Mantik Journal, 9 (4) (2026), ISSN 2685-4236 (Online)

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Analysis of hotel visits in Ambon city using the naive bayes algorithm

Henderina Doren[1], Jermias Victor Manuhutu[2], Yoakhina Nicole Makaruku[3], Wilma Latuny[4], Apritiwi Maimina[5]

[1,2,3,5]Information Systems, Institut Agama Kristen Negeri Ambon, Indonesia

[4]Industrial Engineering, Universitas Pattimura, Indonesia

ARTICLE INFO

Article history:

Received Dec 19, 2025
Revised Jan 13, 2026
Accepted Jan 28, 2026

Keywords:

Classification;
Hotel;
Model Accuracy;
Naïve Bayes;
Online Reviews.

ABSTRACT

The rapid growth of tourism in Ambon City has increased competition among accommodations, necessitating data-driven performance evaluations. Prospective tourists often struggle with unstructured online reviews, while hotel management requires precise insights for improvement. This study aims to systematically classify hotel performance in Ambon City using the Naïve Bayes Algorithm based on reviews from platforms like Agoda and TripAdvisor. Adopting a descriptive quantitative methodology, the study processes and labels performance data as "Good," "Poor," or "Very Good." Findings demonstrate that the Naïve Bayes model is highly effective, achieving 91% accuracy. Evaluation via a Confusion Matrix confirms the model's reliability in predicting majority categories, proving that ratings and reviews are strong satisfaction predictors. While the model faces minor challenges with the "Poor" minority category due to limited data, the study provides strategic value. It offers management guidance for targeted improvements and helps tourists make informed decisions, ultimately enhancing the competitiveness of Ambon's hospitality industry.

Corresponding Author:

Jermias Victor Manuhutu,
Information System,
Institut Agama Kristen Negeri Ambon,
Jalan Dolog, Halong Atas, Halong, Kecamatan Baguala, Kota Ambon, Maluku.
Email: jery.ichigo.manuhutu@gmail.com

1. INTRODUCTION

The rapid development of technology and the internet, particularly social media platforms, has transformed online reviews into a primary source of information for consumers (Filieri, 2016). In the hospitality industry, online reviews play a crucial role in influencing tourists' decision-making processes when selecting hotel accommodations (Xiang et al., 2017). Several studies have shown that reviews on platforms such as

TripAdvisor and Agoda significantly affect hotel reputation, perceived service quality, and booking intentions (Guo et al., 2020; Inan, 2024).

Ambon City, as a central hub of economic and tourism activities in Maluku Province, has experienced a steady increase in visitor arrivals. This growth has driven rapid development in the hotel sector, intensifying competition among accommodation providers. In the context of maritime and coastal tourism, sentiment analysis has been proven to effectively identify the specific needs of visitors and help management improve service standards in these strategic areas (Wulandari & Kurniawan, 2025). Since visitor satisfaction is inherently linked to their stay experiences and perceptions, understanding the factors that shape this satisfaction is essential for hotel management to improve service quality and maintain competitiveness (Hu et al., 2019; Kusuma & Handayani, 2023).

Online reviews contain valuable information in the form of opinions, experiences, and evaluations expressed by guests. These opinions can be systematically extracted through sentiment analysis, which is part of opinion mining—a process that automatically identifies, extracts, and classifies subjective information from textual data (Afif, 2023). Sentiment analysis has been widely applied to hotel review data to identify positive and negative perceptions related to service quality, facilities, price fairness, and location (Guo et al., 2020; Inan, 2024).

This study focuses on sentiment analysis of hotel reviews in Ambon City collected from two major global online travel platforms, Agoda and TripAdvisor, which are among the most influential review platforms in the hospitality industry (Xiang et al., 2017). The analysis considers five key hotel attributes frequently discussed in prior studies: location or accessibility, completeness of facilities, price suitability, numerical review values, and ratings (Filieri, 2016; Guo et al., 2020). These attributes provide a comprehensive representation of hotel performance factors that influence guest satisfaction and booking decisions.

