**Jutisi:** Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index
Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
e-ISSN: 2685-0893

# Comparison of Naive Bayes and SVM Algorithms in Sentiment Analysis for the Optimization of Hotel Operational Services in Central Bangka Regency

DOI: http://dx.doi.org/10.35889/jutisi.v14i2.3107



**Iza Guspian[1*], Mursyid Hasan Basri[2]**
Manajemen, Universitas Bangka Belitung, Bangka, Indonesia
*e-mail *Corresponding Author*: iza@ubb.ac.id

***Abstract***
*Customer reviews significantly influence hotel reputation and service improvement, but the large volume and unstructured nature make manual analysis inefficient. This study compares Naive Bayes and Support Vector Machine (SVM) for sentiment classification of 898 Indonesian-language hotel reviews from two online travel agencies, aiming to support service optimization in Central Bangka Regency. Reviews were preprocessed through cleaning, case folding, tokenization, stopword removal, and stemming, then converted into TF-IDF vectors. Both models performed well: SVM achieved the highest accuracy, precision, and F1-score, while Naive Bayes had the highest recall. This indicates that Naive Bayes is better at detecting more positive reviews, whereas SVM provides more precise and balanced classifications. The findings confirm the applicability of classical machine learning for hotel sentiment analysis and offer a basis for implementing real-time review monitoring systems in the hospitality sector.*
***Keywords:*** *Sentiment Analysis; Naive Bayes; Support Vector Machine; Machine Learning; Hotel Customer Reviews*
**Abstrak**
Ulasan pelanggan sangat memengaruhi reputasi hotel dan strategi peningkatan layanan, namun volume yang besar dan sifatnya yang tidak terstruktur membuat analisis manual menjadi tidak efisien. Penelitian ini membandingkan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) untuk klasifikasi sentimen dari 898 ulasan hotel berbahasa Indonesia yang dikumpulkan dari dua agen perjalanan daring, dengan tujuan mendukung optimasi layanan di Kabupaten Bangka Tengah. Ulasan diproses melalui tahap *cleaning, case folding, tokenization, stopword removal,* dan *stemming*, kemudian diubah menjadi vektor TF-IDF. Kedua model menunjukkan kinerja yang baik: SVM mencapai akurasi, presisi, dan F1-*score* tertinggi, sementara Naive Bayes memiliki recall tertinggi. Hasil ini menunjukkan bahwa Naive Bayes lebih baik dalam mendeteksi ulasan positif, sedangkan SVM memberikan klasifikasi yang lebih presisi dan seimbang. Temuan ini menegaskan penerapan algoritma machine learning klasik untuk analisis sentimen hotel dan menjadi dasar pengembangan sistem pemantauan ulasan secara real-time di sektor perhotelan.
**Kata kunci:** *Analisis Sentimen; Naive Bayes; Support Vector Machine; Machine Learning; Ulasan Pelanggan Hotel*

## 1. Introduction

The hospitality industry plays an essential role in supporting regional economic growth, particularly in areas with significant tourism potential such as Central Bangka Regency. In recent years, there has been a significant increase in the number of domestic and international tourists visiting the region, driven by tourism promotions and infrastructure development. This has spurred the growth of local hospitality businesses striving to offer the best possible service to earn customer trust [1]. In today's digital age, a hotel's image and reputation are no longer built solely through conventional media but also through its presence on e-commerce platforms and customer reviews. These reviews serve as a representation of customer experiences that can directly influence the decisions of potential customers and act as a valuable evaluation tool for hotel managers.

With the increasing use of digital platforms, information about customer experiences with hotel services can be obtained in real time and on a massive scale [2, 3]. Unfortunately, the large volume of reviews and their unstructured format make it difficult for hotel managers to perform manual analysis. Customer reviews, which are subjective and varied, also require an analytical approach that can accurately capture the nuances of sentiment [4-6]. Therefore, machine learning-based sentiment analysis methods offer a relevant and efficient alternative. Sentiment analysis itself is the process of identifying and categorizing users' opinions or attitudes toward a topic, typically classified into positive, negative, or neutral categories. This method has proven effective in helping businesses understand customer perceptions and continuously improve service quality [7].

