Vol. 8, No. 4, August 2026

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Sentiment Analysis of Public Opinion on The Plastic Waste Issue on Social Media X Using TF-IDF, Naïve Bayes, and SVM

Ary Kania Sya'diah[1], Muhammad Arifin[2], Pratomo Setiaji[3]

[1]Information System Department, Postgraduate Program, Muria Kudus University, Kudus, Indonesia

[2,3]Information System Department, Faculty of Engineering, Muria Kudus University, Kudus, Indonesia

Received:
February 25, 2026
Revised:
July 20, 2026
Accepted:
August 4, 2026
Published:
August 29, 2026

Corresponding Author:
Author Name*:
Ary Kania Sya'diah
Email*:
202253007@std.umk.ac.id

DOI:
10.63158/journalisi.v8i4.1782

Abstract. Plastic waste has become a critical environmental challenge due to increasing consumption patterns and inadequate waste management practices. Understanding public perceptions of plastic waste issues is essential for supporting environmental awareness and policy development. Social Media X provides a large-scale and real-time source of public opinions that can be analyzed through sentiment analysis techniques. This study aims to identify public sentiment trends regarding plastic waste issues on Social Media X and compare the performance of Naïve Bayes and Support Vector Machine (SVM) algorithms using TF-IDF feature extraction. The research applies the CRISP-DM framework, including data understanding, data preparation, preprocessing, modeling, evaluation, and visualization stages. A total of 7,277 Indonesian-language posts were collected through web scraping, with 7,275 posts retained after data preparation. The preprocessing process consisted of cleansing, case folding, tokenization, stopword removal, normalization, and lexicon-based sentiment labeling. The dataset was classified into three sentiment categories: positive, negative, and neutral. Model performance was evaluated using Accuracy, Precision, Recall, F1-score, Macro F1-score, Balanced Accuracy, and Stratified 5-Fold Cross-Validation. The results show that SVM achieved better performance than Naïve Bayes, with an accuracy of 83.64%, Macro F1-score of 79.00%, and Balanced Accuracy of 75.94%. These findings indicate that SVM is more effective for sentiment classification of Indonesian plastic waste discussions using TF-IDF-based text representation.

Keywords: Sentiment Analysis; Plastic Waste; Social Media X; TF-IDF; Support Vector Machine

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Plastic waste has become one of the most serious environmental problems worldwide. The high use of plastic in daily life, which is not matched by optimal waste management, leads to the accumulation of plastic waste in various environments, both on land and in waterways. Global plastic waste production continues to increase significantly, and most of this waste is not properly managed, with around 79% of it ending up in the environment or in final disposal sites, causing adverse impacts on the environment, such as soil, water, and marine pollution. [1]. The three types of waste that predominantly dominate the waters and coasts of Indonesia are plastic bags, food packaging, and plastic bottles. In 2021, Indonesia generated 11.6 million tons of plastic waste, or 17% of the total national waste, yet only 9% was successfully recycled. Every day, around 25,000 tons of plastic waste are added, with at least 20% ending up in rivers and coastal waters. [2]. This finding was obtained from a study of 68 locations spread across various regions, including beaches, rivers, and estuaries. [3]. This issue is not only a concern for the government but also requires the attention of the wider community.

With the rapid advancement of information technology, social media has become one of the primary platforms for the public to express opinions on various issues, including plastic waste issues. Social Media X provides real-time public opinions in large volumes, making it a valuable data source for sentiment analysis. Recent studies have also demonstrated the effectiveness of machine learning techniques for analyzing public sentiment on social media across various public issues, including health-related events and emerging societal topics [4]. Previous research by [5] demonstrated that data from Social Media X can be used to identify public perceptions and opinions regarding particular policies or issues. Sentiment analysis is a data mining technique used to identify and classify opinions into positive, negative, and neutral categories [6]. Through sentiment analysis, researchers can identify public opinion trends and understand how people respond to particular issues based on social media data. In this study, positive sentiment refers to opinions that support plastic waste reduction, recycling initiatives, environmental campaigns, or effective waste management. Negative sentiment refers to criticism of plastic pollution, excessive plastic use, or ineffective waste management. Neutral sentiment includes factual information or statements that do not express a clear positive or negative opinion. According to [7], [8], [9], the performance of sentiment

analysis is also influenced by the quality of preprocessing and feature extraction, such as cleaning, case folding, tokenization, stopword removal, normalization, as well as TF-IDF weighting, which generate informative text representations that improve sentiment classification performance.

Previous studies have shown that Support Vector Machine (SVM) generally achieves better performance than Naïve Bayes, particularly for high-dimensional text classification tasks [10]. However, Naïve Bayes is still widely used because it is simple, computationally efficient, and provides satisfactory performance in many text classification applications. More recent studies have also explored deep learning models, such as LSTM and IndoBERT, showing that classification performance is influenced not only by the learning algorithm, but also by preprocessing techniques, feature representation, and the characteristics of the dataset [11].

Although various machine learning and deep learning methods have been applied to sentiment analysis [12], studies specifically investigating public sentiment toward plastic waste on Social Media X remain limited. Existing studies have largely focused on different application domains and frequently relied on accuracy as the primary evaluation metric. Comparative evaluations incorporating lexicon-based sentiment labeling, TF-IDF feature extraction, Macro F1-score, Balanced Accuracy, and Stratified 5-Fold Cross-Validation for Indonesian plastic waste discussions are still scarce. Consequently, further investigation is required to provide a more comprehensive evaluation of traditional machine learning models on this environmental issue.

This study contributes by providing a comprehensive comparative evaluation of Naïve Bayes and Support Vector Machine for Indonesian plastic waste sentiment classification using a manually constructed lexicon-based sentiment labeling approach, TF-IDF feature extraction, and multiple evaluation metrics, including Macro F1-score, Balanced Accuracy, and Stratified 5-Fold Cross-Validation. In addition, this study provides an updated dataset and a descriptive overview of public sentiment regarding plastic waste discussions on Social Media X.