The Naïve Bayes algorithm is employed in this research due to its simplicity, efficiency, and strong performance in sentiment classification tasks (Bustami, 2017; Tripathy et al., 2016). In handling high-dimensional text data, particularly for large-scale datasets, the Multinomial Naïve Bayes variant has proven to be highly effective for classifying visitor review texts on various travel platforms (Utami & Pratama, 2025). Previous studies have demonstrated the effectiveness of Naïve Bayes in hotel review sentiment analysis on platforms such as TripAdvisor and Agoda (Dewi et al., 2025; Duta et al., 2025; Fajar & Hartomo, 2020; Farisi, 2019; Margaretha et al., 2024). Furthermore, recent research in the tourism sector confirms that Naïve Bayes remains highly competitive and reliable for sentiment analysis of priority tourist destinations, even when compared to other robust algorithms such as Random Forest (Hidayat & Santoso, 2024). The accuracy of this algorithm can be further optimized through advanced feature selection and hybrid methods, ensuring reliable results when processing large datasets from platforms like Agoda (Hapsari & Wiguna, 2024). Furthermore, Naïve Bayes has also been successfully applied in various sentiment analysis domains, including online sales reviews, digital services, healthcare, and social media content (Handayani & Rifky, 2025; Hasyim & Darmawan, 2025; Mulyani & Novita, 2022; Rizki et al., 2025). To ensure high classification performance, recent studies emphasize that the stemming stage in data preprocessing directly influences the accuracy of Naïve Bayes in the hotel review context by maintaining feature consistency (Sari & Rahmansyah, 2024).

Despite its advantages, the Naïve Bayes algorithm has limitations, particularly its sensitivity to feature selection. An excessive number of features can increase computational cost and reduce classification accuracy. Recent studies emphasize that careful selection and preprocessing of review attributes, such as stemming, are essential to maintain feature consistency and achieve optimal performance (Sari & Rahmansyah, 2024).

Based on these considerations, this study aims to implement the Naïve Bayes algorithm to classify hotel review sentiments into positive and negative categories using review data from www.agoda.com and www.tripadvisor.com. The use of these platforms is in line with current trends where online review sources, including Google Maps, have become primary benchmarks in analyzing hotel service quality (Putra & Setyanto, 2025). Additionally, recent comparative studies on Google Maps data have shown that the Naïve Bayes model remains highly effective for classifying hotel performance based on real-world user feedback (Pratama & Sari, 2025). The results of this study are expected to provide practical benefits for prospective tourists in selecting suitable accommodations (Jain & Pamula, 2020) and offer strategic insights for hotel management in Ambon City to improve service quality, optimize facilities, and enhance overall guest satisfaction (Hu et al., 2019).

Despite the growing body of research on sentiment analysis in the hospitality sector, studies that apply data mining techniques at the regional level, particularly in emerging tourism destinations such as Ambon City, remain limited. The main novelty of this study lies in its regional-level application of the Naïve Bayes algorithm to systematically classify hotel performance based on online reviews from Agoda and TripAdvisor. By focusing on localized hotel attributes and visitor perceptions, this study provides data-driven insights that bridge academic research and practical decision-making for regional tourism development.

2. RESEARCH METHOD

2.1 Research Type and Approach

This study adopts a descriptive quantitative research approach combined with a machine learning–based classification experiment. The quantitative approach is suitable because the research analyzes structured datasets containing numerical attributes (price, rating, number of reviews) and textual data (review comments) using statistical and computational techniques (Afif, 2023; Putra & Setyanto, 2025). Similar approaches have been widely used in sentiment analysis studies involving hotel and consumer reviews (Guo et al., 2020; Jain & Pamula, 2020).

2.2 Data Source and Dataset

The data used in this study are secondary data obtained from two global online hotel booking and review platforms, Agoda.com and TripAdvisor.com, which are commonly used as data sources in hospitality sentiment analysis research (Inan, 2024; Xiang et al., 2017).

The dataset consists of hotel data located in Ambon City with the following attributes: (a) Hotel Name, (b) Price (Numerical value), (c) Hotel Rating (Numerical value), (d) Number of Reviews (Numerical value), (e) Review Text/Comment (Qualitative data, the main source of sentiment), (f) Category (Class Label): (Good, Moderate/Fair, Poor)

These attributes are consistent with previous studies that examined hotel performance and customer satisfaction using online review data (Filieri, 2016; Guo et al., 2020).