In the field of building systems that classify emotions or sentiments, two popular methods are Naive Bayes and Support Vector Machine (SVM) [8, 9]. Naive Bayes is a straightforward and efficient method that uses probability and Bayes' theorem, and it works well with large amounts of text data. But one of its main issues is that it assumes all features are independent, which isn't always true when dealing with the complicated nature of language. On the other hand, SVM is a type of machine learning that finds the best line or boundary to separate different groups of data in a space with many dimensions. This makes SVM better at handling complex patterns and getting better results in classification [10]. A study by [11] found that SVM performed better, achieving 97% accuracy compared to Naive Bayes' 93% in sentiment analysis on social media posts, showing that SVM is more effective when dealing with difficult and varied language data.

Furthermore, a study by [12] comparing the two algorithms on healthcare sector stock review data shows that SVM has higher performance potential after hyperparameter tuning. In the study, SVM accuracy increased to 85.65% after applying the Grid Search technique, compared to Naive Bayes, which only reached 68.70%. This indicates that SVM is not only superior in terms of algorithmic architecture but also more flexible in terms of adjustment based on data characteristics. Although this study is in a different domain, its approach is relevant for application in the hospitality domain, especially since the type of data involved is similarly customer opinions in the form of free-text reviews. However, the study has not yet examined how each algorithm performs when applied to Indonesian-language review data in the local hospitality services sector.

Based on existing literature reviews, it appears that the majority of previous studies have focused on broader domains such as e-commerce products, stock markets, and social issues on social media [5, 12, 13]. Research specifically highlighting the implementation of sentiment analysis in the local hospitality industry in Indonesia is still very limited. Furthermore, there are few studies that directly examine the effectiveness of Naive Bayes and SVM algorithms on Indonesian-language customer reviews within the context of local language and culture. However, sentence structure, and word choice in Indonesian customer reviews are highly distinctive and can influence classification performance. This highlights an important research gap that needs to be addressed, both in terms of developing suitable models and exploring representative datasets.

This study was carried out to compare the effectiveness of the Naive Bayes and Support Vector Machine algorithms in analyzing sentiment from hotel customer reviews collected from online travel agencies. Data for the review was gathered from two online travel agencies popular among the Indonesian population, concentrating on hotels in Central Bangka Regency. The dataset underwent various data preprocessing stages, such as cleaning, case folding, tokenization, stopwords removal, and stemming. The classification models were subsequently assessed using four primary metrics: accuracy, precision, recall, and F1-score, to deliver a thorough insight into the efficacy of each algorithm. The findings of this research are anticipated to not only add scientifically to the advancement of local sentiment analysis but also provide practical benefits for hotel managers in enhancing services grounded in customer feedback in a methodical and effective way.

With this approach, this study aims to strengthen the decision-making system in hotel service management through the use of analytical technology. On the other hand, this research also contributes to the enrichment of scientific literature in the field of natural language processing in Indonesian, particularly by providing empirical evidence from hotel review data that has rarely been explored. The findings obtained from this study are expected to encourage the development of smarter and more contextual decision support systems in facing customer service challenges in the digital era, while at the same time helping to enhance service quality and improve the competitiveness of the hotel industry in Central Bangka Regency.

## 2. Methodology

This study uses an experimental quantitative approach with a comparative design to compare the performance of two supervised learning algorithms, namely Naive Bayes and Support Vector Machine (SVM), in classifying the sentiment of hotel customer reviews. This study evaluates and compares the accuracy and effectiveness of Naive Bayes and SVM in analyzing Indonesian-language hotel review data

### 2.1. Data Collection

The data for this study were obtained from two online travel agencies in Indonesia, focusing on hotel reviews located in Central Bangka Regency. Data collection was performed using web scraping techniques with the Python programming language. Two libraries were employed: Selenium, to automate browser interaction and dynamically load review pages, and BeautifulSoup, to parse the HTML structure and extract relevant review content. The scraping process was conducted in several steps. First, Selenium was configured with Google Chrome (v139) and Chromedriver of the same version in headless mode to enable automated navigation without a graphical interface. Second, the browser was directed to the target hotel review pages, and a scrolling script was applied repeatedly to ensure all customer reviews were fully loaded. Third, the complete HTML source code of the page was retrieved and parsed using BeautifulSoup to extract review texts along with their contextual elements.

After the reviews were collected, the raw dataset underwent an initial cleaning stage. This stage included removing duplicate entries, filtering out empty reviews, and discarding spam or contextually irrelevant content. A total of 898 valid reviews were obtained that met the inclusion criteria: reviews written in Indonesian, containing complete textual content, and including a star rating as an indicator of user experience. These reviews were then prepared for further preprocessing and sentiment classification.

