Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

Evaluating Lexicon Weighting and Machine Learning Models for Sentiment Classification of Indonesian Mangrove Ecotourism Reviews

Ferdi Chahyadi*[1], Alena Uperiati[2], Risdy Absari Indah Pratiwi [3], Nur Hamid[4]

[1] Informatics Engineering, Universitas Maritim Raja Ali Haji, Indonesia
[2] Software Engineering Technology, Politeknik Negeri Batam, Indonesia
[3] Digital Business, Universitas Maritim Raja Ali Haji, Indonesia
[4] Interdisciplinary Research Center for Smart Mobility and Logistics (IRC-SML), King Fahd University of Petroleum and Minerals, Saudi Arabia

Email: [1]ferdi.chahyadi@umrah.ac.id

Received : Dec 10, 2025; Revised : Dec 18, 2025; Accepted : Dec 21, 2025; Published : Dec 23, 2025

Abstract

Sentiment analysis on ecotourism reviews presents specific challenges due to descriptive writing styles, the use of ambiguous words, and contextual meaning shifts (contextual polarity shift). These characteristics often cause lexicon-based approaches to produce unstable polarity labels. This study aims to evaluate the influence of two lexicon weighting methods, namely Mean Weighting and Summation Weighting, on the initial sentiment labeling of mangrove ecotourism reviews and to assess the performance of machine learning models trained using these labels. The research method includes text preprocessing, lexicon-based scoring using the InSet lexicon, feature extraction with Term Frequency–Inverse Document Frequency (TF–IDF), and the training of two classification algorithms, Support Vector Machine (SVM) and Logistic Regression (LR). The results show that the Mean Weighting method produces more stable polarity scores and higher model performance. The combination of SVM with Mean Weighting achieves the best results with an accuracy of 0.902, macro precision of 0.876, macro recall of 0.819, a macro F1-score of 0.841, and a weighted F1-score of 0.899. Meanwhile, LR with Mean Weighting reaches an accuracy of 0.891 with a similar performance pattern. In contrast, the Summation Weighting method results in lower performance for both algorithms. Error analysis indicates that neutral sentences and ambiguous words such as “bagus” and “ramai” frequently lead to misclassification. These findings highlight that the choice of lexicon weighting method plays a crucial role in improving sentiment classification accuracy and contributes to the development of hybrid approaches in text mining and sentiment analysis for the Indonesian language.

Keywords : *Lexicon Weighting, Logistic Regression, Machine Learning, Mangrove Ecotourism, Sentiment Analysis, Support Vector Machine;*

1. INTRODUCTION

The rapid development of digital platforms such as Google Maps and various travel applications has led to a significant increase in the publication of tourism reviews by visitors. These reviews serve as an important source of information, as they not only describe tourists’ experiences but also reflect perceptions of service quality, environmental conditions, and destination management [1], [2]. In the context of ecotourism, including mangrove areas, visitor opinions are often descriptive in nature and contain a combination of positive and negative evaluations that are not always expressed explicitly [3], [4]. This condition makes sentiment analysis more challenging compared to other types of text [5]. Recent studies have shown that language used in tourism reviews tends to exhibit a high level of ambiguity and is prone to polarity changes depending on sentence context, a phenomenon known as contextual polarity shift [6], [7].

In Indonesian sentiment analysis research, three main approaches have been widely adopted: lexicon based, machine learning based, and hybrid models. The lexicon based approach relies on

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

sentiment dictionaries such as InSet, which contain lists of positive and negative words along with their associated polarity weights, and then aggregates these word scores to determine document level sentiment. Previous studies indicate that InSet is one of the most widely used Indonesian sentiment lexicons, however, most research applies simple aggregation strategies, such as summation or score difference, and rarely discusses in detail how to handle words that potentially exhibit dual polarity [8], [9]. The phenomenon of contextual polarity shift is also a major cause of sentiment labeling inaccuracies in static lexicon based approaches, particularly for descriptive and context-dependent texts [10], [11], [12]. These findings suggest that lexicon weighting strategies play an important role in determining the quality of the resulting sentiment labels [13], [14].

Machine learning approaches, particularly Support Vector Machine (SVM), have been reported to achieve higher accuracy than purely lexicon based methods when labeled data are available. In several public and social issues, SVM combined with TF–IDF features has demonstrated superior accuracy compared to single lexicon-based models [14], [15]. The use of TF–IDF improves text representation by emphasizing words that are more relevant to sentiment, thereby supporting more accurate classification processes [16]. A number of studies have integrated lexicons with machine learning in hybrid approaches, where lexicons are used as feature sources or for initial labeling, while final classification is performed using SVM, Logistic Regression, or other algorithms [17], [18], [19], [20], [21]. Such approaches have been shown to improve classification performance across various domains, including tourism reviews and digital application reviews.

In the Indonesian research context, the InSet Lexicon has been applied across multiple domains. Abdillah et al. (2022) combined InSet and SentiStrength with SVM on social media data and demonstrated that classification performance is strongly influenced by domain characteristics and the static nature of lexicons [13]. Nadira et al. (2023) manually adjusted InSet weights for mobile banking application reviews and achieved high performance; however, the approach was domain-specific and had not been tested across different contexts [22]. Other studies have shown that variations in lexicons and initial labeling mechanisms directly affect accuracy and sentiment class distribution [11], [15]. In addition, several studies emphasize that text weighting schemes such as TF–IDF play a crucial role in classification performance, although they do not explicitly discuss polarity aggregation strategies within Indonesian sentiment lexicons [14].

Based on the above review, several research gaps can be identified. First, there has been no systematic study comparing Mean Weighting and Summation Weighting for words that appear simultaneously in both positive and negative InSet dictionaries, particularly in the context of Indonesian tourism or ecotourism reviews. Second, although hybrid approaches have been widely applied, systematic evaluations of the performance of Support Vector Machine and Logistic Regression on features derived from different lexicon weighting schemes remain limited, especially for mangrove ecotourism reviews on Google Maps.

Therefore, this study aims to design and apply two lexicon weighting schemes, Summation Weighting and Mean Weighting with a specific focus on InSet Lexicon entries that appear in both positive and negative dictionaries, and to compare the effects of these weighting schemes on the distribution and outcomes of sentiment labels for mangrove ecotourism reviews on Google Maps. In addition, this study evaluates and compares the performance of Support Vector Machine and Logistic Regression models trained using sentiment labels generated from each weighting scheme, based on measurable indicators such as accuracy, precision, recall, and F1-score. Through this approach, the study is expected to contribute to the development of more adaptive hybrid lexicon machine learning sentiment analysis methods that are better suited to the linguistic context and domain characteristics of Indonesian ecotourism.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

2. RESEARCH METHODS

The research stages consist of text preprocessing, lexicon weighting to generate polarity scores, and sentiment label determination based on the resulting scores. Lexicon weighting is performed using the InSet sentiment dictionary as the source of word polarity values, where two weighting approaches, Mean Weighting and Summation Weighting are applied to handle cases in which a word appears simultaneously in both the positive and negative dictionaries. After the polarity scores are obtained, sentiment labels are assigned and subsequently used as the basis for training the classification models in the following stage [23].

2.1 Problem Formulation

Lexicon-based sentiment analysis is employed to determine the polarity of a text based on the weights of its constituent words [24]. Let there be a set of texts defined in Equation (1).

$$D = \{d_1, d_2, \dots, d_n\} \quad (1)$$

Each text is represented as a set of tokens, as defined in Equation (2).

$$d_i = \{w_1, w_2, \dots, w_m\} \quad (2)$$

Given two InSet sentiment dictionaries, namely the positive dictionary $L^+$and the negative dictionary L^-, each containing pairs (w, s_w) with sentiment weights $s_w \in \mathbb{R}$ [25]. when a word appears in both dictionaries with different weights s_w^+ dan s_w^-, this study applies two weighting approaches to handle such cases.

