A Sentiment Analysis Study of Banning Single-Use Plastic Bags Based on X Users' Attitude

Zulwelly Murti
Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN)

Soen Steven
Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN)

Riana Y. H. Sinaga
Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN)

Mulyono Mulyono
Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN)

他

https://doi.org/10.5109/7363473

出版情報：Evergreen. 12 (2), pp.755-766, 2025-06. 九州大学グリーンテクノロジー研究教育センター
バージョン：

A Sentiment Analysis Study of Banning Single-Use Plastic Bags Based on X Users' Attitude

Zulwelly Murti[1,*], Soen Steven[1,*], Riana Y. H. Sinaga[1], Mulyono Mulyono[1], Arief A. R. Setiawan[1,2], Geby Otivriyanti[1], Maya L. D. Wardani[1], Nurus S. Laili[1], Anita Yustisia[1], Asep Bustanil Aripin[1,3], Ernie S. A. Soekotjo[1], Muhammad Sudiono[1], Vionita Lukitari[1], Dharmawan Dharmawan[1], Adik A. Soedarsono[4]

[1]Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN), KST BJ Habibie, Building 720 Puspiptek Area, South Tangerang, Banten 15314, Indonesia
[2]Department of Industrial Engineering, University of Nahdatul Ulama Al-Ghazali, Cilacap 53274, Indonesia
[3]Department of Physics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung 40132, Indonesia
[4]Research Center For Process and Manufacturing Industry Technology, National Research and Innovation Agency (BRIN), KST BJ Habibie, Building 720 Puspiptek Area, South Tangerang, Banten 15314, Indonesia

*Author to whom correspondence should be addressed:
Email: zulw001@brin.go.id (ZM); soensteven201194@gmail.com (SS)

(Received August 07, 2024; Revised May 02, 2025; Accepted June 20, 2025)

Abstract: Single-use plastics have become an urgent global environmental concern. The Indonesian government has attempted to introduce a policy that bans the use of single-use plastics as an initial step toward overcoming this issue. However, policy implementation often faces challenges from stakeholders. Therefore, this study aims to analyze public sentiment expressed on the X platform (formerly known as "Twitter") regarding the policy of banning single-use plastic bags in Indonesia. The methods used in this study include data collection, pre-processing (cleaning and transforming), labeling, modeling, and analysis using RapidMiner software. Tweet data were then analyzed using three machine learning methods, i.e., Naïve Bayes, K-Nearest Neighbors (KNN), and Decision Tree. Data were divided into training and test sets with a ratio of 60:40, 70:30, and 80:20. As many as 1038 refined datasets from 2019-2023 with related keywords were obtained. Based on the performance evaluation, the Naïve Bayes algorithm can improve its performance as the amount of training data increases, without overfitting. This algorithm achieves the highest accuracy of 89.73% at an 80:20 ratio. Furthermore, the classification results of the majority (70.6%) of the tweets showed positive support for the policy, 19.6% were negative, and 9.8% were neutral. In other words, the results of this sentiment classification can be used to monitor public responses and formulate environmentally friendly policies that are effective and supported by the majority.

Keywords: algorithm; machine learning; Naïve Bayes; sentiment; single-use plastic

1. Introduction

Single-use plastics are among the biggest environmental challenges facing the world [1–3]. Items such as plastic bags, bottles, straws, food packaging, and disposable cutlery have become an integral part of everyday life (Figure 1). Wastes from single-use plastics are often disposed of as solid waste, adding to the environmental burden and harming human health [4–7]. They not only generate waste that is difficult to decompose but also cause air pollution, global warming, ozone depletion, and climate change since their production still relies on massive fossil resources [8–12]. Many countries worldwide are banning the use of single-use plastics [13–16]. Even in Indonesia, the problem of single-

Cite: Z. Murti et al., "A Sentiment Analysis Study of Banning Single-Use Plastic Bags Based on X Users· Attitude". Evergreen, 12 (02) 755-766 (2025). https://doi.org/10.5109/7363473.

use plastics has become a serious concern. The government introduced a policy of banning the use of single-use plastics as an initial step toward overcoming this environmental crisis [17,18]. Nevertheless, this action has certainly sparked various sentiments. Various parties often face challenges, so implementing this type of policy is not always easy [19].

Fig. 1: Examples of single-use plastics in daily life.