### 2.2. Text Preprocessing

The collected review data is processed through several stages of text preprocessing to clean and normalize the data so that it is ready for use by the classification algorithm. These stages include:

1. Cleaning: Removing irrelevant symbols, numbers, HTML tags, emoticons, and punctuation marks.
2. Case Folding: Converting all letters to lowercase.
3. Tokenization: Splitting sentences into individual words using nltk.tokenize.
4. Stopword Removal: Removing common words that do not carry sentiment meaning (such as “yang,” “dari,” “dan”) using Sastrawi stopwords.
5. Stemming: Converting words to their base form using the Sastrawi library.

These steps refer to the text preprocessing process proven effective in a study by [11, 14], which can also use Indonesian-language data from social media. In this study, preprocessing was implemented in Python 3.11.13 with the Natural Language Toolkit (NLTK v3.9.1) for tokenization, and the Sastrawi library (v1.0.1) for stopword removal and stemming. All reviews were standardized into a uniform text format to minimize noise and ensure consistency across the dataset. This preprocessing pipeline not only simplifies the feature extraction process but also increases the reliability of classification results. By applying these stages, the dataset becomes more representative and suitable for subsequent modeling with Naive Bayes and SVM algorithms.

### 2.3. Feature Extraction

After preprocessing is complete, the text data is transformed into a numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) technique, which is widely used in text mining to measure the relative importance of a word within a document compared to the entire corpus. The TF-IDF weight for term *t* in document *d is* computed as:

$$TF{-}IDF(t,d) = TF(t,d) \cdot \log \frac{N}{DF(t)} \quad (1)$$

Where:
- *TF(t,d)* refers to how often a term t appears within a document d.
- *DF(t)* represents the count of documents in the corpus that include the term t.
- *N* denotes the overall number of documents contained in the dataset.

In this study, the TF-IDF vectorizer produces a sparse matrix with X unique terms (after preprocessing), each representing a feature in the machine learning model. Words that occur frequently in one review but rarely in others are assigned higher weights, enabling the classifier to detect discriminative terms for sentiment analysis. This technique has been shown to be effective in Indonesian sentiment classification tasks [8, 15]. To illustrate the process, consider the review: "*Tolong kamarnya dibuat seperti rumah… dan balonnya dibuat lebih rapi.*" After applying TF-IDF, some of the top-weighted terms from this single review are shown in Table 1.

**Table 1. Illustrative Example of TF-IDF Terms from a Single Review**

| Term | TF-IDF Weight |
|---|---|
| kamar | 0.3852 |
| rumah | 0.3852 |
| balon | 0.2721 |
| rapi | 0.2721 |
| buat | 0.2503 |

This representation allows Naive Bayes and SVM algorithms to operate in a high-dimensional feature space, capturing term importance for sentiment classification. The resulting TF-IDF vectors from all reviews are then used as input for model training and evaluation.

**2.4. Data Splitting and Model Building**

The dataset was separated into two portions, with 80% allocated for training and the testing 20% for testing. This split was carried out using the train_test_split function from the scikit-learn library to ensure balanced evaluation across the models.
Two supervised machine learning algorithms were employed:

1) **Naive Bayes (MultinomialNB)**, a probabilistic classifier particularly suitable for discrete text features. It estimates the probability of a class *Ck* given a feature vector *x = (x1,x2,…,xn)* as:

$$P(C_k|x) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i|C_k)}{P(x)} \quad (2)$$

Here *P(Ck)* denotes the prior probability of class *Ck*, and *P(xi|Ck)* represents the likelihood of feature *xi* given the class. To address zero-probability issues, a smoothing parameter $\alpha = 1$ (Laplace smoothing) was applied.

2) **Support Vector Machine (SVM)**, a margin-based classifier that finds the optimal hyperplane separating data points of different classes. The decision function is:

$$f(x) = \mathrm{sign}(w \cdot x + b) \quad (3)$$

where *w* indicates the weight vector and *b* the bias. The optimization problem solved is:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 \quad (4)$$

A linear kernel was selected due to its strong performance on sparse TF-IDF representations. The regularization parameter *C* was set to 1.0, ensuring a balance between maximizing the margin and reducing classification errors.

Both models were trained using the TF-IDF vectors obtained from the preprocessing stage. The implementation was conducted in a Google Colaboratory environment with Python

Python 3.11.13, scikit-learn, and Sastrawi libraries for text preprocessing. This setup follows the technical approach of [16], which demonstrated the effectiveness of SVM in sentiment classification after optimization, while providing a comparative baseline against Naive Bayes.

