Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews

Sinta Apriliani[1]*, Adhitia Erfina[2], Cecep Warman[3]
Information Systems, Faculty of Computer Engineering and Design[1], [2], [3]
Nusa Putra University
Sukabumi, Jawa Barat, Indonesia
sinta.apriliani_si21@nusaputra.ac.id[1], adhitia.erfina@nusaputra.ac.id[2], cecep.warman@nusaputra.ac.id[3]

***Abstract* – Online reviews significantly shape public perception and play a crucial role in customer decision-making within the hospitality sector. This research aims to conduct aspect-based sentiment analysis on Indonesian five-star hotel reviews using a fine-tuned IndoBERT model. Unlike prior studies that mainly applied IndoBERT to single hotels or small-scale datasets, this study fills that gap by examining 2,499 reviews collected from five luxury hotels in Jakarta. The analysis focuses on five essential service aspects: cleanliness, service quality, room comfort, food & beverages, and core facilities. The IndoBERT-base model was fine-tuned with annotated aspect-sentiment data and assessed using accuracy, precision, recall, F1-score, and confusion matrices. Experimental results show that the model reached 95.28% accuracy with a macro F1-score of 82.44%. Positive sentiment dominated the reviews (81.4%), while neutral and negative sentiments represented 16.9% and 1.7%, respectively. Service, along with food & beverages, received the highest praise, whereas cleanliness and core facilities were more often evaluated neutrally. Aspect and sentiment annotations were carried out semi-automatically using large language models (LLMs) and later validated by human annotators to ensure reliability. These findings highlight IndoBERT's strong capability in aspect-based sentiment classification for Indonesian hotel reviews and provide actionable insights for hotel managers to enhance service quality. Moreover, this study demonstrates both the academic and practical significance of applying fine-tuned Transformer models to real-world customer experience analysis.**

Keywords: Aspect-Based Sentiment Analysis, IndoBERT, Hotel Reviews, Sentiment Classification, Natural Language Processing

I. INTRODUCTION

The hospitality sector, especially in Jakarta, holds a significant position in driving tourism development and contributing to the national economy. As the capital city and a central business hub of Indonesia, Jakarta hosts a wide range of five-star hotels that serve the premium market by offering luxurious services and top-class facilities [1]. Based on 2024 data from Indonesia's Central Bureau of Statistics (BPS, https://www.bps.go.id/id), Jakarta holds the second position after Bali in terms of five-star hotel numbers, with a total of 37 properties.

In today's digital era, online reviews play a crucial role in shaping customer choices and determining the reputation of hospitality businesses. Platforms like Google Reviews, TripAdvisor, and Booking.com enable guests to share their experiences through star ratings and written reviews. Such reviews have a substantial impact on public perception and influence purchasing decisions [1].

To examine customer perspectives expressed in textual reviews, Aspect-Based Sentiment Analysis (ABSA) has become a widely used and relevant approach [2]. Unlike conventional sentiment analysis, ABSA enables the identification and classification of sentiments tied to particular service dimensions, including cleanliness, service quality, room comfort, food and beverages, and core facilities.

Advances in Natural Language Processing (NLP), especially the introduction of Transformer-based architectures such as BERT, have greatly enhanced the effectiveness of sentiment analysis and text classification tasks. IndoBERT, a pre-trained model tailored for Bahasa Indonesia, has achieved strong results across multiple NLP applications [3].

Several previous studies have used IndoBERT alongside other deep learning approaches for Aspect-Based Sentiment Analysis (ABSA) and sentiment classification in Indonesian contexts. For instance, one study compared IndoBERT with a CNN, showing that IndoBERT achieved superior results in sentiment classification, whereas the CNN performed better in aspect identification [4]. IndoBERT has also been widely adopted in tourism reviews [5], student feedback [6], transportation [7], healthcare [8], and e-commerce [9], consistently achieving accuracy and F1 scores above 90%. Beyond these, various models such as LSTM, BiLSTM, GRU, and hybrid BERT-based methods were tested on hotel reviews [10], [11], [12], [13], [14], showing promising results but still constrained by either single-object datasets, limited domains, or smaller-scale applications. These findings confirm the robustness of IndoBERT for Indonesian-language sentiment analysis, yet highlight a gap in multi-hotel, real-world applications using diverse customer-generated reviews.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

Recent research further demonstrates the versatility of IndoBERT. Salma et al. (2025) showed its effectiveness in analyzing informal social media text [15], while Ruslani and Fauzan (2025) achieved over 94% accuracy in e-commerce sentiment classification using IndoBERT combined with IndoT5 for summarization [16]. Similarly, Putra et al. (2025) optimized IndoBERT with the Adam optimizer, reaching above 92% accuracy [17]. In public policy, Firdaus and Trisnawarman (2025) applied IndoBERT Lite Large to 14,618 YouTube comments on the TAPERA housing program, though with lower accuracy (78%) [18]. Merdiansah et al. (2024) used IndoBERT for sentiment on electric vehicles, confirming the importance of domain adaptation [19], while Hakim et al. (2024) applied it to opinions on Indonesia's high-speed rail project but noted challenges in neutral sentiment detection [20].

Despite these advances, a significant gap remains. Most studies focus on single-source datasets, narrow domains, or controlled settings, with limited exploration of diverse, multi-aspect real-world hospitality reviews. In Indonesia's luxury hotel sector, no prior research has applied fine-tuned IndoBERT to aspect-level sentiment analysis across multiple establishments. This gap is critical, as hotel managers increasingly require actionable insights from online reviews to remain competitive in a service-driven market.

