

Sentiment Monitoring of Google Maps Reviews Using Naive Bayes and Support Vector Machine at Aveta Hotel Malioboro

Rinaldi Hamzah^{1*}, Arif Pramudwiatmoko²

¹Student of Computer Science, Faculty of Science & Technology, Universitas Teknologi Yogyakarta

²Department of Engineering, Faculty of Science & Technology, Universitas Teknologi Yogyakarta

Corresponding Author: Rinaldi Hamzah rinaldi.5220411342@student.uty.ac.id

ARTICLE INFO

Keywords: Sentiment Analysis, Google Maps, Naive Bayes, Support Vector Machine, Review Monitoring.

Received : 27, April

Revised : 29, May

Accepted: 30, June

©2026 Hamzah, Pramudwiatmoko:
This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

Manual monitoring of hotel customer reviews on Google Maps is increasingly inefficient due to growing review volumes, particularly in detecting negative sentiment. This study proposes a web-based sentiment monitoring system integrated with Naive Bayes and Support Vector Machine (SVM) for automated sentiment classification of Google Maps reviews at Aveta Hotel Malioboro. A dataset of 2,449 historical reviews was collected via DOM scraping and processed through text preprocessing and TF-IDF feature extraction. SVM with minority-class oversampling achieved the highest performance: accuracy 0.9540, precision 0.9514, recall 0.9540, and F1 Score 0.9522. The implemented system retrieves reviews periodically via SerpAPI, classifies sentiment automatically, and delivers real-time Telegram notifications, enabling structured and data-driven hotel review monitoring

INTRODUCTION

The hospitality industry is a service sector fundamentally shaped by customer perception and satisfaction. In the current digital era, online reviews have become one of the most influential sources of information for prospective guests prior to making accommodation decisions (Caryn et al., 2025). Google Maps, through its integrated review feature, has emerged as one of the most widely used platforms for hotel reviews in Indonesia, given its accessibility and high user adoption rate (Haq & Rachmat, 2020). Reviews posted on this platform directly influence a hotel's digital reputation – positive reviews strengthen public trust and attract new guests, while unaddressed negative reviews can progressively erode the hotel's image and reduce booking intent (Syafirimansyah, 2023). This dynamic makes timely and systematic review monitoring a critical operational necessity for hotel management.

The volume of reviews on Google Maps continues to grow at a pace that renders manual monitoring increasingly impractical. Without an automated mechanism, negative reviews may go undetected for extended periods, preventing hotel management from responding in a timely manner (Litvin et al., 2008). This challenge is directly observed at Aveta Hotel Malioboro, which had accumulated more than 3,430 reviews on Google Maps as of January 2026. The current monitoring process at the hotel relies on manual review checks, with notifications concentrated within a single account, resulting in slow information dissemination and delayed responses – particularly during peak operational periods. The absence of a structured, automated system creates a significant operational gap that can affect both service quality and the hotel's competitive standing.

Sentiment analysis offers a viable solution to this problem by enabling the automatic classification of customer opinions into predefined polarity categories, such as positive and negative (Liu, 2012). Among the available machine learning approaches, Naive Bayes and Support Vector Machine (SVM) have been widely adopted for text-based sentiment classification due to their complementary strengths: Naive Bayes performs efficiently on short, informal texts, while SVM demonstrates robust generalization on high-dimensional feature spaces commonly found in natural language data (Dwitama & Paputungan, 2025). When combined with modern data retrieval techniques and notification infrastructure, these algorithms can serve as the backbone of an operational monitoring system that responds to new reviews in near real-time.

Prior studies on sentiment analysis using Naive Bayes and SVM have consistently demonstrated the effectiveness of these algorithms across various review-based datasets. Handayanto et al. (2021) evaluated both methods on Google Reviews of Summarecon Mall Bekasi and found that SVM achieved higher accuracy than Naive Bayes, supporting its suitability for practical deployment. Alfiyanti & Indra (2024) further explored this comparison on Google Maps reviews of Taman Mini Indonesia Indah (TMII), finding that SVM excels in precision critical for correctly flagging negative reviews while Naive Bayes exhibits stronger recall. Ipmawati et al. (2024) confirmed SVM's practicality in a tourism review context on Google Maps with an accuracy of 87%,

while Sarimole & Kuadrat (2024) reported similar trends on Twitter data, where SVM reached 87.95% accuracy compared to Naive Bayes at 81.65%. Despite these well-established findings, a consistent limitation across all prior studies is their exclusive focus on model evaluation without developing any functional monitoring system. None of the reviewed works integrated the trained model into an end-to-end system capable of automatically retrieving new reviews, classifying sentiment, storing results, and notifying stakeholders in real time.

This gap motivates the present study, which bridges algorithmic evaluation and system implementation by proposing a web-based hotel review monitoring system that integrates Naive Bayes and SVM sentiment classification with SerpAPI-based review retrieval and Telegram Bot notifications. The novelty of this research lies not only in comparing the two algorithms under both original and resampled class distribution scenarios, but also in embedding the best-performing model into a fully operational monitoring system tailored to the hospitality context. Accordingly, this study aims to first to evaluate and compare the performance of Naive Bayes and SVM in classifying sentiment from Google Maps hotel reviews under varying class distribution conditions and secondly to design and implement a web-based monitoring system that automates review retrieval, sentiment classification, and real-time notification delivery to support data-driven hotel review management.

THEORETICAL REVIEW

Sentiment Analysis

Sentiment analysis, also referred to as opinion mining, is a field within natural language processing, computational linguistics, and text mining that systematically analyzes opinions, sentiments, evaluations, and emotions expressed by a speaker or writer toward a particular topic, product, service, or organization (Liu, 2012). This study applies binary document-level sentiment analysis to Indonesian-language Google Maps reviews of Aveta Hotel Malioboro, where the primary objective is to automatically distinguish negative feedback from positive feedback to support timely managerial response.