Naïve Bayes and Support Vector Machine (SVM) were selected because both are widely used algorithms for text classification. Naïve Bayes is computationally efficient, whereas

SVM generally performs better on high-dimensional feature spaces represented by TF-IDF. Therefore, comparing these algorithms helps identify which method is more suitable for sentiment classification of Indonesian plastic waste discussions on Social Media X. Based on the identified research gap, this study aims to analyse public sentiment regarding plastic waste issues on Social Media X and compare the classification performance of both algorithms using TF-IDF feature extraction. The findings provide empirical evidence on the performance of the two algorithms and may serve as a reference for future studies on environmental sentiment analysis.

Based on the identified research gap, this study addresses the following research questions: (1) What is the distribution of positive, neutral, and negative public sentiment regarding plastic waste issues on Social Media X? and (2) Which algorithm, Naïve Bayes or Support Vector Machine (SVM), achieves better performance in sentiment classification using TF-IDF feature extraction?

2. METHODS

This study employs a quantitative approach using text mining techniques based on the CRISP-DM (Cross Industry Standard Process for Data Mining) framework. The framework consists of five stages: Data Understanding, Data Preparation, Modeling, Evaluation, and Visualization [13]. These stages were applied to analyze public sentiment regarding plastic waste issues on Social Media X. This framework was selected because it provides a structured and systematic process for analyzing unstructured text data collected from social media. The research process consists of text preprocessing, as well as feature extraction using TF-IDF, so that the data can be transformed into a numerical representation. The processed data were then classified using the Naïve Bayes and Support Vector Machine (SVM) algorithms. The performance of both models was evaluated using Accuracy, Precision, Recall, F1-score, Macro F1-score, and Balanced Accuracy. In addition, Stratified 5-Fold Cross-Validation was performed to assess the robustness and stability of the proposed models. The evaluation results were used to compare the effectiveness of both algorithms in classifying public sentiment regarding plastic waste issues on Social Media X. The overall research stages are illustrated in Figure 1.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

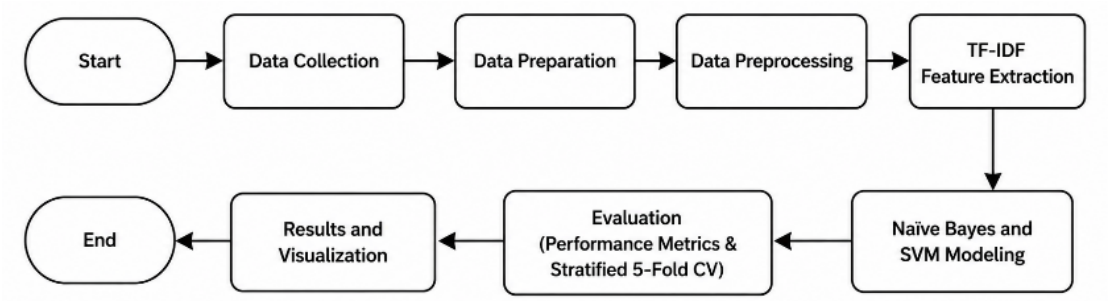

Figure 1. Research Method Flow

2.1 Research Object, Tools and Supporting Devices

The research object in this study includes public opinions on the issue of plastic waste, obtained from posts by Social Media X users related to the topic of plastic waste. The dataset used consists of texts containing public opinions, whether expressing support, criticism, or neutral information. The collected posts were then analyzed to identify public sentiment trends and classified into three categories: positive, negative, and neutral.

This study uses Python 3.12.13, executed in Google Colab. The supporting libraries include Pandas 2.2.2 for data handling, NLTK 3.9.1 for Indonesian stopwords removal, Scikit-Learn for TF-IDF feature extraction and machine learning implementation (Naïve Bayes and Support Vector Machine), OpenPyXL for spreadsheets processing, and Matplotlib and Seaborn for visualizing data. Data were collected using the X API and web scraping techniques. The detailed data collection procedure is described in Section 2.3.

2.2 Data Collection

The data collection stage is an initial process of understanding the data used in this research. At this stage, the Data Understanding process was also carried out to examine the characteristics of the dataset by analyzing the amount of data, data distribution, and the quality of the available data. Data were collected from Social Media X using a combination of the X API and web scraping techniques. The X API was used to retrieve publicly available metadata, while web scraping was used to collect publicly accessible post content related to plastic waste. The data collection was conducted from June 20 to 21, 2026, using keywords related to plastic waste issues, namely "sampah plastik", "limbah plastik", "pencemaran plastik", "plastik sekali pakai", and "sampah plastik di laut". The data collection process successfully collected 7,277 posts, which were subsequently

stored and processed for sentiment analysis. Only posts written in the Indonesian language were retained for further analysis.

During the data preparation stage, records with missing values in the text field were removed to improve data quality. As a result, 2 records were removed, leaving a final dataset of 7,275 posts. Reposts were not treated separately and were retained as regular posts if they contained the specified keywords. In addition, no automatic bot or spam filtering was applied because the publicly available metadata were insufficient to reliably identify automated or spam accounts. This is acknowledged as a limitation of the study, as automated accounts may have influenced the observed sentiment distribution.

0	Hampir setiap hari kita memproduksi sampah, mu...
1	Spring cleaning the kitchen today. Standard la...
2	Tumpukan sampah menggunung kembali terlihat di...
3	melipat plastik sampah yg amburadul, nessa ngo...
4	uda decluttering lemari pun masi penuuhh. 3 ka...

Figure 2. Data Collection Samples (Indonesian Text)

The collected data consisted only of publicly available posts on Social Media X. Personal information such as usernames, profile URLs, and user identifiers was excluded from the dataset to protect user privacy. Data collection and analysis were conducted solely for research purposes in accordance with the platform's publicly accessible content policy.

