researcher used model with the type "indobenchmark/indobert-base-p1" with params 124,5 M. Researchers also used an optimizer, namely AdamW, with a learning rate of 2e-5, a training epoch run of 3 and added some max length which is 128. After that, researcher added some variations of batch size including 16, 32, and 128.

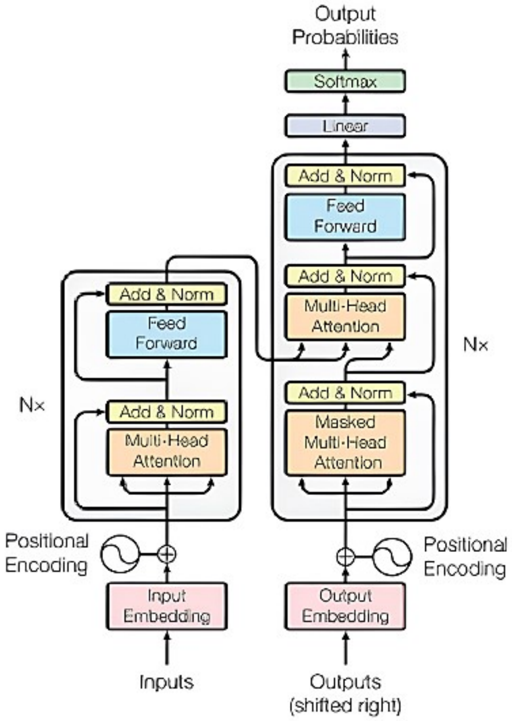


Figure 2. IndoBERT Architercture

2.7 Model Evalution

After passing the modeling stage, the researcher used the confusion matrix as a tool to evaluate the model and classification report to calculate the f1-score value and accuracy of the testing data. Confusion matrix is a tools to measure the accuracy of model by generating precision, recall, and f-1 score [15]. To better understand, researcher provide a table to describe shape of confusion matrix on Table 1.

Table 1. Evaluation Confusion Matriks

	Predicted	
Actual	TP	FN
	FP	TN

Description: True Positive (TP) is the amount of data with positive values and predicted positive. True Negative (TN) is the amount of data with negative values and predicted negative. False Positive (FP) is the amount of data with negative values but predicted positive. False Negative (FN) is the amount of data with positive values but predicted negative. Based on Table 1, the acurrasion of models can be calculated with Equation 1.

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Based on Equation 1, accuracy can calculate by the number of positive data values that are predicted positive and negative data values that are predicted negative divided by the total amount of data in the testing data.

3. RESULTS AND DISCUSSION

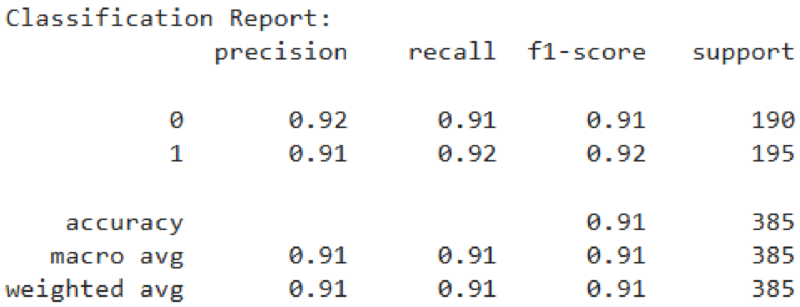
Data that has been scraped on the shopee.co.id website was taken from 1926 shopee users who provided reviews and ratings. After that, the scrapped data will be saved in CSV form and will enter the preprocessing stage where it will be processed to remove noise in the data after going through the stopwords, tokenize, stemming stages. After that the data will be divided or split into two, namely train data and testing data with a ratio of 8:2 respectively. The following is an example of a review table that has been scrapped on the shopee.co.id website on the Table 2.

Table 2. Review Table

No	Review	Rating	Produk Name	Label
1	Pesanan gx sesuai dgn yg d gambar	2	Isi Staples Tembak 8 mm best guard	0
2	Kotak rusak, lem sepatunya gk bngt	1	Sepatu Casual Kets Sport Nike MD Runner / Waffle Trainer Hitam Putih	0
3	Paket mantab langsung dipakai dan hasilnya memuaskan ...recomended nih ..untuk order berikutnya...	5	Staple Gun / Staples Tembak / Staples Jok / Hekter MOLLAR 3 in 1	1
4	Alhamdhulillah pesanan k sudah sampai lebih cepat dari perkiraan, kualitas barang oke banget, sangat memuaskan.	5	Staple Gun / Staples Tembak / Staples Jok / Hekter MOLLAR 3 in 1	1