Google Maps Review

Google Maps Review is a user-generated content feature embedded within Google Maps that allows individuals to submit written reviews and star ratings for places, businesses, and services they have visited (Haq & Rachmat, 2020). As one of the most widely accessed location-based platforms globally, Google Maps serves as a high-visibility channel where customer opinions are publicly visible and directly influence a business's perceived reputation. For hotels, reviews on this platform are particularly impactful because they are indexed by Google Search and accessible without requiring an account, making them a primary reference point for prospective guests during accommodation decision-making (Syafrimansyah, 2023). In this study, Google Maps reviews are selected as the primary data source due to their public accessibility, structured metadata, and direct relevance to hotel reputation management in the Indonesian hospitality context.

Machine Learning

Machine Learning is a branch of artificial intelligence that enables systems to learn patterns from data and make predictions or decisions without being explicitly programmed for each scenario (Muzakir et al., 2024). In the domain of text classification, machine learning algorithms are trained on labeled datasets to recognize linguistic patterns associated with specific categories such as sentiment polarity and subsequently apply the learned patterns to classify new, unseen text. Supervised machine learning approaches, where training labels are provided, are the most commonly adopted paradigm for sentiment classification tasks due to their demonstrated effectiveness across diverse text domains. In this study, supervised machine learning is employed for binary sentiment classification of Indonesian hotel reviews, where labeled training data enable the classifier to distinguish between positive and negative customer feedback in a systematic and reproducible manner.

Text Preprocessing

Text Preprocessing is a foundational step in natural language processing that transforms raw, unstructured text into a clean and structured format suitable for computational analysis (Khofifah et al., 2022). Raw user-generated text, such as hotel reviews, typically contains informal language, typographical errors, slang, punctuation, and irrelevant characters that introduce noise into machine learning models. To address this, preprocessing in this study consists of six sequential stages: (1) Case Folding – converting all text to lowercase to ensure uniformity; (2) Cleaning – removing punctuation, symbols, numbers, and non-alphabetic characters; (3) Tokenization – splitting sentences into individual word tokens; (4) Normalization – standardizing informal expressions and slang words using a custom domain-specific dictionary; (5) Stopword Removal – eliminating high-frequency words that carry no sentiment value; and (6) Stemming – reducing words to their root form using the Sastrawi library for Indonesian text. Each stage contributes to improving feature quality and, consequently, model classification accuracy (Gabriella et al., 2023).

TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF is a statistical feature extraction technique used in text mining to quantify the importance of a term within a document relative to a corpus (Gabriella et al., 2023). This process converts each preprocessed word into a token vector that represents the word. Term Frequency (TF) is the frequency of a word appearing in a document. Inverse Document Frequency (IDF) measures the importance of a word in a document (Ipmawati et al., 2024). The equation for calculating TF-IDF is shown in equation (1):

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \dots\dots (1)$$

Description:

$tf_{i,j}$: number of times word i appears in document j
 df_i : number of documents containing i
 N : total of all documents

TF-IDF was selected over embedding-based representations such as Word2Vec or BERT due to its computational efficiency, transparency, and proven effectiveness in sparse, domain-specific text classification tasks particularly when combined with linear classifiers such as SVM and Naive Bayes on Indonesian review data (Gabriella et al., 2023).

Naive Bayes

Naive Bayes is a probabilistic classifier grounded in Bayes' theorem that estimates the posterior probability of a class given an observed feature vector (Maulana et al., 2024). In text classification, the Multinomial Naive Bayes variant is most commonly applied to term frequency or TF-IDF weighted features, where each term contributes independently to the posterior probability estimation. The conditional probability computation follows Bayes' theorem as defined in equation (2):

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \dots\dots (2)$$

Description:

H = Class hypothesis (positive/negative)

X = Data or document to be classified

P(H|X) = Posterior probability

P(X|H) = Likelihood probability

P(H) = Prior probability

P(X) = Evidence probability

Multinomial Naive Bayes is evaluated as a baseline classifier due to its computational efficiency and established effectiveness in short-text sentiment classification tasks, particularly when training data is limited (Maulana et al., 2024).

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that works by finding the optimal hyperplane that maximally separates two classes in a high-dimensional feature space (Cortes & Vapnik, 1995). The hyperplane is determined by maximizing the margin between the nearest data points of each class, referred to as support vectors (Ipmawati et al., 2024). When the data is not linearly separable, SVM employs a kernel function to map the input data into a higher-dimensional space where linear separation becomes feasible (Cortes & Vapnik, 1995);Ipmawati et al., 2024). The decision function of SVM is formally defined as equation (3):

$$F(x) = sign(w.x + b) \dots\dots (3)$$

Where, F(x) is the prediction function, w is the normal vector of the hyperplane, x is the input feature vector, and b is the bias or intercept. SVM has been widely adopted for text classification tasks due to its effectiveness in handling high-dimensional and sparse feature spaces, such as those produced by

TF-IDF vectorization, while maintaining robust generalization even with limited training data

Oversampling

Oversampling is a data resampling technique used to address class imbalance in supervised learning datasets, wherein one class contains significantly more instances than the other, causing classifiers to develop a bias toward the majority class and resulting in poor detection of the minority class (Chawla et al., 2002; Mujahid et al., 2024). In the context of hotel review sentiment analysis, positive reviews typically far outnumber negative ones, making negative sentiment the minority class. To address this, oversampling synthetically increases the representation of the minority class in the training data, enabling the model to learn a more balanced decision boundary (Chawla et al., 2002). This approach was adopted in the present study because the collected dataset exhibited a severe class imbalance, with positive reviews accounting for approximately 87% of all labeled samples, necessitating synthetic augmentation of the negative class to prevent classifier bias and ensure reliable detection of negative customer feedback.

Web Scraping

Web scraping is an automated data extraction technique that retrieves structured information from web pages by parsing their underlying HTML structure (DwicaHYo & Ratnasari, 2023; Khder, 2021). While web scraping broadly refers to the programmatic collection of data from websites, DOM (Document Object Model) scraping specifically targets the rendered DOM tree of a web page including dynamically loaded content generated by JavaScript making it particularly suitable for platforms such as Google Maps, where review data is not embedded in the static HTML source but rendered dynamically at runtime (Khder, 2021a). DOM scraping was selected in this study specifically for the initial historical data collection phase because Google Maps renders review content dynamically through JavaScript, making static HTML parsing insufficient for comprehensive data retrieval.

