Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Aspect-Based Sentiment Analysis of Google Maps Reviews For Barbershop Competitor and Brand Strategy Analysis Using Machine Learning

Nugroho Kumala Destianto[1], Ryan Randy Suryono*[2], Heni Sulistiani[3]
[1,2,3]Master of Computer Science, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Article Info

Article history:
Received July17, 2026
Revised Aug 26, 2026
Accepted Sep 27, 2026

Keywords:
ABSA,
Google Maps,
Barbershop,
IndoBERT,
Machine learning,
Competitor analysis

ABSTRACT

Customer reviews on Google Maps contain information regarding service experiences but are unstructured and may discuss multiple aspects within a single review. This study applies Aspect-Based Sentiment Analysis (ABSA) to analyze customer perceptions across 23 barbershops and compare the performance of Naive Bayes, Support Vector Machine (SVM), Random Forest, and IndoBERT. A total of 5,418 reviews were collected from the selected barbershops and processed through cleaning, preprocessing, aspect identification, and aspect-level lexicon-based sentiment labeling. After cleaning and deduplication, 5,299 reviews generated 7,457 aspect-level observations classified into positive, neutral, and negative sentiment. Naive Bayes, SVM, and Random Forest used TF-IDF representations, while IndoBERT used contextual transformer-based representations. Evaluation using accuracy, precision, recall, Macro F1-score, and ROC-AUC showed that IndoBERT achieved an accuracy of 98.57%, precision of 93.28%, recall of 93.99%, Macro F1-score of 93.55%, and ROC-AUC of 99.88%. General, Result Quality, and Service were the most frequently identified aspects, while Cleanliness, Comfort, and Service had the highest proportions of positive sentiment. Waiting Time and Price showed relatively higher proportions of negative sentiment. A Wilson Score Lower Bound was used as supplementary descriptive information for comparing the relative positions of the 23 barbershops. Because sentiment labels were automatically generated using a lexicon-based procedure, the classification and competitor-analysis results represent performance and descriptive patterns based on weak labels rather than manually annotated ground truth.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Nugroho Kumala Destianto,
Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia,
Jl. Zainal Abidin Pagaralam No. 9-11, Labuhan Ratu, Bandar Lampung.
Email: nugroho_kumala@teknokrat.ac.id

1. INTRODUCTION

Advances in digital technology have transformed the way customers share their experiences and evaluations of products and services. Customer reviews on digital platforms consist not only of ratings but also describe the experiences, satisfaction levels, and shortcomings customers perceive after using a service. This information can serve as a valuable source for evaluation because it directly reflects customer perceptions. Utilizing review data through text analysis can transform unstructured information into a more actionable format for understanding customer perceptions (Pranatawijaya et al., 2024) One platform that provides such data is Google Maps. In addition to providing location and business information, Google Maps allows users to give ratings and write reviews after visiting a place. These reviews can cover various aspects of the customer experience, making them a potential tool for evaluating service quality. Bahri and Suadaa utilized Google Maps reviews to conduct Aspect-Based Sentiment Analysis (ABSA) on Bromo Tengger Semeru National Park and identify sentiment regarding attractions, facilities, accessibility, and pricing (Bahri & Suadaa, 2023). (Setyawan et al., 2024) also used Google Maps reviews with the IndoBERT model for sentiment analysis at the aspect level and achieved an accuracy of 90.39% and an F1-score of 71.20% based on 5-fold cross-validation.

This situation is particularly relevant to the barbershop business, which relies heavily on customer experience. ABSA is particularly relevant because a single review can address multiple aspects with differing sentiments. For example, a customer may express positive sentiment toward the

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

haircut result while expressing negative sentiment toward the price or waiting time. If reviews are analyzed solely based on a single overall sentiment label, information regarding the strengths and weaknesses of specific service aspects may go unnoticed. ABSA allows sentiment to be associated with specific aspects, thereby generating more targeted information for service evaluation (Apriliani et al., 2025). Customer evaluations of barbershop businesses can cover haircut results, barber service, price, cleanliness, comfort, facilities, location, and waiting time. In this context, each identified aspect needs to be evaluated based on its corresponding local textual context rather than assigning a single sentiment label to the entire review. Indonesian-language reviews also present challenges such as the use of non-standard language, abbreviations, spelling variations, and context-dependent expressions. Febrianto et al. applied ABSA using IndoBERT to visitor reviews of Baturraden through the stages of case folding, cleansing, tokenization, normalization, stemming, and stopword removal. The study achieved an accuracy of 94.61%, a precision of 83.22%, a recall of 96%, and an F1-score of 88.11%, and demonstrated that data imbalance across aspects is one of the challenges in the analysis (Febrianto et al., 24 C.E.).

In addition to transformer-based models, conventional machine learning is still used in text classification with feature representations such as TF-IDF. Naive Bayes, Support Vector Machine (SVM), and Random Forest have different classification characteristics, so their performance may vary across datasets. A sentiment analysis study on Gojek user reviews in Google Play Store used Multinomial Naive Bayes and Logistic Regression with text preprocessing stages, TF-IDF and Count Vectorizer weighting, and SMOTE to address class imbalance (Hidayanah et al., 2026). Another study on 6,000 user reviews of the Jamsostek Mobile app compared SVM and Naive Bayes by applying text preprocessing, TF-IDF weighting, and lexicon-based labeling. The results showed that SVM achieved an accuracy of 88.9%, while Naive Bayes achieved 64.1% (Tria Setyani, 2026). Meanwhile, a sentiment analysis study on Indonesian-language YouTube comments showed that SVM achieved an accuracy of 77.4% and an F1-score of 77.6%, while The development of transformer models offers an alternative by leveraging textual context. IndoBERT is designed to handle the characteristics of the Indonesian language and has been applied to various types of text data. A study on Google Maps reviews for “Rio by The Beach” used IndoBERT and achieved an accuracy of 89.21% and a weighted F1-score of 0.90 (Saputra et al., 2026), while a study on 10,000 TikTok comments related to “Save Raja Ampat” achieved an accuracy of approximately 71% (Eko Putro et al., 2025).

In ABSA, IndoBERT also achieved an accuracy of 91.48% and an F1-score of 71.56%, which is higher than that of SVM, which achieved an accuracy of 89.16% (Nugroho et al., 2025). Meanwhile This indicates that model selection can influence the performance of aspect-based sentiment analysis. ABSA can also be used to compare customer perceptions across brands. Research on Digikala reviews shows that ABSA is capable of identifying satisfaction based on specific aspects while simultaneously comparing service quality with that of competitors (Safiyeh Samadanian, 2024). However, class imbalance remains a concern as it can affect classification performance (Setiawan, 2024). Although ABSA has been applied to Google Maps reviews and various Indonesian-language review datasets, previous research has largely focused on a single object or brand. This approach has demonstrated that ABSA can be used to compare multiple brands, but its application to multiple barbershops simultaneously still has room for improvement. This study addresses this gap by integrating ABSA, comparing conventional machine learning models with transformer models, and conducting a competitor analysis across 23 barbershops. The ABSA results are used to map the strengths and weaknesses of service aspects and to compare the relative positions of competitors using the Wilson Score Lower Bound. This study utilized Google Maps reviews from 23 barbershops with four classification methods: Naive Bayes, SVM, Random Forest, and IndoBERT. Naive Bayes, SVM, and Random Forest employed TF-IDF, SMOTE on the training data, and Stratified Group K-Fold Cross-Validation, while IndoBERT used a 3-fold scheme. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC to obtain a comprehensive comparison of performance.