2.1.1 Method 1 (Mean Weighting)

When a word is found in both the positive and negative dictionaries, the sentiment polarity score is calculated using the average of the two scores [26]. The formulation of the Mean Weighting approach is shown in Equation (3).

$$S_w = \frac{S_w^+ + s_w^-}{2} \quad (3)$$

2.1.2 Method 2 (Summation Weighting)

When a word is found in both the positive and negative dictionaries, the sentiment polarity score is calculated by summing the positive and negative scores [26]. The formulation of the Summation Weighting approach is shown in Equation (4).

$$S_w = s_w^+ + s_w^- \quad (4)$$

2.1.3 Sentiment Polarity Score Computation

The sentiment score of a text is calculated using Equation (5) [27].

$$S_i = \sum_{w \in d_i} S_w \quad (5)$$

The sentiment label is then determined based on a threshold value of $\theta = 0.2$, as defined in Equation (6).

$$Sentimen(d_i) = \begin{cases} Positif, & S_i > \theta \\ Negatif, & S_i < \theta \\ Netral, & -\theta \le S_i \le \theta \end{cases} \quad (6)$$

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

The threshold value θ = 0,2 is selected to reduce the influence of minor polarity fluctuations caused by low-weight sentiment words. Preliminary analysis of the sentiment score distribution indicates that smaller threshold values tend to assign non-neutral labels to reviews with weak or ambiguous sentiment, whereas larger threshold values significantly reduce the proportion of reviews labeled as Neutral [28].

2.2 Research Stages

The research stages are illustrated in Figure 1.

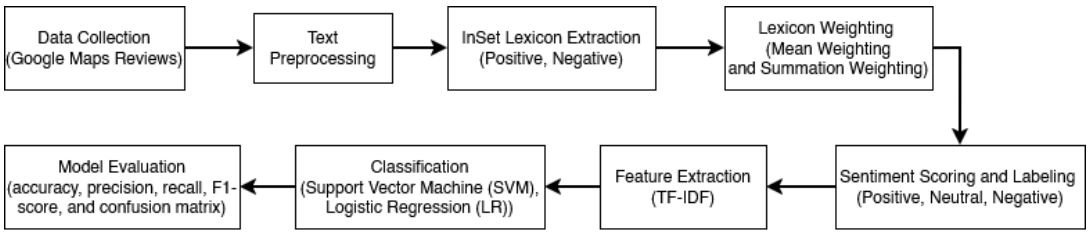

Figure 1. Research Stages

The research workflow shown in Figure 1 describes a sequence of processes starting from data collection to model evaluation. The study begins with the collection of user reviews from Google Maps, followed by a text preprocessing stage consisting of case folding, cleaning, duplicate removal, tokenization, stopword removal, and stemming to produce clean and standardized text. Case folding is performed to convert all text into lowercase format. Duplicate removal aims to eliminate duplicated reviews or review spam. Tokenization is then applied to split sentence strings into individual word units. During this process, a sequence of characters is segmented into word-level tokens. To remove common words that are considered non-informative, such as “itu” (that), “dan” (and), “yang” (which), and “atau” (or), a stopword removal process is applied. Finally, stemming is performed to remove affixes and convert each word into its base form.

Next, the positive and negative InSet sentiment dictionaries are extracted as the basis for computing word polarity weights. To handle words that appear in both dictionaries, two proposed lexicon weighting methods, Mean Weighting and Summation Weighting are applied. These methods generate sentiment scores and polarity labels for each text. Subsequently, features are extracted using TF–IDF before proceeding to the classification stage using two algorithms, namely Support Vector Machine (SVM) and Logistic Regression (LR). The final stage involves evaluating the performance of each model using accuracy, precision, recall, and F1-score metrics to assess the impact of different lexicon weighting methods on classification performance. A confusion matrix is also employed to examine prediction error patterns for each sentiment class.

2.3 Lexicon Weighting Model

Most previous studies have applied simple approaches to handle words that appear in both sentiment dictionaries, such as selecting only one weight or removing ambiguous words altogether [29], [30]. Recent studies, however, indicate that combining word weights can produce more stable and informative polarity representations [31]. Accordingly, this study compares two weight aggregation strategies, Mean Weighting and Summation Weighting to examine their effects on sentiment classification performance. The lexicon weighting process employed in this study is illustrated in Figure 2.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

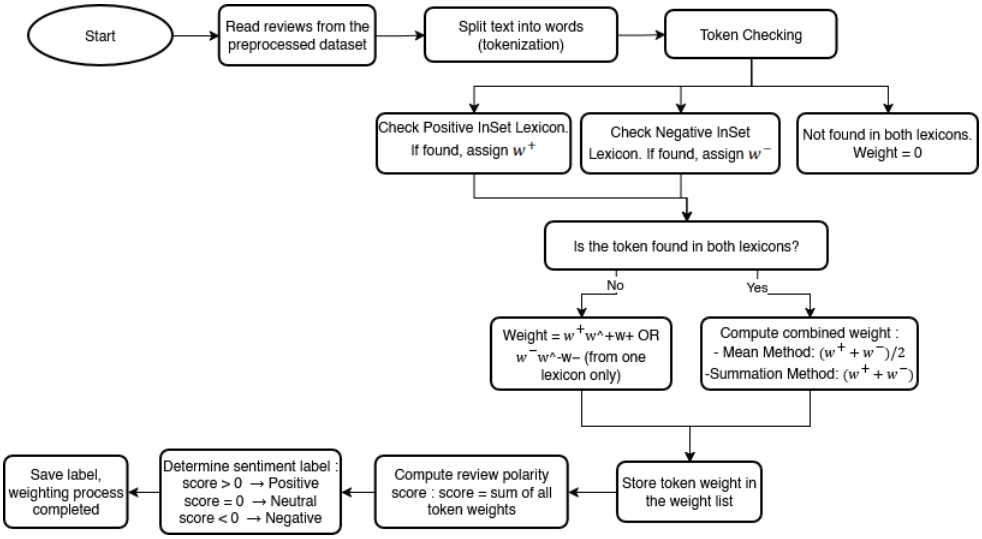

Figure 2. Lexicon Weighting Workflow

The lexicon weighting procedure shown in Figure 2 begins by extracting each review from the preprocessed dataset and segmenting it into word tokens. Each token is then matched against the positive and negative InSet dictionaries. If a word is found in only one of the dictionaries, its corresponding polarity weight is directly applied. However, if the same word appears in both dictionaries, the associated positive weight ($w+$) and negative weight (w-) are processed using the two evaluated weighting approaches, namely the Mean and Summation methods.

The Mean method computes the combined weight as the average of the two values, $(w^+ + w^-)/2$, thereby providing a balancing effect for ambiguous words. In contrast, the Summation method directly adds the two weights $(w^+ + w^-)$, resulting in an accumulated polarity score. Each token weight is then stored and summed to obtain the final polarity score of a review. This score is subsequently used to determine the sentiment label based on predefined labeling rules, where a score greater than zero is classified as Positive, a score equal to zero is classified as Neutral, and a score less than zero is classified as Negative. This process generates automatic sentiment labels for all reviews, which are then used as training data in the classification stage.

2.4 Machine Learning Classification

2.4.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is well known for its superior performance in high-dimensional text data and TF–IDF-based representations due to its ability to maximize the margin between classes [32]. For multiclass classification problems, this study adopts the One-vs-Rest (OvR) strategy, in which a separate binary classifier is trained for each sentiment class.

2.4.2 Logistic Regression (LR)

Logistic Regression (LR) is a linear model that is widely used in sentiment analysis because of its computational efficiency, interpretability, and competitive performance on TF–IDF-based text data. Previous studies have shown that LR serves as a strong baseline for text classification tasks and achieves performance comparable to SVM in many scenarios [33].

2.5 Evaluation Metrics

Model performance is evaluated using several classification metrics commonly employed in sentiment analysis, namely accuracy, precision, recall, and F1-score. These metrics are computed based

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

on the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [34]. Accuracy measures the proportion of correctly classified instances relative to the total number of test samples and is formulated using Equation (7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision measures the degree of correctness of the model in predicting a particular class and is defined using Equation (8).

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall indicates the model's ability to correctly identify all actual instances belonging to a specific class and is formulated using Equation (9).

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

The F1-score represents the harmonic mean of precision and recall, which is used to evaluate the balance between these two metrics and is defined using Equation (10).