2.3 Data Preparation

The data preparation stage was conducted to ensure the quality and relevance of the collected dataset. This stage involved selecting the required attributes, removing incomplete records, eliminating duplicate data, and filtering posts unrelated to plastic waste issues. The purpose of this stage was to obtain a clean and relevant dataset suitable for sentiment analysis. The output of this stage served as the input for the preprocessing process.

2.4 Data Preprocessing

Data preprocessing was conducted to transform raw Social Media X posts into structured and consistent textual data suitable for sentiment classification. This stage aims to

reduce textual noise, improve data quality, and ensure that the extracted features accurately represent the semantic characteristics of the posts. The preprocessing pipeline consisted of several sequential steps, including cleansing, case folding, tokenization, stopword removal, normalization, sentiment labelling, and data splitting. An example of the preprocessing transformation process from raw text into labelled data is presented in Table 1.

The first preprocessing step was cleansing, which involved removing unnecessary elements from the text, including URLs, hashtags, mentions, emojis, numbers, punctuation marks, and other irrelevant characters. Although some of these elements may contain sentiment-related information, they were excluded to maintain consistency across the dataset and reduce textual noise during the feature extraction process. The second step was case folding, where all characters in the text were converted into lowercase format. This process was applied to eliminate differences between uppercase and lowercase representations of the same word, ensuring that identical words were treated as the same feature during subsequent analysis. The third step was tokenization, which separated sentences into individual tokens or words. The resulting tokens served as the basic textual units for further preprocessing and feature extraction. This step enables the classification model to analyze the contribution of each word independently.

The fourth step was stopword removal, which eliminated common words that provide limited contribution to sentiment classification. Words with low semantic importance, such as frequently occurring functional words, were removed to reduce the dimensionality of the text representation and allow the model to focus on more informative terms.

The fifth step was normalization, which converted non-standard words into their standard forms to improve consistency within the dataset. This step was particularly important for Social Media X data, where informal expressions, abbreviations, and slang words are frequently used. For example, informal words such as “gak” and “udah” were normalized into their standard forms “tidak” and “sudah”. This process helped reduce vocabulary variation caused by informal writing styles.

The sixth step was sentiment labelling, which was performed using a lexicon-based approach. A sentiment lexicon was manually developed specifically for this study based on Indonesian words commonly associated with plastic waste discussions on Social Media X. The lexicon consisted of positive and negative sentiment words, where each word was assigned a sentiment score ranging from 0 to 2 according to its polarity and intensity. Positive sentiment words included terms such as "baik", "bagus", "peduli", "bersih", "lestari", and "apresiasi", whereas negative sentiment words included "limbah", "pencemaran", "bahaya", "merusak", "krisis", and "polusi". The lexicon was constructed specifically for this research and was not adapted from existing Indonesian sentiment lexicons.

The sentiment score of each post was calculated by matching the preprocessed words with the developed lexicon and subtracting the total negative sentiment score from the total positive sentiment score. Posts with sentiment scores greater than or equal to 2 were categorized as positive, while posts with scores lower than or equal to −2 were categorized as negative. Posts with scores between −1 and 1 were classified as neutral. These threshold values were selected to minimize the effect of isolated sentiment words and ensure that sentiment labels represented stronger overall opinions rather than individual word occurrences. However, this lexicon-based approach has limitations because it relies on direct word matching and does not explicitly capture contextual meaning, negation expressions, sarcasm, or irony. Indonesian slang and abbreviated terms were partially addressed during the normalization process to improve the robustness of sentiment identification.

The final preprocessing stage was data splitting, where the labelled dataset was divided into training and testing subsets using a stratified 80:20 ratio. Stratified splitting was applied to preserve the original sentiment class distribution in both subsets. The final dataset consisted of 5,820 posts in the training set and 1,455 posts in the testing set. The training data were utilized for developing and optimizing the sentiment classification models, while the testing data were used to evaluate the final model performance. The complete example of each preprocessing transformation, from the original Social Media X post to the final sentiment label, is illustrated in Table 1.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 1. Example of Preprocessing Stage

Preprocessing Stage	Text Transformation
Original Text	Rayain Hari Penyu Sedunia sebetulnya gak perlu muluk-muluk kok. Mulai dari hal kecil kayak ngurangin penggunaan sampah plastik, itu udah ngebantu banget buat jagain rumah mereka di laut.
Cleansing	Rayain Hari Penyu Sedunia sebetulnya gak perlu muluk-muluk kok. Mulai dari hal kecil kayak ngurangin penggunaan sampah plastik itu udah ngebantu banget buat jagain rumah mereka di laut
Case Folding	rayain hari penyu sedunia sebetulnya gak perlu mulukmuluk kok mulai dari hal kecil kayak ngurangin penggunaan sampah plastik itu udah ngebantu banget buat jagain rumah mereka di laut
Tokenization	['rayain', 'hari', 'penyu', 'sedunia', 'sebetulnya', 'gak', 'perlu', 'mulukmuluk', 'kok', 'mulai', 'dari', 'hal', 'kecil', 'kayak', 'ngurangin', 'penggunaan', 'sampah', 'plastik', 'itu', 'udah', 'ngebantu', 'banget', 'buat', 'jagain', 'rumah', 'mereka', 'di', 'laut']
Stopword Removal	['rayain', 'penyu', 'sedunia', 'gak', 'mulukmuluk', 'kayak', 'ngurangin', 'penggunaan', 'sampah', 'plastik', 'udah', 'ngebantu', 'banget', 'jagain', 'rumah', 'laut']
Normalization	['rayain', 'penyu', 'sedunia', 'tidak', 'mulukmuluk', 'kayak', 'ngurangin', 'penggunaan', 'sampah', 'plastik', 'sudah', 'ngebantu', 'banget', 'jagain', 'rumah', 'laut']
Labelling	Neutral