SerpAPI

SerpAPI is a third-party API service that provides structured access to Google Search Engine Result Pages (SERPs), including Google Maps reviews, without requiring direct browser automation or exposure to Google's anti-scraping mechanisms. Web scraping is particularly useful in situations where data is not provided in a machine-readable format such as JSON or XML, and structured data extraction via automated tools enables near real-time data collection from web sources (Khder, 2021b). The service returns review data in a standardized JSON format, facilitating seamless integration into backend systems for programmatic processing (Khder, 2021). In this study, SerpAPI is integrated within the monitoring system's periodic scheduler implemented using a cron job mechanism to automatically retrieve the latest customer reviews of Aveta Hotel Malioboro at defined intervals, ensuring the review database remains continuously updated without manual intervention (Adila et al., 2022). SerpAPI was selected over direct DOM scraping for the real-time monitoring phase because its managed infrastructure eliminates the risk of IP blocking and

CAPTCHA interruptions that would compromise the reliability of an automated monitoring workflow operating continuously in a production environment.

Telegram Bot

Telegram Bot is an automated messaging agent built on the Telegram Bot API that enables applications to send and receive messages programmatically through the Telegram messaging platform (Andriani & Sa'di, 2024; Egamberdiev et al., 2024). Unlike conventional email notifications, Telegram Bot messages are delivered instantly to subscribers' mobile devices via push notifications, ensuring time-sensitive information reaches recipients without delay (Prakosa et al., 2023). Bots can be configured to broadcast messages to multiple subscribers simultaneously, making them an effective medium for distributing monitoring alerts to a dispersed hotel management team (Andriani & Sa'di, 2024). Telegram Bot was selected as the notification channel in this study because it offers instant push notification delivery, supports simultaneous broadcast to multiple subscribers, and does not require recipients to access a separate web dashboard making it particularly suitable for hotel management teams who require immediate awareness of negative customer feedback while performing on-site operational duties (Andriani & Sa'di, 2024; Prakosa et al., 2023).

After reviewing the theory and previous research, the conceptual framework of this study can be illustrated in the following figure, which depicts the relationship between the problem, technological solutions, and the expected outcomes:

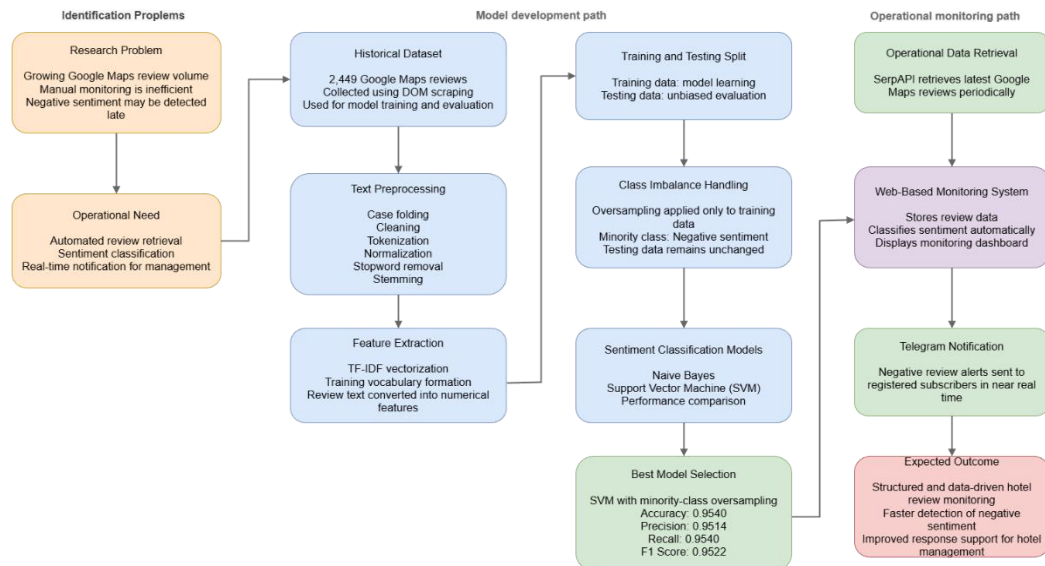


Figure 1. Research Conceptual Framework

Based on the theoretical studies and previous research reviewed, it can be concluded that the growing volume of online customer reviews on platforms such as Google Maps presents both an opportunity and a challenge for hotel management. While user-generated reviews contain valuable insights into customer satisfaction, the manual monitoring of such reviews is time-consuming, inconsistent, and prone to delayed responses. Most existing studies have addressed sentiment analysis and review monitoring as separate concerns,

without integrating automated data collection, real-time classification, and instant notification into a single unified workflow. Therefore, this study combines automated data retrieval via SerpAPI, sentiment classification using Naive Bayes and Support Vector Machine (SVM) with TF-IDF feature extraction and oversampling for class imbalance, and real-time alert delivery via Telegram Bot into a single integrated monitoring system, aimed at producing a more responsive and data-driven solution for hotel reputation management.

METHODOLOGY

Data Collection

A total of 2,449 historical customer reviews of Aveta Hotel Malioboro were collected from Google Maps spanning February 2020 to May 2025, using DOM (Document Object Model) scraping via browser-based automation. This technique was selected because Google Maps renders review content dynamically through JavaScript, making it inaccessible via static HTML parsing (Dwicahyo & Ratnasari, 2023). The collected data includes user names, timestamps, star ratings, and review text.

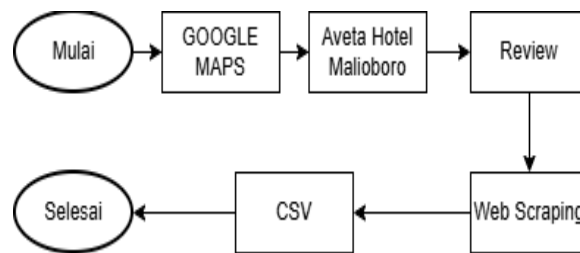


Figure 1. Historical Data Scraping Flow

Data Labeling

Sentiment labeling was performed automatically using a pre-trained Indonesian RoBERTa-based transformer model via the Hugging Face library. This approach produced consistent binary labels positive (1) and negative (0) without manual annotation effort. Of the 2,449 reviews, 2,152 were labeled positive, 250 negative, and 47 remained unlabeled due to highly informal or incomplete text, yielding 2,402 valid labeled samples for model development.