Analyzing public sentiment towards single-use plastic bans is useful in understanding public response and support. Sentiment data can be obtained from the X platform, formerly known as Twitter. It is one of the most active and diverse social media platforms and provides a rich source of data to analyze public views, feelings, and opinions regarding these issues [20]. The main goal of sentiment analysis is to identify an author's attitude toward a particular topic, whether positive, negative, or neutral [21]. Sentiment analysis is the process of extracting subjective information from text. This is an important area in machine learning and natural language processing [22,23]. Machine learning is a branch of artificial intelligence that allows systems to automatically learn and improve from experience without explicit programming [24]. It focuses on developing algorithms and techniques that enable computers to learn from data, recognize patterns, and make predictions [25,26].

On that basis, this study intends to analyze public sentiment toward this policy in Indonesia using data from the X platform. The objective of this study is also to evaluate the performance of the aforementioned machine learning algorithms in classifying these sentiments. The novelty lies in the integration of social media-based sentiment analysis with machine learning techniques. This not only offers a deeper understanding of public responses but also bridges the use of digital technologies in environmental policy analysis in Indonesia.

The outline of this study begins with the background and research objectives of sentiment analysis concerning the prohibition of single-use plastic bags. The next section outlines the methodological framework. The following section explains the results of sentiment analysis and discusses the research findings based on the performance of each algorithm. The final section reveals the conclusions and further recommendations.

2. Methodology

This section details the procedures and tools for analyzing public sentiment toward the single-use plastic ban policy in Indonesia. In general, the methodological framework in this study consisted of data collection, pre-processing (cleaning and transforming), labeling, modeling, and analysis, as shown in Figure 2. Data collection employed Tweet Harvest v2.6.1 software, whereas the other subsequent processes utilized RapidMiner software due to its powerful data analysis and machine learning modeling tools [27,28].

2.1. Sample size determination

Two approaches were used to preliminarily determine the sample, depending on the type of population. The chosen population was the total Indonesian population in 2024 (larger population) and the number of Indonesian people using the X platform until 2024 (smaller population). The confidence levels were tested at 90%, 95%, and 99% because they are commonly used in engineering studies.

Cochran's method, as written in Eq. (1) and Eq. (2), was utilized for large populations, whereas Yamane's method, Eq. (3), was utilized for small populations [29,30].

$$n_0 = \frac{z^2 p(1-p)}{\alpha^2} \tag{1}$$

$$n = \frac{n_0}{1+(n_0-1)N} \tag{2}$$

$$n = \frac{N}{1+N(\alpha)^2} \tag{3}$$

where n_0 is the first estimation of sample size, z is the axis value of two-tailed z distribution depending on the confidence level, p is the predicted proportion sample in the population, α is the significance level or margin of error, n is the sample size, and N is the population size.

2.2. Data collection

The X platform was used as the data source to understand people's views on the ban on single-use plastics. Tweet data were limited to X users located in Indonesia or using the Indonesian language. This coverage provided a broad perspective on public sentiment, but might not fully represent the views of individuals who did not use X as their social media platform. Tweets relevant to this topic were collected using the Indonesian keyword "kantong plastik sekali pakai" (meaning "single-use plastic bags") in the period from 2019 to 2023. The code involved in data collection is listed in Table 1.

