Comparison of IndoBERT and SVM Algorithm to Perform Aspect Based Sentiment Analysis using Hierarchical Dirichlet Process

Sheila Prima Octarini[1], Alfi Yusrotis Zakiyyah[2], Kartika Purwandari[3*]
[1-3]Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
sheila.octarini@binus.ac.id, alfi.zakiyyah@binus.ac.id,
kartika.purwandari@binus.ac.id

*Correspondence: kartika.purwandari@binus.ac.id

Abstract – *Analyzing the performance of SVM and IndoBERT models for aspect-based sentiment analysis on fashion reviews in Tokopedia E-Commerce. This study employs the SMOTE technique due to the imbalance in the original data. Aspect determination using the Hierarchical Dirichlet Process (HDP) model yields satisfactory results with an adequate coherence score. The comparison between SVM and IndoBERT methods for aspect-based sentiment analysis shows that SVM is superior. IndoBERT achieved an accuracy of 87%, precision of 91%, recall of 93%, and F1-Score of 92%, while SVM attained an accuracy of 96%, precision of 100%, recall of 92%, and F1- Score of 96%. Therefore, the SVM model was chosen for implementation on a website that allows users to view aspect-based sentiment analysis on products in E-Commerce. The HDP model effectively grouped related terms into aspects such as "Material," "Shipping," and "Colour," enhancing interpretability in sentiment classification. The resulting website enables users to analyze product sentiments interactively, providing actionable insights for both sellers and customers to assess product quality and service satisfaction more efficiently.*

Keywords: *Hierarchical Dirichlet Process; SVM; IndoBERT; SMOTE; Aspect Based Sentiment Analysis*

I. INTRODUCTION

As the digital age progresses, more people prefer online shopping. The convenience of browsing various products, comparing prices, and the ease of shopping online are some of the advantages of online (Parengkuan & Nurhasanah, 2021). Additionally, consumers benefit from being able to review product feedback or reviews, which greatly influences purchasing decisions. Especially relevant in fashion products, where the quality presented on online stores may differ from the actual item in terms of size, design, and material. This study aims to assist consumers in minimizing the time spent reviewing product reviews by using Aspect-based Sentiment Analysis (ABSA).

Sentiment analysis, also known as opinion mining, is a technique in natural language processing (NLP) used to determine sentiment or opinions within a (Goldberg, 2017). Sentiment analysis can be categorized into three levels: document, sentence, and aspect (Liu, 2020). At the aspect level, documents are analyzed based on predefined aspects or entities.

In this study, aspect identification will use the Hierarchical Dirichlet Process (HDP) method. HDP is a statistical model used to group textual data into various naturally occurring topics or aspects. The HDP method is an improvement over the LDA method, as HDP delivers better performance in topic modeling, especially when dealing with continuously growing text data (Maulani et al., 2024). In previous research, the HDP model achieved higher accuracy compared to LDA and avoided overfitting.

A study by M.P. Geetha et al. in 2021 showed that the BERT (Bidirectional Encoder Representations from Transformers) method outperformed Naive Bayes, Long Short-term Memory (LSTM), and Support Vector Machine (SVM). SVM followed with higher recall, precision, and F1- score compared to LSTM and Naive Bayes. This motivated the author to further explore the BERT and SVM methods.

IndoBERT is a variant of BERT designed for Indonesian language (Taufiq Dwi Purnomo & Joko Sutopo, 2024). It processes text in two main phases: pre-training and fine-tuning (Koto et al., 2020). In pre-training, the model is trained on tasks like predicting missing words (masked language model) or predicting the relationship between two sentences. The second phase, fine-tuning, involves sentiment analysis using labeled data relevant to the problem at hand. In this study, aspect-based sentiment analysis focuses on positive and negative sentiments for each aspect.