**2.5. Model Performance Evaluation**

The effectiveness of the two algorithms was assessed using four commonly applied classification metrics [17, 18]:

1. Accuracy: The proportion of correctly classified instances to the total number of instances in the test set.
2. Precision: Indicates the capability of the model to identify positive reviews accurately, calculated as True Positive / (True Positive + False Positive).
3. Recall: Reflects the extent to which the model successfully detects all actual positive reviews, measured as True Positive / (True Positive + False Negative).
4. F1-Score: Represents the harmonic average of precision and recall, offering a balanced measure of classification performance.

These four metrics were selected because they provide a balanced and comprehensive evaluation of sentiment classification. Accuracy offers an overall measure of correctness, precision is crucial to avoid misclassifying negative feedback as positive, recall ensures that as many relevant positive reviews as possible are detected, and F1-score balances both precision and recall. In addition, a confusion matrix was used to analyze the distribution of correct and incorrect predictions, providing deeper insight into false positives and false negatives. This helps identify trade-offs between precision and recall, which are important for determining whether Naive Bayes or SVM is more suitable for real-world hotel review monitoring. Evaluation results are presented in tables and graphs to facilitate interpretation and direct comparison between the two algorithms.

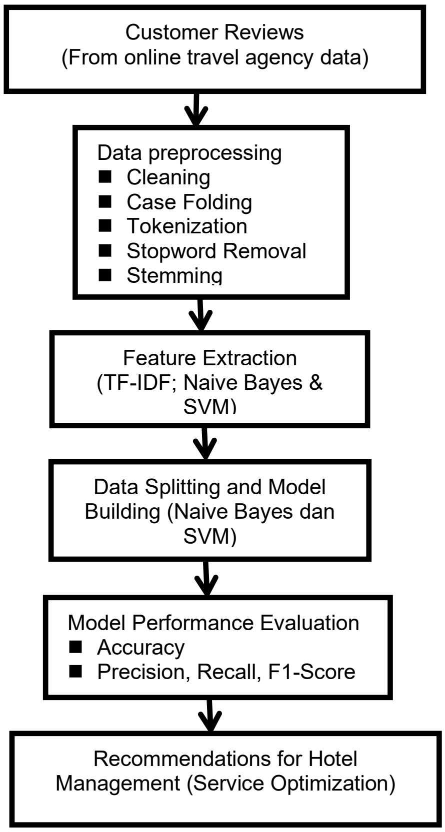


Figure 1. Research Methodology Flowchart

Figure 1 presents the overall research workflow, summarizing the main stages from data collection, preprocessing, feature extraction, and model development to performance evaluation and formulation of recommendations for hotel service optimization.

## 3. Results

### 3.1 Data Collection

The dataset used in this study consists of 898 hotel customer reviews sourced from two popular online travel agencies in Indonesia. Data were collected using web scraping techniques with Python's BeautifulSoup and Selenium libraries. The selection criteria included:

1. Reviews written in Bahasa Indonesia.
2. Reviews containing complete textual content (not empty).
3. Reviews accompanied by a star rating as an indicator of customer experience, which was standardized into a 1–5 star scale because one platform used a 1–10 scale and the other used a 1–5 scale.

The first 8 reviews from the data collection were kept intact. The remaining data were adjusted to maintain a proportion of 69.6% positive (625 reviews) and 30.4% negative (273 reviews). Table 2 shows a sample of the first 8 reviews.

**Table 2. Sample of Hotel Review Dataset**

| No | Platform | Hotel | Rating | Review | Sentiment |
|---|---|---|---|---|---|
| 1 | Online Travel Agency A | B | 4 | Tidak menyediakan air mineral untuk tamu lagi. Sangat disayangkan. | Negative |
| 2 | Online Travel Agency A | B | 5 | Pelayanannya bagus dan menu sarapannya lengkap. | Positive |
| 3 | Online Travel Agency A | C | 2 | Tolong kamarnya dibuat seperti rumah... dan banyak bunga biar indah. | Negative |
| 4 | Online Travel Agency B | F | 5 | Kamarnya keren-keren, nyaman dan cocok untuk keluarga. | Positive |
| 5 | Online Travel Agency B | E | 5 | Bagusssssssssssssssssssssssssssss sssssssssssssssss... | Positive |
| 6 | Online Travel Agency B | F | 5 | Nyaman bersih good. Parking area is well maintained. | Positive |
| 7 | Online Travel Agency B | A | 5 | Semua OK, tdk ada kendala. Lokasi dekat bandara. | Positive |
| 8 | Online Travel Agency A | D | 3 | Hotel lama tidak terawat dengan baik. | Negative |