2.5 Feature Extraction TF-IDF

After preprocessing was completed, TF-IDF (Term Frequency–Inverse Document Frequency) was applied for feature extraction. [14]. This method is used to represent text data in a numerical form that can be further processed using machine learning algorithms, as shown in Equation 1.

$$TF - IDF = TF(t,d) \times IDF(t) \quad (1)$$

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

TF-IDF assigns a weight to each term based on its frequency within a document and its distribution across the document collection, allowing important terms to receive higher weights while reducing the influence of common terms [15]. Before feature extraction, the dataset was divided into training and testing sets using a stratified 80:20 split to preserve the original class distribution. The TF-IDF vectorizer was configured with max_features = 5000, ngram_range = (1,1), min_df = 2, max_df = 0.95, and L2 normalization. The vectorizer was fitted only on the training dataset using the fit_transform() function and subsequently applied to the testing dataset using the transform() function to prevent data leakage. The resulting TF-IDF vectors consisted of 5,000 features, which were used as numerical representations of the text data in the classification stage.

2.6 Classification Models

This stage focuses on developing sentiment classification models using the Naïve Bayes and Support Vector Machine (SVM) algorithms. Naïve Bayes was selected because of its computational efficiency, whereas SVM was selected for its ability to handle high-dimensional text data effectively [16]. The TF-IDF feature vectors generated in the previous stage were used as input for both classification models. The Naïve Bayes classifier was implemented using the Multinomial Naïve Bayes (MultinomialNB) algorithm, which is widely used for text classification because of its probabilistic learning mechanism and suitability for document representation using term-frequency features [17]. Meanwhile, the Support Vector Machine classifier was implemented using LinearSVC with random_state = 42 and max_iter = 5000. Since SVM constructs an optimal hyperplane with the maximum margin between classes [18]. LinearSVC is well suited for high-dimensional sparse text data [19]. Both models were trained using the training dataset and evaluated on the testing dataset to classify sentiments into three categories: positive, negative, and neutral. The classification results produced by the Naïve Bayes and SVM models were then compared to determine which algorithm provided better performance for sentiment classification of plastic waste discussions on Social Media X.

2.7 Evaluation

The evaluation phase aims to measure the model's performance using a confusion matrix and the metrics as shown in Equation 2 to 7:

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (2)$$

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \quad (3)$$

$$Recall = \frac{\text{TP}}{\text{TP+FN}} \quad (4)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$Macro\ F1 = \frac{F1\ negative + F1\ neutral + F1\ positive}{3} \quad (6)$$

$$Balanced\ Accuracy = \frac{Recall\ negative + Recall\ neutral + Recall\ positive}{3} \quad (7)$$

In addition to the train–test split evaluation, stratified 5-Fold Cross-Validation was performed to assess the robustness and stability of the classification models. The dataset was divided into five folds while preserving the original class distribution in each fold. In each iteration, four folds were used for training and the remaining fold was used for testing until every fold had been used once as the testing set. The mean and standard deviation of Accuracy, Macro F1-score, and Balanced Accuracy were then calculated to evaluate the consistency of both models across different data partitions.

2.8 Visualization

The visualization stage is the final phase that presents the sentiment analysis results and model evaluation in an informative and easily understandable form. The results are visualized using tables, bar charts, confusion matrices, sentiment distribution charts, Word Cloud visualizations, and model performance comparison graphs. In addition, the results of Stratified 5-Fold Cross-Validation are presented in tabular form to illustrate the average performance and standard deviation of Accuracy, Macro F1-score, and Balanced Accuracy for both classification models. These visualizations provide a comprehensive overview of sentiment distribution, the dominant terms appearing in the dataset, the comparative performance of the Naïve Bayes and Support Vector Machine (SVM) models, and the consistency of their performance across different data partitions. The obtained results are then interpreted to understand public opinion trends regarding plastic waste issues on Social Media X.

3. RESULTS AND DISCUSSION

3.1. Sentiment Distribution

After the preprocessing and lexicon-based labeling stages, the dataset consisted of 7,275 posts, including 3,462 neutral, 3,250 negative, and 563 positive sentiments. The distribution indicates that positive sentiment represents the minority class, while neutral and negative sentiments dominate the dataset. This class imbalance may influence the performance of classification models, particularly their ability to recognize the minority class. Therefore, evaluation metrics beyond overall accuracy, such as Macro F1-score and Balanced Accuracy, were included to provide a more comprehensive assessment of model performance.