To bridge this research gap, the present study employs a fine-tuned IndoBERT model for ABSA, trained on 2,499 reviews collected from five luxury hotels in Jakarta. This work makes two main contributions: (1) delivering data-driven insights that can assist hotel management in enhancing service quality, and (2) expanding the application of fine-tuned Transformer models for Indonesian ABSA in a practical, multi-hotel setting. To our knowledge, this is the first study to apply fine-tuned IndoBERT for multi-aspect sentiment analysis across several five-star hotels in Indonesia, thereby addressing a key gap in hospitality data analysis.

II. METHODOLOGY

This research adopts a quantitative approach using Aspect-Based Sentiment Analysis (ABSA) to categorize sentiments associated with specific service aspects in Indonesian-language reviews of five-star hotels. The primary technique applied was fine-tuning the IndoBERT model, chosen for its robust ability to capture the contextual nuances of Bahasa Indonesia. All experimental processes were carried out in Python on the Google Colaboratory platform, utilizing essential NLP libraries, including Hugging Face Transformers.

A. Research Flow

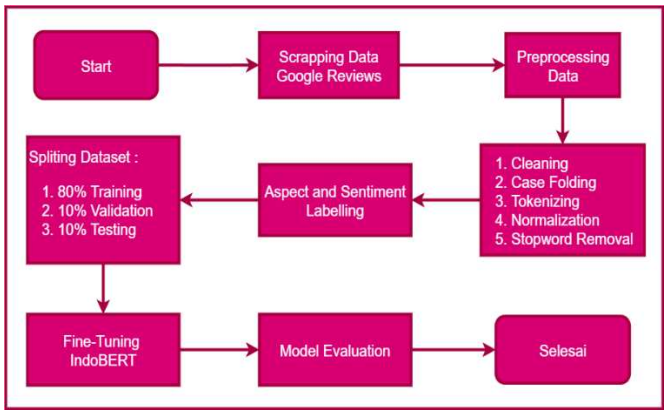

Fig. 1. Research Flow

The research followed a structured workflow, as illustrated in Figure 1. The major stages included: data scraping, text preprocessing, aspect and sentiment annotation, dataset splitting, model fine-tuning, evaluation, and result interpretation.

B. Data Collection

Customer reviews were scraped from Google Reviews using the SerpAPI service. An initial set of 2,500 reviews was collected from five leading five-star hotels in Jakarta: Fairmont Jakarta, Four Seasons Jakarta, The Langham Jakarta, Raffles Jakarta, and The Grove Suites by Grand Aston. After filtering out non-Indonesian entries and empty reviews, 2,499 valid reviews were retained for further processing.

C. Data Preprocessing

Preprocessing is a vital stage in text analysis that converts raw text into a structured and interpretable format, ensuring it is ready for subsequent modeling [20]. In this study, several preprocessing steps were applied. First, a cleaning process was performed to remove unnecessary elements, such as symbols, emojis, numbers, and URLs. Next, case folding was performed by converting all text to lowercase for consistency. The text was then tokenized into individual word units, followed by normalization to correct informal or misspelled words using a manually curated dictionary. Lastly, stopwords were handled selectively—retained during model training to preserve contextual meaning but excluded during exploratory visualization.

D. Aspect and Sentiment Labelling

Each review was annotated with one or more of five service aspects: cleanliness, service, room comfort, food & beverages, and core facilities. For each aspect, the sentiment was categorized into three classes: positive, neutral, or negative. The labeling process was conducted semi-automatically with the assistance of large language models (LLMs) such as ChatGPT and Gemini, and subsequently verified manually to maintain accuracy and consistency.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

E. Dataset Splitting

The annotated dataset was partitioned into three subsets for model training and evaluation. Specifically, 80% of the data was used for training, 10% for validation, and the remaining 10% for testing. This distribution was intended to promote effective learning, reduce the risk of overfitting, and enable reliable assessment on unseen samples.

F. Fine-Tuning IndoBERT

The IndoBERT-base model was fine-tuned to handle multi-label sentiment classification across different aspects. A classification layer was added on top of the model, employing a sigmoid activation function to generate probability scores for each element–sentiment pair. Training was carried out for 20 epochs using the AdamW optimizer with a learning rate of 1e-5. To prevent overfitting and improve generalization, early stopping was applied based on validation loss.

G. Evaluation Metrics

Model performance was assessed using commonly adopted classification metrics: accuracy, precision, recall, and F1-score. The evaluation was performed individually for each aspect. Additionally, confusion matrices and classification reports were produced to gain more detailed insights into the model's predictive performance across sentiment categories.

III. RESULT AND DISCUSSION

This section outlines the research findings, starting with the characteristics of the prepared dataset and then presenting the performance results of the fine-tuned IndoBERT model. The discussion further examines the model's effectiveness across various aspects and sentiment categories, as well as patterns observed across different hotels.

A. Preprocessing Result

An initial dataset of 2,500 raw hotel reviews was processed to prepare it for sentiment classification. During preprocessing, one review was removed due to empty or corrupted content, resulting in 2,499 valid entries. The preprocessing pipeline included several stages designed to clean and standardize the data for model input. Table I summarizes the results of each step.