### 3.2 Text Preprocessing

The review texts underwent several preprocessing steps to clean and normalize the data before further analysis. These steps included cleaning, case folding, tokenization, stopword removal, and stemming. This process ensured that the text was standardized and free from noise before feature extraction. Table 3 and Table 4 show a sample review at each stage of preprocessing:

**Table 3. Text Preprocessing Stages of Online Travel Agency A**

| Stage | Output |
|---|---|
| Original | Tidak menyediakan air mineral untuk tamu lagi. Sangat disayangkan. |
| Cleaning | Tidak menyediakan air mineral untuk tamu lagi Sangat disayangkan |
| Case Folding | tidak menyediakan air mineral untuk tamu lagi sangat disayangkan |
| Tokenization | ['tidak', 'menyediakan', 'air', 'mineral', 'untuk', 'tamu', 'lagi', 'sangat', 'disayangkan'] |
| Stopword Removal | ['menyediakan', 'air', 'mineral', 'tamu', 'sangat', 'disayangkan'] |
| Stemming | ['sedia', 'air', 'mineral', 'tamu', 'sangat', 'sayang'] |

**Table 4. Text Preprocessing Stages of Online Travel Agency B**

| Stage | Output |
|---|---|
| Original | Kamarnya keren-keren, nyaman dan cocok untuk keluarga. Saya pasti akan kembali lagi ke sini. |
| Cleaning | Kamarnya keren keren nyaman dan cocok untuk keluarga Saya pasti akan kembali lagi ke sini |
| Case Folding | kamarnya keren keren nyaman dan cocok untuk keluarga saya pasti akan kembali lagi ke sini |
| Tokenization | ['kamarnya', 'keren', 'keren', 'nyaman', 'dan', 'cocok', 'untuk', 'keluarga', 'saya', 'pasti', 'akan', 'kembali', 'lagi', 'ke', 'sini'] |
| Stopword Removal | ['kamarnya', 'keren', 'keren', 'nyaman', 'cocok', 'keluarga', 'sini'] |
| Stemming | ['kamar', 'keren', 'keren', 'nyaman', 'cocok', 'keluarga', 'sini'] |

### 3.3 Feature Extraction

Feature extraction was performed using Term Frequency-Inverse Document Frequency (TF-IDF) to transform the text into high-dimensional numerical vectors. A sample review from the dataset *"Tidak menyediakan air mineral untuk tamu lagi.. dan itu sangat menyedihkan.. setidaknya berikan seperti sebelumnya dan naikkan harga kamar jika tidak mampu menyediakan air mineral itu"* was processed, and the resulting top TF-IDF terms are presented in Table 5:

**Table 5. Sample Output of Top 5 TF-IDF Terms from a Single Review**

| Term | TF-IDF Weight |
|---|---|
| air | 0.4241 |
| mineral | 0.4241 |
| sedia | 0.4241 |
| tamu | 0.4241 |
| lagi | 0.3999 |

This approach has been shown effective for Indonesian sentiment classification tasks [15].

### 3.4 Data Splitting and Model Building

The dataset was split into 80% training data and 20% testing data using scikit-learn's train_test_split function. Two supervised learning models were developed:

1. Naive Bayes (MultinomialNB): A probabilistic model suitable for discrete features such as TF-IDF, with Laplace smoothing parameter α=1.
2. Support Vector Machine (LinearSVC): A margin-based classifier with a linear kernel. Regularization parameter C was set to 1.0.

Both models were trained on TF-IDF features and evaluated on the test set using accuracy, precision, recall, and F1-score. The classification reports for each model are shown in Tables 6 and 7.

**Table 6. Classification Report of Naive Bayes**

| Sentiment | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negatif | 0.93 | 0.75 | 0.83 | 55 |
| Positif | 0.90 | 0.98 | 0.93 | 125 |
| Accuracy | | | 0.91 | 180 |

**Table 7. Classification Report of SVM (LinearSVC)**

| Sentiment | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negatif | 0.91 | 0.87 | 0.89 | 55 |
| Positif | 0.94 | 0.96 | 0.95 | 125 |
| Accuracy | | | 0.93 | 180 |

### 3.5 Model Performance Evaluation

The performance of the Naive Bayes and SVM models was evaluated using accuracy, precision, recall, and F1-score. The results are presented in Table 8.