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks (Blanco et al., 2023), (Sarang, 2023). It works by finding a hyperplane that maximally separates two classes in the data. The main goal of SVM is to find the best separating line that maximizes the margin between two data classes (Gaye et al., 2021). In SVM, some key data points from each class, known as support vectors, are used. The hyperplane is positioned in a way that maximizes the distance between support vectors from different classes. SVM is not limited to two-dimensional data; the kernel trick can be used to map data to higher dimensions, allowing for the separation of non-linearly separable data in the original space (Cortes & Vapnik, 2019).

In this research, IndoBERT and SVM models will be compared to analyze aspect-based sentiment in fashion products on the Tokopedia marketplace. The results of this study are expected to help consumers analyze reviews based on predefined aspects with the support of the HDP statistical model, showing the percentage of positive and negative opinions for each aspect.

Another study by Geetha and Renuka (Geetha & Renuka, 2021; Brownlee, 2020) focused on improving the performance of aspect-based sentiment analysis using a fine-tuned BERT Base Uncased model. They compared several classification methods, including Naive Bayes, LSTM, SVM, and BERT. Among these, the BERT model achieved the highest precision (88.09%) and recall (86.22%), demonstrating its effectiveness in handling nuanced sentiment aspects. However, SVM also showed promising results with a precision of 82.68% and recall of 84.31%, indicating that while deep learning models like BERT are powerful, traditional machine learning models like SVM still hold value in certain contexts.

Budiman et al. (Budiman et al., 2024) conducted a comparative study on the classification performance of BERT and IndoBERT using self-reported COVID-19 statuses on social media. The findings revealed that IndoBERT outperformed BERT, achieving a higher accuracy rate of 94% compared to BERT's 82%. This suggests that IndoBERT may be better suited for sentiment analysis in Indonesian-language datasets, given its fine-tuning on specific language characteristics.

The analysis of sentiment using various models has been a topic of considerable research in recent years. In a study by Merdiansah (Merdiansah et al., 2024), sentiment analysis of Indonesian users concerning electric vehicles was conducted using an IndoBERT model integrated with IndoNLU. This model achieved remarkable performance, with training data accuracy, precision, recall, and F1-score all at 100%. On validation data, it also showed impressive results, with 98% accuracy, precision, and recall, indicating the robustness of the IndoBERT model in sentiment analysis tasks.

Maulani et al. (Maulani et al., 2024) approached the classification of reviews from Google Play by employing the Hierarchical Dirichlet Process (HDP) and Latent Dirichlet Allocation (LDA). Their research demonstrated that HDP outperformed LDA in feature extraction, highlighting HDP's ability to handle complex and evolving topics more effectively than traditional topic modeling methods.

Pavithra and Savitha (Pavithra & Savitha 2024) extended the exploration of topic modeling by comparing different methods such as LDA, HDP, Non-Negative Matrix Factorization (NMF), and BERTopic on research papers. The study showed varied coherence scores across these methods, with LDA achieving a score of 1.7859 and HDP at 2.6523, indicating that each method has different strengths in capturing the underlying topics in textual data.

Overall, these studies emphasize the effectiveness of using advanced models like IndoBERT and HDP in sentiment analysis and topic modeling, especially in the context of Indonesian language and e-commerce reviews. The comparison of different models also highlights that the choice of model can significantly impact the performance outcomes, with IndoBERT and HDP standing out for their robust capabilities in dealing with complex textual data.

II. METHODS

This study utilizes data from the e-commerce platform, Tokopedia. The data collected includes customer reviews from Hana Fashion Shop, obtained through web scraping. A total of 1,000 reviews were gathered, with 100 reviews each from the 10 best-selling products at the store.

Each review was labeled to be used in the training process for classification. The review attributes reflect customer experiences regarding their satisfaction with shopping at Hana Fashion Shop. Labeling was conducted based on product ratings, where ratings of 1-3 were categorized as negative, and ratings of 4-5 as positive.