TABLE I. EXAMPLE PREPROCESSING

Cleaning	"✨Hotel terbaik!! 👍" → "Hotel terbaik"
Case Folding	"Hotel Terbaik" → "hotel terbaik"
Tokenization	"Hotel Terbaik" → "hotel terbaik"
Normalization	"bgt" → "banget"; "nginap" → "menginap"
Stopword Removal	"dan", "yang", "di" retained for semantic structure

B. Aspect and Sentiment Annotation

Each review was manually labeled with one or more of the five predefined aspects. Sentiment polarity (positive, neutral, or negative) was assigned for each aspect present in the review. Labeling was done using a semi-automated approach: ChatGPT and Gemini LLMs provided initial suggestions, which were then manually verified for accuracy and consistency.

The resulting annotated dataset consisted of 12,495 aspect-level sentiment entries (2,499 reviews × up to 5 aspects). Distribution across sentiment classes was highly imbalanced:

TABLE II. ASPECT AND SENTIMENT ANNOTATION

Aspect	**Label**		
	Positive	**Neutral**	**Negative**
Cleanliness	1.984	443	72
Service	2.142	292	65
Room Comfort	2.046	372	81
Food & Beverages	2.068	371	60
Core Facilities	2.030	387	82
Totally	**10.270**	**1.865**	**360**

This distribution confirms the dominance of positive sentiment and the relative scarcity of negative expressions, which is typical for publicly posted hotel reviews.

C. Dataset Spliting

For balanced training and evaluation, the annotated dataset was split into training, validation, and test sets at 80:10:10. The split was performed at the aspect–sentiment level, treating each sentiment-aspect pair in a review as a separate entry. This approach enabled the model to learn and assess sentiment classification for each aspect independently.

TABLE III. DATASET SPLITING

Dataset	**Aspect**	**Pos**	**Neut**	**Neg**
Training 80%	Cleanliness	1.586	353	60
	Service	1.723	224	52
	Room Comfort	1.633	301	65
	Food & Beverages	1.653	297	49
	Core Facilities	1.627	307	65
Validation 10%	Cleanliness	201	42	7
	Service	208	36	6
	Room Comfort	209	32	9
	Food & Beverages	209	34	7
	Core Facilities	205	35	10
Testing 10%	Cleanliness	197	48	5
	Service	211	32	7
	Room Comfort	204	39	7
	Food & Beverages	206	40	4
	Core Facilities	198	45	7
Totally		**10.270**	**1.865**	**360**

Table III illustrates the distribution of data across the five hotel service aspects and the three sentiment classes (positive,

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

neutral, and negative) within each subset. As indicated, the dataset is skewed, with positive sentiments overwhelmingly represented in all aspects. This imbalance poses a challenge for the model, especially in effectively capturing the characteristics of the underrepresented negative class.

D. Fine-Tuning IndoBERT

The IndoBERT-base model was fine-tuned on the training set for 20 epochs, utilizing the AdamW optimizer with a learning rate of 1e-5. Early stopping, guided by validation loss, was employed to minimize overfitting. During training, both accuracy and loss on the training and validation sets were consistently tracked to monitor model performance.

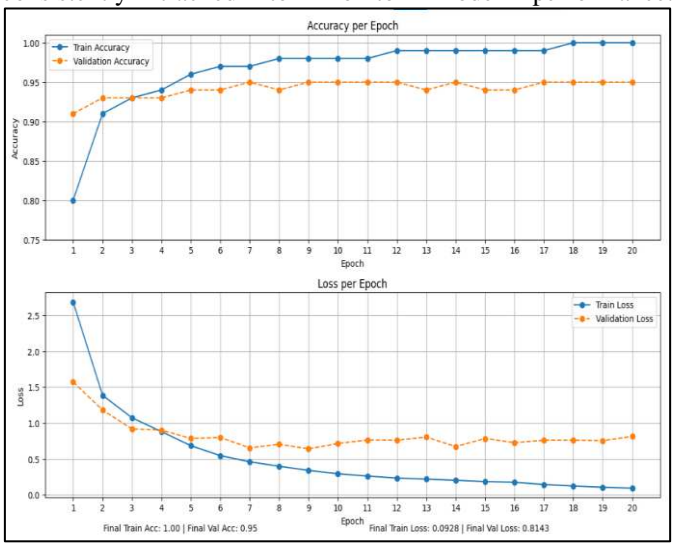

Fig. 2. Model Performance Graph

The upper graph in Figure 2 shows that training accuracy steadily increased and reached 100% by epoch 20. In comparison, validation accuracy plateaued at approximately 95%, suggesting a stable training process with no signs of overfitting. The lower graph illustrates a consistent decline in training loss to 0.0928 and a fluctuating yet controlled validation loss that peaked at 0.8143. These trends indicate that the model successfully converged and maintained generalization capability during training.

E. Evaluation Metrics

Upon completion of training, the model was assessed on the reserved test set, which included 1,264 aspect–sentiment instances. The evaluation utilized four standard classification metrics—accuracy, precision, recall, and F1-score—with macro-averaging applied to address class imbalance. The model's test-time prediction loss was also measured.

The evaluation results indicated that the model attained a test accuracy of 0.95. The macro-averaged F1-score was 0.82, accompanied by a recall of 0.80 and a precision of 0.86. Furthermore, the overall test loss was recorded at 1.0177.