**Table 8. Model Performance Evaluation**

| Metric | Naive Bayes | SVM (Linear) |
|---|---|---|
| Accuracy | 0.9056 | 0.9333 |
| Precision | 0.8971 | 0.9449 |
| Recall | 0.9760 | 0.9600 |
| F1-score | 0.9349 | 0.9524 |

Both models demonstrated high classification performance, with SVM achieving the highest accuracy (93.33%), precision (0.9449), and F1-score (0.9524). This indicates that SVM provides more balanced predictions across sentiment categories. Naive Bayes, on the other hand, achieved the highest recall (0.9760), meaning it was more effective in capturing almost all actual positive reviews, although with slightly lower precision compared to SVM. These results suggest that SVM is preferable when consistent and balanced classification is required, while Naive Bayes is advantageous in scenarios where maximizing the detection of positive sentiment is the priority.

### 3.6 Visualization and Interpretation

Figure 2 presents the evaluation results of Naive Bayes and SVM (Linear) models. Figure 2a shows the console output of model evaluation across four metrics: accuracy, precision, recall, and F1-score. To provide a clearer comparison, Figure 2b visualizes the same results in the form of a bar chart. Both models performed strongly; however, SVM demonstrated superior overall balance in accuracy, precision, and F1-score, indicating more consistent sentiment classification. In contrast, Naive Bayes achieved higher recall, reflecting its advantage in detecting a larger proportion of positive reviews, though with a slight reduction in precision compared to SVM.

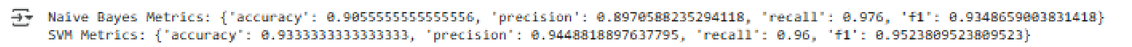


**Figure 2a. Console output of model evaluation (Naive Bayes and SVM)**

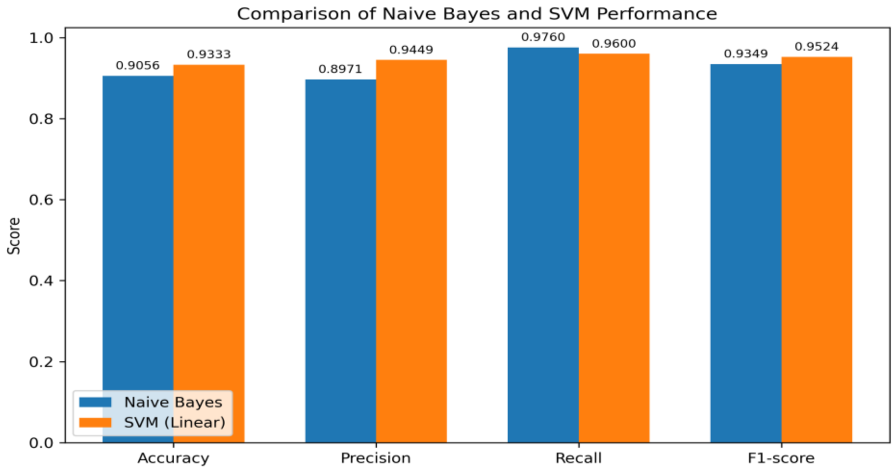


**Figure 2b. Comparison of Naive Bayes and SVM performance**

The confusion matrices of Naive Bayes and SVM provide deeper insights into the distribution of correct and incorrect predictions beyond the summary statistics. As shown in Figure 3a, the Naive Bayes classifier correctly identified 122 out of 125 positive reviews, resulting in a recall of 0.98. However, it misclassified 14 negative reviews as positive, which reduced its precision to 0.90. This indicates that while Naive Bayes is very effective in capturing almost all positive reviews, it is less reliable when distinguishing negative sentiments. Such behavior reflects the model's probabilistic nature, which favors sensitivity to the majority class but at the expense of generating more false positives.

In contrast, the SVM model exhibited a more balanced distribution of predictions, as illustrated in Figure 3b. It correctly classified 48 out of 55 negative reviews and 120 out of 125 positive reviews, which significantly reduced the number of false positives compared to Naive Bayes. Consequently, SVM achieved higher precision (0.94) while maintaining a strong recall of 0.96. This balanced trade-off confirms that SVM is more effective for real-world applications where accurate detection of negative reviews is critical for improving hotel services. Together, these confusion matrices visually demonstrate the contrasting strengths of the two algorithms: Naive Bayes is better suited when the goal is to capture as many positive reviews as possible, whereas SVM is more reliable for ensuring balanced sentiment classification.