Text Preprocessing

Text preprocessing was applied to clean and standardize raw review text prior to feature extraction. The pipeline consisted of six sequential stages: (1) Case Folding converting all text to lowercase; (2) Cleaning removing punctuation, numbers, and non-alphabetic characters; (3) Tokenization splitting sentences into individual tokens; (4) Normalization – standardizing informal language using a custom hotel-domain slang dictionary; (5) Stopword Removal eliminating non-informative words while retaining negation terms such as *tidak*, *bukan*, and *jangan* that carry significant sentiment value; and (6) Stemming reducing words to their root form using the Sastrawi library (Khofifah et al., 2022).

Feature Extraction

Feature extraction was conducted using TF-IDF with `ngram_range=(1,3)`, `max_df=0.9`, `min_df=2`, and `sublinear_tf=True`, producing a vocabulary of 5,122

features (Gabriella et al., 2023). To prevent data leakage, the TF-IDF vectorizer was fitted exclusively on the training data, with the test data transformed using the resulting vocabulary and IDF weights. The dataset was partitioned using an 80:20 stratified split, yielding 1,740 training samples (1,592 positive, 148 negative) and 435 test samples (398 positive, 37 negative).

Oversampling

Given the pronounced class imbalance – positive reviews comprising 87.9% of the dataset random oversampling with replacement was applied exclusively to the TF-IDF-transformed training data to upsample the negative class from 148 to 1,592 samples. The test set was deliberately left in its original distribution to ensure evaluation reflects real-world conditions (Mujahid et al., 2024; Carvalho et al., 2025).

Model Training and Evaluation

Two classification models were trained: Multinomial Naive Bayes and Linear Support Vector Machine (LinearSVC). Hyperparameter optimization was conducted using GridSearchCV with 5-fold Stratified K-Fold Cross Validation and macro F1-score as the optimization metric, ensuring class balance was maintained across each fold (Alfiyanti & Indra, 2024). Two experimental scenarios were evaluated: (1) original class distribution, and (2) oversampled training data. Optimal parameters were $\alpha=1.0$ and $C=0.05$ for the original scenario, and $\alpha=0.01$ and $C=2$ for the oversampling scenario. Model performance was assessed using accuracy, precision, recall, and F1-score.

System Design

The system was designed using four UML diagrams, Flowchart Figure 2, Use Case Figure 3, Activity Figure 4, ERD Figure 5 Data and Figure 6 Database Tabel to define the system workflow, actor interactions, activity sequences, and database structure respectively.

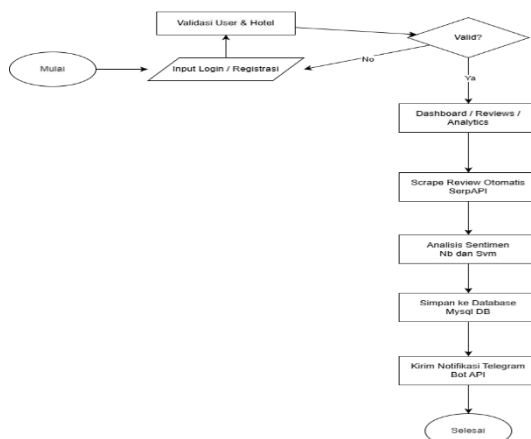


Figure 2. System Flowchart Diagram



Figure 3. Use Case Diagram

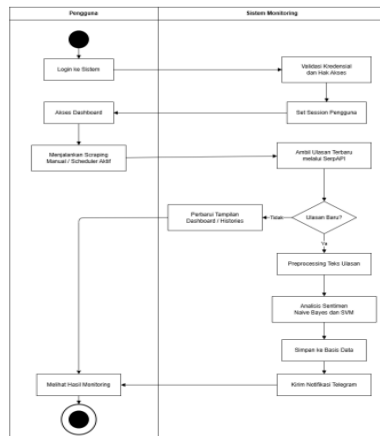


Figure 4. Activity Diagram

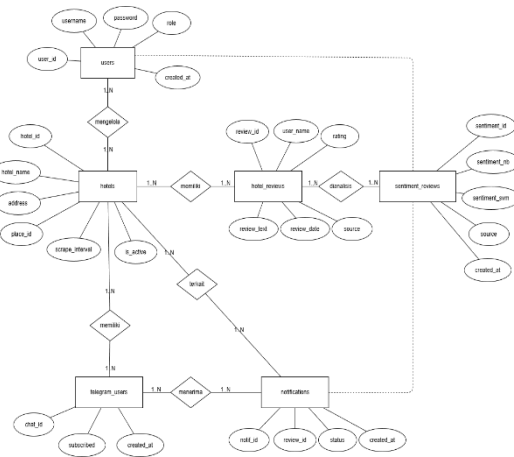


Figure 5. Entity Relationship Diagram

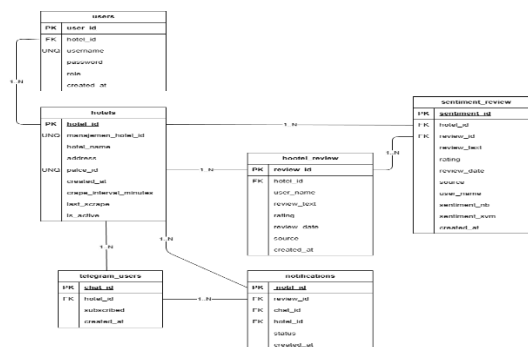


Figure 6. Database Table Relationship Design

System Implementation

The best-performing model was integrated into a web-based monitoring system developed using Flask, MySQL, APScheduler, SerpAPI, and Telegram Bot API. SerpAPI retrieves the latest Google Maps reviews at periodic intervals, the trained model classifies their sentiment, MySQL stores review data and classification results, and the Telegram Bot delivers automated notifications to all registered subscribers (Egamberdiev et al., 2024; Maududy & Nursamsi, 2023). Finally, black-box functional testing was conducted to verify that all system components operated as designed.