The data then cleaned under a preprocessing, which included:

1) Cleaning – Removing duplicate or empty data, numbers, and emoticon. Also turning all the text to lowercase.
2) Normalization – Correcting misspelled words or abbreviations to their original form.
3) Stopword removal – Eliminating common words that don't provide significant information for text analysis.
4) Tokenization – Breaking down the text into individual words or terms.
5) Stemming – Reducing words to their base form by removing suffixes or affixes.

Table 1. Example of Data after Preprocessing

Input	Output
istri | suka | nyaman | dipakai	istri suka nyaman pakai
sumpah | bagus | banget | bahanya | ga | kaleng | kaleng | hana | fashion | murah | murah | berat | badan | kilogram | badan | sentimeter | haha	sumpah bagus banget bahanya ga kaleng hana fashion murah berat badan kilogram badan sentimeter haha

The Hierarchical Dirichlet Process (HDP) was employed to help identify topics used as aspects in the aspect-based sentiment analysis for the data that has been through preprocessing process. Wordcloud will be used to visualize words that have been clustered. Figure 1 until Figure 3 shows the topics obtained using HDP Method.

Figure 1. Wordcloud of topic 1 from HDP Process

Figure 2. Wordcloud of topic 2 from HDP Process

Figure 3. Wordcloud of topic 3 from HDP Process

Since HDP can only assists in categorizing related words within the data, the researcher must manually define the topics that will serve as segments. From the analysis, several aspects, such as "Bahan" means materials, "Pengiriman" means delivery, and "Warna" means color, were identified as appropriate representations of the words found in the word cloud. The words within the wordcloud will be used as boundaries for each aspect, and additional words from other topics that are relevant will also be considered to further refine the aspect boundaries.

Once the aspects were determined, classification was performed using the Support Vector Machine (SVM) and IndoBERT models. Initially, the review attributes were weighted using Term Frequency (TF-IDF), considering only term frequency. This resulted in a classification model that was later applied during the testing phase, where reviews were categorized into different aspects and analyzed for positive or negative sentiment.

III. RESULT AND DISCUSSION

Due to the imbalanced sentiment in the data, where most of the data is positive in sentiment, the researcher employed SMOTE to address class imbalance.

3.1 SVM And IndoBERT

The SVM model analyzes sentiment across each aspect, displaying the percentage of both negative and positive sentiments. The bar chart visualization, as shown in Figure 4, reveals the sentiment distribution for different aspects.

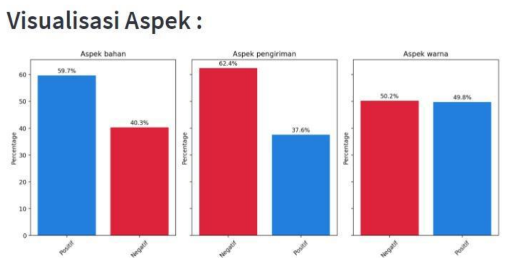

Figure 4. Aspect-based Sentiment Analysis Result using SVM

When applying the SVM model with SMOTE, the resulting confusion matrix showed that out of 162 samples, there were 129 true positive (TP) predictions, 0 false positive (FP), 140 true negative (TN), and 12 false negative (FN) predictions.

For the “Material” aspect, 59.7% of sentiments are positive, while 40.3% are negative. The “Shipping” aspect has 63.4% positive and 37.6% negative sentiments, while the “Colour” aspect shows 50.2% positive and 49.8% negative sentiments.

In contrast, when using the IndoBERT model with SMOTE, the confusion matrix for 160 samples reveals 12 true positive (TP) predictions, 142 false positive (FP), 3 true negative (TN), and 3 false negative (FN) predictions.

3.2 Comparison

Both models (SVM and IndoBERT) need to be compared to determine the best-performing model for this research. Model evaluation will be conducted by calculating accuracy, precision, recall, and F1-score. The best-performing model will be implemented on the website for users to conduct aspect-based sentiment analysis on fashion products.