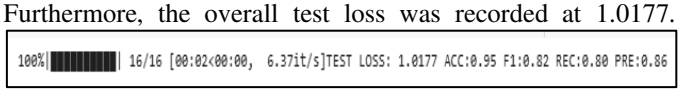

Fig. 3. Evalution Models

The findings indicate that the model achieved strong performance in predicting sentiment across multiple aspects of Indonesian hotel reviews. The high precision score indicates a low false-positive rate. In contrast, the lower recall suggests that some accurate sentiment labels—especially those from the minority negative class—were not fully captured. Even so, the high F1-score, combined with the low loss value, demonstrates that the model generalized well to unseen data and effectively performed multi-label sentiment classification.

In addition to the IndoBERT fine-tuned model, baseline methods from prior studies were considered for comparison. For example, Pratama et al. [4] reported that CNN achieved an accuracy of 89.83% on hotel reviews, outperforming IndoBERT (77.22%) for aspect mapping. Similarly, Cahyaningtyas et al. [4] found CNN and LSTM architectures reaching around 90% accuracy in sentiment and aspect classification. Traditional machine learning methods such as SVM have also shown strong performance, with Munir et al. [14] achieving 95% accuracy on hotel reviews using TF-IDF + SVM. Despite these results, the IndoBERT model in this study achieved 95.28% accuracy with a macro F1-score of 82.44%, demonstrating that fine-tuned Transformer models are more robust in handling multi-aspect sentiment classification across multiple hotels in real-world contexts.

To address class imbalance, class weighting was applied during training. This strategy ensured that the minority class (negative sentiment) was given higher importance, thereby reducing bias toward the majority (positive) class. Although negative sentiment performance remained lower, this mitigation improved recall and F1-score compared to training without class weighting.

F. Confusion Matrix

In this research, confusion matrices were used to assess the model's ability to classify aspect-based sentiment across five hotel service dimensions: cleanliness, service, room comfort, food & beverages, and core facilities. Each aspect was evaluated separately, with sentiment categorized into three classes: positive, neutral, and hostile.

1) Confusion Matrix Aspect Cleanliness

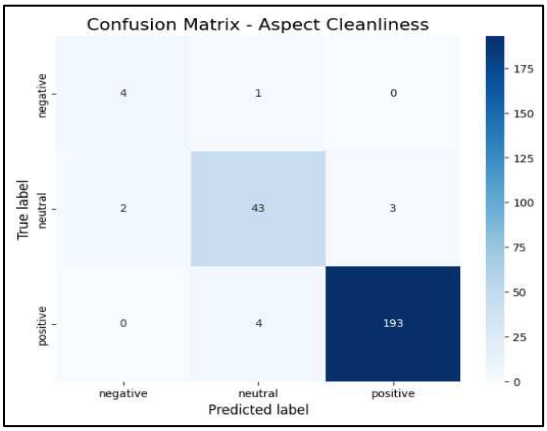

Fig 4. Confusion Matrix Aspect Cleanliness

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

The confusion matrix for the Cleanliness aspect (Figure 4) shows that the model performed well at classifying positive sentiment, correctly identifying 193 reviews. Only four positive reviews were misclassified as neutral, and none as negative. For neutral sentiment, 43 reviews were correctly predicted, while two were misclassified as negative and three as positive. Meanwhile, out of five negative reviews, four were correctly classified and one was misclassified as neutral.

2) Confusion Matrix Aspect Service

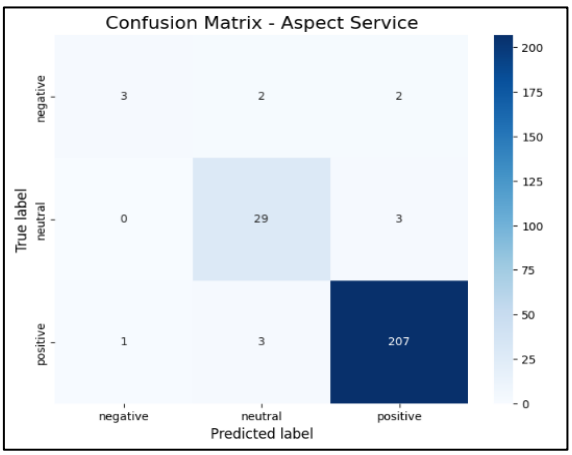

Fig 5. Confusion Matrix Aspect Service

The confusion matrix for the Service aspect (Figure 5) indicates strong performance in identifying positive sentiment, with 207 reviews correctly classified. Only four positive reviews were misclassified—three as neutral and one as negative. For neutral sentiment, the model correctly predicted 29 reviews, with three misclassified as positive. From the seven negative reviews, three were correctly classified, while two were labeled as neutral and two as positive.

3) Confusion Matrix Aspect Room Comfort

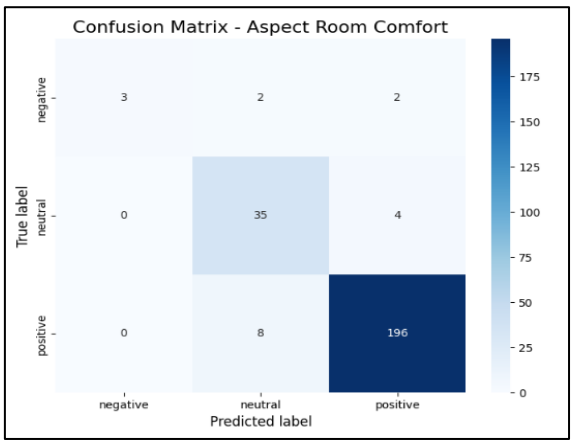

Fig 6. Confusion Matrix Aspect Room Comfort

The confusion matrix for the Room Comfort aspect (Figure 6) demonstrates that the model predicted positive sentiment with high accuracy, achieving 196 correct predictions. However, eight positive reviews were misclassified as neutral, and none as negative. For neutral sentiment, 35 entries were correctly predicted, while four were misclassified as positive. Out of seven negative reviews, the model correctly identified three, while two were misclassified as neutral and two as positive.