RESULTS

Data Collection and Labeling Results

The DOM scraping process successfully collected 2,449 historical customer reviews of Aveta Hotel Malioboro from Google Maps, spanning from February 20, 2020 to May 25, 2025. The dataset includes user names, review timestamps, star ratings, and review text. Figure 10 shows a sample of the collected data.

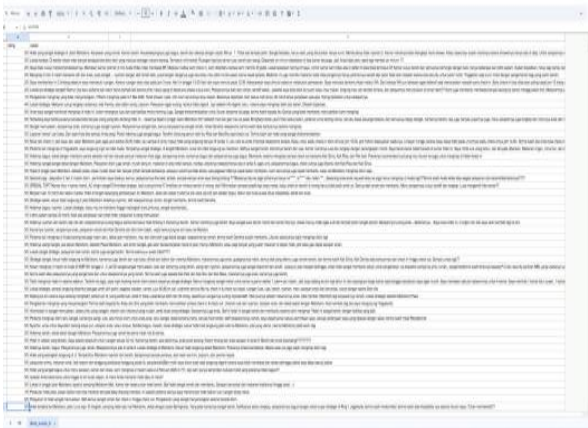


Figure 7. Collected Historical Review Data

Automatic sentiment labeling produced the class distribution shown in Figure 8. The dataset is dominated by positive sentiment (2,152 reviews; 87.9%), with 250 negative reviews (10.2%) and 47 unlabeled samples (1.9%). This significant class imbalance was a key factor in the experimental design

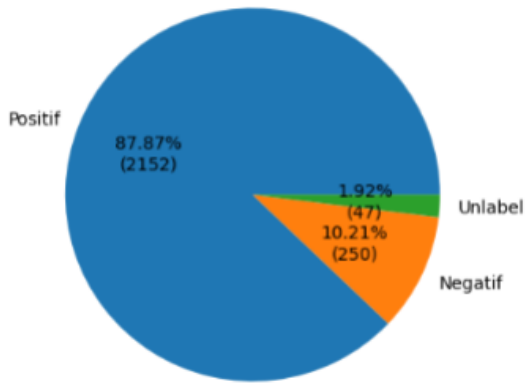


Figure 8. Sentiment Labeling Results Distribution

Preprocessing Results

The preprocessing pipeline successfully transformed raw review text into clean, structured tokens. Table 1 presents representative preprocessing results across review samples (index 21), demonstrating the progressive transformation from raw text through all six stages.

Tabel 1. Sample Preprocessing Results of Review Data

Stage	Sample - Positive Review (Index 21)
1) Raw Text	the hotel is good, comfortable. strategic location, if you want to go to Malioboro, just step open the door, highly recommended...
2) Case Folding	the hotel is good, comfortable. strategic location, if you want to go to Malioboro, just step open the door, highly recommended...

3) Cleaning	The hotel is nice, comfortable, strategic location, if you want to go to Malioboro, just step open the door, highly recommended
4) Tokenization	['hotel', 'good', 'comfortable', 'location', 'strategic', 'if', 'want', 'to', 'malioboro', 'stay', 'step', 'open', 'door', 'very', 'recommended']
5) Normalization	['hotel', 'good', 'comfortable', 'location', 'easy to find', 'if', 'want', 'to', 'malioboro', 'stay', 'step', 'open', 'door', 'very', 'recommended']
6) Stopword Removal	['hotel', 'good', 'comfortable', 'location', 'easy', 'find', 'malioboro', 'stay', 'step', 'open', 'door', 'recommendation']
7) Stemming	['hotel', 'good', 'comfortable', 'location', 'easy', 'meet', 'malioboro', 'stay', 'lang', 'open', 'door', 'recommendation']

The results demonstrate that each preprocessing stage contributes to progressively reducing noise and improving feature quality. Notably, the normalization stage effectively standardized informal language common in Google Maps reviews, while the retention of negation words during stopword removal preserved critical sentiment-bearing tokens such as 'tidak' (not) and 'kurang' (less/insufficient).

Class Distribution and Feature Extraction

After preprocessing and excluding unlabeled data, the class distribution showed a significant imbalance as illustrated in Figure 9. The positive class dominates at 87.9%, making this a canonical imbalanced classification problem.

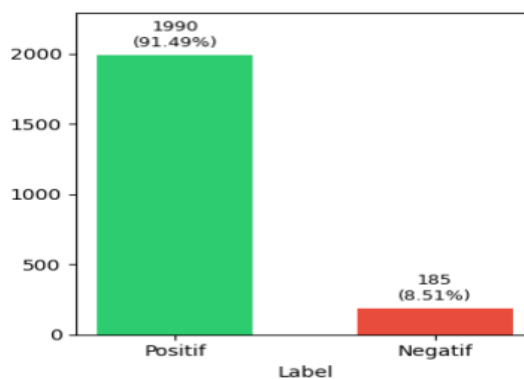


Figure 9. Class Distribution After Preprocessing

Data was split using an 80:20 ratio, producing 1,740 training samples (1,592 positive, 148 negative) and 435 test samples (398 positive, 37 negative). Figure 13 shows the train-test distribution for the original scenario, while Figure 14 shows the distribution after oversampling was applied to the training set only.