The comparison at Table II indicates that the SVM model with SMOTE outperforms the IndoBERT model, achieving an accuracy of 96%, precision of 100%, recall of 92%, and an F1-score of 96%. Although IndoBERT showed better recall, SVM dominated in overall performance with higher scores. Therefore, the researcher will implement the SVM model on the website to allow users to perform aspect-based sentiment analysis on fashion products.

Table 2 Result of Comparison Model using SVM And IndoBERT

Model	Accuracy	Precision	Recall	F1 Score
SVM	96%	100%	98%	96%
IndoBERT	90%	92%	92%	95%

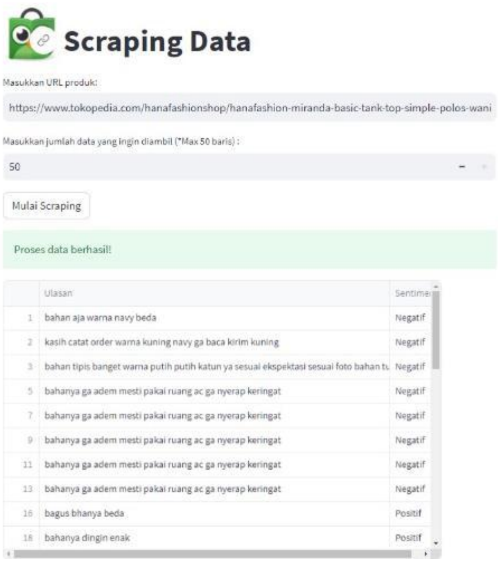

Figure 5. User menu (Inserting link)

3.3 Interface

On the interface page, there are several available menus. These include the user menu on figure 5, which allows users to perform aspect-based sentiment analysis on new products just by inserting the link of the product. At figure 6, the user can see the results of ABSA of the product from the link that has been inserted.

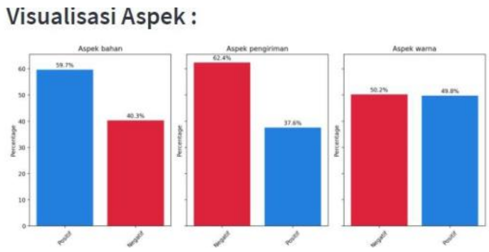

Figure 6. User menu (ABSA result)

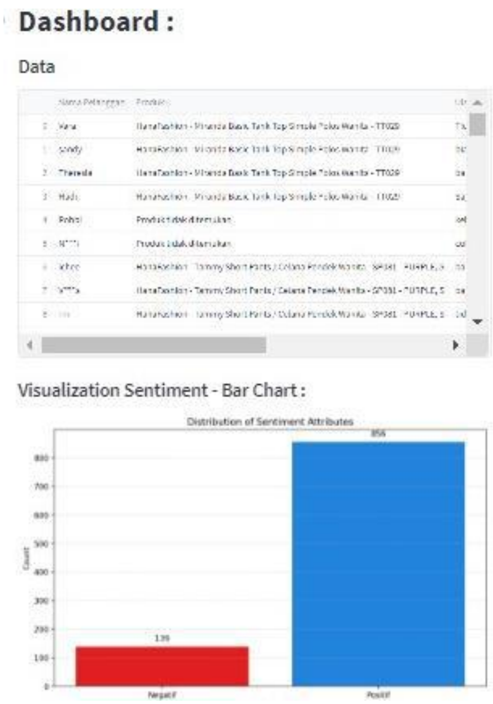

Figure 7. Dashboard menu (dataset and sentiment bar chart)

At the dashboard menu on figure 7 providing an overview of the application based on the data that have been obtained by the author. The data scraping menu from figure 9 is meant for collecting review information from products in Hana Fashion Shop on Tokopedia. User could insert any desired product link from the shop to collect the reviews of the product. There is no differences method that used for scrapping product reviews from the User menu and Scrapping menu. The only differences is the user menu automatically shown the ABSA results meanwhile the scrapping menu only pulled the reviews without doing any other steps.