4) Confusion Matrix Aspect Food & Beverages

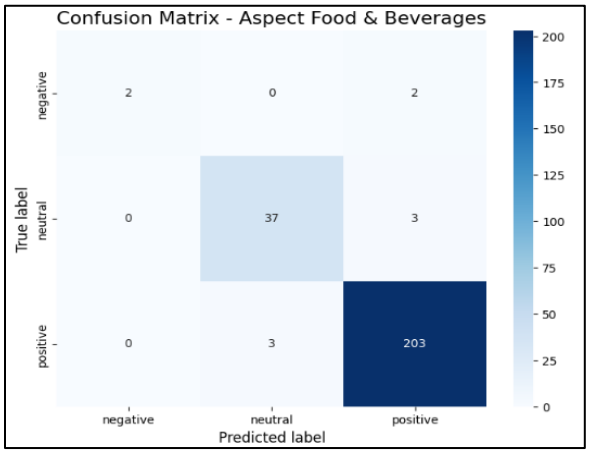

Fig 7. Confusion Matrix Aspect Food & Beverages

The confusion matrix for the Food & Beverages aspect (Figure 7) shows very accurate predictions for positive sentiment, with 203 reviews correctly identified. Three positive reviews were misclassified as neutral, and none as negative. For neutral sentiment, 37 entries were correctly predicted, while three were misclassified as positive. Out of four negative reviews, the model correctly identified two, with two misclassified as positive.

5) Confusion Matrix Aspect Core Facilities

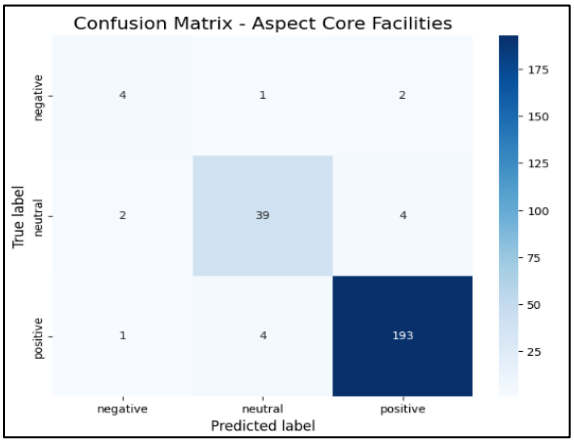

Fig 8. Confusion Matrix Aspect Core Facilities

The confusion matrix for the Core Facilities aspect (Figure 8) illustrates that the model predicted positive sentiment very accurately, with 193 correct predictions. Four positive reviews were misclassified as neutral and one as negative. For neutral sentiment, 39 entries were correctly predicted, while four were misclassified as positive and two as negative. Out of seven negative reviews, the model correctly identified four, with one misclassified as neutral and two as positive.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

Overall, the confusion matrices indicate that the model excels in identifying positive sentiment across all five hotel service aspects. Minor misclassifications occurred mainly between positive and neutral classes. Further fine-tuning could improve accuracy in distinguishing subtle expressions and mixed feelings.

G. Classification Report

The classification report is an evaluation tool that presents precision, recall, F1-score, and support for each sentiment class. Precision reflects the correctness of predictions; recall measures the model's ability to capture relevant cases; the F1-score is the harmonic mean of precision and recall; and support denotes the number of actual samples. In this research, the classification report was used to assess the model's performance in categorizing aspect-based sentiment (positive, neutral, negative) across five service dimensions, offering an overview of its accuracy and reliability.

1) Classification Report Aspect Cleanliness

Classification Report Aspect Cleanliness

	precision	recall	f1-score	support
negative	0.6667	0.8000	0.7273	5
neutral	0.8958	0.8958	0.8958	48
positive	0.9847	0.9797	0.9822	197
accuracy			0.9600	250
macro avg	0.8491	0.8918	0.8684	250
weighted avg	0.9613	0.9600	0.9605	250

Fig 9. Classification Report Aspect Cleanliness

As shown in Fig. 9, the model performed very well in predicting sentiment related to cleanliness, reaching an overall accuracy of 96.00%. The highest performance was obtained in the positive class with an F1-score of 0.9822, followed by the neutral class at 0.8958. The negative class, however, showed weaker results with an F1-score of 0.7273. The macro averages for precision, recall, and F1-score were 0.8491, 0.8918, and 0.8684, respectively, while the weighted averages were 0.9613, 0.9600, and 0.9605, respectively.