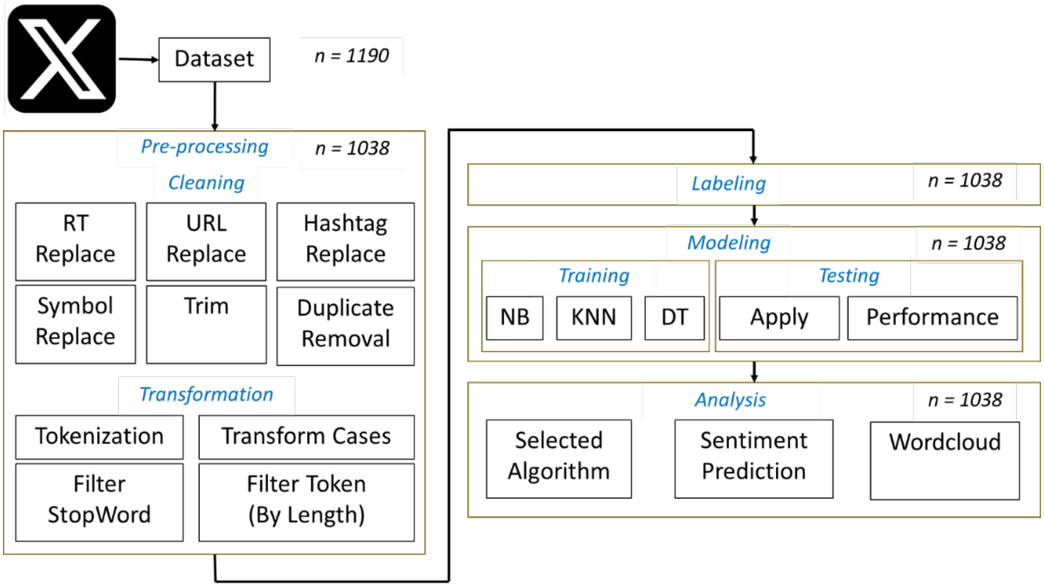

Fig. 2: Methodological framework

Table 1: Code for data collection

```
#@title X Auth Token
X_auth_token = '5f860e266c147a014f12b3117e00068d6b6946db'
# Import required Python package
!pip install pandas
# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-repo.gpg.key | sudo gpg --
dearmor -o /etc/apt/keyrings/nodesource.gpg
!NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo tee
/etc/apt/sources.list.d/nodesource.list
!sudo apt-get update
!sudo apt-get install nodejs -y
# Crawl Data
filename = 'Plastik.csv'
search_keyword = 'Kantong Plastik Sekali Pakai until:2019-01-01 since:2023-12-31
lang:id'
limit = 1500
!npx --yes tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" -l {limit} --
token {X_auth_token}
import pandas as pd
# Specify the path to your CSV file
file_path = f"tweets-data/{filename}"
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(file_path, delimiter=",")
# Display the DataFrame
display(df)
# Check the length of the DataFrame
num_rows = len(df)
print(f"The number of rows in the DataFrame is: {num_rows}")
```

2.3. Data pre-processing

The collected data underwent a series of pre-processing stages, i.e., data cleaning and transformation.

2.3.1. Data cleaning

Initially, data cleaning was employed to remove special characters, links, and irrelevant elements. Afterward, case folding was conducted to convert all the letters in the text to lowercase letters for a more consistent analysis. Tokenization was also performed to divide the text into individual words and to apply the filtering process to remove words that did not make a significant contribution to sentiment analysis. The RapidMiner operators used for data cleaning are given in Figure 3.

2.3.2. Data transformation

The process document from the data operator generates word vectors from string attributes using the term frequency-inverse document frequency $(TF - IDF)$, which indicates the importance of a word in a particular document [31]. TF, calculated based on Eq. (4), refers to how often a word appears in a document compared to the total number of words used in the document. This parameter considers all words in a document to be equally important [32].

On the other side, IDF enhances the number of unique words present in the document and decreases the value of commonly used terms. It is used to measure the importance of how many times a term or word appears [33]. The calculation follows Eq. (5).

$$TF = T/D \tag{4}$$
$$IDF(T) = N/D \tag{5}$$

where T states how many times a term or word appears, D is the number of words in the document, and N is the total number of words.

Zero cannot be in the denominator, hence, adding 1 to it avoids division by zero, resulting in Eq. (6). The final formula in Eq. (7) was applied to measure the importance of a word in a collection of documents.

$$IDF = \frac{log\, N}{DF+1} \tag{6}$$
$$TF - IDF = TF \times \frac{log\, N}{DF+1} \tag{7}$$

where DF is the number of occurrences of term T in the document.

The document processing operator from the data performs several processes to prepare the dataset for use in developing models for sentiment analysis. The processes applied to the dataset were tokenize, transform cases, filter stopwords, and filter tokens by length. The tokenization divides tweets into sequences of tokens and terms. It also removes punctuation and white spaces from the tweets [34].

The transformation case converts all uppercase letters to lowercase letters and vice versa. The researcher chose to change to lowercase in the parameters section [35].

The connecting words in tweets were removed using the Indonesian StopWords Filter Operator. Words such as is, then, so, then, although, only, a, but, that, although, namely, that is, which, and others, were also eliminated [31]. Filter tokens by length explore all tokenized terms and filter out words that are shorter or longer than a specified number of characters. Researchers used a minimum of 4 and a maximum of 25 characters per word. The RapidMiner operators used for data transformation are schemed in Figure 4.