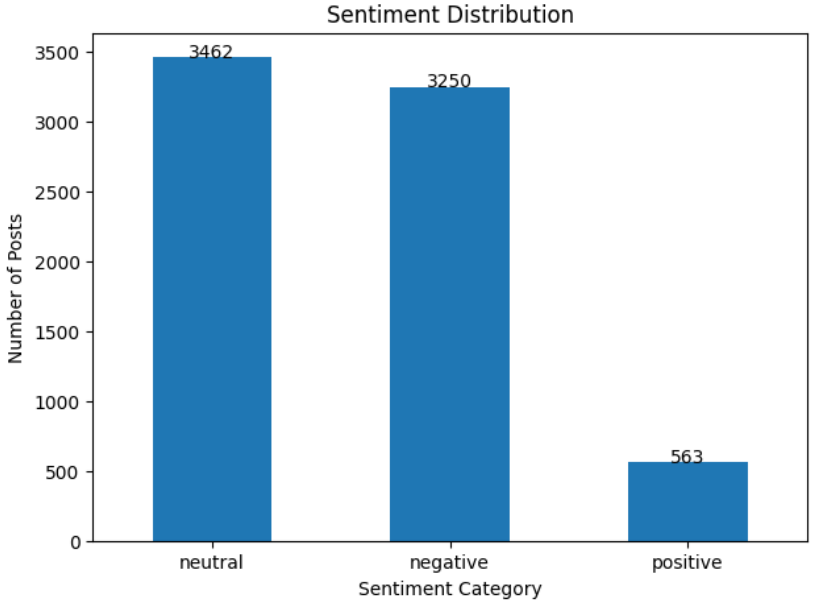

Figure 3. Sentiment Distribution

3.2. Classification Performance of Naïve Bayes

The Naïve Bayes model was employed to classify sentiment related to plastic waste issues into three categories: positive, neutral, and negative. Based on the evaluation results, the model achieved an accuracy of 68.66%, precision of 64.25%, recall of 68.66%, and F1-score of 65.89%. To provide a more reliable evaluation for the imbalanced dataset, Macro F1-score and Balanced Accuracy were also calculated. The model obtained a Macro F1-score of 47.63% and a Balanced Accuracy of 49.49%, indicating that its overall performance across all sentiment classes was considerably lower than suggested by the accuracy metric alone.

Table 2. Analysis Performance of Naïve Bayes

	Precision	Recall	F1-Score	Support
Negative	0.76	0.68	0.71	650
Neutral	0.64	0.81	0.71	692
Positive	0.00	0.00	0.00	113

According to the classification report, the negative sentiment class obtained a precision of 0.76, recall of 0.68, and F1-score of 0.71. Similarly, the neutral sentiment class achieved a precision of 0.64, recall of 0.81, and F1-score of 0.71. However, the positive sentiment class recorded precision, recall, and F1-score values of 0.00, indicating that the model failed to correctly identify positive sentiment instances in the testing dataset.

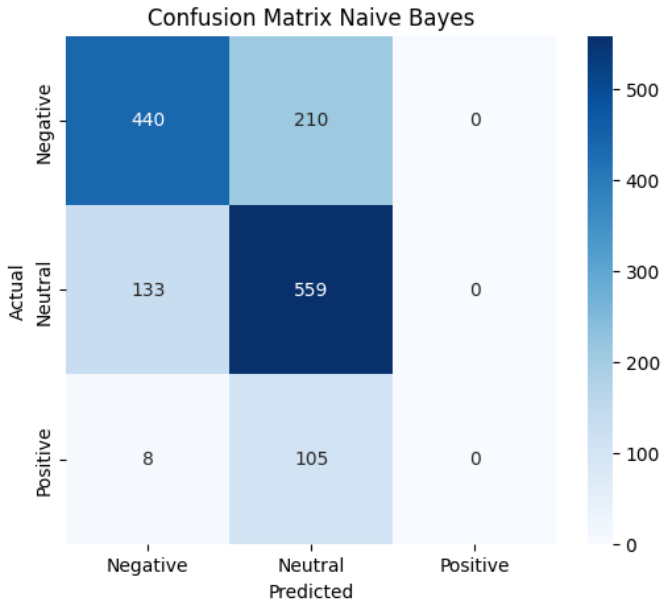

Figure 4. Confusion Matrix of Naïve Bayes

The confusion matrix shown in Figure 4 further illustrates the classification performance of the Naïve Bayes model. A total of 440 negative posts and 559 neutral posts were correctly classified. In contrast, none of the 113 positive posts were correctly predicted. Most positive posts were misclassified as neutral sentiment, indicating that the classifier was strongly influenced by the dominant sentiment classes. These results indicate that the Naïve Bayes model was substantially affected by the class imbalance in the dataset. Although the model achieved an overall accuracy of 68.66%, its Macro F1-score of 0.4763 and Balanced Accuracy of 0.4949 reveal that the classifier did not perform equally well

across all sentiment classes. This discrepancy occurred because the positive class contained only 563 of the 7,275 posts, making it considerably more difficult for the model to learn its characteristics. Consequently, the classifier tended to favor the majority classes (negative and neutral), resulting in poor detection of minority-class sentiments.

3.3. Classification Performance of Support Vector Machine (SVM)

The Support Vector Machine (SVM) model demonstrated strong performance in classifying sentiment related to plastic waste issues into positive, neutral, and negative categories. Based on the evaluation results, the SVM model achieved an accuracy of 83.64%, precision of 83.69%, recall of 83.64%, and F1-score of 83.41%. To provide a more comprehensive evaluation for the imbalanced dataset, Macro F1-score and Balanced Accuracy were also calculated. The SVM model obtained a Macro F1-score of 79.00% and a Balanced Accuracy of 75.94%, indicating that it maintained good classification performance across all sentiment classes, including the minority positive class.

Table 3. Analysis Performance of Support Vector Machine

	Precision	Recall	F1-Score	Support
Negative	0.85	0.87	0.86	650
Neutral	0.82	0.85	0.83	692
Positive	0.85	0.56	0.67	113

According to the classification report, the negative sentiment class obtained the highest performance with a precision of 0.85, recall of 0.87, and F1-score of 0.86. The neutral sentiment class also showed satisfactory results, achieving a precision of 0.82, recall of 0.85, and F1-score of 0.83. Meanwhile, the positive sentiment class achieved a precision of 0.85, recall of 0.56, and F1-score of 0.67. The lower recall indicates that several positive posts were still misclassified into other sentiment categories.

The confusion matrix presented in Figure 5 further confirms the effectiveness of the SVM model. Most negative and neutral sentiment posts were correctly classified, as indicated by the high values on the main diagonal of the matrix. A total of 568 negative posts, 586 neutral posts, and 63 positive posts were correctly classified. Although several positive posts were still misclassified as neutral sentiment, the SVM model successfully recognized a substantial portion of the minority class, unlike the Naïve Bayes model,

which failed to identify any positive instances. Overall, the SVM model demonstrated strong classification performance despite the imbalanced sentiment distribution. The Macro F1-score 0.7900 and Balanced Accuracy 0.7594 indicate that the model maintained balanced performance across all sentiment classes rather than relying solely on the majority classes. Compared with Naïve Bayes, SVM showed a considerably better ability to classify minority-class sentiments, making it more suitable for sentiment analysis of plastic waste discussions represented by high-dimensional TF-IDF features.