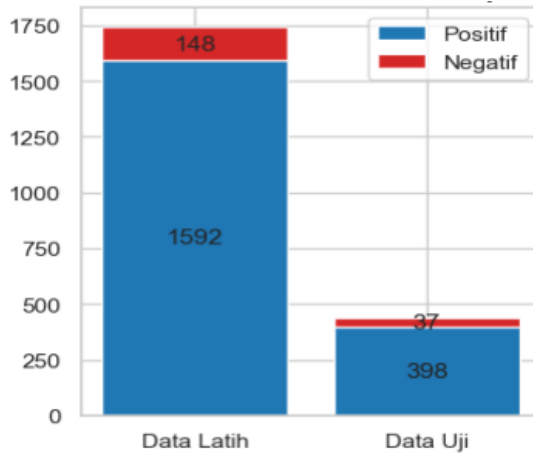


Figure 10. Training and Test Data Distribution (Original Scenario)

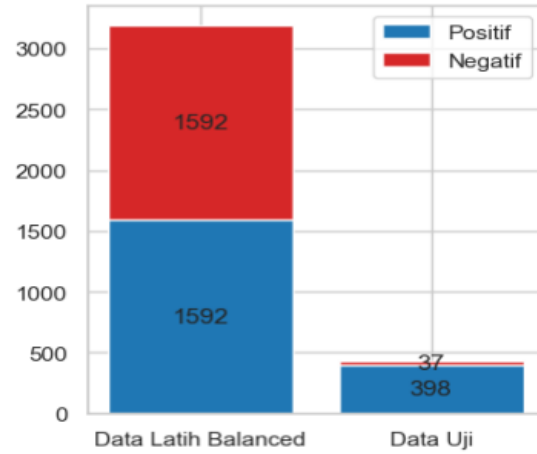


Figure 11. Training and Test Data Distribution (After Resampling)

TF-IDF feature extraction with $ngram_range=(1,3)$ produced a vocabulary of 5,122 features. Hyperparameter tuning via GridSearchCV yielded optimal parameters as follows: for the original scenario, Naive Bayes $\alpha=1.0$ and SVM $C=0.05$; for the resampling scenario, Naive Bayes $\alpha=0.01$ and SVM $C=2$.

Model Evaluation Results

Model evaluation was conducted under two experimental scenarios. The confusion matrix evaluation used accuracy, weighted precision, weighted recall, and weighted F1 Score, as formulated in Equations (4)–(7):

$$\text{Accuracy} = (TP+TN)/(TP+FP+TN+FN) \times 100\% \dots\dots(4)$$

$$\text{Precision} = TP/(FP+TP) \times 100\% \dots\dots(5)$$

$$\text{Recall} = TP/(FN+TP) \times 100\% \dots\dots(6)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \times 100\% \dots\dots(7)$$

Tabel 2. Confusion Matrix

	Actual: Negative	Actual: Positive
Predicted: Negative	TN (True Negative)	FP (False Positive)
Predicted: Positive	FN (False Negative)	TP (True Positive)

Tabel 3. Model Evaluation Results — Original Class Distribution

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.9494	0.9451	0.9494	0.9449
Support Vector Machine	0.9425	0.9534	0.9425	0.9463

Tabel 4. Model Evaluation Results – Resampled Training Data

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	0.9402	0.9450	0.9402	0.9422
Support Vector Machine	0.9540	0.9514	0.9540	0.9522

SVM with resampled training data achieved the highest overall performance across all four metrics. Based on these results, the SVM resampling model was selected as the best-performing classifier and integrated into the operational monitoring system.

System Implementation Results

The web-based monitoring system was successfully implemented and provides five main interface pages. The Dashboard page Figure 12 is the main monitoring interface, presenting a real-time summary including hotel data, latest reviews, sentiment distribution charts, weekly trend comparisons, average star ratings, top sentiment keywords, and an integrated scheduler control panel

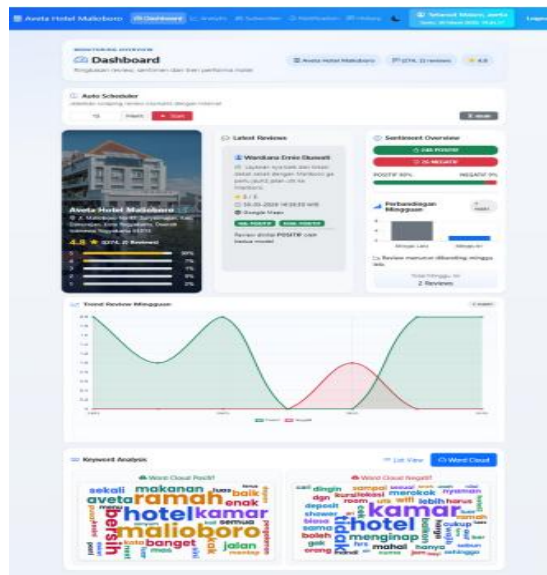


Figure 12. Dashboard Page Implementation

The Analytics page Figure 13 provides comprehensive statistical summaries and temporal trend visualizations, including daily sentiment distribution, weekly comparison charts, and average rating trends and Subscriber Management page Figure 14 enables addition and removal of Telegram Bot subscribers linked to specific hotels, with control over the Telegram Bot's active status.



Figure 13. Analytics Page Implementation



Figure 14. Subscriber Management Page Implementation

The Notification Log page Figure 15 displays the complete history of automated Telegram notifications, including review_id, subscriber chat_id, delivery status, and timestamps and Review Histories page Figure 16 presents a searchable, filterable table of all processed reviews with classification results from both models, enabling direct comparison of Naive Bayes and SVM outputs.

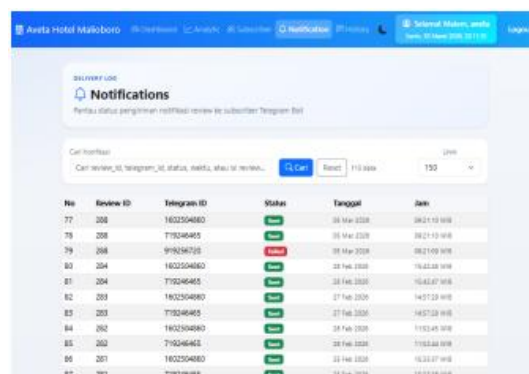


Figure 15. Notification Log Page Implementation

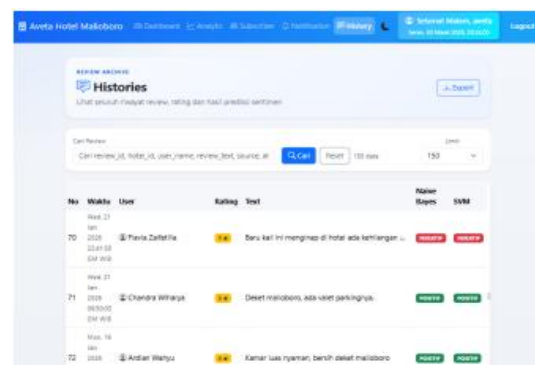


Figure 16. Review Histories Page Implementation

Functional Testing Results

Black-box functional testing was conducted on twelve core system features: user registration and login, hotel-based data isolation, periodic review scraping via SerpAPI, duplicate review detection, sentiment classification using both models, database storage, dashboard and histories display, automated scheduler control, Telegram Bot subscriber registration, notification delivery, and notification log recording.