The contribution of this study lies in the application of ABSA results to the analysis of barbershop competitors. The analysis was used to identify service aspects that constitute the strengths and weaknesses of each barbershop and to compare them across competitors. The Wilson Score Lower Bound was used as supplementary information in the ranking based on the proportion of positive sentiment and the amount of data, with lexicon-based labels serving as the basis for descriptive analysis. This study aims to analyze sentiment based on specific aspects, compare the performance of

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

the four models, and identify positive and negative service aspects as a basis for service evaluation and brand strategy development.

2. RESEARCH METHODOLOGY

2.1 Research Design

This study uses an experimental approach to perform Aspect-Based Sentiment Analysis (ABSA) on Google Maps reviews of barbershops in Bandar Lampung. The research objects consist of 23 barbershops that met the minimum criterion of 40 customer reviews. These 23 barbershops represent barbershops in the Bandar Lampung area that met the specified review-availability criterion. The minimum number of reviews was used as a descriptive data-availability criterion to ensure that each barbershop provided sufficient review data for aspect-level analysis and comparison. The research data consisted of customer reviews obtained from Google Maps. Data collection was conducted in August 2026 through a crawling process using the Easy Scraper tool. The crawling process was performed on the Google Maps pages of the selected 23 barbershops to collect available customer review data. The collected reviews were then combined into a single dataset for subsequent analysis. The initial crawling process resulted in 5,418 review records. The collected data were subsequently examined for data completeness and duplicate records before being used in the analysis.

The research stages consist of data collection, data preprocessing, aspect identification, aspect-specific sentiment labeling, data splitting, classification modeling, and model evaluation and analysis. Four classification models were used in this study: Naive Bayes, Support Vector Machine (SVM), Random Forest, and IndoBERT. Naive Bayes, SVM, and Random Forest represent conventional machine learning approaches, while IndoBERT represents a transformer-based approach. The use of customer review data as a source for ABSA has been applied in previous studies to identify sentiment based on service-related aspects (Setiawan, 2024) (Kontonatsios et al., 2023); making this type of data relevant for analyzing customer perceptions of barbershop services. The collected reviews were subjected to text preprocessing, including cleaning, case folding, normalization, tokenization, stopword removal, and stemming. After preprocessing, barbershop service aspects were identified using predefined aspect keywords. For each identified aspect, the local textual context surrounding the aspect keyword was extracted and used to determine the sentiment associated with that aspect. Therefore, when a review contained multiple aspects, each aspect could have its own sentiment label based on its corresponding local context. The resulting aspect-level data were then divided into training and validation sets using Stratified Group K-Fold Cross-Validation based on the original review identifier (id_ulasan). This grouping ensured that aspect-level observations originating from the same review were not distributed across different folds. The grouped splitting strategy was applied to reduce the risk of data leakage resulting from multiple aspect observations originating from the same original review. The aspect-level data were subsequently used to train and evaluate Naive Bayes, SVM, Random Forest, and IndoBERT models. Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix. The resulting model performance and aspect-level sentiment distributions were then analyzed to describe customer perceptions and compare the service aspects of the 23 barbershops. Figure 1 illustrates the sequential research stages used in this study, beginning with data collection and preprocessing, followed by aspect identification and aspect-specific sentiment labeling, data splitting, classification modeling, and evaluation and analysis.

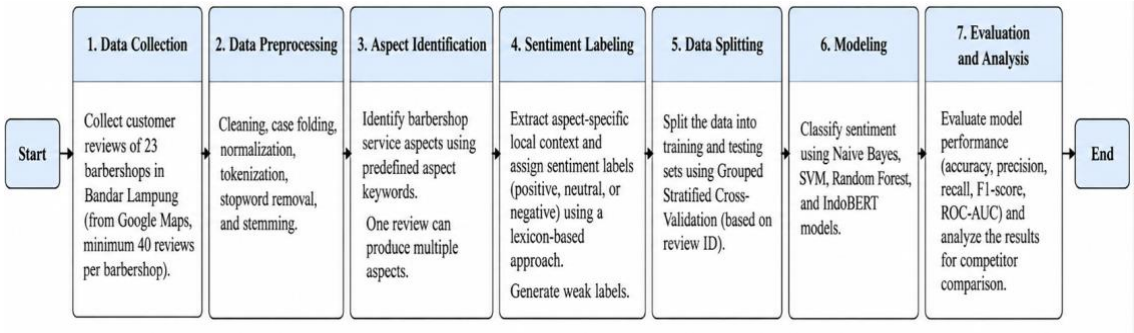


Figure 1. Research Stages

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

As shown in Figure 1, the research began with the collection of Google Maps reviews from the selected barbershops, followed by text preprocessing. The preprocessed reviews were then used for aspect identification and aspect-level context formation, followed by aspect-specific sentiment labeling. The resulting aspect-level data were divided using grouped cross-validation and used for classification with Naive Bayes, SVM, Random Forest, and IndoBERT. Finally, model performance, aspect-level sentiment distributions, and inter-barbershop comparisons were analyzed.

2.2 Pre-processing and Aspect Data Formation

After the review data were collected, text preprocessing was performed to obtain a standardized representation suitable for aspect identification and sentiment analysis. The preprocessing stages consisted of cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Cleaning was performed to remove unnecessary characters and non-informative elements from the review text. Case folding converted all alphabetic characters into lowercase form. Normalization was used to standardize non-standard words, abbreviations, and spelling variations commonly found in Indonesian-language reviews. Tokenization divided the review text into individual tokens, followed by stopword removal to eliminate words with limited contribution to the analysis. Finally, stemming reduced inflected words to their base forms. These preprocessing stages follow the text-processing procedure applied in a previous ABSA study using TF-IDF and SVM (Nurul Hidayati et al., 2023). After preprocessing, aspect identification was performed using predefined aspect keywords. The keywords represented service aspects relevant to barbershop customer experiences, namely quality of results, general impressions, service, comfort, cleanliness, price, waiting time, facilities, and location. The keyword-based approach was used to identify the service aspects explicitly mentioned in each review. A single review could contain more than one service aspect. Therefore, each identified aspect was converted into a separate aspect-level observation while retaining the original review identifier (id_ulasan). For example, a review containing information about haircut results, service, and price could generate three separate aspect-level records. This approach allowed multiple service aspects discussed within the same review to be represented separately in the analysis.