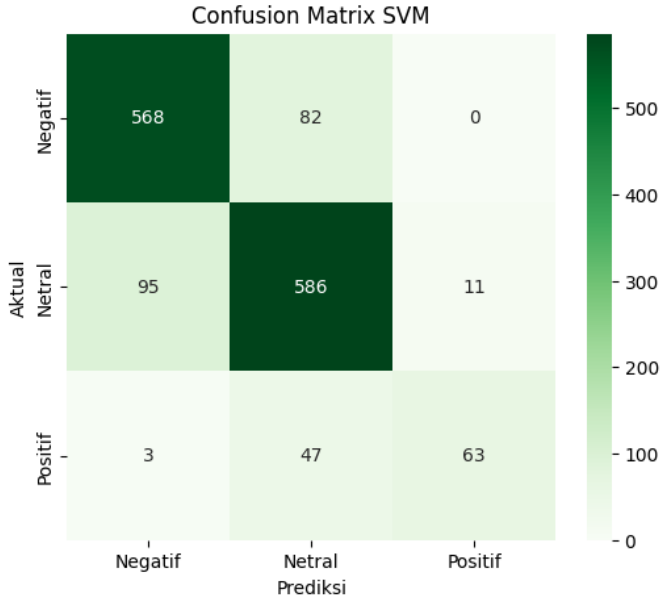

Figure 5. Confusion Matrix of Support Vector Machine

3.4. Comparison Testing Algorithm

To evaluate the effectiveness of the classification algorithms, the performance of Naïve Bayes and Support Vector Machine (SVM) was compared using accuracy, precision, recall, and F1-score metrics, Macro F1-score, and Balanced Accuracy. Since the dataset exhibits class imbalance, Macro F1-score and Balanced Accuracy were included to provide a more reliable assessment of each model's performance across all sentiment classes. The comparison results are presented in Table 4.

Table 4. Comparison Analysis of Naïve Bayes and Support Vector Machine

Metric	Naïve Bayes	SVM
Accuracy	0.686598	0.836426
Precision	0.642508	0.836912

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Metric	Naïve Bayes	SVM
Recall	0.686598	0.836426
F1-Score	0.658898	0.834126
Macro F1-Score	0.476262	0.789998
Balanced Accuracy	0.494908	0.759396

Table 4 shows that the SVM model outperformed the Naïve Bayes model across all evaluation metrics. SVM achieved an accuracy of 83.64%, compared with 68.66% for Naïve Bayes. Similar improvements were observed in precision, recall, and F1-score. More importantly, SVM achieved a substantially higher Macro F1-score (79.00%) and Balanced Accuracy (75.94%), whereas Naïve Bayes obtained only 47.63% and 49.49%, respectively. The large difference in Macro F1-score indicates that SVM provided more balanced classification performance across all sentiment classes, including the minority positive class. Likewise, the higher Balanced Accuracy demonstrates that SVM classified each sentiment category more consistently, while the Naïve Bayes model was heavily influenced by the majority classes. These findings suggest that relying solely on overall accuracy could overestimate model performance when class imbalance exists. Figure 6 further illustrates the comparison between the two classification models. The results clearly show that SVM consistently achieved higher performance across all evaluation metrics, confirming that it is more suitable than Naïve Bayes for sentiment classification of plastic waste discussions on Social Media X.

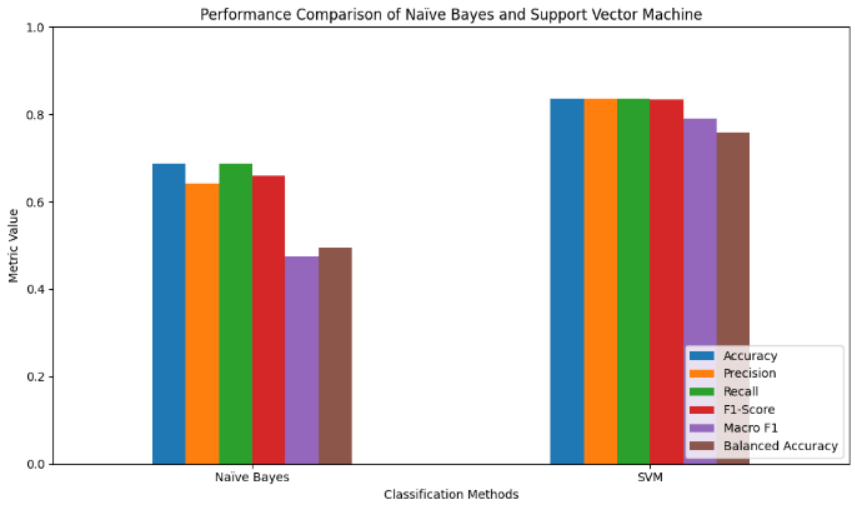

Figure 6. Performance Comparison of Naïve Bayes and Support Vector Machine

3.5. Stratified 5-Fold Cross-Validation

To further evaluate the robustness and stability of the proposed models, stratified 5-fold cross-validation was performed. Stratified sampling was applied to preserve the original class distribution in each fold. The average performance and standard deviation of each evaluation metric are presented in Table 5.

Table 5. Stratified 5-Fold Cross-Validation Results

Model	Accuracy (Mean ± SD)	Macro F1 (Mean ± SD)	Balanced Accuracy (Mean ± SD)
NB	0.6942 ± 0.0096	0.4838 ± 0.0061	0.5013 ± 0.0065
SVM	0.8341 ± 0.0145	0.7745 ±0.0183	0.7442 ± 0.0147

The cross-validation results further confirm that SVM consistently outperformed Naïve Bayes across all evaluation metrics. In addition, the relatively small standard deviation values indicate that both models produced stable performance across different folds. However, SVM consistently achieved higher performance, indicating better robustness and generalization capability than Naïve Bayes.