Table 5. Black-Box Functional Testing Results

No.	Feature Tested	Expected Result	Status
1	User Registration & Login	New account created and authenticated	✓ Pass
2	Hotel Data Isolation	Dashboard only shows data for the logged-in hotel	✓ Pass
3	Periodic Review Scraping via SerpAPI	Latest reviews retrieved automatically on schedule	✓ Pass
4	Duplicate Review Detection	Previously stored reviews are not reprocessed	✓ Pass
5	Naive Bayes Sentiment Classification	Review classified as POSITIF or NEGATIF	✓ Pass
6	SVM Sentiment Classification	Review classified as POSITIF or NEGATIF	✓ Pass
7	Database Storage	Classification results saved to MySQL database	✓ Pass
8	Dashboard Display	Real-time sentiment summary displayed	✓ Pass
9	Review Histories Display	All processed reviews listed with classification results	✓ Pass
10	Scheduler Control	Start/stop scheduler from dashboard interface	✓ Pass
11	Telegram Subscriber Registration	Subscriber chat_id linked to hotel	✓ Pass
12	Telegram Notification Delivery	Notification sent to all active subscribers	✓ Pass

All twelve features achieved a Successful status, confirming that the complete system workflow operates as designed from user authentication to automated monitoring and notification delivery.

DISCUSSION

The evaluation results indicate that neither model is absolutely superior across all experimental scenarios. In the original class distribution scenario, Naive Bayes achieved a higher accuracy (0.9494) than SVM (0.9425). However, SVM demonstrated higher precision (0.9534) and F1 Score (0.9463), reflecting its greater ability to minimize false positives in negative review detection. This pattern aligns with Alfiyanti & Indra (2024), who found that in their TMII review study, Naive Bayes excelled in recall while SVM was superior in precision – a consistent pattern

suggesting that algorithm selection should be guided by the specific operational priority of the monitoring use case.

In the resampling scenario, contrasting effects were observed. Naive Bayes experienced a decline in accuracy from 0.9494 to 0.9402 and in F1 Score from 0.9449 to 0.9422, indicating that synthetic augmentation of training data disrupted its feature independence assumption. This outcome is consistent with the nature of Multinomial Naive Bayes, which is sensitive to changes in term frequency distributions introduced by oversampling. In contrast, SVM in the resampling scenario achieved the highest overall performance with accuracy 0.9540, precision 0.9514, recall 0.9540, and F1 Score 0.9522. This improvement aligns with findings by Handayanto et al. (2021) and Maulana et al. (2024), both of whom concluded that SVM generally outperforms Naive Bayes in Google review sentiment tasks. The improvement can be attributed to SVM's margin-maximizing structure, which benefits from a more balanced class distribution in training data, allowing it to better learn the decision boundary separating positive and negative classes.

Comparing with related studies, Sarimole & Kuadrat (2024) reported SVM accuracy of 87.95% and Naive Bayes of 81.65% on Twitter data, while Ipmawati et al. (2024) reported SVM accuracy of 87% on Google Maps tourism reviews. This study achieves notably higher accuracy (95.40% for SVM with resampling) on hotel review data. The performance improvement is attributable to the domain-specific preprocessing pipeline, including the custom normalization dictionary tailored for hotel-related informal language, retention of negation words during stopword removal, and the use of StratifiedKfold hyperparameter optimization with F1-macro scoring.

From a system implementation perspective, the monitoring system successfully operationalizes the trained model in a practical hospitality context. The integration of SerpAPI with APScheduler enables near real-time monitoring without continuous manual intervention. The Telegram Bot integration distributes monitoring alerts to multiple subscribers simultaneously, directly addressing the centralized-notification limitation identified in the existing manual system. The role-based access control and hotel-based data isolation ensure both data security and operational relevance for multi-property management scenarios. All twelve functional features passed black-box testing, confirming end-to-end system reliability from authentication to automated monitoring.

CONCLUSIONS AND RECOMMENDATIONS

This study draws two main conclusions. First, the performance comparison of Naive Bayes and SVM demonstrates that both algorithms are viable sentiment classifiers for Google Maps hotel reviews. The best overall performance was achieved by SVM in the resampled training data scenario, with accuracy 0.9540, precision 0.9514, recall 0.9540, and F1 Score 0.9522. The resampling approach proved more beneficial for SVM than for Naive Bayes, making SVM with minority-class oversampling the recommended model for this classification task. The superiority of SVM is consistent with prior studies and can be attributed to its structural advantage in handling high-dimensional and imbalanced text classification problems.

Second, this research successfully designed and implemented a web-based hotel review monitoring system integrating sentiment analysis and a Telegram Bot. The system automatically retrieves the latest reviews via SerpAPI, performs sentiment classification using both trained models, stores results in a MySQL database, displays information through an interactive dashboard and analytics pages, manages Telegram subscribers, and delivers automated notifications. Operating in a near real-time manner through periodic scheduler-based scraping, the system provides practical value to hotel management by enabling faster, more structured, and data-driven review monitoring that transforms raw customer feedback into actionable operational intelligence.

FURTHER STUDY

Future research is recommended to address two primary limitations of this study. First, the dataset should be expanded by incorporating reviews from multiple hotels and diverse review platforms to improve data diversity and enhance the generalizability of the classification models, as the current study is limited to a single hotel source which may constrain the model's applicability across broader hospitality contexts. Second, future work may explore the integration of transformer-based classifiers such as IndoBERT, which are specifically pre-trained on Indonesian text corpora, to evaluate whether deep learning approaches can yield superior sentiment classification performance compared to the traditional machine learning methods employed in this study.