To ensure that sentiment was assigned specifically to the relevant aspect, an aspect-specific local textual context was extracted for each identified aspect. The local context consisted of the words surrounding the identified aspect keyword within the preprocessed review. Consequently, different aspects identified in the same review could have different textual contexts. Sentiment was then determined independently from each aspect-specific local context rather than assigning a single overall sentiment label to the entire review. For example, when a customer review expressed a positive opinion about the haircut result but a negative opinion about the price, the result-quality aspect was evaluated using the local context surrounding the result-related expression, while the price aspect was evaluated using its corresponding local context. Therefore, the positive sentiment associated with the haircut result was not automatically assigned to the price aspect. This procedure maintained the relationship between each identified aspect and the sentiment expressed toward that aspect. Sentiment labeling was performed using a lexicon-based approach with three sentiment classes: positive, neutral, and negative. The sentiment polarity was determined based on sentiment-bearing terms identified within the aspect-specific local context. A positive label represented a favorable expression toward the corresponding aspect, whereas a negative label represented an unfavorable expression. A neutral label was assigned when the aspect-specific context did not provide a sufficiently clear positive or negative polarity according to the lexicon-based scoring procedure. The resulting sentiment labels were used as the target classes for the classification models. The sentiment labels generated through this procedure were considered weak labels because they were automatically generated using a sentiment lexicon rather than obtained through manual annotation by human annotators. Therefore, these labels were not treated as manually annotated ground truth. The classification results of Naive Bayes, SVM, Random Forest, and IndoBERT were consequently interpreted as the models' performance against the automatically generated weak-label classes. After aspect identification and sentiment labeling, the resulting aspect-level dataset contained the original review identifier, barbershop identifier, identified aspect, aspect-specific local context, and sentiment label. The id_ulasan field was retained as a grouping variable because one review could generate multiple aspect-level observations. During data splitting, all aspect-level observations originating from the same id_ulasan were kept within the same training or validation fold. This grouped splitting strategy was applied to reduce the risk of

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

information leakage caused by multiple observations originating from the same original review and to provide a more reliable estimate of model performance.

2.3 Modeling and Evaluation

Modeling was performed using two approaches: conventional machine learning and transformer-based models. The conventional models consisted of Naive Bayes, Support Vector Machine (SVM), and Random Forest, while the transformer-based model used IndoBERT. For the conventional models, the aspect-specific textual context was converted into numerical features using Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF assigns weights to terms based on their frequency in a document and their relative importance across the corpus. The resulting TF-IDF representation was used as input for Naive Bayes, SVM, and Random Forest. Naive Bayes is used as a probabilistic classification model with TF-IDF features (Azizah et al., 2024). SVM was used to determine the decision boundary between sentiment classes and is suitable for text data with a high-dimensional feature space. Hidayati et al. applied SVM with TF-IDF for ABSA classification and evaluated the model using K-fold cross-validation (Ipmawati et al., 2024). Random Forest was used as an ensemble learning method that combines multiple decision trees to generate the final prediction. Previous research has also applied Random Forest with SMOTE to address class imbalance in sentiment analysis (Mualfah et al., 2022). Unlike the conventional models, IndoBERT is a transformer-based model developed for Indonesian-language processing that uses contextual representations to capture relationships between words. The use of transformer-based models represents an approach for handling contextual information in Indonesian-language sentiment analysis (Geni et al., 2023). Class imbalance in the conventional machine learning models was addressed using the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE generates synthetic samples for minority classes to improve the balance of the training data. In this study, SMOTE was applied only to the training portion of each cross-validation fold, while the validation data were not oversampled. Applying SMOTE only to the training data prevents information from the validation data from being used during the oversampling process and reduces the risk of data leakage. The application of SMOTE for handling class imbalance in sentiment analysis has also been reported in previous studies (Suandi dkk., 2024 ; Patrycia Dewi dkk., 2024).

Model evaluation was performed using Stratified Group K-Fold Cross-Validation. This evaluation strategy maintains the distribution of sentiment classes across folds while ensuring that aspect-level observations originating from the same id_ulasan remain within the same fold (Ridson Al Farizal P, 2023). This grouping strategy is important because one review can generate multiple aspect-level observations. Naive Bayes, SVM, and Random Forest were evaluated using 5-fold cross-validation, while IndoBERT was evaluated using 3-fold cross-validation. For the conventional models, SMOTE was applied independently to the training data within each fold. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy measures the proportion of correctly classified observations among all observations. Precision measures the proportion of correctly predicted observations among observations predicted as a particular class, whereas recall measures the ability of the model to correctly identify observations belonging to a particular class. The F1-score represents the harmonic mean of precision and recall. Because this study involves three sentiment classes, namely positive, neutral, and negative, precision, recall, and F1-score were calculated for each class and summarized using the macro average. Per-class evaluation was included to provide a more detailed assessment of model performance, particularly for the minority negative class.

ROC-AUC was used to evaluate the ability of the models to distinguish among the three sentiment classes. Because the classification problem is multiclass, ROC-AUC was calculated using the One-vs-Rest (OvR) approach, in which each sentiment class is evaluated against the remaining classes. The reported ROC-AUC represents the macro-averaged performance across the three sentiment classes. In addition, a confusion matrix was generated to examine the number and distribution of correct and incorrect predictions for each sentiment class. The sentiment labels used as the target classes were generated through the lexicon-based aspect-level labeling procedure described in Section 2.2. Therefore, the evaluation results represent model performance against automatically generated weak labels rather than performance against manually annotated ground truth. This distinction is considered when interpreting the classification results.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

2.4 Sentiment Analysis and Competitor Analysis

The evaluation results of Naive Bayes, SVM, Random Forest, and IndoBERT were compared based on accuracy, precision, recall, Macro F1-score, and ROC-AUC to examine differences in classification performance. In addition to model evaluation, the aspect-level sentiment data were analyzed to describe customer perceptions of barbershop services. The analysis was conducted by examining the distribution of positive, neutral, and negative sentiment across each identified service aspect and barbershop. The aspect-level analysis covered nine service aspects: result quality, general impressions, service, comfort, cleanliness, price, waiting time, facilities, and location. The distribution of sentiment for each aspect was examined to identify aspects associated with relatively more positive, neutral, or negative customer perceptions. Because the sentiment labels were generated automatically using a lexicon-based approach, these distributions were interpreted as descriptive representations of customer perceptions based on the collected reviews rather than as manually validated ground-truth measurements. Competitor analysis was conducted by comparing the aspect-level sentiment distributions across the 23 barbershops. Each barbershop was examined based on the number and proportion of positive, neutral, and negative aspect-level observations. This comparison was used to describe differences in customer perceptions and to identify service aspects that may represent relative strengths or areas requiring attention for each barbershop. To support the comparative analysis, the Wilson Score Lower Bound was calculated using a 95% confidence level. The Wilson Score considers both the proportion of positive sentiment and the amount of available data, providing a more conservative measure than ranking solely according to the proportion of positive sentiment. In the competitor analysis, a minimum number of aspect-level observations was applied as a descriptive filtering criterion to ensure that the comparison was based on barbershops with sufficient available data. This minimum observation criterion was not interpreted as a statistical significance threshold.

The Wilson Score analysis considered positive sentiment as the favorable outcome, while neutral and negative observations were retained as part of the total number of observations used in the calculation. The competitor comparison therefore considered both the proportion of positive sentiment and the amount of available aspect-level data. The resulting Wilson Score values were used as supplementary descriptive information for comparing the relative position of the barbershops. It should be emphasized that the competitor analysis and Wilson Score calculation used the lexicon-based aspect-level sentiment labels as the basis for descriptive analysis rather than direct IndoBERT predictions. Thus, model evaluation was used to assess the classification performance of the four models against the weak-label classes, whereas competitor analysis was used to describe customer perceptions based on the aspect-level sentiment data. The results were subsequently used to identify service aspects that may contribute to differences in customer perceptions across the competing barbershops.