2.3 Data Collection Technique

Data collection was conducted using a structured manual documentation technique. The researcher directly observed and recorded hotel information from Agoda and TripAdvisor pages without using automated web scraping tools. This approach ensured data accuracy and allowed visual verification of each sample, as applied in several prior studies using manual review extraction (Agustinus & Antonius, 2018).

The recorded attributes include overall rating scores, review texts containing positive and negative opinions, room prices for specific periods, hotel location information, and facility completeness. Facility completeness was assessed based on the availability of essential amenities such as Wi-Fi, swimming pools, restaurants, and parking facilities (Hendradjaya & Wijaya, 2017).

2.4 Data Analysis Technique

The data analysis process follows a systematic sentiment analysis approach for consumer reviews, which includes: (Jain & Pamula, 2020): Data Collection: Gathering data from Agoda and TripAdvisor using manual documentation or scraping techniques. Data Preprocessing includes: (a) Data Cleaning: Removing empty and duplicate data. (b) Data Transformation: Converting data types into numerical formats. (c) Labeling: conducted by categorizing hotels based on their average review scores, which is widely recognized as a proxy for hotel performance and customer satisfaction in hospitality studies. Online ratings represent an aggregate customer evaluation of service quality, facilities, and overall experience. Thresholds (>4.5 = Good, 3.5–4.4 = Average, <3.5 = Poor) are applied to distinguish high-performing, average-performing, and below-average hotels, a categorization commonly used in tourism and hospitality analytics

Application of the Naïve Bayes Algorithm. The steps include: (a) Splitting the dataset into training and testing data. (b) Calculating class and feature probabilities using the Gaussian Naïve Bayes model. (c) Evaluating classification performance based on accuracy, as implemented in similar studies (Afif, 2023; Dey et al., 2016).

3. RESULTS AND DISCUSSIONS

3.1 Data Processing

Based on the compilation of the obtained data, cleaning and pre-processing of the data were subsequently performed. The data cleaning and processing stage is a crucial step to ensure the quality and readiness of the data before implementation in the Naïve Bayes Algorithm. Whitespace Removal and Text Consistency: The strip() and lowercase() functions were applied to text-type columns (e.g., Location, Amenities) to eliminate superfluous whitespace and standardize the letter format, ensuring identical entries are not treated as distinct.

3.2 Selected feature coefficients from the linear regression model

The linear regression analysis is included as a supporting analysis to provide interpretative insights into the influence of price-, location-, and facility-related features. While the main focus of this study is hotel performance classification using the Naïve Bayes algorithm, the regression results help explain the contribution of individual features that underpin the classification outcomes.

The primary goal of the modeling process was to identify key predictors that influence hotel pricing. The results of the linear regression analysis, showing the magnitude and direction of influence for each selected feature on the target variable (Hotel Price), can be observed in Figure 3.1.

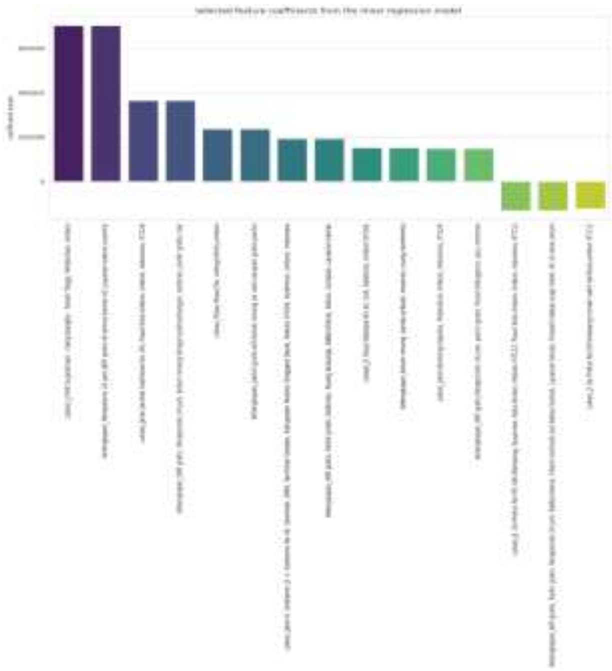

Figure 3.1 Selected feature coefficients from the linear regression model

This visualization aims to measure and represent the magnitude of influence (regression coefficient) of each predictor feature (hotel location, amenity completeness) on the target variable, which is Hotel Price. The goal is to identify the dominant factors that statistically and significantly affect market pricing in the study area.