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (10)$$

In addition to these metrics, this study also employs a confusion matrix to analyze prediction error patterns for each sentiment class. The confusion matrix provides information on the number of correct and incorrect predictions for each class, enabling the evaluation of model sensitivity to specific classes, identification of frequently confused classes, and detailed assessment of classification error distribution [35]. The evaluation is conducted for each combination of lexicon weighting methods (Mean and Summation) and classification algorithms (SVM and Logistic Regression).

3. RESULTS

3.1. Data Preprocessing

Table 1. Examples of Review Translation Results

Original Review Text	Translated Review Text
A good place for a photoshoot either you just want to snap some sceneries or posing on the decks.	Tempat yang bagus untuk pemotretan, baik Anda hanya ingin mengambil beberapa pemandangan atau berpose di dek.
Вход 10к с человека. Место супер топ! Дорожки хорошие, гулять интересно, прохладно. Есть парковка, есть пара местных кафе	Tiket masuk 10k per orang. Tempat super top! Jalannya bagus, menarik untuk dilalui, dan sejuk. Ada tempat parkir, ada beberapa kafe lokal
ちょうど雨が降りだして　ちょっと残念な景色是した。	Hujan baru saja mulai turun, membuat pemandangannya agak mengecewakan.
جميل جدا وهادئ واجواء منعشه انصح بزيارته	Sangat indah, tenang dan mempunyai suasana yang menyegarkan. Saya sarankan mengunjunginya

The dataset used in this study was derived from user reviews on Google Maps related to various mangrove ecotourism locations across Indonesia. Data collection was conducted using the Apify platform, a cloud-based web scraping and data extraction tool. A total of 55,784 review records were

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

initially retrieved, not all records contained textual reviews. After filtering, only reviews that included textual content were retained, resulting in 28,368 usable reviews.

Although the collected data included several attributes, this study focused exclusively on the textual review field for subsequent analysis. The reviews were written in multiple languages. To ensure consistency during preprocessing, all reviews were translated into Indonesian using the deep_translator library in Python. Examples of the original and translated reviews are presented in **Kesalahan! Sumber referensi tidak ditemukan.**

The subsequent preprocessing stage includes case folding, text cleaning, duplicate removal, tokenization, stopword removal, and stemming. This sequence of processes aims to reduce lexical variation, eliminate non-informative elements, and suppress linguistic noise commonly found in online reviews. As a result, the preprocessed text becomes more consistent and relevant for both lexicon-based sentiment score computation and TF–IDF feature extraction. Examples of the preprocessing results are presented in Table 2.

Table 2. Examples of Review Preprocessing Results

Original Review Text	Casefolding	Tokenisasi	Stopword Removal	Stemming
Tempat yang bagus untuk pemotretan, baik Anda hanya ingin mengambil beberapa pemandangan atau berpose di dek.	tempat yang bagus untuk pemotretan baik anda hanya ingin mengambil beberapa pemandangan atau berpose di dek	['tempat', 'yang', 'bagus', 'untuk', 'pemotretan', 'baik', 'anda', 'hanya', 'ingin', 'mengambil', 'beberapa', 'pemandangan', 'atau', 'berpose', 'di', 'dek']	['tempat', 'bagus', 'pemotretan', 'baik', 'mengambil', 'beberapa', 'pemandangan', 'berpose']	['tempat', 'bagus', 'potret', 'baik', 'ambil', 'beberapa', 'pandang', 'pose']
Hujan baru saja mulai turun, membuat pemandangannya agak mengecewakan.	hujan baru saja mulai turun membuat pemandangannya agak mengecewakan	['hujan', 'baru', 'saja', 'mulai', 'turun', 'membuat', 'pemandangannya', 'agak', 'mengecewakan']	['hujan', 'baru', 'mulai', 'turun', 'membuat', 'pemandangannya', 'mengecewakan']	['hujan', 'baru', 'mulai', 'turun', 'buat', 'pandang', 'kecewa']
Sangat indah, tenang dan mempunyai suasana yang menyegarkan. Saya sarankan mengunjunginya	sangat indah tenang dan mempunyai suasana yang menyegarkan saya sarankan mengunjunginya	['sangat', 'indah', 'tenang', 'dan', 'mempunyai', 'suasana', 'yang', 'menyegarkan', 'saya', 'sarankan', 'mengunjunginya']	['sangat', 'indah', 'tenang', 'mempunyai', 'suasana', 'menyegarkan', 'sarankan', 'mengunjunginya']	['sangat', 'indah', 'tenang', 'punya', 'suasana', 'segar', 'saran', 'unjung']

Overall, the preprocessing stage plays a crucial role in improving the quality of text representation. The normalization of word forms and the removal of irrelevant elements help reduce bias in lexicon based sentiment score calculation, while also producing more stable and informative TF–IDF features for subsequent classification stages.

3.2. Dataset dan Distribusi Label Sentimen

After the text preprocessing stage which includes case folding, duplicate removal, tokenization, stopword removal, stemming, and text normalization, a total of 23,485 cleaned reviews were obtained. These processed texts were then used as the basis for sentiment label assignment using the InSet lexicon-

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

based approach. Sentiment labeling was performed using two lexicon weighting schemes, namely Method 1 (Mean Weighting) and Method 2 (Summation Weighting). For both methods, sentiment polarity was calculated by aggregating the lexicon scores of each word appearing in a review. The distribution of sentiment label counts generated by both methods is presented in Table 3, while the comparison of their percentage proportions is illustrated in Figure 3.

Table 3. Sentiment Label Distribution

Sentiment Class	Method	
	Mean Weighting	Summation Weighting
Positif	9195	8786
Negatif	12095	12107
Netral	2195	2592

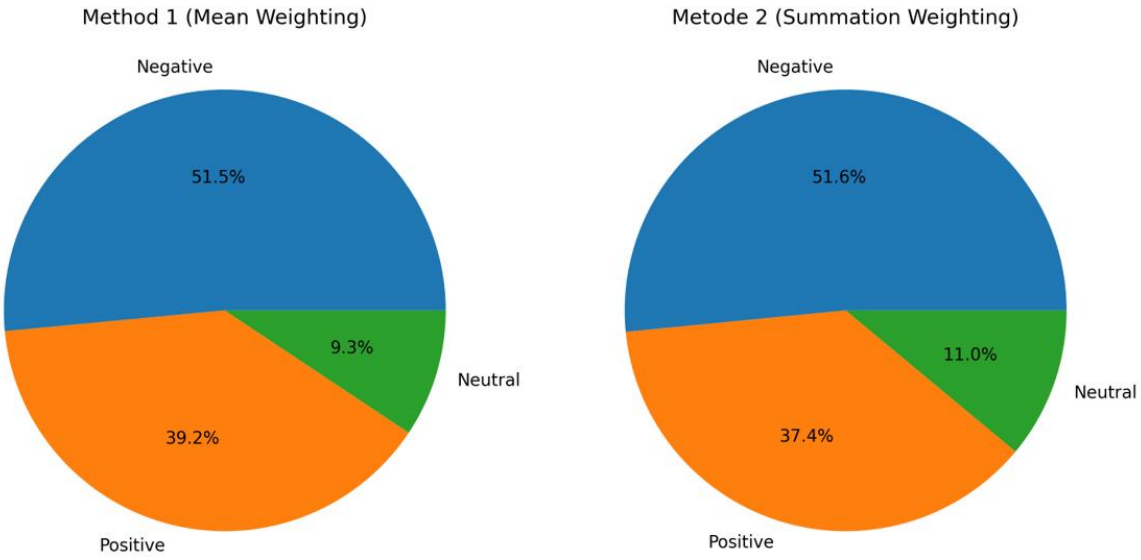

Figure 3. Percentage Distribution of Sentiment Labels for Both Weighting Methods

Based on these results, the Negative class dominates the dataset under both weighting methods. Using Method 1, the proportion of Negative sentiment reaches 51.5%, followed by Positive sentiment at 39.2% and Neutral sentiment at 9.3%. A relatively similar distribution pattern is also produced by Method 2, although slight differences in class proportions are observed due to differences in the weighting mechanisms applied.

The dominance of the Negative class under both methods reflects the characteristics of mangrove ecotourism reviews, where users tend to express dissatisfaction more explicitly regarding specific aspects such as facilities, cleanliness, accessibility, or environmental conditions. In contrast, the relatively small proportion of the Neutral class indicates that most reviews contain evaluative elements, even when they are conveyed in a descriptive manner. This pattern suggests that the ecotourism domain is characterized by opinions that are not entirely neutral, implying that lexicon-based labeling strategies can directly influence the resulting class distribution.