2.4. Data labeling

In order to train a machine learning algorithm, tweet data should be manually labeled with sentiments, as outlined in Table 2. This means that each tweet is labeled as positive, negative, or neutral according to the expression of the sentiment contained within it. Consequently, it is important to develop models that can automatically classify sentiments [36].

2.5. Data Modeling: Training and testing

Three machine learning algorithms were applied, i.e., Naïve Bayes, K-Nearest Neighbors (KNN), and Decision Tree. The Naïve Bayes algorithm has been widely used for sentiment classification due to its computational efficiency and reliable performance on short-text data, including tweets [37]. KNN and Decision Tree are also favored in practical sentiment analysis studies because of their ease of implementation and interpretability, particularly when computational resources are limited [38,39]. In summary, these algorithms have merits in terms of transparency, interpretability, and suitability to the nature and size of the data set, especially in the context of resource-efficient processing using RapidMiner [40].

Besides, the k-fold cross-validation was not used because this study focused on evaluating model performance using holdout validation at various data split ratios. Although k-fold cross-validation offers more robust performance estimates, the holdout validation approach is more suitable for comparative analysis of different training-test proportions. Furthermore, holdout validation is superior in providing prompt estimation with minimal computational cost [41].

The dataset was divided into training and testing data with different ratios of 60:40, 70:30, and 80:20. The training data were used to train the algorithm model, whereas test data were used to test the model's performance. This data sharing aided in evaluating the model's ability to predict sentiments with varying accuracy. The RapidMiner operators used in the modeling process are revealed in Figure 5.

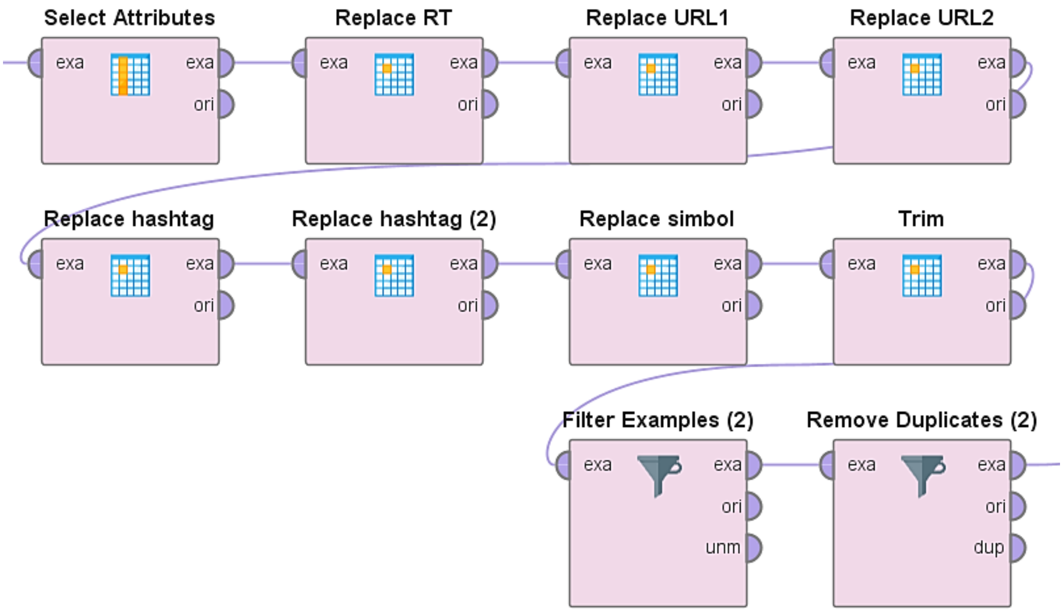

Fig. 3: RapidMiner’s operators used for data cleaning

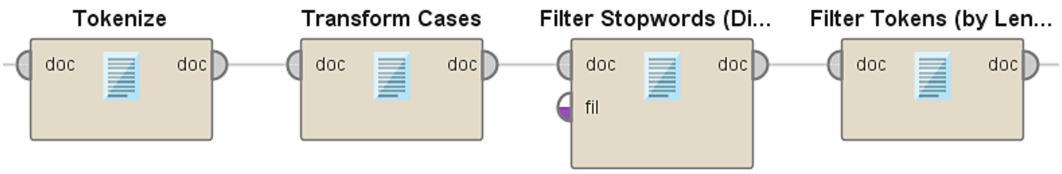

Fig. 4: RapidMiner’s operators used for data transformation

Table 2: Data labeling

Text	Sentiment
Come on, everyone, stop using single-use plastic. If you go to minimarkets or stalls, bring your own shopping bag, and bring a Tumblr instead of buying bottled water.	Positive
It’s true, I only use it once at home, I even use it to store all sorts of things, the small ones are usually for storing chili sauce or fried onions, and the big ones are usually for storing vegetables or even for storing plastic bags like this.	Negative
When they arrived in front of the main gate, several temple volunteer officers asked residents who were dressed in traditional clothing to hand over plastic bags and disposable plastic bags for the offerings they brought.	Neutral