3.1 Analysis

In the first training, researchers employed the IndoBERT model using the "indobenchmark/indobert-base-p1" configuration. The batch size was set to 16, with a maximum sequence length of 128. The study utilized the AdamW optimizer with a learning rate of 2e-5. After three training epochs, the model achieved an accuracy of 91%. The confusion matrix for this training showed that for the negative label (label 0), 172 instances were correctly predicted, while 18 were incorrectly predicted. For the positive label (label 1), 180 instances were correctly predicted, and 15 were incorrectly predicted, resulting in a total of 385 testing samples. These results are depicted in Figures 3 and 4.



```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.91      0.91       190
           1       0.91      0.92      0.92       195

    accuracy                           0.91       385
   macro avg       0.91      0.91      0.91       385
weighted avg       0.91      0.91      0.91       385
```

Figure 3. Classification Report

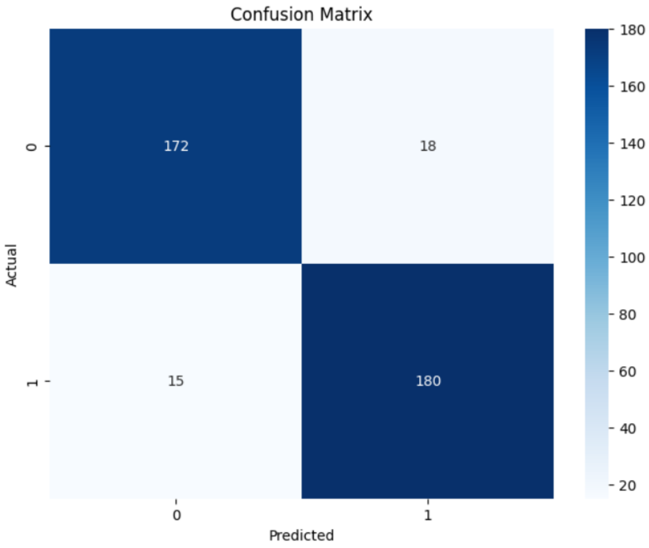
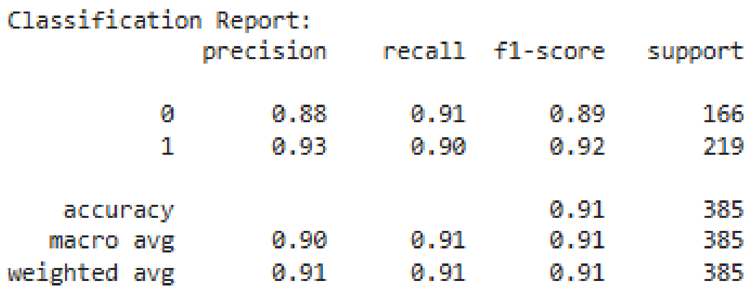


Figure 4. Confusion Matriks

In the second training, researcher still used the same model, same method just like in the first training. But in the second training, researcher change the batch size to 32. After carring out 3 training of epochs, the accuracy score was 91% also same with accuracy in the first training. But in the confusion matrix showing the different. With the confusion matrix based on second training showing the negative label or label 0 predicts 151 data are correctly predicted and 15 data are predicted incorrectly. Meanwhile, the positive label or label 1 predicts 198 data are correctly and 21 data are predicted incorrectly. So, this second training is almost as same as first training. This data can be prove based on Figure 5 and Figure 6.



```
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.91      0.89       166
           1       0.93      0.90      0.92       219

    accuracy                           0.91       385
   macro avg       0.90      0.91      0.91       385
weighted avg       0.91      0.91      0.91       385
```

Figure 5. Classification Report

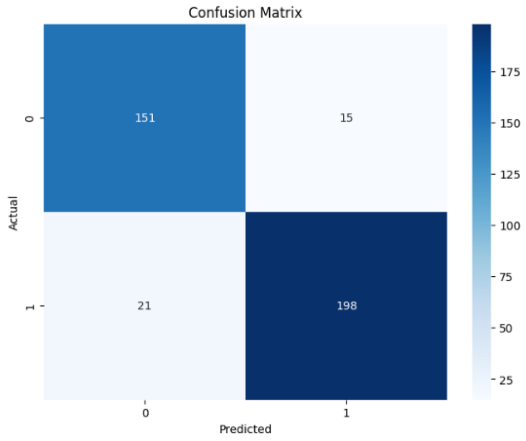
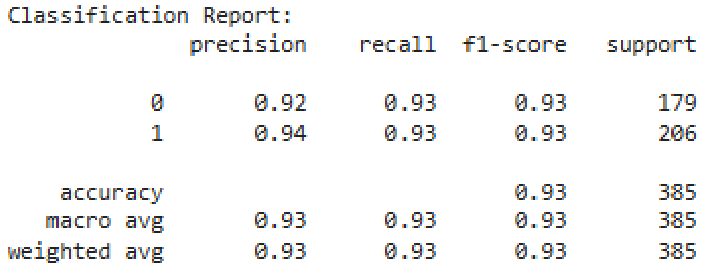