3.3 Hotel price variation The "Hotel Price Variation"

Descriptive statistical analysis of the Hotel Price variable shows a non-normal distribution. The frequency distribution of hotel prices per night can be visually observed in Figure 3.2.

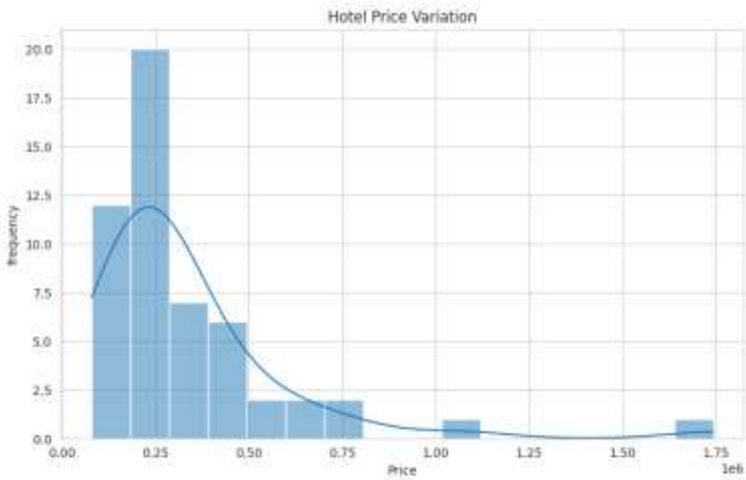

Figure 3.2 Hotel price variation The "Hotel Price Variation"

Based on Figure 3.2, the Hotel Price Variation graph is a combination of a histogram and a density curve (Kernel Density Estimation - KDE), which shows the frequency distribution of hotel prices. This graph indicates that the majority of visits (the highest frequency) are concentrated in the low price range (below Rp 300,000). Statistically, this price distribution shows a right-skewed shape, which means there are extreme values in the form of several very expensive hotels with low visit frequencies (outliers). In the context of Naïve Bayes analysis, the nature of this data requires further processing, such as discretization or transformation, before being used in modeling.

3.4 Hotel Satisfaction Level This figure presents the Model Accuracy

The performance of the Naïve Bayes classification model is evaluated based on its ability to accurately predict the hotel satisfaction categories. The distribution of the satisfaction categories used, as well as the main success metric of the model, can be observed in Figure 3.3.

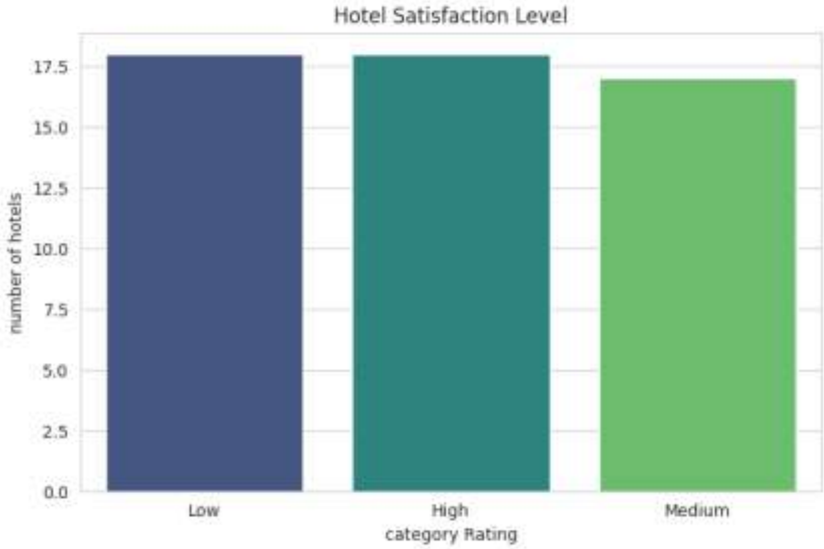

Figure 3. Hotel Satisfaction Level

This figure presents the Model Accuracy metric, which is the main indicator of the success of the Naïve Bayes classification model in predicting hotel performance, with a single value of 0.91. This value means that your model achieves a 91% success rate in classifying each hotel into its correct performance category (Good, Less Than Good, or Very Good). Essentially, this visualization directly confirms that Ratings and Reviews are very strong and reliable predictors for assessing the competitive status of hotels in the Ambon market.