2) Classification Report Aspect Service

Classification Report Aspect Service

	precision	recall	f1-score	support
negative	0.7500	0.4286	0.5455	7
neutral	0.8529	0.9062	0.8788	32
positive	0.9764	0.9810	0.9787	211
accuracy			0.9560	250
macro avg	0.8598	0.7720	0.8010	250
weighted avg	0.9543	0.9560	0.9538	250

Fig 10. Classification Report Aspect Service

As illustrated in Fig. 10, the model achieved strong performance in classifying sentiment for the service aspect, with an overall accuracy of 95.60%. The positive class achieved the best result with an F1-score of 0.9787 from 211 samples, followed by the neutral class with an F1-score of 0.8788 from 32 samples. The negative class performed the weakest, reaching an F1-score of 0.5455, despite having relatively high precision (0.7500). This was primarily due to its low recall (0.4286) and the limited number of data points (7). The macro averages for precision, recall, and F1-score were 0.8598, 0.7720, and 0.8010, respectively, whereas the weighted averages were 0.9543, 0.9560, and 0.9538, respectively.

3) Classification Report Aspect Room Comfort

Classification Report Aspect Room Comfort

	precision	recall	f1-score	support
negative	1.0000	0.4286	0.6000	7
neutral	0.7778	0.8974	0.8333	39
positive	0.9703	0.9608	0.9655	204
accuracy			0.9360	250
macro avg	0.9160	0.7623	0.7996	250
weighted avg	0.9411	0.9360	0.9347	250

Fig 11. Classification Report Aspect Room Comfort

Figure 11 shows that the model delivered good performance on the room comfort aspect, with an overall accuracy of 93.60%. The positive class achieved the best result with an F1-score of 0.9655 from 204 entries, followed by the neutral class at 0.8333 from 39 entries. The negative class showed the lowest performance, scoring an F1-score of 0.6000, despite perfect precision (1.0000). The low recall (0.4286) and small sample size (7) were contributing factors. Macro averages for precision, recall, and F1-score were 0.9160, 0.7623, and 0.7996, respectively, while the weighted averages were 0.9411, 0.9360, and 0.9347, respectively.

4) Classification Report Aspect Food & Beverages

Classification Report Aspect Food & Beverages

	precision	recall	f1-score	support
negative	1.0000	0.5000	0.6667	4
neutral	0.9250	0.9250	0.9250	40
positive	0.9760	0.9854	0.9807	206
accuracy			0.9680	250
macro avg	0.9670	0.8035	0.8574	250
weighted avg	0.9682	0.9680	0.9667	250

Fig 12. Classification Report: Food & Beverages

As shown in Fig. 12, the model achieved excellent performance on food and beverage reviews, attaining an overall accuracy of 96.80%. The positive class achieved the highest F1-score of 0.9807 from 206 data points, followed by the neutral class with an F1-score of 0.9250 from 40 data points. The negative class recorded the weakest performance at 0.6667, despite perfect precision (1.0000), due to its low recall (0.5000) and the minimal dataset (4). The macro averages for precision, recall, and F1-score were 0.9670, 0.8035, and 0.8574, respectively, while the weighted averages were 0.9682, 0.9680, and 0.9667, respectively.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

5) Classification Report Aspect Core Facilities

Classification Report Aspect Core Facilities

	precision	recall	f1-score	support
negative	0.5714	0.5714	0.5714	7
neutral	0.8864	0.8667	0.8764	45
positive	0.9698	0.9747	0.9723	198
accuracy			0.9440	250
macro avg	0.8092	0.8043	0.8067	250
weighted avg	0.9437	0.9440	0.9438	250

Fig 13. Classification Report: Aspect Core Facilities

According to Fig. 13, the model achieved an overall accuracy of 94.40% in predicting sentiment related to core facilities. The positive class performed best with an F1-score of 0.9723 from 198 samples, followed by the neutral class with an F1-score of 0.8764 from 45 samples. The weakest results were observed in the negative class, which achieved F1-scores of 0.5714 for both precision and recall, due to its small sample size (7). The macro averages for precision, recall, and F1-score were 0.8092, 0.8043, and 0.8067, whereas the weighted averages reached 0.9437, 0.9440, and 0.9438.

H. Sentiment Distribution Analysis

The overall sentiment distribution analysis aims to understand customers' general perception of five-star hotels. The sentiments are categorized as positive, neutral, or negative based on sentiment classification of customer reviews.

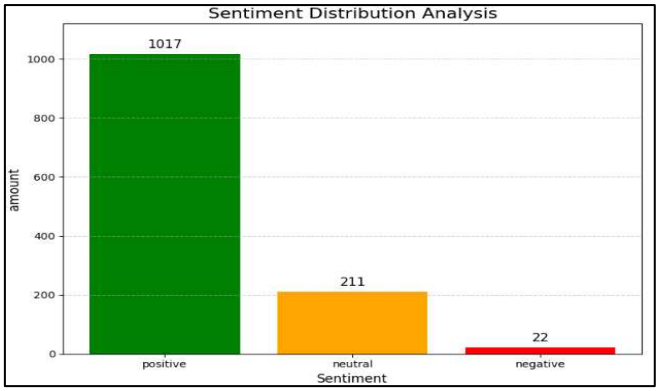

Fig 14. Sentiment Distribution Analysis

Based on the overall sentiment distribution, positive sentiment dominates with 1,017 reviews, followed by 211 neutral and 22 negative. This dominance indicates that customers are generally satisfied with the services and facilities provided by the five-star hotels studied.

The findings suggest that service quality meets or exceeds customer expectations across many aspects. However, neutral and negative sentiments should not be overlooked. Neutral sentiments may reflect ordinary experiences or a lack of strong impressions, while negative sentiments, though minimal, require further investigation to identify sources of dissatisfaction. This analysis provides an initial overview of customer satisfaction and serves as a foundation for a more in-depth study of each service aspect.