Differences between Mean Weighting (Method 1) and Summation Weighting (Method 2) are evident in the resulting sentiment label distributions, which are directly influenced by the mathematical formulations of each method. In Summation Weighting, sentiment polarity scores are computed by

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

directly summing word weights, causing the score magnitude to increase as the number of sentiment-bearing words in a review grows. Consequently, reviews containing multiple sentiment words tend to receive larger absolute scores and are more frequently classified into extreme classes (Positive or Negative), leading to a reduced proportion of Neutral labels. In contrast, Mean Weighting calculates the average of word weights when polarity conflicts occur, thereby normalizing the influence of the number of sentiment words and producing a more stable and proportional label distribution. This pattern indicates that lexicon weighting particularly the Mean approach, is more effective for short to medium length reviews containing ambiguous words, as it reduces text length bias and yields more representative sentiment labels. The resulting label distribution at this stage forms an essential foundation for the training and evaluation of classification models in subsequent stages.

3.3. Training and Testing Data Split

To ensure a reliable model development process, the sentiment-labeled dataset was divided into training and testing sets using an 80:20 ratio. This split was performed using a stratified sampling strategy to maintain the original label proportions in both the training and testing sets. By doing so, each sentiment class (Negative, Neutral, and Positive) remains proportionally represented, preventing the model from being biased toward any particular class. The data split results for Method 1 (Mean Weighting) are presented in Table 4.

Table 4. Training and Testing Data Distribution for Method 1 (Mean Weighting)

Label	Training Data		Testing Data	
	Number of Samples	Percentage (%)	Number of Samples	Percentage (%)
Negatif	9676	51,50%	2419	51,5%
Netral	1756	9,35%	439	9,35%
Positif	7356	39,15%	1839	39,15%

As shown in Table 4, the proportions of Negative, Positive, and Neutral classes in the training and testing sets remain consistent with the original dataset distribution, where the Negative class continues to dominate, followed by the Positive class, and the Neutral class representing the smallest proportion. This consistency indicates that the stratification process successfully preserves the label distribution structure, ensuring that model evaluation results are not affected by shifts in class composition between the training and testing stages. The training and testing data distribution for Method 2 (Summation Weighting) is presented in Table 5.

Table 5. Training and Testing Data Distribution for Method 2 (Summation Weighting)

Label	Training Data		Testing Data	
	Number of Samples	Percentage (%)	Number of Samples	Percentage (%)
Negatif	9700	51,63%	2407	51,25%
Netral	2055	10,94%	537	11,43%
Positif	7033	37,43%	1753	37,32%

The resulting distribution pattern exhibits a similar trend to that of Method 1, although minor differences in percentage values are observed due to variations in the weighting mechanism applied during the initial labeling stage. Nevertheless, the class proportions in both the training and testing sets remain consistently maintained, allowing for an objective and fair comparison of model performance between Method 1 and Method 2.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

Overall, the results of the training and testing data split demonstrate that the use of stratified sampling provides a solid foundation for training classification models. With balanced and consistent label distributions across both weighting methods, performance differences observed in subsequent evaluation stages can be directly attributed to the characteristics of the lexicon weighting strategies and classification algorithms employed, rather than to differences in data structure.

3.4. TF–IDF Feature Analysis

TF–IDF feature analysis was conducted to identify lexical characteristics that distinguish reviews labeled as Negative, Neutral, and Positive under each lexicon weighting method. TF–IDF representation was employed because it emphasizes informative terms within a review while reducing the influence of common words that frequently appear across documents. As such, TF–IDF features provide a relevant statistical basis for sentiment classification. The distribution of the highest weighted TF–IDF features for Method 1 (Mean Weighting) is presented in Figure 4, which illustrates the top 20 dominant features for each sentiment class.

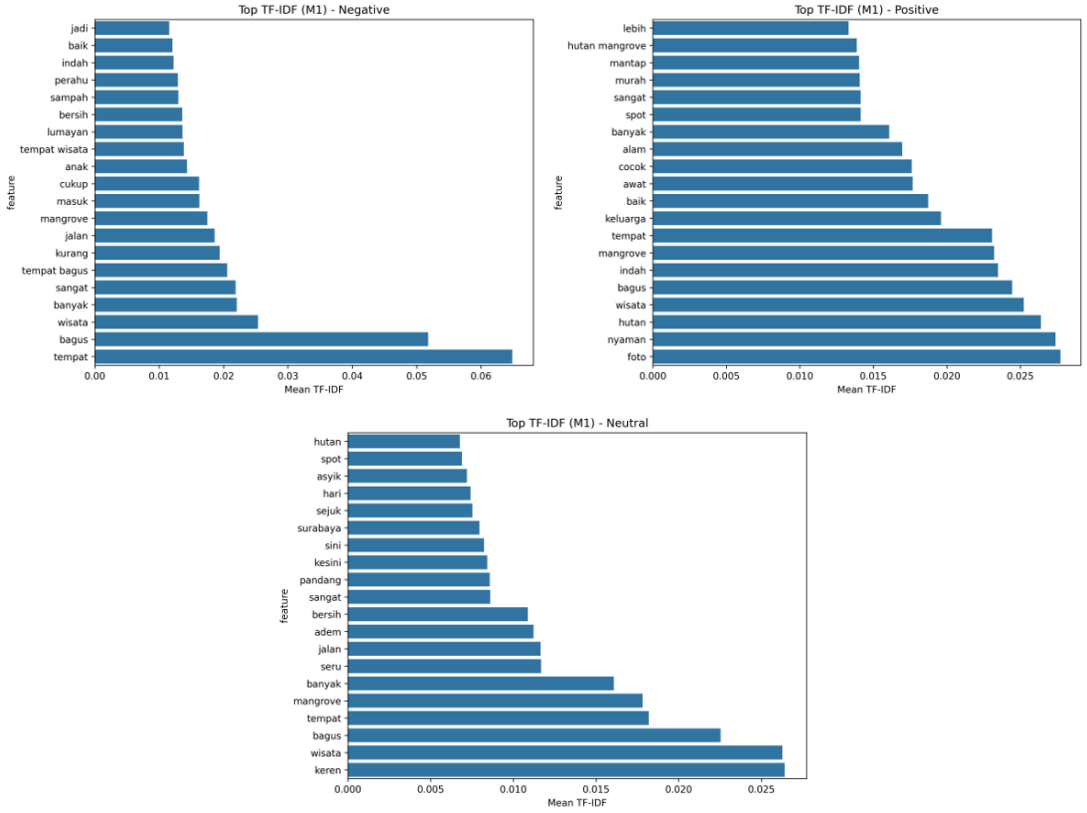

Figure 4. Top TF–IDF Features for Method 1

Based on Figure 4, the Negative class is characterized by dominant features such as “tempat” (place), “kurang” (lacking), “banyak” (many), and “sampah” (trash). Although some of these words are lexically neutral, their contextual usage in the reviews reflects negative evaluations of facilities and environmental conditions perceived as inadequate. This pattern indicates that negative reviews in the mangrove ecotourism domain tend to focus on the physical condition of locations and site management.

In contrast, the Positive class is dominated by high-weight TF–IDF features such as “indah” (beautiful), “nyaman” (comfortable), “hutan” (forest), and “foto” (photo), which reflect visual appeal, comfort, and natural attractions. The dominance of these features suggests that positive reviews emphasize aesthetic aspects and recreational experiences. Meanwhile, the Neutral class is characterized

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

by terms such as “wisata” (tourism), “mangrove”, and “bagus” (good), which are largely informative and do not explicitly convey emotional orientation. This lexical pattern explains why the Neutral class is more difficult to distinguish during classification. The TF–IDF feature distribution for Method 2 (Summation Weighting) is shown in Figure 5, which presents the dominant features for each sentiment class.