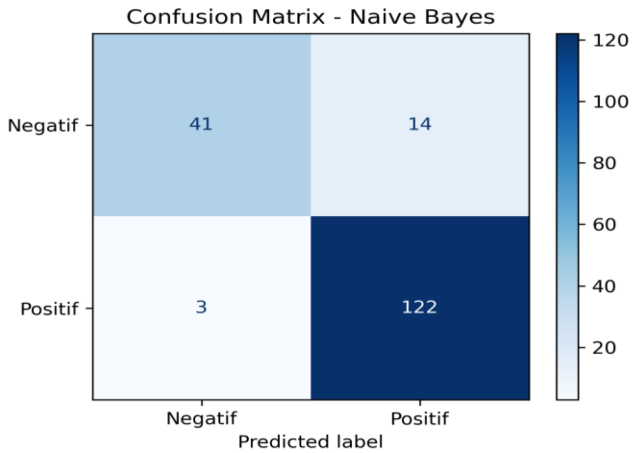


**Figure 3a. Confusion Matrix Naive Bayes**

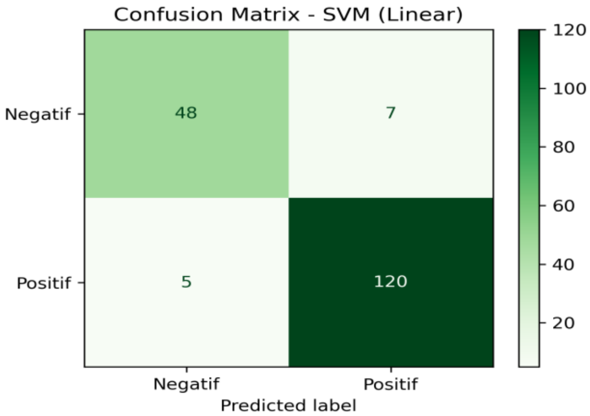

**Figure 3b. Confusion Matrix SVM (linear)**

### 3.7 Contextual Discussion

The findings of this study confirm that both Naive Bayes and SVM (Linear) models are effective for sentiment classification of Indonesian hotel reviews, with SVM achieving higher accuracy, precision, and F1-score, and Naive Bayes achieving higher recall. This trade-off indicates that SVM delivers more balanced predictions, while Naive Bayes is more sensitive in detecting positive reviews. Conceptually, this can be explained by the characteristics of each algorithm: Naive Bayes uses probabilistic reasoning based on word occurrence frequencies, making it less strict and therefore able to capture almost all positive reviews (high recall), but at the cost of producing more false positives, which lowers its precision. In contrast, SVM optimizes a separating hyperplane with maximum margin in the TF-IDF feature space, which makes it more selective in assigning reviews to the positive class. This results in higher precision and F1-score, although some actual positive reviews may be missed, leading to a slightly lower recall. These results align with prior research [11], which reported similar performance patterns between Naive Bayes and SVM in sentiment classification for social media data. Likewise, [12] and [19] highlighted that SVM's margin optimization can improve balanced classification outcomes in high-dimensional spaces such as TF-IDF vectors.

From an operational perspective, these findings have direct implications for optimizing hotel services in Central Bangka Regency. When the priority is to identify as many positive reviews as possible for promotional purposes, Naive Bayes may be preferred due to its higher recall. Conversely, when hotel managers require a model that consistently delivers accurate and precise classifications across sentiment categories for example, in monitoring guest feedback to address negative experiences promptly, SVM is the better choice. Implementing these models in real-time review monitoring systems can enable hotel management to make data-driven service improvements, enhance staff responsiveness, and strengthen brand reputation.

From a research perspective, future work could explore integrating these classical algorithms with advanced deep learning architectures such as BERT or LSTM, as suggested by [3], to further improve semantic understanding of reviews. Expanding the dataset to multiple regions and including code-mixed or multilingual feedback would also allow testing the models' generalizability, as recommended by [8]. Ultimately, the integration of sentiment analysis into operational decision-making systems can help hotels in Central Bangka Regency proactively address service gaps and leverage positive feedback for strategic growth.