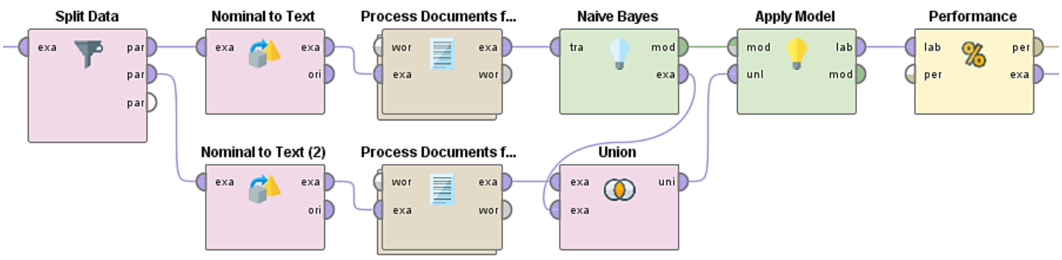

Fig. 5: RapidMiner’s operators used for data modeling

3. Results and discussion

3.1. Results of sample size adequacy test

Figure 6 demonstrates the trends that reflect fluctuations in online conversations regarding environmental regulation. From a total of 1190 tweets, this cleaning process resulted in 1038 tweets. The first term of 2019 recorded the highest peak with 390 tweets, which then decreased to 237 tweets in the first term of 2020. The highest peak occurred in the

second term of 2020, with 466 tweets, reflecting a significant response to this issue in Indonesian society.
However, conversations drastically decreased in the first and second halves of 2021, with 74 and 18 tweets, respectively. Although there was a positive alteration in the first and second terms of 2022, with 95 and 100 tweets in a successive term, it is still below the peak in 2020. The first and second terms of 2023 recorded 64 and 95 tweets, successively, indicating that the ban on single-use plastics in Indonesia remains problematic. Attention, even though the intensity of the conversation did not reach peak levels.
The Indonesian population in 2024 is recorded at 279,298,049 [42], while the number of Indonesian people who used X as their active social media is 24.69 million [43].
The sample size determination for both large and small populations using Cochran's method and Yamane's method is 68-100 for a confidence interval of 90%, 385-400 for a confidence interval of 95%, and 10,000-13,507 for a confidence interval of 99%, as served in Table 3.
The implication is that a sample size of 1038 tweets can represent the entire X users in Indonesia. In other words, it still complies with a meaningful statistical basis at a confidence level of 96.5% (margin of error of 3.5%). In the context of social media, where active users are more likely to engage in public policy discussions, this sample offers valuable insights into the opinions of this influential group [44]. This approach also increases the possibility of capturing public opinion across a broad spectrum [45].