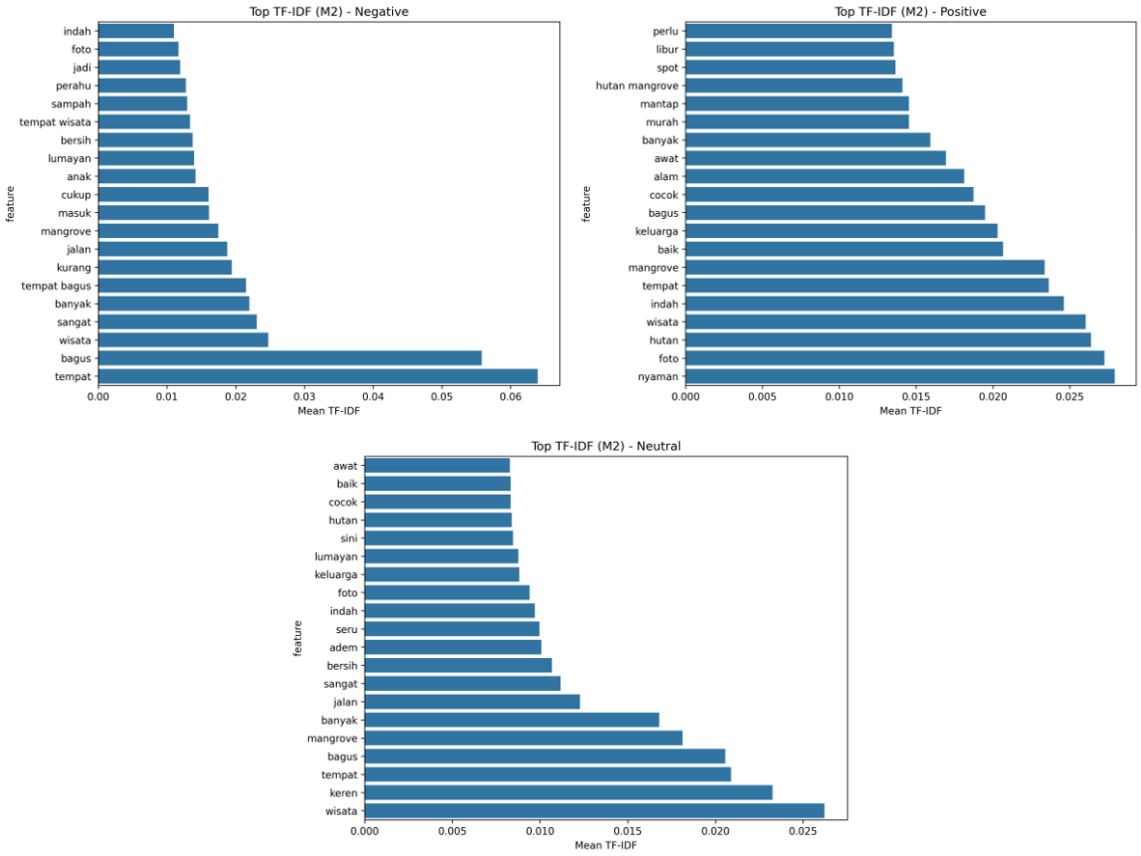

Figure 5. Top TF–IDF Features for Method 2

The results in Figure 5 indicate that the dominant features identified using Method 2 are largely similar to those obtained with Method 1. This similarity confirms that differences in lexicon weighting mechanisms (Mean and Summation) do not substantially alter the statistical structure of the vocabulary in the dataset. Consequently, differences in classification performance observed during evaluation are more strongly influenced by the quality and stability of the sentiment labels generated by each weighting method rather than by changes in TF–IDF feature distributions.

To provide a visual overview of class separability, dimensionality reduction was performed using two-dimensional Principal Component Analysis (PCA) on the TF–IDF representations, as shown in Figure 6.

The visualization in Figure 6 reveals a tendency for sentiment-based clustering, although overlaps between classes remain evident. The Negative class forms a relatively denser region compared to the Positive and Neutral classes, indicating more consistent lexical patterns in negative reviews. In contrast, substantial overlap between the Neutral class and other classes reflects lexical similarity, which contributes to misclassification in the Neutral class and is further discussed in the confusion matrix analysis.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

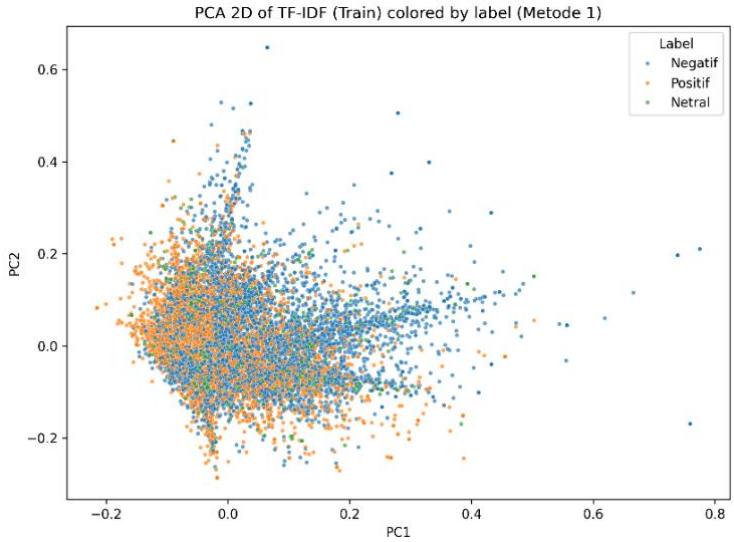

Figure 6. Two-Dimensional PCA Visualization of TF–IDF Features (Method 1)

Overall, the TF–IDF feature analysis demonstrates that the dominant features align with the semantic characteristics of mangrove ecotourism reviews. However, vocabulary overlap across sentiment classes, particularly involving ambiguous words, limits perfect class separation and represents an important factor influencing classification performance in the evaluation stage.

3.5. Evaluation

Model performance evaluation was conducted by comparing Support Vector Machine (SVM) and Logistic Regression (LR) under two lexicon weighting approaches, namely Method 1 (Mean Weighting) and Method 2 (Summation Weighting). Performance comparison focused on accuracy, macro F1-score, weighted F1-score, and prediction error patterns observed through confusion matrices. All experiments employed TF–IDF feature representations and stratified train test splits to ensure consistent class distributions.

3.5.1. Model Combination Evaluation

All models were trained using TF–IDF feature representations and evaluated on stratified test data. A summary of the overall performance for each model combination is presented in Table 6, using evaluation metrics including accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score.

Table 6. Performance Evaluation of Model Combinations

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	F1-Score (Wighted)
SVM + Lexicon Mean (Metode 1)	0,902	0,876	0,819	0,841	0,899
SVM + Lexicon Summation (Metode 2)	0,888	0,859	0,779	0,804	0,88
Logistic Regression + Lexicon Mean (Metode 1)	0,891	0,878	0,792	0,822	0,886
Logistic Regression + Lexicon Summation (Metode 2)	0,876	0,867	0,757	0,786	0,865

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

Based on Table 6, the combination of SVM with Method 1 (Mean Weighting) demonstrates the most superior and consistent performance across all evaluation metrics, achieving an accuracy of 0.902, a macro F1-score of 0.841, and a weighted F1-score of 0.899. The relatively balanced macro precision and macro recall values (0.876 and 0.819, respectively) indicate that this model not only performs well overall but also maintains balanced classification performance across sentiment classes, including the minority Neutral class. These results suggest that normalizing lexicon scores through averaging produces a more stable label structure that can be more effectively learned by a margin-based classifier such as SVM.

In contrast, when SVM is combined with Method 2 (Summation Weighting), a performance degradation is observed across all major metrics. Notably, the macro F1-score decreases from 0.841 to 0.804, and macro recall declines from 0.819 to 0.779. This reduction indicates an increase in misclassification for certain classes, particularly the Neutral class. The decline can be attributed to the tendency of the Summation method to generate polarity scores with larger magnitudes, which increases the likelihood of assigning reviews to extreme classes (Positive or Negative) and reduces model sensitivity to ambiguous or informative reviews.

A similar pattern is observed for Logistic Regression. The combination of LR with Method 1 achieves an accuracy of 0.891 and a macro F1-score of 0.822, outperforming LR combined with Method 2, which yields a macro F1-score of only 0.786. Although LR generally exhibits lower performance than SVM, these results confirm that lexicon weighting schemes play a critical role in shaping sentiment label quality, even for linear classifiers. Mean Weighting enables LR to learn TF–IDF feature distributions more proportionally, whereas Summation Weighting tends to amplify bias toward majority classes.