3.5 Top 10 hotel completeness in the city of Ambon This Confusion

To identify priorities and competitive facilities, an analysis of the frequency of hotel facility occurrences in Ambon was conducted. The ten most frequently available hotel facilities in the dataset can be seen in Figure 3.4.

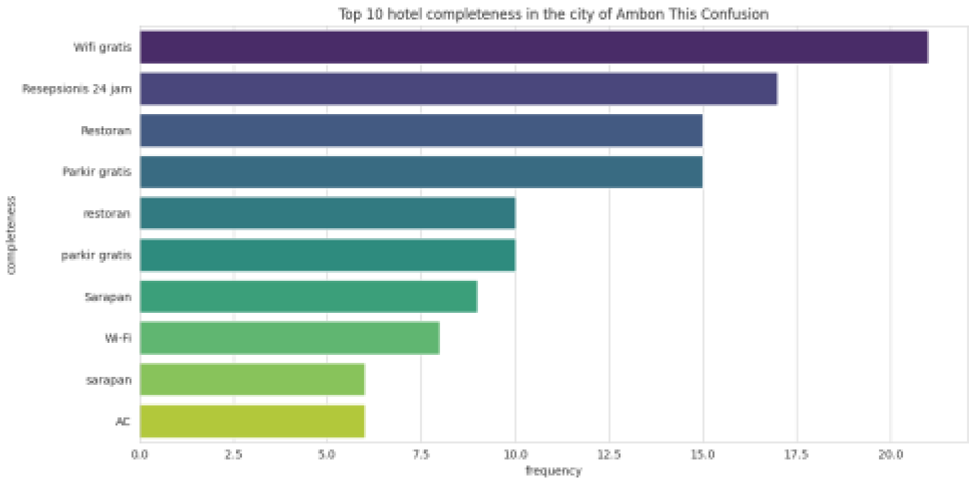

Figure 4 Top 10 hotel completeness in the city of Ambon This Confusion

Matrix serves as a detailed evaluation of the Naïve Bayes classification model, visually comparing the model's predictions with the actual hotel performance. The main diagonal of the matrix confirms the success of the majority of predictions, with 47 'Good' and 40 'Very Good' hotels being classified correctly. However, the matrix highlights that the model faces challenges in the minority category of 'Poor'; of the total 8 'Poor' hotels, 2 were incorrectly predicted as 'Good' (Row 1, Column 0), indicating the model's tendency to be slightly optimistic and its difficulty in accurately identifying the lowest-performing hotels due to the lack of data samples in that category. Nevertheless, this high true positive rate ( ) provides a strong validation basis for the model's 91% accuracy.

3.6 Hotel location profile This visualization is a heatmap of the Confusion

Descriptive statistical analysis of the Hotel Price variable is an initial step to understand the characteristics of the data. The frequency distribution and variation of hotel prices per night in the dataset can be visualized in Figure 3.5.

Figure 3.5 Hotel location profile

This visualization is a heatmap of the Confusion Matrix, which visually measures the accuracy of the Naïve Bayes classification model by comparing the model's predictions (X-axis) against the actual hotel performance (Y-axis). The darkest numbers on the main diagonal confirm the model's dominant success, accurately classifying 47 'Good' hotels and 40 'Very Good' hotels. However, this heatmap also highlights a major weakness in the minority category 'Poor' (Row 1), where 2 hotels that actually performed poorly were incorrectly predicted as 'Good' (Column 0). Although there are other minor errors at the boundary between 'Good' and 'Very Good', the dominant dark color on the diagonal indicates that the model is highly reliable overall (with 91% accuracy) in identifying high-performing hotels.

3.7 Confusion matrix for Naïve BayesClassifier

The performance of the Naïve Bayes classification model is evaluated in depth for each hotel performance category. The matrix that presents the prediction results compared with the actual values, known as the Confusion Matrix, is shown in Figure 6.