## 4. Conclusions

This study compared the performance of Naive Bayes and Support Vector Machine (SVM) algorithms in classifying sentiment from Indonesian-language hotel reviews collected from online travel agencies. The dataset consisted of 898 reviews (69.6% positive and 30.4% negative) that underwent crawling, selection, and preprocessing steps including cleaning, case folding, tokenization, stopword removal, and stemming. Using TF-IDF feature extraction and an 80:20 train-test split, SVM achieved the highest accuracy, precision, and F1-score, while Naive Bayes obtained the highest recall. These results indicate that SVM provides more balanced and precise sentiment classification performance, whereas Naive Bayes is more effective in maximizing the detection of positive reviews

From an operational service management perspective, these findings provide clear guidance for hotel managers in Central Bangka Regency. Naive Bayes can be prioritized when the primary objective is to detect as many positive reviews as possible, such as for marketing campaigns or promotional content. Conversely, SVM is better suited for consistent and balanced sentiment classification, making it valuable for real-time monitoring of guest feedback to identify service strengths and address weaknesses promptly. The integration of these models into hotel operational systems can support data-driven decision-making, improve service quality, and enhance customer satisfaction.

For future research, expanding the dataset to include multiple regions, code-mixed language, and multilingual reviews will help evaluate model generalizability. The use of advanced models such as BERT and LSTM can also be explored to improve contextual understanding in sentiment analysis. Furthermore, incorporating hyperparameter optimization and ensemble learning methods may yield more robust and adaptive classification systems. Such advancements can strengthen the role of sentiment analysis as a decision-support tool not only in the hospitality sector but also across other service industries that rely heavily on customer feedback.

**Acknowledgements**

The author would like to thank Universitas Bangka Belitung through the Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) for supporting this research under the 2025 Peneliti Muda (PM) scheme.

**References**

[1] K. Z. Zhang, H. Xu, S. Zhao, and Y. Yu, "Online reviews and impulse buying behavior: the role of browsing and impulsiveness," *Internet Research,* vol. 28, no. 3, pp. 522-543, 2018.

[2] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA: Morgan & Claypool Publishers, 2012.

[3] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences,* vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[4] L. Davoodi, J. Mezei, and M. Heikkilä, "Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms," *Electron Commer Res,* pp. 1-43, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s10660-025-09948-4.

[5] M. Ghiassi, J. Skinner, and D. Zimbra, "Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network," *Expert Systems with Applications,* vol. 40, no. 16, pp. 6266-6282, 2013.

[6] T. T. Thet, J. C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews using grammatical relations," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, pp. 812-817, doi: 10.1145/1774088.1774248.

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval,* vol. 2, no. 1–2, pp. 1-135, 2008.

[8] A. M. Rahat, A. Kahir, and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 2019: IEEE, pp. 266-270, doi: 10.1109/smart46866.2019.9117512.

[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal,* vol. 5, no. 4, pp. 1093-1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[10] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.

[11] F. Millennianita, U. Athiyah, and A. W. Muhammad, "Comparison of Naïve Bayes Classifier and Support Vector Machine Methods for Sentiment Classification of Responses to Bullying Cases on Twitter," *Journal of Mechatronics and Artificial Intelligence,* vol. 1, no. 1, pp. 11-26, 2024. [Online]. Available: http://ejournal.upi.edu/index.php/jmai/.

[12] K. Chong and N. Shah, "Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non Hyperparameter Tuned Healthcare Stock Market

Sentiment Analysis," *International Journal of Advanced Computer Science and Applications,* vol. 13, no. 12, pp. 90-94, 2022, doi: 10.14569/IJACSA.2022.0131213.
[13] K. H. Yoo and U. Gretzel, "Use and creation of social media by travellers," in *Social Media in Travel, Tourism and Hospitality*. London: Routledge, 2016, pp. 189-206.
[14] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," *Computational and Mathematical Organization Theory,* vol. 25, no. 3, pp. 319-335, 2018, doi: 10.1007/s10588-018-9266-8.
[15] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA Jurnal,* vol. 10, no. 2, pp. 71-76, 2020, doi: 10.32664/smatika.v10i02.455.
[16] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 2020: IEEE, pp. 217-220.
[17] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, "Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis," *Jurnal Informatika: Jurnal Pengembangan IT,* vol. 6, no. 1, pp. 56-60, 2021, doi: 10.30591/jpit.v6i1.3245.
[18] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Systems with Applications,* vol. 57, pp. 117-126, 2016, doi: 10.1016/j.eswa.2016.03.028.
[19] D. Diana, R.E. Indrajit, & E. Dazki, "Komparasi Algoritma NaÃ¯ ve Bayes, Logistic Regression Dan Support Vector Machine pada Klasifikasi File Application Package Kit Android Malware. Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi, vol. 11, no. 1, pp. 109-122, 2022.