Overall, the results presented in Table 6 confirm that Mean Weighting is more effective than Summation Weighting, particularly when used as the basis for sentiment labeling in supervised classification. Mean based lexicon weighting is better suited to ecotourism reviews, which are typically short to medium in length and often contain words with ambiguous polarity, as it reduces the dominance effect of sentiment word frequency. These findings also indicate that SVM achieves optimal performance when trained on a stable labeling structure, making the combination of SVM and Mean Weighting the most effective approach in this study.

3.5.2. Class Performance and Confusion Matrix Analysis

Confusion matrix analysis provides a detailed view of the prediction error patterns produced by each model. Figure 7 presents the confusion matrices for the SVM based models, while Figure 8 illustrates the results for the Logistic Regression (LR) models.

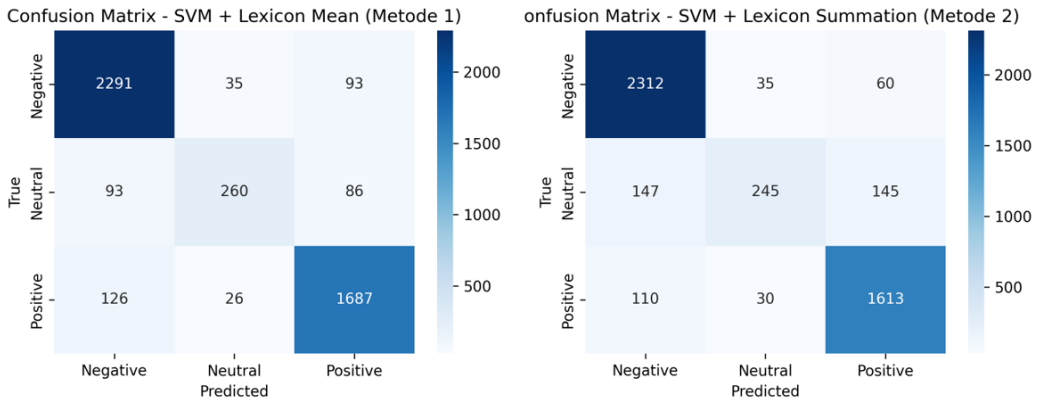

Figure 7. Confusion Matrix of the SVM Classifier

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

As shown in Figure 7, the combination of SVM with Method 1 (Mean Weighting) achieves a high number of correct predictions for the Negative and Positive classes, accompanied by relatively low cross class misclassification. Although the Neutral class remains the most challenging to predict accurately, this combination yields a higher number of correct Neutral predictions compared to the other model configurations. This pattern supports the results reported in Table 6, where this model achieves the highest macro F1-score and weighted F1-score. These findings indicate that Mean Weighting produces a more balanced sentiment label structure, enabling SVM to construct more consistent decision boundaries across classes.

In contrast, SVM combined with Method 2 (Summation Weighting), as depicted in Figure 7, exhibits an increase in cross class misclassification, particularly between the Negative and Positive classes, along with a noticeable decline in prediction accuracy for the Neutral class. This observation is consistent with the reduction in macro F1-score reported in Table 6, suggesting that score accumulation in the Summation method amplifies extreme polarity values and reduces the model's sensitivity to neutral or ambiguous reviews.

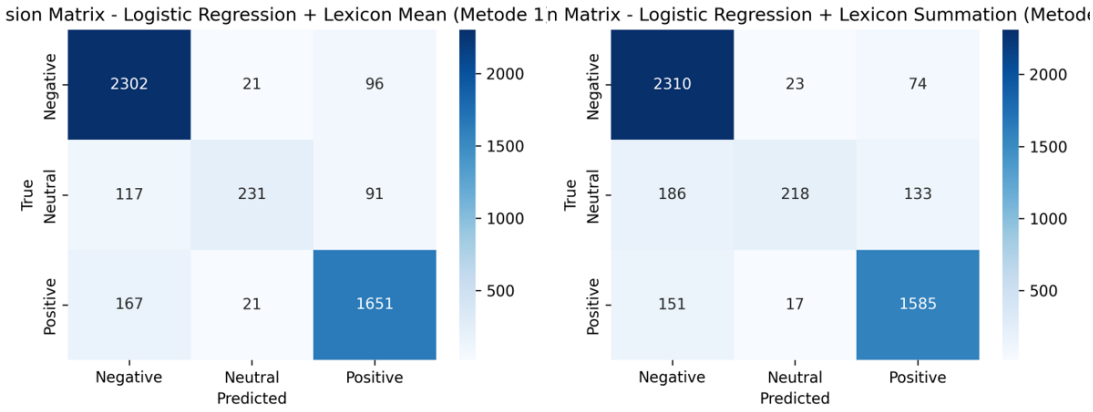

Figure 8. Confusion Matrix of the Logistic Regression Classifier

A similar trend is observed for Logistic Regression, as illustrated in Figure 8. LR combined with Method 1 maintains relatively stable performance on the majority classes but produces more misclassifications in the Neutral class compared to SVM. Meanwhile, the combination of LR with Method 2 results in the highest overall error rate, particularly for the Neutral class, which is frequently misclassified as either Negative or Positive. This outcome highlights the higher sensitivity of linear models to imbalanced sentiment score distributions, especially when lexicon weighting is based on summation.

Overall, the confusion matrix analysis presented in Figure 7 and Figure 8 confirms that the combination of SVM and Mean Weighting delivers the most consistent performance across all sentiment classes. This superiority is reflected not only in aggregate evaluation metrics but also in more controlled and interpretable misclassification patterns, particularly in reducing cross class confusion among the primary sentiment categories.

3.5.3. Error Analysis

Error analysis was conducted to identify misclassification patterns occurring in the best performing model, namely SVM with Method 1 (Mean Weighting), and to compare these patterns with other model configurations. This analysis aims to reveal sources of classification errors that are not fully

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

captured by aggregate evaluation metrics. A summary of the total number of misclassified instances for each model is presented in Table 7.

Table 7. Summary of Classification Errors

Model	Number of Misclassifications
SVM	459
Logistic Regression	513

Based on Table 7, the SVM model with Mean Weighting produces the lowest number of misclassifications compared to the other configurations. This result is consistent with its superior accuracy and F1-score reported in Table 6. The noticeable difference in error counts between Method 1 and Method 2 across both algorithms indicates that the stability of sentiment labels generated by the lexicon weighting scheme plays a direct role in reducing prediction errors, rather than the classification algorithm alone. To further investigate the nature of these errors, Table 8 presents representative examples of misclassified reviews produced by the best model (SVM + Mean Weighting).

The examples in Table 8 demonstrate that misclassifications are not solely caused by individual word weights, but also by the limitations of TF-IDF feature representation in capturing contextual relationships between words. Reviews containing positive terms such as “bagus” (good) or “ramai” (crowded) may still be labeled as negative when these words appear within an overall negative evaluative context, for instance in sentences that include complaints about facilities or cleanliness.

Table 8. Examples of Misclassified Reviews

Review Text	Actual Label	Predicted Label
“bagus tapi kotor”	Negatif	Positif
“cukup bagus”	Netral	Positif
“ramai dan sesak”	Negatif	Netral

These findings indicate the presence of contextual polarity shift, where word level polarity does not necessarily align with the sentiment of the entire sentence. This phenomenon constitutes a major source of classification errors, particularly for the Neutral class and in cross class predictions between Positive and Negative sentiments. Although the Mean Weighting method helps balance the influence of opposing polarity scores for ambiguous words, the Bag-of-Words based representation used by TF-IDF remains unable to explicitly model semantic dependencies and sentence structure, making certain misclassifications unavoidable.

Overall, this error analysis highlights that improvements in model performance are not determined solely by the choice of classification algorithm or lexicon weighting strategy, but are also constrained by the expressive capacity of the text representation. These results reinforce earlier findings that the combination of Mean Weighting and SVM yields the most stable performance, while still exhibiting limitations in handling complex linguistic context and semantic ambiguity.