3. RESULTS AND DISCUSSION

3.1 Data Collection and Exploration Results

The dataset used in this study consisted of customer reviews collected from Google Maps for 23 selected barbershops in Bandar Lampung. Data collection was conducted in August 2026 through a crawling process using the Easy Scraper tool. The crawling process was performed on the Google Maps pages of the selected barbershops to collect available customer review records. The initial crawling process resulted in 5,418 review records. The initial dataset was examined based on attribute structure, data completeness, duplicate records, and the distribution of review counts across the selected barbershops. The collected dataset contained several attributes, including review identifier, barbershop name, reviewer information, review text, review date or time, star rating, and other source-related attributes. The review text was used as the primary textual information for aspect-based sentiment analysis, while the barbershop name was retained as the competitor identifier and the review identifier was retained to identify the original review from which aspect-level observations were generated. Table 1 presents the structure and function of the attributes contained in the initial dataset.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Table 1. Initial Dataset Structure

No.	Attribute	Description
1	review_id	Review identifier
2	barbershop_name	Barbershop name
3	reviewer_name	Reviewer name
4	review	Customer review text
5	review_time	Review date/time
6	star_rating	Customer rating
7	price_range	Price range information
8	reviewer_profile_url	Reviewer profile URL
9	source_file	Data source
10	duplicate_within_barbershop	Duplicate data indicator

Based on Table 1, the review text was the primary attribute used for the analysis, while the barbershop name was retained for competitor comparison. The review identifier was also retained because a single review could generate multiple aspect-level observations. The star_rating attribute was examined as a potential supplementary weak-label validation reference. However, the collected rating data could not be reliably used for this purpose because valid and consistently associated rating values were not available for the review records used in the final analysis. Therefore, the star rating was not used as an independent sentiment label or validation reference. The distribution of review counts across the 23 barbershops was uneven. Barbershop 1 had the highest number of collected reviews, while several other barbershops had substantially fewer reviews. This difference in the amount of available review data was considered in the subsequent competitor analysis. Figure 2 presents the distribution of review counts across the 23 barbershops.

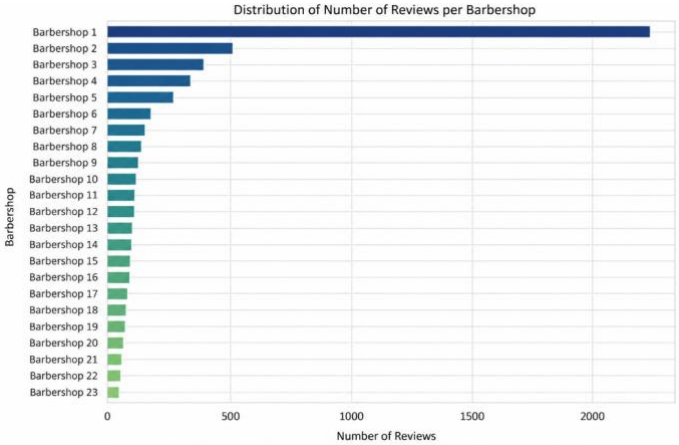


Figure 2. Review Count Distribution Across 23 Barbershops

Figure 2 illustrates the differences in the number of collected reviews among the 23 barbershops.The unequal number of reviews indicates that each competitor had a different amount of available data. Therefore, comparisons based solely on sentiment proportions may be affected by differences in sample size. To complement the aspect-level sentiment comparison, the Wilson Score Lower Bound was subsequently used as supplementary descriptive information by considering both the proportion of positive sentiment and the amount of available aspect-level data.

3.2 Pre-processing Results

The preprocessing stage was conducted to transform the review text into a more uniform format before it was used for aspect identification and sentiment classification. The steps applied included cleaning, case folding, normalization, tokenization, stopword removal, and stemming. These steps were used to reduce unnecessary characters, standardize the text format, handle non-standard words, and produce text more suitable for subsequent analysis. During the cleaning stage, 34 data points were found to be empty because they lacked usable text, so these data points were excluded. Next, a

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

duplication check was performed based on the combination of barbershop names and review text. The check identified 85 duplicate data points, which were then deleted to prevent the same reviews from being counted repeatedly. After all preprocessing steps were completed, data organization was performed at the aspect level. Table 2 shows the changes in the number of data points at each processing stage.

Table 2. Data Count Changes After Pre-processing

Stage	Number
Initial dataset	5,418
Empty data after cleaning	34
Dataset after cleaning	5,384
Duplicates removed	85
Final dataset	**5,299**

Based on Table 2, out of the initial 5,418 reviews, 5,299 were deemed usable for the next stage of analysis. Empty data and duplicates were removed to maintain data quality and avoid duplicate information that could affect the results of aspect identification and sentiment classification.

3.3 Aspect Identification Results

Aspect identification was performed using predefined keywords representing the service aspects relevant to barbershop customer experiences. Each review could contain more than one aspect, and each identified aspect was converted into a separate aspect-level observation while retaining the original review identifier. In the revised procedure, the sentiment associated with each aspect was subsequently determined from its corresponding local textual context rather than from a single overall sentiment label assigned to the review. From the 5,299 reviews retained after preprocessing and deduplication, 3,379 reviews (63.8%) contained at least one identified specific service aspect. The aspect identification process generated 7,457 aspect-level observations. The distribution of the identified aspects is presented in Table 3.

Table 3. Aspect Distribution in Reviews

Aspect	Number	Percentage
General	1,920	25.7%
Result Quality	1,792	24.0%
Service	1,549	20.8%
Comfort	1,218	16.3%
Cleanliness	430	5.8%
Price	257	3.4%
Waiting Time	140	1.9%
Facilities	106	1.4%
Location	45	0.6%
Total	**7,457**	**100%**

Based on Table 3, the General aspect had the highest number of observations, with 1,920 records (25.7%), followed by Result Quality with 1,792 records (24.0%) and Service with 1,549 records (20.8%). Comfort accounted for 1,218 observations (16.3%), while Cleanliness, Price, Waiting Time, Facilities, and Location accounted for smaller proportions of the aspect-level data. The distribution indicates that customer reviews most frequently contained information related to general impressions, haircut results, service, and comfort. Meanwhile, Location and Facilities were mentioned less frequently. The resulting aspect-level structure enables sentiment to be analyzed according to the specific service aspects discussed in each review rather than treating the entire review as a single sentiment observation.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

3.4 Sentiment Labeling Results

After aspect identification, sentiment labeling was performed independently for each aspect-level observation using a lexicon-based approach. The sentiment was determined from the local textual context surrounding the identified aspect keyword. Therefore, when a review contained multiple aspects, each aspect could receive a different sentiment label according to its corresponding context. The three sentiment classes used in the analysis were positive, neutral, and negative. The resulting labels were automatically generated and are therefore considered weak labels rather than ground truth obtained through manual annotation. The sentiment distribution at the aspect level is presented in Table 4.