ACKNOWLEDGMENT

The author expresses sincere gratitude to Dr.Eng. Arif Pramudwiatmoko, S.T., M.Eng. as the supervising lecturer for guidance, suggestions, and invaluable support throughout the research process.

REFERENCES

- Adila, N., Sembiring, F., & Jatmiko, W. (2022). Implementation of Web Scraping for Journal Data Collection on the SINTA Website. *Sinkron*, 7(4), 2478–2485. <https://doi.org/10.33395/sinkron.v7i4.11576>
- Alfiyanti, D., & Indra. (2024). Penerapan Algoritma Multinomial Naïve Bayes dan Support Vector Machine (SVM) untuk Analisis Sentimen terhadap Ulasan Google Maps di Taman Mini Indonesia Indah (TMII). *Jurnal SIMETRIS*, 15(1).
- Andriani, R., & Sa'di, A. (2024). Implementasi Notifikasi Bot Telegram pada Sistem Monitoring Perangkat Jaringan. *SISTEMASI: Jurnal Sistem Informasi*, 13(1). <http://sistemasi.ftik.unisi.ac.id>
- Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: a review from a data perspective. *Journal of Big Data*, 12(1). <https://doi.org/10.1186/s40537-025-01119-4>
- Caryn, F., Alya, D., Christin, N., Kailash Khemlani, B., Egary, S., Agustinus, J., & Soraya, K. (2025). Pengaruh Ulasan Daring Terhadap Keputusan Tamu

- Dalam Pemesanan Kamar Hotel. *Jurnal Garuda Pengabdian Kepada Masyarakat*, 3(2), 66–74. <https://doi.org/10.55537/gabdimas.v3i2.1169>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297.
- DwicaHYo, K., & Ratnasari, C. I. (2023). Perbandingan Metode Web Scraping Dalam Pengambilan Data: Kajian Literatur. *Automata*, 4(2).
- Dwitama, R. P., & Paputungan, I. V. (2025). Analisis Perbandingan Naive Bayes dan SVM dalam Klasifikasi Gender di Platform Twitter. *Edusaintek: Jurnal Pendidikan, Sains dan Teknologi*, 12(3), 2025–1637. <https://doi.org/10.47668/edusaintek.v12i3.1811>
- Egamberdiev, K., Khidirova, N., & Islomov, A. (2024). Creation of an Automated Information System Using Python and the Django Framework with a Telegram Bot. *Big Data and Computing Visions*, 4(3), 219–226. <https://doi.org/10.22105/bdcv.2024.482114.1210>
- Gabriella, Y. A. P., Ginting, G. L., & Panjaitan, M. (2023). Optimasi Penerimaan Siswa Baru Dengan Penerapan Algoritma Text Mining Dan Tf-Idf. *Journal of Computing and Informatics Research*, 2(3), 110–117. <https://doi.org/10.47065/comforch.v2i3.941>
- Handayanto, Rahmadya, T., Herlawati, H., Atika, P. D., Khasanah, F. N., Yusuf, A. Y. P., & Septia, D. Y. (2021). Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine. *Jurnal Komtika (Komputasi dan Informatika)*, 5(2), 153–163. <https://doi.org/10.31603/komtika.v5i2.6280>
- Haq, F. U., & Rachmat, H. (2020). Penggunaan Google Review Sebagai Penilaian Kepuasan Pengunjung dalam Parawisata. *Journal of Sustainable Tourism Research*, 2(1), 61–72. <https://doi.org/https://doi.org/10.24198/tornare.v2i1.25826>
- Ipmawati, J., Saifulloh, & Kusnawi. (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>
- Khder, M. A. (2021a). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, 13(3), 144–168. <https://doi.org/10.15849/ijasca.211128.11>
- Khder, M. A. (2021b). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft*

- Computing and Its Applications*, 13(3), 144–168.
<https://doi.org/10.15849/ijasca.211128.11>
- Khofifah, W., Rahayu, D. N., & Yusuf, A. M. (2022). Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps. *Jurnal Interkom: Jurnal Publikasi Ilmiah Bidang Teknologi Informasi Dan Komunikasi*, 16(4), 28–38. <https://doi.org/10.35969/interkom.v16i4.192>
- Litvin, S. W., Goldsmith, R., & Bing, P. (2008). Electronic Word of Mouth in Hospitality and Tourism Management. *Tourism Management*, 29(29), 458–468.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Maududy, R., & Nursamsi, D. R. (2023). Pengembangan Real-Time Monitoring dan Data Logging Berbasis Web Pada Proses Robot Painting untuk Meningkatkan Efisiensi Produksi. *Informatics And Digital Expert (Index)*, 5(2), 89–94. <https://e-journal.unper.ac.id/index.php/informatics>
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384. <https://doi.org/10.57152/malcom.v4i2.1206>
- Mujahid, M., Kina, E., Rustam, F., Villar, M. G., Alvarado, E. S., Diez, I. D. L. T., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00943-4>
- Muzakir, A., Adi, K., & Kusumaningrum, R. (2024). *Penerapan Konsep Machine Learning & Deep Learning*. Asosiasi Dosen Sistem Informasi Indonesia.
- Prakosa, B. A., Hendrawan, A. H., Setiadi, I. H., Ritzkal, Riawan, I., & Riana, F. (2023). Implementation of a Real-Time Wi-Fi Voucher Notification System Utilizing Telegram API Bot. *Ingenierie Des Systemes d'Information*, 28(6), 1587–1596. <https://doi.org/10.18280/isi.280615>
- Sarimole, F. M., & Kuadrat. (2024). Analisis Sentimen Terhadap Aplikasi Satu Sehat Pada Twitter Menggunakan Algoritma Naive Bayes Dan Support Vector Machine. *Jurnal Sains Dan Teknologi*, 5(3), 783–790. <https://doi.org/10.55338/saintek.v5i1.2702>
- Syafrimansyah. (2023). Peningkatan Kinerja Hotel melalui Pelayanan Unggul: Penelitian tentang Pengaruh Pelayanan Pelanggan yang Baik terhadap Jumlah Tamu Hotel. *nnovative: Journal Of Social Science Research*, (1), 537–546.