3.6. Discussion

The comparison results indicate that the Support Vector Machine (SVM) outperformed Naïve Bayes in classifying sentiments related to plastic waste discussions on Social Media X. This finding suggests that SVM is better suited for handling high-dimensional TF-IDF feature representations, enabling it to distinguish sentiment classes more effectively than Naïve Bayes. The superior performance of SVM was consistently observed across all evaluation metrics.

In addition to the standard evaluation metrics, this study also employed Macro F1-score and Balanced Accuracy to provide a more reliable assessment of model performance on the imbalanced dataset. The SVM model achieved a Macro F1-score of 77.45% and a Balanced Accuracy of 74.42%, whereas Naïve Bayes obtained only 48.38% and 50.13%, respectively. These results indicate that although Naïve Bayes achieved a moderate overall accuracy, its performance across all sentiment classes was considerably lower. In contrast, SVM was able to classify the three sentiment classes more consistently, making it more suitable for handling imbalanced sentiment data.

One possible explanation for the superior performance of SVM is its ability to construct an optimal decision boundary that separates sentiment classes more effectively. his characteristic makes SVM particularly suitable for high-dimensional sparse text representations such as TF-IDF vectors [19]. Text data represented by TF-IDF vectors typically contain thousands of features and sparse values. In this study, the TF-IDF process generated 5,000 features, resulting in a high-dimensional feature space. SVM is well known for its ability to handle such data and identify the most informative boundaries between classes, leading to better classification performance.

These findings are consistent with previous studies reporting that Support Vector Machine generally outperforms Naïve Bayes in text classification tasks using TF-IDF feature representations [4], [10]. This superiority is largely attributed to SVM's ability to construct an optimal decision boundary that effectively separates sentiment classes in high-dimensional and sparse text data.

In contrast, the Naïve Bayes algorithm produced lower classification performance, particularly in identifying positive sentiment. The classification report showed that the model failed to correctly predict positive sentiment instances, resulting in precision, recall, and F1-score values of 0.00 for the positive class. This limitation may be related to the independence assumption used by Naïve Bayes, which assumes that features are independent of one another. In practice, words appearing in sentiment-related texts often have complex relationships that cannot be fully captured by this assumption. Consequently, Naïve Bayes was more sensitive to class imbalance and minority-class distributions than SVM in this study.

Another factor that influenced the classification results was the imbalance in sentiment distribution. The dataset consisted of 3,462 neutral posts, 3,250 negative posts, and only 563 positive posts. The relatively small number of positive sentiment samples made it more difficult for the models, especially Naïve Bayes, to learn the characteristics of the minority class. Consequently, many positive sentiment posts were misclassified as neutral sentiment. This finding is also reflected in the low Macro F1-score and Balanced Accuracy obtained by Naïve Bayes, which reveal that high overall accuracy alone is insufficient to evaluate classification performance on imbalanced datasets.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

To further evaluate the robustness of the classification models, stratified 5-fold cross-validation was performed. The results showed that SVM consistently outperformed Naïve Bayes across different data partitions. SVM achieved a mean accuracy of 83.41% ± 1.45%, a Macro F1-score of 77.45% ± 1.83%, and a Balanced Accuracy of 74.42% ± 1.47%. Meanwhile, Naïve Bayes obtained a mean accuracy of 69.42% ± 0.96%, a Macro F1-score of 48.38% ± 0.61%, and a Balanced Accuracy of 50.13% ± 0.65%. The relatively small standard deviation values indicate that both models produced stable results across different folds, although SVM consistently achieved higher performance. These findings confirm that the superiority of SVM was not dependent on a single train-test split but remained consistent under repeated validation.

The sentiment distribution obtained in this study also provides insight into public opinion regarding plastic waste issues on Social Media X. Neutral sentiment dominated the dataset, followed by negative sentiment, while positive sentiment accounted for only a small proportion of the collected posts. This finding suggests that discussions related to plastic waste on Social Media X are generally focused on sharing information, expressing concerns, and discussing environmental problems rather than expressing positive opinions. The high proportion of negative sentiment also indicates that many users perceive plastic waste as a serious environmental issue requiring greater attention and action.

Figure 7. Word Cloud of Posts Related to Plastic Waste Issues on Social Media X

The Word Cloud visualization provides an overview of the most frequently occurring words in posts related to plastic waste issues on Social Media X. The visualization results show that words such as "sampah," "plastik," "lingkungan," and other terms related to

waste management appear with high frequency. This condition indicates that public attention is primarily focused on environmental problems and efforts to address plastic waste. This finding aligns with the results of sentiment classification, which are dominated by neutral and negative sentiments. The high frequency of informative and factual words suggests that most posts were intended to share information or discuss ongoing environmental issues. In addition, the frequent occurrence of words related to pollution and environmental impacts reflects public concern about plastic waste, which also contributes to the relatively high proportion of negative sentiment in the dataset. The Word Cloud was generated solely for descriptive visualization and was not used as an input feature in the classification models. It should be noted that the Word Cloud only illustrates the frequency of words appearing in the dataset and does not represent sentiment polarity, semantic relationships, or discussion topics. Therefore, this visualization serves only as a complementary analysis and should not be interpreted as direct evidence of public sentiment.

Despite producing satisfactory results, this study has several limitations. First, sentiment labeling was performed using a lexicon-based approach without manual validation by independent annotators, which may affect the reliability of the assigned sentiment labels. Second, the dataset exhibited class imbalance, particularly in the positive sentiment category, which may have influenced the performance of the classification models. Third, the analysis was limited to posts collected from Social Media X and therefore may not fully represent public opinions expressed on other social media platforms.