Table 4. Sentiment Distribution at Aspect Data Level

Sentiment	Number	Percentage
Positive	5,882	78.9%
Neutral	1,427	19.1%
Negative	148	2.0%
Total	**7,457**	**100%**

Based on Table 4, positive sentiment dominated the aspect-level data, accounting for 5,882 observations (78.9%). Neutral sentiment accounted for 1,427 observations (19.1%), while negative sentiment accounted for only 148 observations (2.0%). The distribution indicates a substantial class imbalance, with positive sentiment representing the majority class and negative sentiment representing the minority class. The imbalance was considered during the classification stage. For the conventional machine learning models, SMOTE was applied only to the training data within each cross-validation fold, while the validation data were retained in their original distribution. This procedure was used to reduce the influence of class imbalance during model training without introducing validation-data information into the oversampling process. Because the sentiment labels were generated automatically using a lexicon-based procedure, the classification results should be interpreted as model performance against weak labels rather than against manually annotated ground truth. The sentiment distribution should likewise be interpreted as a descriptive representation of the automatically labeled aspect-level review data.

3.5 Classification Results Using Machine Learning and IndoBERT

Classification was performed using Naive Bayes, Linear Support Vector Machine (SVM), Random Forest, and IndoBERT. Naive Bayes, SVM, and Random Forest used TF-IDF and 5-fold Stratified Group K-Fold Cross-Validation, while IndoBERT used 3-fold Stratified Group K-Fold Cross-Validation. The id_ulasan field was retained as the grouping key to prevent aspect-level observations from the same review from being distributed across different folds. SMOTE was applied only to the training data within each fold. The comparison of model performance is presented in Table 5.

Table 5. Model Performance Comparison

Model	Accuracy	Precision	Recall	Macro F1	ROC-AUC
Naive Bayes	86.56%	62.63%	77.66%	65.91%	94.72%
Random Forest	94.86%	86.97%	77.26%	79.94%	97.45%
SVM Linear	93.95%	86.32%	81.37%	83.50%	97.24%
IndoBERT	98.57%	93.28%	93.99%	93.55%	99.88%

Based on Table 5, IndoBERT achieved an accuracy of 98.57%, precision of 93.28%, recall of 93.99%, Macro F1-score of 93.55%, and ROC-AUC of 99.88%. Among the conventional models, Linear SVM obtained a Macro F1-score of 83.50%, followed by Random Forest at 79.94% and Naive Bayes at 65.91%. The results were evaluated using macro-averaged metrics because the sentiment classes were imbalanced. ROC-AUC was calculated using the One-vs-Rest (OvR) approach and macro-averaged across the three sentiment classes. Since the sentiment labels were automatically generated using a lexicon-based approach, the evaluation results represent model performance against weak labels rather than manually annotated ground truth. To further examine the effect of class

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

imbalance, precision, recall, and F1-score were also evaluated separately for the negative, neutral, and positive classes. The per-class performance is presented in Table 6.

Table 6. Per-Class Classification Performance

Model	Class	Precision	Recall	F1-score
Naive Bayes	Negative	22.85%	68.24%	34.24%
	Neutral	68.03%	72.18%	70.04%
	Positive	96.78%	90.51%	93.54%
SVM Linear	Negative	74.80%	64.19%	69.09%
	Neutral	88.86%	81.64%	85.10%
	Positive	95.48%	97.70%	96.58%
Random Forest	Negative	75.90%	42.57%	54.55%
	Neutral	85.74%	91.45%	88.50%
	Positive	97.52%	97.02%	97.27%
IndoBERT	Negative	83.55%	85.81%	84.67%
	Neutral	96.90%	96.36%	96.63%
	Positive	99.35%	99.42%	99.39%

The per-class results in Table 6 show that the negative class was generally more difficult to classify than the neutral and positive classes. This pattern is particularly evident for Naive Bayes and Random Forest, which obtained lower F1-scores for the negative class than for the other sentiment classes. IndoBERT achieved a negative-class precision of 83.55%, recall of 85.81%, and F1-score of 84.67%. For the neutral class, IndoBERT achieved a precision of 96.90%, recall of 96.36%, and F1-score of 96.63%. The positive class obtained the highest performance, with a precision of 99.35%, recall of 99.42%, and F1-score of 99.39%. These results provide a more detailed assessment of model performance than overall accuracy alone, particularly because the negative class accounted for only 2.0% of the 7,457 aspect-level observations. The per-class metrics therefore indicate how the models performed on both the majority and minority sentiment classes. Nevertheless, because the target labels were generated automatically using a lexicon-based procedure, these metrics should be interpreted as performance against weak labels rather than manually annotated ground truth.

3.6 Confusion Matrix Analysis

The confusion matrix was used to examine the classification patterns of Naive Bayes, Linear SVM, Random Forest, and IndoBERT by comparing actual and predicted sentiment classes. Figure 3 presents the aggregate confusion matrices across all cross-validation folds.

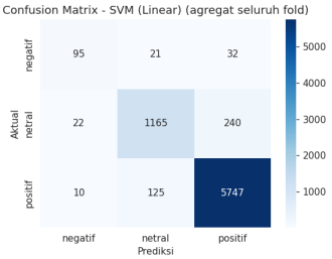


Figure (a)

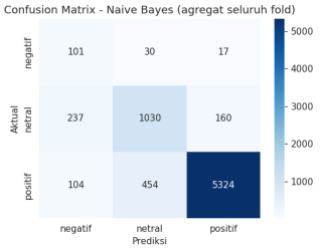


Figure (b)

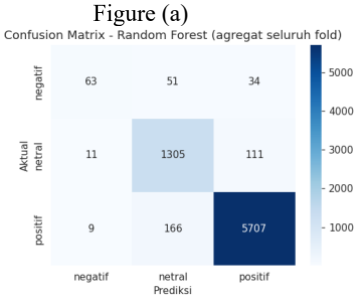


Figure (c)

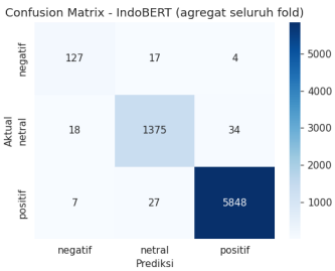


Figure (d)

Figure 3. Aggregate Confusion Matrix Across All Folds: (a) Naive Bayes, (b) SVM, (c) Random Forest, and (d) IndoBERT

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Based on Figure 3, all models correctly classified most observations in the positive class, which represents the majority class (78.9%). SVM correctly classified 95 negative, 1,165 neutral, and 5,747 positive observations. Naive Bayes correctly classified 101 negative, 1,030 neutral, and 5,324 positive observations. Random Forest correctly classified 63 negative, 1,305 neutral, and 5,707 positive observations. IndoBERT correctly classified 127 negative, 1,375 neutral, and 5,848 positive observations. The confusion matrices show that IndoBERT produced the highest number of correct classifications across the three sentiment classes. This result is consistent with its per-class F1-scores of 84.67% for the negative class, 96.63% for the neutral class, and 99.39% for the positive class. The negative class remained more challenging to classify because it represented only 2.0% of the aspect-level data. Since the sentiment labels were automatically generated using a lexicon-based procedure, the confusion matrix results represent model performance against weak labels rather than manually annotated ground truth.

3.7 *Aspect-Based Sentiment Analysis*

Aspect-level sentiment analysis was conducted to examine positive, neutral, and negative sentiment across the nine identified service aspects. Sentiment was determined independently for each aspect using the local textual context surrounding the identified aspect keyword. The resulting labels were automatically generated using a lexicon-based procedure and are therefore considered weak labels rather than manually annotated ground truth.