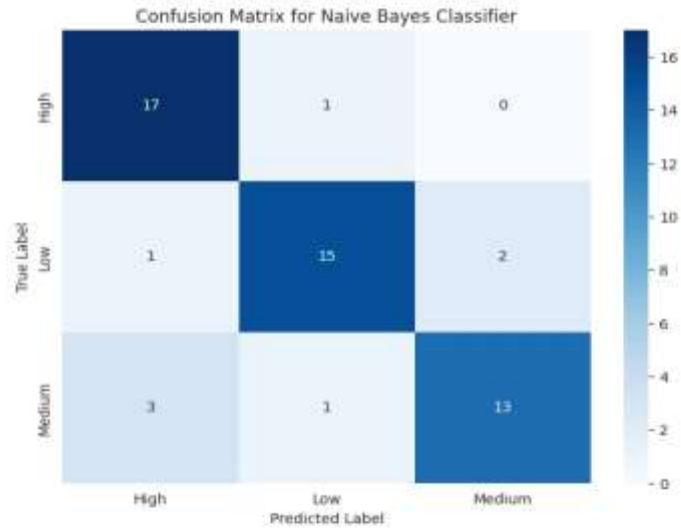

Figure 6 Confusion matrix for Naïve BayesClassifier

This classification chart presents the performance of the Naïve Bayes model in depth for each hotel performance category. The 'Good' (0) category shows the strongest performance (F1-Score 0.95), followed by 'Very Good' (2) with an F1-Score of 0.89. Conversely, the 'Poor' (1) category, which has the smallest number of samples (support) (only 8 hotels), recorded the lowest performance (F1-Score 0.67), mainly due to the low Precision value (0.60), indicating that the model often misclassifies hotels from other categories as 'Poor'. Overall, the model achieved an Accuracy of 0.91 (91%), which is the weighted average of all metrics, confirming the model's general effectiveness in classifying hotel performance, although challenges in the minority category of region of origin may have contributed to the pass rate.

4. Discussion

The findings from this analysis, supported by market segmentation and classification evaluation (Naïve Bayes), indicate clear differentiation in the factors that attract visitors and determine hotel performance in Ambon City. The evaluation of the Naïve Bayes Classifier model shows very high performance, with an Accuracy of 91% [Figure 4] and an F1-Score of 0.95 for the 'Good' category [Classification Figure]. The model's dominant success in accurately classifying 47 'Good' hotels and 40 'Very Good' hotels [Figure 7: Confusion Matrix] confirms that Ratings and Reviews are very strong and reliable predictors for assessing the competitive status of hotels. This superior predictive performance reinforces the Expectancy-Disconfirmation Theory (Oliver, 1980), which states that satisfaction (reflected in high ratings) is the result of hotel performance exceeding visitors' initial expectations, and these expectations are empirically measured through existing features.

However, the Confusion Matrix highlights challenges in the minority category 'Poor' (only 8 samples), where 2 hotels were incorrectly predicted as 'Good' [Figure 7], resulting in the lowest Precision (0.60). This phenomenon suggests the existence of an Imbalanced Class issue in the data, which can inherently cause the classification model to be biased towards the majority classes ('Good' and 'Very Good'). Based on these classification results, it was observed that the model showed a lower performance in identifying the 'Poor' category compared to the 'Excellent' or 'Good' categories, likely due to the limited number of negative reviews in the dataset. To address this issue of class imbalance, future research could implement optimization techniques such as Synthetic Minority Over-sampling Technique (SMOTE) to enhance the model's sensitivity towards minority classes (Ramadhan & Lestari, 2025).

ards the majority classes ('Good' and 'Very Good'). The strong positive coefficients on certain features, such as the combination of a specific location (Jl. WR Supratman) with the availability of air conditioning, room service, and laundry [Figure 1: Selected Feature Coefficients], indicate that these features are the main drivers that increase a hotel's competitive status. This is in line with Central Place Theory (Christaller, 1933), which places accessibility and logistical convenience as fundamental determinants of service success. Hotels with the most basic amenities (free Wi-Fi, 24-hour reception, restaurant) [Figure 5] may be classified as 'Good', but hotels that achieve the 'Very Good' category and the highest scores are those that offer superior services in strategic locations.