3.5.4. Words Appearing in Both InSet Lexicon Dictionaries

Based on the processing results using the InSet Lexicon, which consists of a positive dictionary and a negative dictionary, one word was identified as appearing in both dictionaries. Several examples of words that appear in both dictionaries are presented in Table 9.

The results in Table 9 show that ambiguous words produce different score values when calculated using Method 1 (Mean Weighting) and Method 2 (Summation Weighting). In the Summation method, positive and negative weights are directly summed, resulting in larger polarity values. In contrast, the

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

Mean method computes the average of the two weights, producing more proportional scores. This difference in weighting mechanisms directly affects the sentiment score at the token level and the accumulation of sentiment scores at the overall review level.

Table 9. Words Appearing in Both Positive and Negative InSet Lexicons

Ambiguous Word	Positive Weight	Negative Weight	Mean Weight (M1)	Summation Weight (M2)
bagus	+3	–1	+1	+2
indah	+4	–2	+1	+2
ramai	+1	–1	0	0
bersih	+2	–1	+0.5	+1

These findings are consistent with the results of the previously conducted error analysis, where words such as “bagus”, “indah”, and “ramai” frequently appeared in misclassification cases. When such words occur in sentences containing negative evaluations, the Summation method tends to amplify the contribution of certain polarity scores, thereby increasing the likelihood of reviews being classified into extreme classes. Conversely, the Mean method produces more stable values, reducing the dominance of a single type of weight in determining the final label. Differences in lexicon weighting strategies directly influence the stability of sentiment scores at the word level, which subsequently affects the quality of review labeling. These findings reinforce earlier results indicating that Mean Weighting is more suitable for reviews containing words with dual meanings, particularly in the ecotourism context, which is rich in descriptive expressions and subjective evaluations. Nevertheless, although Mean Weighting produces more consistent weighting, the limitations of Bag of Words–based approaches in capturing sentence context remain a source of error that cannot be fully resolved.

4. DISCUSSION

This study demonstrates that the combination of the Mean Weighting lexicon strategy and the Support Vector Machine (SVM) algorithm achieves the best overall performance, with an accuracy of 0.902 and a macro F1-score of 0.841. These findings confirm that a lexicon aggregation strategy that normalizes word polarity values is more effective in handling lexical ambiguity and the phenomenon of contextual polarity shift in Indonesian review texts. This pattern is consistent with the findings reported by Abdillah et al. [13] and Alfauzan et al. [14], who emphasized that sentiment label stability has a direct impact on improving the performance of SVM based models.

Compared with the studies by Khairani et al. [20] and Noviani et al. [18], the results of this study further support the conclusion that integrating lexicon-based labeling with machine learning techniques can improve sentiment classification accuracy, particularly when using TF–IDF features. This study extends previous work by explicitly evaluating two lexicon weighting schemes, namely Mean Weighting and Summation Weighting, which have not been extensively examined in the context of the Indonesian language. The evaluation results indicate that Mean Weighting outperforms Summation Weighting because it produces a more proportional and stable label distribution. Such stability facilitates SVM in constructing optimal decision boundaries between sentiment classes.

When compared with the studies by Aulia et al. [12] and Lubihana and B. Y. [16], which applied purely lexicon based approaches in the tourism domain, this research demonstrates a significant performance improvement through the adoption of a hybrid approach combining lexicon-based labeling, TF–IDF feature representation, and SVM classification. Similar findings were reported by Syahrul and Fatharani [17], who observed an accuracy improvement of approximately 5–10% using hybrid lexicon

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

machine learning models. Therefore, this study reinforces existing evidence that hybrid approaches can effectively mitigate the limitations of static lexicons in representing contextual semantics.

From a theoretical perspective, these findings support the views of Wang et al. [29] and Jiang et al. [31], who argued that polarity aggregation strategies that account for weight balance can enhance semantic coherence in sentiment representations. The Mean Weighting approach is shown to provide a stabilizing effect on data with high ambiguity, such as ecotourism reviews, where the same word may convey positive or negative sentiment depending on its contextual usage.

From a methodological standpoint, the primary contribution of this study lies in the systematic evaluation of two lexicon weighting schemes for Indonesian sentiment analysis through an adaptive enhancement of the InSet lexicon weighting mechanism. This aspect has not been previously examined in a structured manner. The proposed approach contributes to the development of more context-aware hybrid sentiment analysis frameworks by demonstrating that the combination of Mean Weighting and SVM yields more stable and accurate results than single-lexicon models or simple summation-based weighting strategies.

5. CONCLUSION

This study confirms that the combination of the Mean Weighting lexicon strategy and the Support Vector Machine (SVM) algorithm represents the most effective approach for sentiment analysis of Indonesian-language ecotourism reviews. This approach successfully addresses the limitations of traditional lexicon-based methods by normalizing the polarity values of ambiguous words, thereby producing a more proportional label distribution and improving model performance stability. With an accuracy of 0.902 and a macro F1-score of 0.841, the SVM model combined with Mean Weighting demonstrates a strong ability to consistently classify sentiment in texts characterized by high semantic ambiguity.

The main contribution of this research lies in the systematic development and evaluation of two Indonesian lexicon weighting schemes through the application of Mean Weighting and Summation Weighting. The findings provide a new methodological foundation for the development of hybrid sentiment analysis approaches that integrate lexicon weighting, statistical feature representations (TF–IDF), and machine learning algorithms. In this regard, the study strengthens the basis for advancing Natural Language Processing (NLP) technologies in local language contexts with high semantic diversity, such as the Indonesian language.

For future research, the proposed model can be extended by integrating deep learning approaches and contextual word embedding representations, such as Bidirectional Encoder Representations from Transformers (BERT) or IndoBERT. These approaches are expected to capture contextual meaning between words more effectively, reduce errors caused by contextual polarity shifts, and improve model generalization across different domains and writing styles. Such integration may also broaden the applicability of the model to tourism recommendation systems, public opinion analysis, and other sentiment-based applications within the Indonesian NLP ecosystem.

REFERENCES

[1] A. R. Alaei, S. Becken, and B. Stantic, “Sentiment Analysis in Tourism: Capitalizing on Big Data,” *J Travel Res*, vol. 58, no. 2, pp. 175–191, Feb. 2019, doi: 10.1177/0047287517747753.

[2] J. P. Mellinas and M. Sicilia, “Comparing Google reviews and TripAdvisor to help researchers select the more appropriate information source,” *Consumer Behavior in Tourism and Hospitality*, vol. 19, no. 4, pp. 646–655, Nov. 2024, doi: 10.1108/CBTH-01-2024-0039.

[3] Eka, T. Saputra, and Wasiah Sufi, “PENGELOLAAN KAWASAN EKOWISATA HUTAN MANGROVE,” *Multidisciplinary Indonesian Center Journal (MICJO)*, vol. 1, no. 4, pp. 1806–1812, Oct. 2024, doi: 10.62567/micjo.v1i4.205.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

[4] A. Novita and E. Mukhtar, "Review Article: Mangrove Ecotourism Development Potential," *International Journal of Progressive Sciences and Technologies*, vol. 46, no. 2, p. 653, Sep. 2024, doi: 10.52155/ijpsat.v46.2.6534.

[5] S. A. Parvin, M. Sumathi, and C. Mohan, "Challenges of Sentiment Analysis - A Survey," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, Jun. 2021, pp. 781–786. doi: 10.1109/ICOEI51242.2021.9453026.

[6] S. Almatarneh, I. A. Almatarneh, G. Samara, M. Aljaidi, A. Alamleh, and A. Abuawad, "Polarity Classification of Hotel Reviews: Lexicon-Based Method," in *2022 International Arab Conference on Information Technology (ACIT)*, IEEE, Nov. 2022, pp. 1–4. doi: 10.1109/ACIT57182.2022.9994180.

[7] O. Kellert, C. Gómez-Rodríguez, and M. Uz Zaman, "Unveiling factors influencing judgment variation in sentiment analysis with natural language processing and statistics," *PLoS One*, vol. 19, no. 5, pp. 1–19, May 2024, doi: 10.1371/journal.pone.0304201.