Table 7. Sentiment Distribution by Aspect

Aspect	Negative	Neutral	Positive
Facilities	2.8%	10.4%	86.8%
Price	8.6%	14.0%	77.4%
Cleanliness	0.2%	0.7%	99.1%
Comfort	0.2%	1.3%	98.5%
Result Quality	2.7%	16.5%	80.7%
Location	0.0%	37.8%	62.2%
Service	1.7%	6.7%	91.5%
General	1.4%	48.2%	50.4%
Waiting Time	12.9%	12.9%	74.3%

Based on Table 7, positive sentiment dominated most service aspects. Cleanliness had the highest positive proportion at 99.1%, followed by Comfort at 98.5% and Service at 91.5%. In contrast, Waiting Time had the highest negative proportion at 12.9%, followed by Price at 8.6%. The General aspect showed a relatively balanced distribution, with 50.4% positive and 48.2% neutral sentiment. These results indicate variations in customer perceptions across service aspects. However, because the sentiment labels were automatically generated using a lexicon-based procedure, the results should be interpreted as descriptive evidence from the collected reviews rather than as manually validated sentiment measurements.

3.8 Inter-Barbershop Sentiment and Competitor Analysis

The inter-barbershop analysis was conducted to describe differences in customer sentiment across the 23 barbershops. Figure 4 presents the proportions of negative, neutral, and positive sentiment for each barbershop.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

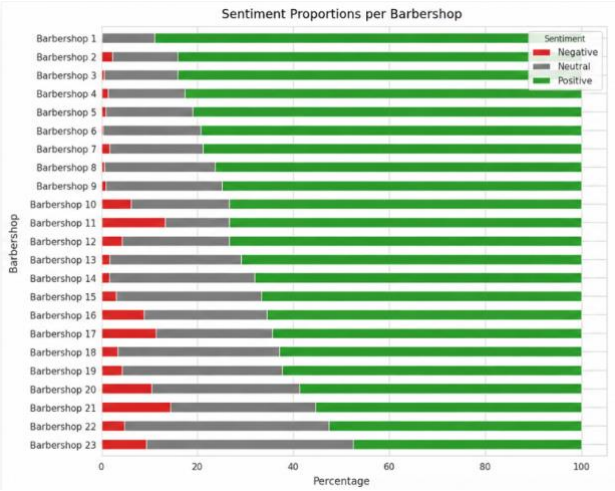


Figure 4. Sentiment Proportions for Each Barbershop

Based on Figure 4, Most barbershops were dominated by positive sentiment, although the proportions of positive, neutral, and negative sentiment varied across competitors. To examine differences at the aspect level, net sentiment was calculated as the difference between the number of positive and negative observations.

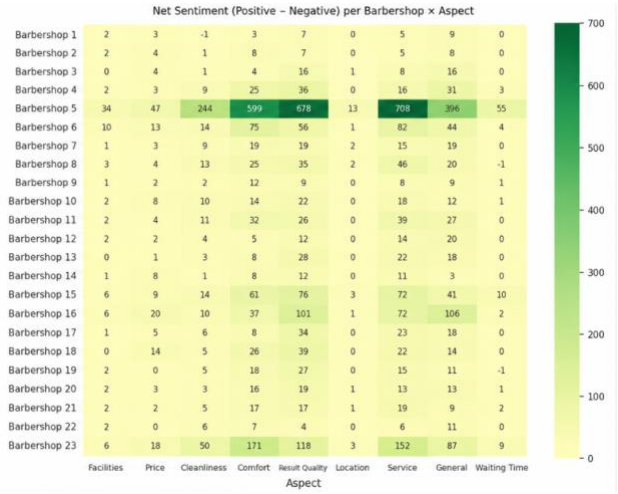


Figure 5. Net Sentiment Heatmap of Barbershops by Aspect

Figure 5 shows variations in net sentiment across barbershops and service aspects. Barbershop 5 showed relatively high net sentiment for Service (708), Result Quality (678), and Comfort (599), while Barbershop 23 showed high values for Comfort (171), Service (152), and Result Quality (118). Negative net sentiment values were also observed, including −1 for Cleanliness in Barbershop 1 and −1 for Waiting Time in Barbershops 8 and 19. Because net sentiment is calculated from the difference between positive and negative observations, its magnitude is also influenced by the amount of available aspect-level data. Therefore, the heatmap is interpreted as a descriptive comparison of sentiment direction and volume rather than as a direct measure of service quality. Overall, the results indicate that customer perceptions varied across barbershops and service aspects.

3.9 Relative Strengths and Weaknesses of Barbershops

The relative strengths and weaknesses of each barbershop were analyzed using the net sentiment rate for each aspect. The net sentiment rate was calculated from the difference between the proportion of positive and negative sentiment for an aspect. To avoid interpreting very small numbers of observations, aspects with fewer than 10 aspect-level observations within a barbershop were excluded from the strength and weakness comparison. This minimum of 10 observations was used only as a descriptive filtering criterion and was not considered a statistical significance threshold.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Table 8. Relative Strengths and Weaknesses of Barbershops

Barbershop	Number of Reviews	Strongest Aspect	Weakest Aspect
Barbershop 1	2,293	Cleanliness	General
Barbershop 2	480	Price	General
Barbershop 3	359	Price	General
Barbershop 4	296	Cleanliness	Price
Barbershop 5	253	Cleanliness	General
Barbershop 6	155	Comfort	General
Barbershop 7	134	Cleanliness	General
Barbershop 8	115	Cleanliness	General
Barbershop 9	110	Price	General
Barbershop 10	99	Comfort	General
Barbershop 11	96	Service	General
Barbershop 12	96	Comfort	Price
Barbershop 13	95	Comfort	General
Barbershop 14	90	Service	General
Barbershop 15	87	Service	General
Barbershop 16	85	Service	General
Barbershop 17	81	Service	General
Barbershop 18	80	Comfort	General
Barbershop 19	80	Cleanliness	General
Barbershop 20	67	Service	Result Quality
Barbershop 21	58	Comfort	Result Quality
Barbershop 22	46	Result Quality	General
Barbershop 23	44	Result Quality	General

Based on Table 8, the identified strengths and weaknesses varied across the 23 barbershops. Cleanliness, Comfort, Service, Price, and Result Quality appeared as the strongest aspects across different barbershops. Meanwhile, General was frequently identified as the weakest aspect, although Price and Result Quality were also identified as the weakest aspects for several barbershops. The results should be interpreted as relative descriptive patterns within each barbershop rather than as absolute measures of service quality. The identified strongest aspect represents the aspect with the highest net sentiment rate among aspects meeting the minimum observation criterion, while the weakest aspect represents the aspect with the lowest net sentiment rate. The number of reviews also varied substantially among the barbershops. Therefore, comparisons of strengths and weaknesses should consider the amount of available review data. Barbershops with fewer observations may produce more variable descriptive patterns than barbershops with larger amounts of review data.

3.10 Barbershop Comparison Using Wilson Score

To complement the aspect-level comparison, the Wilson Score Lower Bound was used to provide a conservative descriptive comparison among the 23 barbershops. Positive sentiment was treated as the favorable outcome, while neutral and negative observations were retained in the total number of aspect-level observations. The calculation used a 95% confidence level. Table 9 presents the number of positive, neutral, and negative aspect-level observations together with the resulting Wilson Score Lower Bound.