I. Dominant Positive Aspect for Hotel

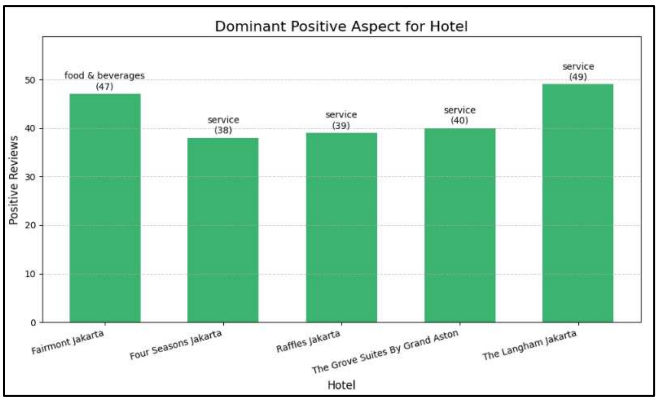

Fig 15. Dominant Positive Aspect

The bar chart in Figure 15 illustrates the dominant positive aspect for each five-star hotel. At Fairmont Jakarta, the food and beverage aspect leads with 47 positive reviews. At Four Seasons Jakarta, service dominates, with 38 positive reviews. Similarly, service is the most praised aspect at Raffles Jakarta and The Grove Suites by Grand Aston, with 39 and 40 positive reviews, respectively. The Langham Jakarta also highlights its service, with 49 positive reviews. Overall, the findings show that service is the most frequently praised aspect across most hotels, except at Fairmont Jakarta, where food and beverages lead in positive sentiment.

J. Dominant Neutral Aspect for Hotel

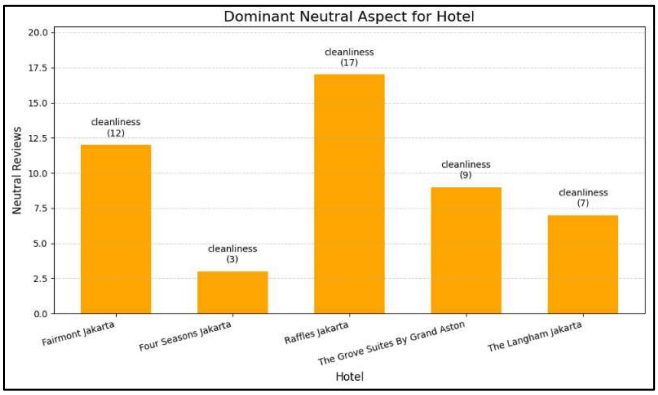

Fig 16. Dominant Neutral Aspect

The bar chart in Figure 16 presents the dominant neutral aspect for each five-star hotel. At Fairmont Jakarta, the cleanliness aspect leads with 12 neutral reviews. Four Seasons Jakarta also shows cleanliness as the dominant neutral aspect with three reviews, while Raffles Jakarta records 17 neutral reviews for cleanliness, making it the most noted aspect. The Grove Suites by Grand Aston and The Langham Jakarta also highlight cleanliness as the leading neutral aspect, with 9 and 7 reviews, respectively. These findings suggest that cleanliness consistently receives neutral feedback across all hotels, indicating areas where customer expectations are met but not significantly exceeded.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

K. *Dominant Negative Aspect for Hotel*

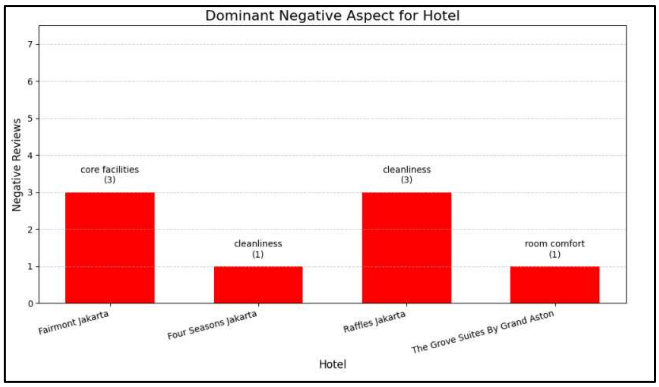

Fig 17. Dominant Negative Aspect for Hotel

The bar chart in Figure 17 displays the dominant negative aspect for each five-star hotel. At Fairmont Jakarta, the core facilities received the most negative feedback, with 3 reviews. Cleanliness emerges as the primary concern at Four Seasons Jakarta and Raffles Jakarta, with 1 and 3 negative reviews, respectively. At The Grove Suites by Grand Aston, room comfort is the main issue, with one negative review recorded. Meanwhile, The Langham Jakarta shows no dominant negative aspect in the chart, indicating the absence of significant negative feedback. Overall, while negative reviews are relatively few, the most recurring concerns are related to cleanliness and core facilities, signaling areas that may require managerial attention.

IV. CONCLUSION

This research implemented aspect-based sentiment analysis (ABSA) with a fine-tuned IndoBERT model to examine customer reviews of five-star hotels in Jakarta. The model attained an overall accuracy of 95.28%, with macro-level precision, recall, and F1 Scores of 0.86, 0.80, and 0.82, respectively, demonstrating strong classification performance across sentiment classes. The analysis revealed that positive sentiment was predominant, especially in service and food & beverages, while neutral opinions were most common in cleanliness. Although limited in proportion, negative sentiment was mainly associated with core facilities and room comfort.