Figure 8. Dashboard menu (wordcloud negative and positive words)

Figure 9. Scrapping data menu

From figure 10, the merge data menu is shown. It can help user to combine multiple datasets obtained from the data scraping menu. Additionally, there is a dataset menu for viewing the merged datasets at figure 11.

Figure 10. Merge data menu

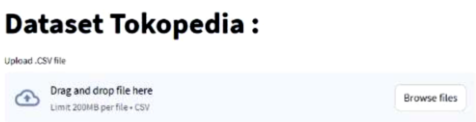

Figure 11. Dataset menu

A preprocessing menu shown at figure 12 can be used to prepare the data for analysis where all the results of all the preprocessing process like cleaning, normalizing, stopword removal, tokenization, and stemming.

Figure 12. Preprocessing data menu

After the preprocessing process, at figure 13, users can use the visualization desired data on visualization menu to see the amount of negative and positive reviews (figure 14). Other than that, this menu also shows the wordcloud visualization of the negatives and positives words shown at figure 15.

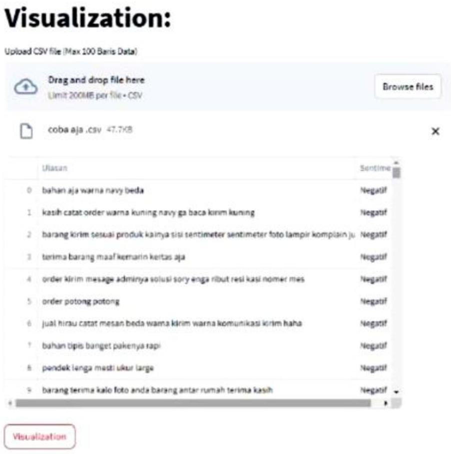

Figure 13. Visualization menu (to insert file)

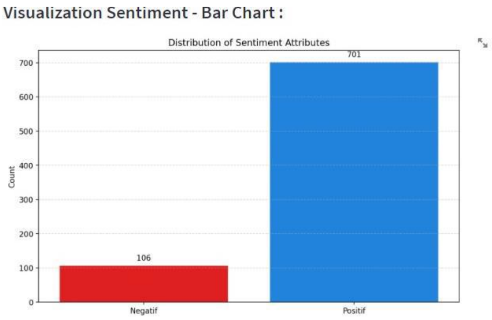

Figure 14. Visualization menu (to see the negative and positives words)

Figure 15. Visualization menu (to see the number of negative and positives data)

Both the SVM and IndoBERT menus for viewing the results of ABSA from training data using the Support Vector Machine and IndoBERT (figure 16) algorithms. Users can insert the dataset that has been through the preprocessing menu and see the results like on figure 6 with an addition of details of confusion matrix to see the accuracy, recall, precision, and F-1 score from the model.

Figure 16. Training IndoBERT model menu

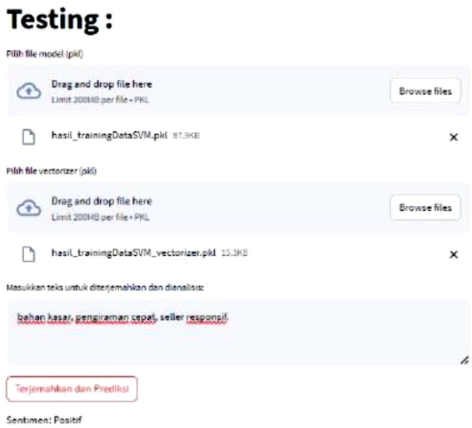

Figure 17. Testing menu

Lastly, the testing menu on figure 17 is used to test the trained model. User can check any sentences sentiment using SVM and IndoBERT. With this range of menus, users can easily access various features and steps within the application to analyze data and evaluate models.