Table 9. Wilson Score Lower Bound Comparison of Barbershops

Barbershop	Number of Aspect-Level Observations	Positive	Neutral	Negative	Wilson Score
Barbershop 2	693	616	75	2	0.8633
Barbershop 1	3,332	2,791	524	17	0.8247
Barbershop 5	369	304	60	5	0.7817

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Barbershop	Number of Aspect-Level Observations	Positive	Neutral	Negative	Wilson Score
Barbershop 8	180	151	25	4	0.7782
Barbershop 3	452	357	93	2	0.7499
Barbershop 16	119	96	22	1	0.7266
Barbershop 7	183	144	36	3	0.7220
Barbershop 9	159	121	37	1	0.6891
Barbershop 4	425	310	97	18	0.6853
Barbershop 19	118	88	29	1	0.6603
Barbershop 18	115	84	24	7	0.6429
Barbershop 13	126	89	35	2	0.6217
Barbershop 21	74	54	10	10	0.6191
Barbershop 6	197	131	60	6	0.5965
Barbershop 11	121	82	37	2	0.5901
Barbershop 10	140	90	34	16	0.5606
Barbershop 17	104	68	27	9	0.5584
Barbershop 12	122	76	41	5	0.5344
Barbershop 22	59	37	20	2	0.4995
Barbershop 23	58	34	18	6	0.4580
Barbershop 20	89	49	27	13	0.4473
Barbershop 15	105	55	45	5	0.4291
Barbershop 14	117	55	51	11	0.3820

The Wilson Score values provide a relative descriptive comparison that considers both the proportion of positive observations and the amount of available data. The results show variation in the relative position of the 23 barbershops based on the aspect-level sentiment data. Barbershop 2 had a Wilson Score of 0.8633, followed by Barbershop 1 with 0.8247 and Barbershop 5 with 0.7817. At the other end of the comparison, Barbershop 14 had a Wilson Score of 0.3820, while Barbershop 15 and Barbershop 20 had values of 0.4291 and 0.4473, respectively. The Wilson Score should not be interpreted as a direct measure of overall service quality or as a causal indicator of business performance. Instead, it provides supplementary descriptive information for comparing the relative position of the barbershops based on the automatically generated aspect-level sentiment labels and the amount of available aspect-level data. It is also important to note that the Wilson Score analysis was based on lexicon-generated weak labels rather than direct predictions from IndoBERT. Therefore, the results reflect the descriptive characteristics of the collected and automatically labeled review data. The inclusion of neutral observations in the total sample ensures that the Wilson Score calculation accounts for all aspect-level observations rather than considering only positive and negative observations.

3.11 Discussion

The results show that Aspect-Based Sentiment Analysis (ABSA) can identify customer perceptions based on specific barbershop service aspects. From 5,418 initial reviews collected from 23 barbershops, 5,299 reviews remained after cleaning and deduplication. Aspect identification subsequently generated 7,457 aspect-level observations, indicating that a single review could contain multiple service aspects. Among the classification models, IndoBERT achieved the highest performance, with an accuracy of 98.57%, Macro F1-score of 93.55%, and ROC-AUC of 99.88%. Among the conventional models, Linear SVM achieved the highest Macro F1-score at 83.50%, followed by Random Forest at 79.94% and Naive Bayes at 65.91%. These differences indicate that contextual representation from IndoBERT provided stronger classification performance on the research dataset than the TF-IDF-based conventional models. This finding is consistent with previous studies reporting strong performance of IndoBERT for Indonesian-language ABSA, although performance may vary according to dataset characteristics, class distribution, and labeling methods.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

At the aspect level, General was the most frequently identified aspect (25.7%), followed by Result Quality (24.0%) and Service (20.8%). Based on sentiment proportions, Cleanliness (99.1%), Comfort (98.5%), and Service (91.5%) had the highest positive proportions. In contrast, Waiting Time (12.9%) and Price (8.6%) had relatively higher negative proportions. These results indicate that customer perceptions vary across service dimensions, with Waiting Time and Price showing relatively more negative expressions in the collected reviews. The inter-barbershop analysis also showed variations in sentiment across service aspects. The net sentiment heatmap indicated different patterns among competitors, although its magnitude was influenced by the amount of available aspect-level data. Therefore, the heatmap provides a descriptive comparison of sentiment direction and volume rather than a direct measure of service quality. The Wilson Score Lower Bound was used as supplementary descriptive information by considering the proportion of positive sentiment and the amount of available data. The analysis included positive, neutral, and negative observations in the total number of aspect-level data.

The findings can provide supporting information for evaluating service aspects and identifying areas that receive relatively positive or negative customer perceptions. However, several limitations should be considered. The sentiment labels were automatically generated using a lexicon-based procedure and therefore represent weak labels rather than manually annotated ground truth. The number of reviews also varied across barbershops, while positive sentiment dominated the dataset. Some aspects had relatively few observations, which may limit the stability of comparisons. Therefore, the competitor analysis and Wilson Score results should be interpreted as a descriptive snapshot of customer perceptions based on the available review data rather than as an absolute measure of barbershop quality. Overall, the findings indicate that IndoBERT provided the highest classification performance, while ABSA enabled more detailed analysis of customer perceptions across service aspects and barbershops. The combination of classification, aspect-level sentiment analysis, competitor comparison, and Wilson Score provides complementary information for service evaluation and customer-oriented brand strategy.

4. CONCLUSION

This study applied Aspect-Based Sentiment Analysis (ABSA) to Google Maps reviews from 23 barbershops to identify customer perceptions based on service aspects and compare the performance of Naive Bayes, Linear SVM, Random Forest, and IndoBERT. From 5,418 initial reviews, 5,299 reviews were retained after preprocessing and deduplication, producing 7,457 aspect-level observations. The classification results showed that IndoBERT achieved an accuracy of 98.57%, precision of 93.28%, recall of 93.99%, Macro F1-score of 93.55%, and ROC-AUC of 99.88%. Among the conventional models, Linear SVM achieved the highest Macro F1-score at 83.50%, followed by Random Forest at 79.94% and Naive Bayes at 65.91%.

At the aspect level, General was the most frequently identified aspect (25.7%), followed by Result Quality (24.0%) and Service (20.8%). Cleanliness (99.1%), Comfort (98.5%), and Service (91.5%) had the highest positive sentiment proportions, while Waiting Time (12.9%) and Price (8.6%) showed relatively higher negative sentiment proportions. The inter-barbershop analysis also indicated variations in customer sentiment across service aspects. The Wilson Score Lower Bound was used as supplementary descriptive information by considering the proportion of positive sentiment and the amount of available aspect-level data. These results provide supporting information for evaluating customer perceptions and identifying service aspects that may require attention across competing barbershops. However, the findings are subject to several limitations. The sentiment labels were automatically generated using a lexicon-based approach and therefore represent weak labels rather than manually annotated ground truth. Differences in the amount of review data across barbershops, the imbalance among sentiment classes, and the relatively small number of observations for some aspects should also be considered when interpreting the results. Therefore, the competitor analysis and Wilson Score results should be interpreted as descriptive information based on the available review data rather than as absolute measures of barbershop quality. Future research may improve the analysis by using manually annotated sentiment data, expanding and balancing the dataset, and applying more advanced language models or validated annotation procedures to strengthen the reliability of aspect-level sentiment and competitor analysis.