Overall, the strong positive correlation between Rating and Satisfaction (even Hotel Satisfaction Levels [Figure 4] but dominated by the 'Good' and 'Very Good' categories) is supported by a reliable classification model. This analysis indicates that hotel management focused on optimizing value (Price-Quality) and providing premium services in strategic locations has the greatest opportunity to achieve the 'Very Good' performance category and attract visits.

5. CONCLUSION

The research on the analysis of hotel visits in Ambon City using the Naïve Bayes algorithm has proven the effectiveness of this classification method in modeling the performance of the local hospitality industry. The validity of the classification model is demonstrated by the Naïve Bayes algorithm's success in classifying hotel performance based on rating, review, location, completeness, and price attributes with a high accuracy rate of 91%, which confirms the model's reliability as an efficient tool for analysis based on online review data. The findings of this study demonstrate that sentiment analysis can serve as a powerful tool for hotel management to evaluate their performance objectively. Furthermore, integrating sentiment analysis with specific facility evaluations provides strategic insights that allow hotel management to identify which aspects of their service require immediate improvement to remain competitive (Kusuma & Handayani, 2023). Furthermore, the rating and review variables were identified as the strongest and most significant predictors in determining hotel performance categories, underscoring the importance of guest perceptions and experiences as the main key to hotel competitiveness in Ambon. Based on the collective analysis of visualization and feature coefficients, it is concluded that improving guest satisfaction must be focused on responsive online review management, enhancing facility standards appropriate to expectations, and setting prices proportional to the value of the service provided.
Nevertheless, this study has several limitations that open opportunities for future research. Further studies are recommended to improve the model’s sensitivity and inferential power by applying data balancing techniques, such as Synthetic Minority Over-sampling Technique (SMOTE), to address the class imbalance issue in the minority “Poor” category. In addition, comparative analysis using other classification algorithms,

including Support Vector Machine, Random Forest, or Logistic Regression, would provide a more comprehensive assessment of model robustness. Finally, expanding the scope of analysis to include multiple cities would enhance the generalizability of the model and allow cross-regional comparisons of hotel performance patterns, thereby strengthening the contribution of this study to regional tourism and data mining literature.

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to Vallens Suitela, Jenifer Latuhihin, Juan Nacikit, and Wulan Rumagia for their assistance, support, and cooperation throughout the process of preparing this journal. Their contributions in terms of time, effort, and ideas have been invaluable in completing this journal successfully. May all the help that has been given be rewarded accordingly.

REFERENCES

Afif, A. (2023). Analisis metode klasifikasi data Naïve Bayes dan SVM dalam menentukan keunikan hotel. *Jurnal Teknologi Komputer Dan Informasi*, *11*(1), 38–46.

Agustinus, B. W., & Antonius, C. R. (2018). Jurnal Ilmiah Teknologi Informasi Asia. *Jurnal Ilmiah Teknologi Informasi Asia*.

Bustami. (2017). Penerapan algoritma Naïve Bayes untuk mengklasifikasi data nasabah asuransi. *Jurnal Informatika*, *6*(2), 97–104.

Dewi, N. N. T. P., Pitanatri, P. D. S., & Adyatma, P. (2025). Analisis sentimen dengan klasifikasi Naïve Bayes pada ulasan Tripadvisor di Luxury Resort untuk strategi peningkatan hunian kamar. *Jurnal Sosial Humaniora Dan Pendidikan*, *4*(1), 788–800.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using Naive Bayes and K-NN classifier. *I.J. Information Engineering and Electronic Business*, *8*, 54–62.

Duta, P. K. A., Pitanatri, P. D. S., & Loanata, C. P. (2025). Big data-based sentiment analysis on TripAdvisor reviews using Naïve Bayes classification. *Media Wisata*, *23*(1), 92–107.

Fajar, M. A., & Hartomo, W. (2020). Analisis sentimen review hotel menggunakan algoritma Naïve Bayes Classifier. *Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, *9*(2), 129–137.

Farisi, A. A. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics*.

Filieri, R. (2016). What makes an online review helpful? *International Journal of Hospitality Management*, *53*, 113–123. https://doi.org/10.1016/j.ijhm.2015.12.001

Guo, Y., Zhang, C., & Lu, S. (2020). Hotel review sentiment analysis based on machine learning: A comparative study. *Journal of Hospitality and Tourism Technology*, *11*(3), 517–531.