Figure 6. Confusion Matrix

In the third training, also with the same model and method also get a adjust in batch size using 128. After carring out 3 training of epochs, the accuracy was 93%. Just little bit higher then a first and second training. This can be prove by the confusion matrix which shows on label negative or label 0 can predicts 167 data are correctly and predicts 12 data are incorrectly. Meanwhile, the positive label or label 1 predicts 191 data are correctly and 15 data are incorrectly. This data can be seen on Figure 7 and Figure 8.



```
Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.93      0.93       179
           1       0.94      0.93      0.93       206

    accuracy                           0.93       385
   macro avg       0.93      0.93      0.93       385
weighted avg       0.93      0.93      0.93       385
```

Figure 7. Classification Report

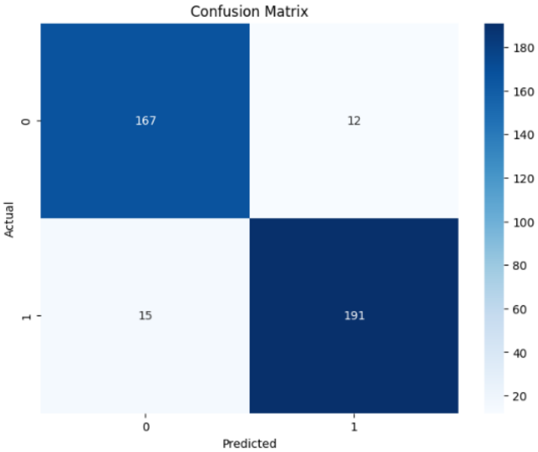


Figure 8. Confusion Matrix

To facilitate a comparison of the three training sessions under different conditions, refer to Table 3.

Table 3. Comparison Table

Conditions					Accuracy
Ratio Dataset	Max Length	Optimizer	Batch Size	Learning Rate	
8:2	128	AdamW	16	2e-5	91%
8:2	128	AdamW	32	2e-5	91%
8:2	128	AdamW	128	2e-5	93%

3.2 Discussion

The results of this research demonstrate the effectiveness of the IndoBERT model in performing sentiment analysis on Shopee product reviews. Across three training sessions with varying batch sizes, the model consistently achieved high accuracy, peaking at 93% when the batch size was increased to 128. This outcome highlights IndoBERT's capability to handle the complexities of sentiment analysis in the Indonesian language, particularly within the context of e-commerce product reviews, where diverse expressions and linguistic nuances are prevalent. The model's consistent performance across different settings suggests its robustness and adaptability to changes in batch sizes while maintaining a high level of predictive accuracy.

The first two training sessions, with batch sizes of 16 and 32, both resulted in an accuracy of 91%. This suggests that the model's performance was relatively stable regardless of these initial batch sizes. However, the confusion matrices for these two settings showed slight variations in the correct and incorrect predictions for both the positive and negative labels. These differences indicate that while the overall accuracy remained the same, the distribution of correctly and incorrectly classified instances varied. This observation could imply that the model's sensitivity to certain patterns or data characteristics might be influenced by the batch size, although these variations did not significantly impact the overall accuracy.

The third training session, where the batch size was increased to 128, resulted in a marginally higher accuracy of 93%. This improvement, although slight, suggests that a larger batch size may help the model generalize better across the dataset. The confusion matrix for this training run shows fewer incorrect predictions for both positive and negative labels compared to the previous sessions, which indicates that increasing the batch size contributed to better overall performance. This result aligns with the hypothesis that larger batch sizes can help stabilize the learning process by providing a more comprehensive representation of the data in each training iteration. However, the improvement is modest, suggesting that beyond a certain point, further increases in batch size may yield diminishing returns in accuracy.

The findings suggest that the IndoBERT model is highly effective for sentiment analysis tasks within the specific domain of Shopee product reviews, achieving high accuracy with relatively low training epochs and modest computational requirements. The consistent performance across different batch sizes underscores the model's flexibility and reliability, which are crucial for practical applications in real-world scenarios where computational resources and time might be constrained. Additionally, the results demonstrate that IndoBERT, even with its base configuration, is capable of producing reliable sentiment predictions, making it a valuable tool for e-commerce platforms looking to automate and scale their sentiment analysis efforts.

The study successfully achieved its aim of implementing and evaluating the IndoBERT model for sentiment analysis on Shopee product reviews. The results indicate that IndoBERT can deliver high performance with minimal tuning, suggesting its potential for broader application in other domains requiring sentiment analysis in the Indonesian language. Future research could explore further optimization strategies, such as experimenting with different learning rates, optimizers, or data augmentation techniques, to potentially enhance the model's performance even further. Additionally, examining the model's behavior with different datasets and exploring its capacity to handle more diverse or imbalanced datasets could provide deeper insights into its generalizability and robustness.