[8] A. Rufaida, A. Permanasari, and N. Setiawan, "Lexicon-Based Sentiment Analysis Using Inset Dictionary: A Systematic Literature Review," in *Proceedings of the 5th International Conference on Applied Engineering, ICAE 2022, 5 October 2022, Batam, Indonesia*, EAI, 2023. doi: 10.4108/eai.5-10-2022.2327474.

[9] F. T. Saputra, S. H. Wijaya, Y. Nurhadryani, and Defina, "Lexicon Addition Effect on Lexicon-Based of Indonesian Sentiment Analysis on Twitter," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Nov. 2020, pp. 136–141. doi: 10.1109/ICIMCIS51567.2020.9354269.

[10] F. Akbar, . Hadiyanto, and C. E. Widodo, "Sentiment Analysis of Data on Google Maps Reviews Regarding Tourism on Keraton Kasepuhan Cirebon Using the Lexicon Based Method," in *Proceedings of the 3rd International Conference on Advanced Information Scientific Development*, SCITEPRESS - Science and Technology Publications, 2023, pp. 19–24. doi: 10.5220/0012440100003848.

[11] S. A. S. Mola, T. Widiastuti, R. V. K. I. O. Roma, A. S. Karnyoto, and B. Pardamean, "Sentiment Analysis: Indonesia Netflix User's Comment Using Multiple Lexicon-Based Dictionaries," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, IEEE, Dec. 2024, pp. 630–635. doi: 10.1109/ICICYTA64807.2024.10912916.

[12] M. A. Aulia, B. Solihah, and A. Zuhdi, "Sentiment Analysis and Topic Modeling of Tourist Reviews on Bali Island Attractions on Tripadvisor Using Lexicon-Based Method and Latent Dirichlet Allocation (LDA)," *Intelmatics*, vol. 5, no. 1, pp. 1–7, Feb. 2025, doi: 10.25105/v5i1.17619.

[13] W. F. Abdillah, A. Premana, and R. M. H. Bhakti, "Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 02, pp. 160–170, Nov. 2021, doi: 10.46772/intech.v3i02.556.

[14] M. F. Alfauzan, Y. Sibaroni, and F. Fitriyani, "Sentiment Classification of Fuel Price Rise in Economic Aspects Using Lexicon and SVM Method," *sinkron*, vol. 8, no. 4, pp. 2526–2536, Oct. 2023, doi: 10.33395/sinkron.v8i4.12851.

[15] T. Hendrawati, N. L. W. S. R. Ginantra, and C. M. Saiman, "Analisis Sentimen Larangan Impor Pakaian Bekas Menggunakan Metode Support Vectore Machine dan Lexicon Based," *TEMATIK*, vol. 11, no. 1, pp. 56–64, Jun. 2024, doi: 10.38204/tematik.v11i1.1890.

[16] E. Lubihana and B. Y., "Design of a Tourism Recommendation System Based on Sentiment Analysis with Lexicon LSTM," in *2022 International Symposium on Electronics and Smart Devices (ISESD)*, IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/ISESD56103.2022.9980738.

[17] E. Syahrul and D. Fatharani, "HYBRID SENTIMENT ANALYSIS OF MAXIM APP USERS USING SUPPORT VECTOR MACHINE AND LEXICON-BASED APPROACH," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, Oct. 2025, doi: 10.23960/jitet.v13i3S1.8148.

[18] E. F. Noviani, D. Purwitasari, and R. W. Sholikah, "Sentiment Analysis of Indonesian Temple Reviews Using Lexicon-Based Features and Stochastic Gradient Descent," in *2023 International*

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

Conference on Information Technology and Computing (ICITCOM), IEEE, Dec. 2023, pp. 232–237. doi: 10.1109/ICITCOM60176.2023.10442938.

[19] A. May Nggiri, F. Hariadi, and N. Berlian Uly, "Analysis of Visitor Sentiment to Matayangu Waterfall Tourism in Central Sumba Regency Using Naïve Bayes," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 5, no. 1, pp. 397–404, Oct. 2025, doi: 10.59934/jaiea.v5i1.1333.

[20] D. Khairani, A. Setiawan, and S. U. Masruroh, "Enhancing Understanding of Public Sentiment on Twitter Using SVM and Lexicon Methods," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*, IEEE, Aug. 2024, pp. 1–5. doi: 10.1109/ICCIT62134.2024.10701213.

[21] M. K. Anam, T. A. Fitri, A. Agustin, L. Lusiana, M. B. Firdaus, and A. T. Nurhuda, "Sentiment Analysis for Online Learning using The Lexicon-Based Method and The Support Vector Machine Algorithm," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 290–302, Aug. 2023, doi: 10.33096/ilkom.v15i2.1590.290-302.

[22] A. Nadira, N. Y. Setiawan, and W. Purnomo, "ANALISIS SENTIMEN PADA ULASAN APLIKASI MOBILE BANKING MENGGUNAKAN METODE NAÏVE BAYES DENGAN KAMUS INSET," *INDEXIA*, vol. 5, no. 01, p. 35, Apr. 2023, doi: 10.30587/indexia.v5i01.5138.

[23] N. B. Bahadure *et al.*, "Comparative Analysis of Polarity of Text-based Sentiment Analysis," in *2024 3rd International Conference for Innovation in Technology (INOCON)*, IEEE, Mar. 2024, pp. 1–5. doi: 10.1109/INOCON60754.2024.10512041.

[24] T. P. Sahu and S. Khandekar, "A Machine Learning-Based Lexicon Approach for Sentiment Analysis," in *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, IGI Global, 2022, pp. 836–851. doi: 10.4018/978-1-6684-6303-1.ch044.

[25] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," in *2017 International Conference on Asian Language Processing (IALP)*, IEEE, Dec. 2017, pp. 391–394. doi: 10.1109/IALP.2017.8300625.

[26] D. H. Abd, A. R. Abbas, and A. T. Sadiq, "Analyzing sentiment system to specify polarity by lexicon-based," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 283–289, Feb. 2021, doi: 10.11591/eei.v10i1.2471.

[27] M. D. Almeida, V. M. Maia, R. Tommasetti, and R. de O. Leite, "Sentiment analysis based on a social media customised dictionary," *MethodsX*, vol. 8, p. 101449, 2021, doi: 10.1016/j.mex.2021.101449.

[28] S. Sazzed, "Understanding Linguistic Variations in Neutral and Strongly Opinionated Reviews," in *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, Nassau, Bahamas: IEEE, Dec. 2022, pp. 1512–1516. doi: 10.1109/ICMLA55696.2022.00237.

[29] S. Wang, G. Lv, S. Mazumder, and B. Liu, "Detecting Domain Polarity-Changes of Words in a Sentiment Lexicon," in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2021, pp. 3657–3668. doi: 10.18653/v1/2021.findings-acl.320.

[30] Y. Wang, F. Yin, J. Liu, and M. Tosato, "Automatic construction of domain sentiment lexicon for semantic disambiguation," *Multimed Tools Appl*, vol. 79, no. 31–32, pp. 22355–22373, Aug. 2020, doi: 10.1007/s11042-020-09030-1.

[31] Z. Jiang, Y. Zhang, C. Liu, J. Chen, J. Zhao, and K. Liu, "Interpreting Sentiment Composition with Latent Semantic Tree," in *Findings of the Association for Computational Linguistics: ACL 2023*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2023, pp. 7464–7478. doi: 10.18653/v1/2023.findings-acl.471.

[32] Kavyasri. G, "Margin Maximization of Text Classification based on Support Vector Machine," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 12, pp. 789–792, Dec. 2023, doi: 10.22214/ijraset.2023.57420.

[33] D. M. Ulya, J. Juhari, R. E. Yuliana, and M. Jamhuri, "Reliable and Efficient Sentiment Analysis on IMDb with Logistic Regression," *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 10, no. 2, pp. 821–834, Aug. 2025, doi: 10.18860/cauchy.v10i2.33809.

Vol. 6, No. 6, December 2025, Page. 5679-5698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5563

[34] R. Kansal and C. Diwaker, "Efficiency Determination of Various Machine Learning Techniques for Sentiment Analysis on Social Media Platforms," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25584–25589, Aug. 2025, doi: 10.48084/etasr.11158.

[35] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel)*, vol. 12, no. 7, p. 113, Jul. 2024, doi: 10.3390/technologies12070113.