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to all parties who supported and contributed to the completion of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest related to this research or its publication.

DATA AVAILABILITY

The data used in this study consist of publicly available Google Maps reviews collected from 23 barbershops. The data are available upon reasonable request to the corresponding author, subject to the applicable terms and conditions of the data source.

AUTHOR CONTRIBUTIONS

Nugroho Kumala Destianto: Conceptualization; Data Curation; Data Labeling; Methodology; Software; Writing–Original Draft; Visualization.**Ryan Randy Suryono**: Conceptualization; Methodology; Supervision; Validation; Writing–Review & Editing.**Heni Sulistiani**: Methodology; Supervision; Validation; Writing–Review & Editing.

REFERENCES (11 PT)

Apriliani, S., Erfina, A., & Warman, C. (2025). Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, *14*(4), 437–445. https://doi.org/10.32736/sisfokom.v14i4.2491

Azizah, H., Syuhada, F., & Sa'adati, Y. (2024). SENTIMEN ANALISIS TEMPAT WISATA BERDASARKAN ULASAN GOOGLE MAPS MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Bukit Merese) Sentiment Analysis of Tourist Destinations Based on Google Maps Reviews Using the Naïve Bayes Method (Case Study of Bukit Merese). *SainsTech Innovation Journal*, *7*(2), 467–475. https://doi.org/https://doi.org/10.37824/sij.v7i2.2024.753

Bahri, C. A., & Suadaa, L. H. (2023). Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *17*(1), 79. https://doi.org/10.22146/ijccs.77354

Eko Putro, D., Juarsa, D., Putra Hermana, B., Bagastian, B., & Sulistiani, H. (2025). Analisis Sentimen Publik terhadap 'Save Raja Ampat' di Media Sosial Menggunakan Model IndoBERT. *Bulletin of Computer Science Research*, *5*(5), 1067–1075. https://doi.org/10.47065/bulletincsr.v5i5.621

Febrianto, D. C., Fitriani, M. A., Afrad, M., & Khadija, M. A. (24M). ASPECT BASED SENTIMENT ANALYSIS MENGGUNAKAN INDOBERT MODEL TERHADAP REVIEW PENGUNJUNG OBJEK WISATA BATURRADEN. *Melek IT: Information Technology Journal*, *10*(2), 157–166. https://doi.org/https://doi.org/10.30742/melekitjournal.v10i2.358

Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using Bert Language Models. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, *9*(3), 746–757. https://doi.org/10.26555/jiteki.v9i3.26490

Hidayanah, N., Fitriani, S., Winadya Permadani, I., & Randy Suryono, R. (2026). Analisis Sentimen Ulasan Pengguna Aplikasi Gojek di Google Play Store Menggunakan Metode Multinomial Naive Bayes dan Logistic Regression. *Jurnal Ilmu Komputer, Teknologi Dan Informasi*, *4*(2), 134–142. https://doi.org/10.62866/jurikti.v4i2.333

Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 247–256. https://doi.org/10.57152/malcom.v4i1.1066

Kontonatsios, G., Clive, J., Harrison, G., Metcalfe, T., Sliwiak, P., Tahir, H., & Ghose, A. (2023). FABSA: An aspect-based sentiment analysis dataset of user reviews. *Neurocomputing*, *562*. https://doi.org/10.1016/j.neucom.2023.126867

Mualfah, D., Fadila, W., & Firdaus, R. (2022). Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest. *Jurnal CoSciTech (Computer*

Vol. 6, No. 2 - December 2026, pp. 101-117
DOI: 10.59395/jitp.v6i2.231

Science and Information Technology), *3*(2), 107–113. https://doi.org/10.37859/coscitech.v3i2.3912

Nugroho, R., Azka, N., & Sayudha Rendra Graha, W. (2025). Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *9*(2), 495–505. https://doi.org/https://doi.org/10.35870/jtik.v9i2.3340

Nurul Hidayati, Faqih Hamami, & Riska Yanu Fa'rifah. (2023). Aspect-Based Sentiment Analysis On FLIP Application Reviews (Play Store) Using Support Vector Machine (SVM) Algorithm. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *7*(1), 183–197. https://doi.org/10.31289/jite.v7i1.9768

Patrycia Dewi, N. P. E., Yanti, C. P., & Yusa, I. M. M. (2024). Enchancing K-NN Performance With SMOTE for Sentiment Analysis of Streaming App Reviews. *Building of Informatics, Technology and Science (BITS)*, *6*(3), 1966–1976. https://doi.org/10.47065/bits.v6i3.6190

Pranatawijaya, V. H., Sari, N. N. K., Rahman, R. A., Christian, E., & Geges, S. (2024). Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews. *Journal of Information Systems Engineering and Business Intelligence*, *10*(3), 328–339. https://doi.org/10.20473/jisebi.10.3.328-339

Ridson Al Farizal P, W. A. A. N. (2023). Analisis Sentimen Berbasis Aspek Pada Ulasan Google Maps (Aspect-Based Sentiment Analysis on Google Maps Reviews: A Case Study of Tourism Sector Recovery Strategies Post Covid-19 Pandemic in West Papua Province). *Seminar Nasional Official Statistics, 1*(1), 297–308. https://doi.org/10.34123/semnasoffstat.v2025i1.2483

Safiyeh Samadanian, M. K. (2024). Aspect-Based Sentiment Analysis of After-Sales Service Quality: A Case Study of Snowa and Competitors Using Digikala Reviews. *International Conference on Information and Knowledge Technology (IKT)*, 1–5. https://doi.org/10.1109/IKT65497.2024.10892728

Saputra, Y. B., Yusuf Inonu, O., Destianto, N. K., & Suryono, R. R. (2026). Analisis Sentimen Ulasan Pengunjung Rio by The Beach Menggunakan Model IndoBERT. *Jurnal Pustaka Robot Sister (Jurnal Pusat Akses Kajian Robotika, Sistem Tertanam, Dan Sistem Terdistribusi)*, *4*(2), 147–153. https://doi.org/10.55382/jurnalpustakarobotsister.v4i2.2204

Setiawan, B. (2024). A Review of Sentiment Analysis Applications in Indonesia Between 2023-2024. *Journal Information Engineering and Educational Technology*, *8*(2), 71–83. https://doi.org/https://doi.org/10.26740/jieet.v8n2.p71-83

Setyawan, A. R., Suadaa, L. H., & Yuniarto, B. (2024). Sistemasi: Jurnal Sistem Informasi Aspect-Based Sentiment Analysis using Adaptive Aspect on Tourist Reviews in Jakarta. *Sistemasi: Jurnal Sistem Informasi*, *13*(5), 2456–2466. http://sistemasi.ftik.unisi.ac.id

Suandi, F., Anam, M. K., Firdaus, M. B., Fadli, S., Lathifah, L., Yumami, E., Saleh, A., & Hasibuan, A. Z. (2024). Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning Algorithms. *International Conference on Applied Engineering*, 126–138. https://doi.org/10.2991/978-94-6463-620-8_10

Tria Setyani, K. S. H. N. H. R. R. S. (2026). Analisis Sentimen Pengguna Aplikasi Jamsostek Mobile Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *6*(1), 373–384. https://doi.org/10.57152/malcom.v6i1.2526