4. CONCLUSION

Based on the evaluation results presented, the IndoBERT model demonstrates strong performance in classifying sentiment in Shopee product reviews, achieving a maximum accuracy of 93% across three training conditions with varying batch sizes. This outcome confirms that the IndoBERT model is highly effective for sentiment analysis tasks in the Indonesian language, maintaining high accuracy with different parameter settings. The consistent results across different configurations suggest that the model is robust and adaptable, making it a valuable tool for similar applications in e-commerce platforms. The findings also highlight the positive impact of using the IndoBERT model for sentiment analysis, showcasing its ability to deliver reliable performance with minimal adjustments. However, to further enhance the model's effectiveness, future research should consider addressing potential issues related to noise within the dataset. Additionally, experimenting with various versions of IndoBERT or incorporating newer models could provide further improvements in accuracy and generalizability. By refining these aspects, future studies could unlock even greater potential for the IndoBERT model in sentiment analysis, ensuring its continued relevance and utility in the ever-evolving landscape of natural language processing and machine learning applications.

REFRENCES

[1] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.

[2] N. A. Wardah and H. Harti, "Pengaruh Gaya Hidup Berbelanja Dan Promosi Penjualan Terhadap Pembelian Impulsif Avoskin Di Shopee," *Ecobisma (Jurnal Ekon. Bisnis Dan Manajemen)*, vol. 8, no. 2, pp. 145–166, 2021, doi: 10.36987/ecobi.v8i2.2090.

[3] D. N. Sari, D. N. Sari, F. Adelia, F. Rosdiana, B. B. Butar, and M. Hariyanto, "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier," *JIKA (Jurnal Inform.*, vol. 4, no. 3, p. 109, 2020, doi: 10.31000/jika.v4i3.3086.

[4] T. Bey Kusuma, I. Komang, and A. Mogi, "Implementasi BERT pada Analisis Sentimen Ulasan Destinasi Wisata Bali," *J. Elektron. Ilmu Komput. Udayana*, vol. 12, no. 2, pp. 409–420, 2023.

[5] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[6] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and Resources for Evaluating

Indonesian Natural Language Generation," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021, doi: 10.18653/v1/2021.emnlp-main.699.

[7] E. P. A. Akhmad, "Analisis Sentimen Ulasan Aplikasi DLU Ferry Pada Google Play Store Menggunakan Bidirectional Encoder Representations from Transformers," *J. Apl. Pelayaran Dan Kepelabuhanan*, vol. 13, no. 2, pp. 104–112, 2023, doi: 10.30649/japk.v13i2.94.

[8] W. Nurfitri and A. Chowanda, "Analisis Sentimen Pada Kasus Positif Covid-19 Berdasarkan Pemberitaan Media Di Indonesia Menggunakan Indobert," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 580, 2024, doi: 10.35889/progresif.v20i1.1897.

[9] K. Cindy Pradhisa and 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.

[10] Ardiansyah, Adika Sri Widagdo, Krisna Nuresa Qodri, F. E. N. Saputro, and Nisrina Akbar Rizky P, "Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT," *J. Fasilkom*, vol. 13, no. 02, pp. 326–333, 2023, doi: 10.37859/jf.v13i02.5170.

[11] P. F. Supriyadi and Y. Sibaroni, "Xiaomi Smartphone Sentiment Analysis on Twitter Social Media Using IndoBERT," *J. Ris. Komputer)*, vol. 10, no. 1, pp. 2407–389, 2023, doi: 10.30865/jurikom.v10i1.5540.

[12] R. Maulana Arrasyid, D. Enggar Putera, and A. Yunizar Pratama Yusuf, "Analisis Sentimen Review Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Tekno Kompak*, vol. 18, no. 2, pp. 1–12, 2021.

[13] A. Prabowo and F. Indra Sanjaya, "Penerapan Metode Transfer Learning Pada Indobert Untuk Analisis Sentimen Teks Bahasa Jawa Ngoko Lugu," *J. Sist. Inf. dan Sist. Komput.*, vol. 9, no. 2, pp. 205–217, 2024, [Online]. Available: http://e-jurnal.stmikbinsa.ac.id/simkom

[14] B. Juarto and Yulianto, "Indonesian News Classification Using IndoBert," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 2, pp. 454–460, 2023.

[15] T. Hidayati and B. F. Putra, "Implementasi Deep Learning Untuk Image Classification menggunakan Convolutional Neural Network Pada Citra Wayang (Studi Kasus : SDN Leuwibatu 03)," *Sci. Sacra J. Sains, Teknol. dan Masy.*, vol. 4, no. 1, pp. 1–7, 2024.