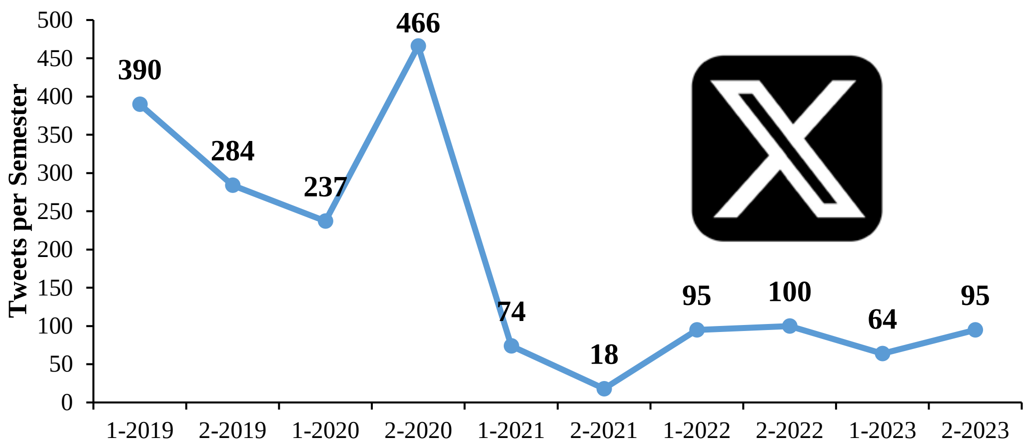

Fig. 6: Frequency of tweets from 2019-2023

Table 3: Results of sample size determination

Parameter	Confidence Level 90%	Confidence Level 95%	Confidence Level 99%
z, two-tailed	-1.64	-1.96	-2.33
p	0.5	0.5	0.5
α	0.10	0.05	0.01
n_0	68	385	13,515
N_1, Indonesian Population (Larger Population)	279,798,049	279,798,049	279,798,049
N_2, Indonesian X Platform Users (Smaller Population)	24,690,000	24,690,000	24,690,000
n_1, Cochran's Method	68	385	13,514
n_2, Cochran's Method	68	385	13,507
n_1, Yamane's Method	100	400	10,000
n_2, Yamane's Method	100	400	10,000

3.2. Performance of Naïve Bayes, KNN, and Decision Tree algorithms

Naïve Bayes is a machine learning algorithm based on Bayes' theorem. This algorithm assumes independence between the predictor variables. This simplicity and independence assumption make Naïve Bayes fast and efficient in classifying data on a large scale. Naïve Bayes is suitable for text classification and simple pattern recognition [46].
Sentiment classification is carried out by calculating the posterior probabilities $P(C|X)$, where C is the sentiment class (positive, negative, or neutral), and X represents the set of features (words or phrases). The independence assumption among features significantly reduces the computational complexity and enhances the efficiency. In

this study, the multinomial Naïve Bayes variant was employed, as it is well suited for discrete features such as word frequencies.

KNN is a machine learning algorithm that classifies objects based on the nearest-neighbor data in a vector space. It does not use any model to match the new data with the training set. In contrast, KNN is based only on the distance from the test data to the training data and is suitable for nonlinear cases and when the data distribution is unknown [47].

Other than that, Decision Tree is a machine learning algorithm that models problem solutions into a tree structure and divides data into nodes based on their attributes. It is easy to interpret and does not require a complex data preparation process. This algorithm is frequently used for data classification and prediction [48].

Based on the machine learning algorithm performance data provided, Naïve Bayes showed the most optimal results among the three algorithms, as summarized in Table 4. Naïve Bayes was able to consistently increase its prediction accuracy from 85.63% at a train-to-test data ratio of 60:40 to 87.99% at a ratio of 70:30 and reached the highest accuracy of 89.73% when the training-data ratio was altered to 80:20.

Table 4: Model performance under several holdout variations

Algorithm Models	60:40 Ratio	70:30 Ratio	80:20 Ratio
Naïve Bayes	85.63%	87.99%	89.73%
KNN	68.38%	79.16%	78.95%
Decision Tree	76.28%	78.95%	79.26%

The increasing amount of training data by Naïve Bayes indicates that the algorithm does not experience overfitting. Naïve Bayes' ability to model the probability of correlation between predictor variables and target variables also makes it suitable for data classification [49]. Meanwhile, the performance of KNN becomes poor when the training data ratio escalates from 70:30 to 80:20, where the accuracy alleviates from 79.16% to 78.95%. This implies that KNN starts to experience overfitting because it adjusts too much to noise and outliers in the increasingly large training data. KNN's high dependency on numerical data also influences its performance less than optimal for classification [47].

For sentiment classification, a new tweet is determined based on the majority class among the 'k' closest labeled tweets, using distance metrics such as the Euclidean distance in vector space. This study used $k = 5$ as the default, and $TF - IDF$ features to calculate distances. Unlike Naïve Bayes, KNN does not involve an explicit training phase, making it computationally expensive during inference and prone to overfitting when dealing with noisy social media data.

The decision tree algorithm is in the middle position between Naïve Bayes and KNN in terms of prediction accuracy. This algorithm constructs a tree structure in which nodes represent features, branches represent decision rules, and leaves represent sentiment classes. The tree recursively splits the dataset based on the feature values that yield the highest information gain. In RapidMiner, the Gini index is used as the splitting criterion [21]. Decision Trees are interpretable and provide a visual understanding of how certain keywords influence sentiment classification, but they are prone to overfitting unless pruned properly [26,50]. Although the decision tree accuracy continues to signify as the training data increases, its value is still below that of Naïve Bayes for all ratio comparisons. The limited complexity of the decision tree model is a factor in its performance, which does not surpass Naïve Bayes.