Handayani, H., & Rifky, R. M. (2025). Implementasi Naïve Bayes untuk menganalisis sentimen terhadap perilaku pengguna dan kualitas layanan digital radio. *Jurnal Ilmiah Informatika Dan Komputer*, *13*(1), 18–26.

Hapsari, R. D., & Wiguna, A. S. (2024). Analisis Sentimen Ulasan Pengguna Agoda Menggunakan Algoritma Naïve Bayes Berbasis Particle Swarm Optimization. *Jurnal Sistem Informasi Dan Bisnis*, *12*(1), 15–25.

Hasyim, A. N., & Darmawan, I. (2025). Penerapan algoritma Naive Bayes untuk analisis sentimen pengunjung di Pantai Kejawanan. *Jurnal Penelitian Informatika*, *13*(1), 1–7.

Hendradjaya, B., & Wijaya, Y. (2017). Desain visualisasi data. *Jurnal Ilmiah Teknologi Informasi Asia*.

Hidayat, R., & Santoso, B. (2024). Analisis Sentimen Destinasi Wisata Prioritas Menggunakan Perbandingan Naïve Bayes dan Random Forest. *Jurnal Media Informatika Budidarma*, *8*(3), 1540–1549.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analytics using structural topic model. *Economic Modelling*, *83*, 108–120.

Inan, H. E. (2024). *Comparison of machine learning algorithms for classification of hotel reviews: Sentiment analysis of TripAdvisor reviews*.

Jain, P. K., & Pamula, R. (2020). *A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews*.

Kusuma, W. A., & Handayani, T. (2023). Analisis Sentimen Berbasis Aspek Menggunakan Naïve Bayes untuk Evaluasi Fasilitas Hotel di Bali. *Jurnal Sistem Informasi Dan Bisnis*, *11*(2), 89–98.
Margaretha, I. T., Umar, E., & Adis, A. (2024). Analisis sentimen review hotel menggunakan algoritma Naïve Bayes pada Ella Hotel Tambolaka. *Journal of Informatics and Business*, *2*(3), 346–356.
Mulyani, S., & Novita, R. (2022). Implementation of the Naïve Bayes classifier algorithm for classification of community sentiment about depression on YouTube. *Jurnal Teknik Informatika*, *3*(5), 1355–1361.
Pratama, M. A., & Sari, K. (2025). Comparison of Sentiment Analysis on Hotel Reviews Using Naïve Bayes and Support Vector Machine on Google Maps Data. *Journal of Data Science and Artificial Intelligence*, *3*(1), 40–52.
Putra, M. R., & Setyanto, A. (2025). Analisis Sentimen Ulasan Pengguna Google Maps Terhadap Kualitas Layanan Hotel Menggunakan Algoritma Naïve Bayes. *Jurnal Teknologi Informasi Dan Komunikasi*, *14*(1), 45–56.
Ramadhan, G., & Lestari, D. P. (2025). Optimasi Akurasi Sentimen Review Hotel Menggunakan SMOTE pada Algoritma Naïve Bayes. *Jurnal Ilmu Komputer Dan Informasi*, *18*(1), 12–22.
Rizki, A. F., Prihartono, W., & Rohman, F. (2025). Analisis sentimen ulasan Google Maps Rumah Sakit Khalishah di Cirebon dengan algoritma Naive Bayes. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(2).
Sari, I. P., & Rahmansyah, A. (2024). Perbandingan Pengaruh Stemming Terhadap Akurasi Algoritma Naïve Bayes pada Sentimen Review Perhotelan. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *6*(2), 112–120.
Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using machine learning techniques. *Procedia Computer Science*, *57*, 821–829.
Utami, S. F., & Pratama, Y. (2025). Implementasi Multinomial Naïve Bayes untuk Klasifikasi Big Data Ulasan Pengunjung pada Platform Traveloka. *Jurnal Sains Dan Teknologi*, *13*(1), 200–212.
Wulandari, S., & Kurniawan, D. (2025). Analisis Sentimen Pengunjung Hotel di Kawasan Wisata Bahari Menggunakan Klasifikasi Naïve Bayes Berbasis Lexicon. *Jurnal Teknologi Informasi*, *14*(2), 110–121.
Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51–65.