Compared with baseline methods such as CNN, LSTM, and SVM, the proposed fine-tuned IndoBERT demonstrates superior robustness in handling multi-aspect sentiment classification across multiple hotels. To address class imbalance, class weighting was applied, which improved recall and F1-score for the minority negative class.

By explicitly analyzing reviews across multiple five-star hotels, this study addresses the gap in prior research that focused mainly on single datasets or domains. The findings provide actionable insights for hotel management to enhance service quality, particularly in cleanliness and facilities, and demonstrate the effectiveness of Transformer-based models for real-world sentiment analysis in the Indonesian hospitality sector.

REFERENCES

[1] E. Wibowo and I. Pratama, "Analisis Sentimen Terhadap Ulasan Hotel Melalui Platform Google Review Menggunakan Metode Stacking," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 4, pp. 774–784, 2024, doi: https://doi.org/10.47233/jteksis.v6i4.1475.

[2] H. Chyntia Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 1702–1708, 2022, doi: https://doi.org/10.36802/jnanaloka.2022.v3-no1-33-38.

[3] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019, doi: https://doi.org/10.48550/arXiv.1810.04805.

[4] F. P. Pratama, Indriati, and A. Ridok, "Perbandingan Metode IndoBERT Dengan CNN Untuk Analisis Sentimen Berbasis Aspek Terhadap Ulasan Pelanggan (Studi Kasus : Hotel Indonesia Kempinski Jakarta Pada Website Travel Agent Tiket . Com)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 1, pp. 1–10, 2025, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14292

[5] C. A. Bahri and L. H. Suadaa, "Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 17, no. 1, p. 79, 2023, doi: https://doi.org/10.22146/ijccs.77354.

[6] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: https://doi.org/10.3390/app15010172.

[7] A. K. Febriany, E. Dyar Wahyuni, and R. Permatasari, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Indrive Menggunakan Bidirectional Encoder Representations From Transformers (Bert)," *J. Ilm. Wahana Pendidik.*, vol. 10, no. 20, pp. 105–115, 2024, doi: https://doi.org/10.5281/zenodo.14263982.

[8] Tarwoto, R. Nugroho, N. Azka, and W. S. R. Graha, "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 495–505, 2025, doi: https://doi.org/10.35870/jtik.v9i2.3340.

[9] D. Nuryadi *et al.*, "Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3577–3583, 2025, doi: 10.36040/jati.v9i2.13204.

[10] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, pp. 391–403, 2022, doi: https://doi.org/10.26555/ijain.v8i3.691.

[11] S. Cahyaningtyas, D. Hatta Fudholi, and A. Fathan Hidayatullah, "Deep Learning for Aspect-Based Sentiment Analysis on Indonesian Hotels Reviews," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 6, no. 3, pp. 239–248, 2021, doi: https://doi.org/10.22219/kinetik.v6i3.1300.

[12] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *J. Buana Inform.*, vol. 14, no. 02, pp. 107–116, 2023, doi: https://doi.org/10.24002/jbi.v14i02.7244.

[13] A. Nadeem *et al.*, "Resolving ambiguity in natural language for enhancement of aspect-based sentiment analysis of hotel reviews," *PeerJ Comput. Sci.*, vol. 11, pp. 1–29, 2025, doi: 10.7717/PEERJ-CS.2635.

[14] A. Munir, E. P. Atika, and A. D. Indraswari, "Analisis Sentimen pada review hotel menggunakan metode pembobotan dan klasifikasi," *Jnanaloka*, vol. 3, no. 1, pp. 33–38, 2022, doi: 10.36802/jnanaloka.2022.v3-no1-33-38.

[15] T. D. Salma, M. F. Kurniawan, R. Darmawan, and A. Basri, "Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Nusantara pada Perayaan Kemerdekaan Indonesia yang Pertama," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 757–764, 2025, doi: https://doi.org/10.35870/jtik.v9i2.3535.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025

[16] R. Aulia *et al.*, "Pengembangan Model Transformer untuk Analisis Sentimen dan Ringkasan Ulasan E-commerce Berbahasa Indonesia Abstrak," in *Seminar Nasional Inovasi Vokasi (SNIP)*, 2025, pp. 1213–1224. [Online]. Available: https://prosiding.pnj.ac.id/index.php/sniv/article/view/4188

[17] A. Putra, Ismarmiaty, and Apriani, "Pengukuran Tingkat Akurasi Pada Ulasan E-Commerce Menggunakan Metode INDOBERT Dengan Optimizer Adam," *J. Komputer, Inf. dan Teknol.*, vol. 5, no. 1, p. 14, 2025, doi: https://doi.org/10.53697/jkomitek.v5i1.2448.

[18] M. P. Firdaus and D. Trisnawarman, "Analisis Sentimen Publik terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 359–368, 2025, doi: https://doi.org/10.57152/malcom.v5i1.1744.

[19] 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: https://doi.org/10.55338/jikomsi.v7i1.2895.

[20] G. Hakim, T. N. Fatyanosa, and A. W. Widodo, "Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 10, pp. 1–10, 2024, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14291.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i4.2491, Copyright ©2025
Submitted: August 5, 2025, Revised: August 26, 2025, Accepted: September 1, 2025, Published: October 15, 2025