Future studies are recommended to use larger and more balanced datasets, apply data balancing techniques such as SMOTE to address class imbalance, and compare traditional machine learning algorithms with more advanced approaches such as Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT). Future research is also encouraged to incorporate manually validated sentiment labels or hybrid labeling approaches to improve label reliability. These improvements may contribute to more accurate sentiment classification and provide deeper insights into public perceptions of plastic waste issues.

4. CONCLUSION

This study analyzed public sentiment regarding plastic waste issues on Social Media X using the Naïve Bayes and Support Vector Machine (SVM) classification algorithms. A total of 7,277 posts were collected. But only 7,275 posts were used after data pre-processing. The experimental results demonstrated that SVM consistently outperformed Naïve Bayes across all evaluation metrics and exhibited more stable performance during Stratified 5-Fold Cross-Validation. Therefore, SVM can be considered the best model for this sentiment analysis. In addition, the sentiment distribution showed that neutral sentiment was the dominant class, followed by negative sentiment, whereas positive sentiment accounted for the smallest proportion of the dataset. These findings indicate that SVM is a more effective method than Naïve Bayes for sentiment classification of Indonesian plastic waste discussions on Social Media X. Positive sentiment is still very difficult to classify using Naïve Bayes modelling, and further research is needed to determine the most effective approach. This study contributes a comprehensive comparative evaluation of two widely used machine learning algorithms using lexicon-based sentiment labeling and were not manually validated, TF-IDF feature extraction, and multiple evaluation metrics. However, this study has several limitations, including the use of lexicon-based sentiment labeling without manual validation, class imbalance, one-platform data, no manual annotation, the absence of automatic bot or spam filtering, no topic modelling, and the use of data collected only from Social Media X. Therefore, the findings should be interpreted within the scope of these limitations. Nevertheless, the findings provide useful insights into public perceptions of plastic waste issues and may support researchers, policymakers, and environmental stakeholders in understanding public opinion and developing more effective environmental awareness and waste management strategies. Future research should consider incorporating manually annotated datasets, data balancing techniques such as SMOTE, transformer-based models such as BERT, topic modelling, and data collected from multiple social media platforms to improve classification performance and enhance the generalizability of the results.

REFERENCES

[1] R. Rinanda *et al.*, "Plastic Waste Management: A Bibliometric Analysis (1992–2022)," *Sustainability (Switzerland)*, vol. 15, no. 24, Dec. 2023, doi: 10.3390/su152416840.

[2] N. F. Pambudi, S. M. S. M. K. Samarakoon, T. M. Simatupang, R. M. C. Ratnayake, and N. B. Mulyono, "Risk management for the circular economy business model sustainability of reduce, reuse, and recycling in plastic waste management," *Discover Sustainability*, vol. 6, no. 1, Dec. 2025, doi: 10.1007/s43621-025-02134-4.

[3] M. R. Kelly, M. R. Cordova, S. Jobling, and R. C. Thompson, "Meta-analysis of the spatial distribution and composition of plastic macro-debris in Indonesia," *Reg. Stud. Mar. Sci.*, vol. 90, Dec. 2025, doi: 10.1016/j.rsma.2025.104460.

[4] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

[5] I. P. Rahayu, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 296, Dec. 2022, doi: 10.30865/json.v4i2.5381.

[6] N. Dwi Husna Sadikin and S. Susanti, "Analisis Sentimen Publik Terhadap Kampanye Pengurangan Sampah Plastik Menggunakan Algoritma Naïve Bayes," *JURNAL FASILKOM*, vol. 15, no. 2, pp. 202–212, Aug. 2025, doi: 10.37859/jf.v15i2.9574.

[7] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.

[8] K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr. 2023, doi: 10.3390/app13074550.

[9] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment analysis for customer review: Case study of Traveloka," *Procedia Comput. Sci.*, vol. 216, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.

[10] M. Z. Ali, Ehsan-Ul-Haq, S. Rauf, K. Javed, and S. Hussain, "Improving Hate Speech Detection of Urdu Tweets Using Sentiment Analysis," *IEEE Access*, vol. 9, pp. 84296–84305, 2021, doi: 10.1109/ACCESS.2021.3087827.

[11] M. Errami, M. A. Ouassil, R. Rachidi, B. Cherradi, S. Hamida, and A. Raihani, "Sentiment Analysis on Moroccan Dialect based on ML and Social Media Content Detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 415–425, 2023, doi: 10.14569/IJACSA.2023.0140347.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[12] E. R. Nababan, I. Made, D. Maysanjaya, and G. S. Mahendra, "Analisis Sentimen Destinasi Wisata Kabupaten Karangasem pada Ulasan Digital Menggunakan Model IndoBERT," *TEKNOMATIKA*, vol. 16, no. 01, pp. 25–37, Mar. 2026, doi: 10.61423/3gf4n981.

[13] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[14] M. Das, S. K., and P. J. A. Alphonse, "A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset," Aug. 2023, Accessed: Jul. 24, 2026. [Online]. Available: http://arxiv.org/abs/2308.04037

[15] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, 1988.

[16] N. E. Salassa, A. Patanduk, A. Yusupa, and Y. D. Y. Rindengan, "Comparative Analysis of Naive Bayes and Support Vector Machine for Sentiment Classification of Indonesian-Language Mobile Application Reviews on Google Play Store," *Jurnal Ilmiah Informatika dan Komputer*, vol. 3, no. 1, pp. 20–25, Jun. 2026, doi: 10.69533/informatech.volume3number1.515.

[17] A. Mccallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI Conference on Artificial Intelligence*, pp. 41–48, 1998, Accessed: Jul. 31, 2026. [Online]. Available: https://cdn.aaai.org/Workshops/1998/WS-98-05/WS98-05-007.pdf

[18] C. Cortes, V. Vapnik, and L. Saitta, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[19] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.