IV. CONCLUSION

Based on the results of the research conducted, the following points can be concluded:

a) Data extraction from Tokopedia resulted in 1000 data points, which were reduced to 811 after the preprocessing phase. 80% of this data was used for training, while the remaining 20% was used for testing.
b) SMOTE was employed in this study because the original data from Hana Fashion Shop was relatively imbalanced, with more positive sentiment data compared to negative sentiment.
c) Aspect determination using the HDP model provided satisfactory results. The word grouping worked well, aiding the author in determining the topics used as aspects ("Material", "Shipping", "Colour").

The implementation of the SVM and IndoBERT methods went smoothly. A comparison between SVM and achieving an accuracy of 96%, precision of 100%, recall achieved an accuracy of 87%, precision of 91%, recall of 93%, and an F1 score of 92%. Based on these results, the SVM model was chosen for implementation on the website, allowing users to view Analysis-based Sentiment Analysis for products in Hana Fashion Shop on Tokopedia.

REFERENCES

Brownlee, J. (2020). Imbalanced classification with Python: Better metrics, balance skewed classes, and apply cost-sensitive learning. Machine Learning Mastery.

Blanco, V., Japón, A., & Puerto, J. (2023). Multiclass optimal classification trees with SVM-splits. Machine Learning, 112(12), 4905–4928. https://doi.org/10.1007/s10994-023-06366-1

Budiman, I., Faisal, M. R., Faridhah, A., Farmadi, A., Mazdadi, M. I., Saragih, T. H., & Abadi, F. (2024). Classification performance comparison of BERT and IndoBERT on self-report of COVID-19 status on social media. Journal of Computer Sciences Institute, 30, 61–67. https://doi.org/10.35784/jcsi.5564

Cortes, C., & Vapnik, V. (2019). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. Mathematical Problems in Engineering, 2021, 1–9. https://doi.org/10.1155/2021/5594899

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

Goldberg, Y. (2017). Neural network methods for natural language processing. Morgan & Claypool Publishers. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Koto, F., Lau, J. H., & Baldwin, T. (2020). IndoBERT: A pre-trained language model for Indonesian. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 757–770).

Liu, B. (2020). Introduction. In Sentiment analysis: Mining opinions, sentiments, and emotions (pp. 1–17). Cambridge University Press.

Maulani, I., Fatichah, C., & Wijaya, A. Y. (2024). Klasifikasi ulasan berdasarkan divisi pada Google Play menggunakan metode Hierarchical Dirichlet Process dan metode Ensemble [*Review classification based on divisions on Google Play using the Hierarchical Dirichlet Process and Ensemble methods*]. ILKOMNIKA: Journal of Computer Science and Applied Informatics, 6(1), 30–42. https://doi.org/10.28926/ilkomnika.v6i1.596

Merdiansah, R., Siska, & Ridha, A. A. (2024). Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT [*Sentiment analysis of Indonesian X users regarding electric vehicles using IndoBERT*]. Jurnal Informatika Komputer, 9(2), 100–110. https://ejournal.sisfokomtek.org/index.php/jikom/article/view/2895/2044

Parengkuan, S., & Nurhasanah, N. (2021). Analisis komparatif preferensi konsumen dalam belanja online [*Comparative analysis of consumer preferences in online shopping*]. Jurnal Ekonomi: Journal of Economic, 12(2). https://doi.org/10.47007/jeko.v12i02.4345

Pavithra, C. B., & Savitha, J. (2024). Topic modeling for evolving textual data using LDA, HDP, NMF, BERTopic, and DTM with a focus on research papers. Journal of Machine Learning Research, 12(1), 45–57.

Sarang, P. (2023). Support Vector Machines (pp. 153–165). https://doi.org/10.1007/978-3-031-02363-7_8

Taufiq Dwi Purnomo, & Joko Sutopo. (2024). Comparison Of Pre-Trained Bert-Based Transformer Models for Regional Language Text Sentiment Analysis in Indonesia. International Journal Science and Technology, 3(3), 11–21. https://doi.org/10.56127/ijst.v3i3.1739