3.3. The Naïve Bayes as the best algorithm

The Naïve Bayes algorithm is a classification method based on Bayes' theorem with the assumption of "purity" or "simplicity" (naïve) of the relationship between the features used in classification [51]. This algorithm is often used in text analysis, pattern recognition, and other data classifications. There are several steps in the Naïve Bayes algorithm calculation process and formula, i.e., training data, class probabilities, feature probabilities, posterior probability, and classification [49,50].

The training data were used to train the Naïve Bayes model. This data must contain samples that have been labeled (for example, data that has been labeled as "positive" or "negative" or "neutral"). The probability of each class appearing in the training data is calculated. This was performed by counting the number of times each class appeared in the training data and dividing it by the total amount of training data.

$$P(C) = \frac{Number\ of\ samples\ with\ a\ certain\ class}{Total\ number\ of\ samples\ in\ the\ training\ data}$$

The probability of occurrence of each feature in each class was calculated. This involves counting the number of times a particular feature appeared in each class.

$$P(Xi|C) = \frac{Number\ of\ samples\ with\ a\ particular\ feature\ in\ a\ particular\ class}{Number\ of\ samples\ with\ a\ certain\ class}$$

After calculating the class and feature probabilities within each class, the posterior probability for each class was calculated using the input data (features).

$$P(C|X) = \frac{P(C) \cdot P(X1|C) \cdot P(X2|C) \cdot \ldots \cdot P(Xn|C)}{P(X)}$$

where X is the input feature vector (e.g. a classified text document).

The class with the highest posterior probability is selected as the class generated by the Naïve Bayes model. The Naïve Bayes algorithm is simple and efficient but has a fairly strong assumption that all features in the input data are independent of each other, which is often not met in the real world [37]. Nevertheless, this algorithm often provides good results in many cases, particularly in text classification, such as sentiment analysis.

Based on the three confusion matrix tables provided in Table 5, the overall performance of Naïve Bayes improves with an increasing proportion of training data. This is reflected in the overall accuracy, which continued to intensify from 85.63% (60:40 ratio) to 89.73% (80:20 ratio). Specifically, the increase in Naïve Bayes performance is also marked by improvements in the precision and recall values for almost all classes. For example, the precision and recall of the positive class continue to enhance as the amount of training data increases. The positive precision changed from 90.06% to 92.75% and 95.98%, while the positive recall signified from 90.98% to 91.27% and 90.39%, respectively.

Table 5: Confusion matrix for the Naïve Bayes algorithm

	Negative	**Positive**	**Neutral**	**Class Recall**
60:40 Ratio				
Actual Negative	625	36	26	90.98%
Actual Positive	42	143	6	74.87%
Actual Neutral	27	3	66	68.75%
Class Precision	90.06%	78.57%	67.35%	
70:30 Ratio				
Actual Negative	627	36	24	91.27%
Actual Positive	29	156	6	81.68%
Actual Neutral	20	2	74	77.08%
Class Precision	92.75%	80.41%	71.15%	
80:20 Ratio				
Actual Negative	168	17	6	87.96%
Actual Positive	36	621	30	90.39%
Actual Neutral	2	9	85	88.54%
Class Precision	81.55%	95.98%	70.25%	

The same phenomenon also occurs in the negative and neutral classes, although it does not always consistently increase. For example, the negative recall value decreased slightly from 81.68% (70:30) to 87.96% (80:20), although it was still better than 74.87% (60:40). This trend reflects that the Naïve Bayes model can utilize increasing training data to improve its prediction ability without overfitting. However, the performance in minority classes, such as neutral, is still not optimal, and there is still room for further improvement.

Most X users appreciated the policy of banning single-use plastics, as seen by 70.6% of their positive opinions. This indicates a high level of environmental awareness. Nevertheless, the government needs to pay attention to 19.6% of negative opinions. This can be affected by industrial and business players. Socialization and discussion are needed to reduce this negative opinion. In addition, 9.8% of the neutral opinions still needed to be converted to support the policy. Massive education and information are important to increase community support.

The frequencies of the words and their associated classifications are tabulated in Table 6. In the meantime, Figure 7 illustrates the visual representation of word clouds for positive, neutral, and negative classifications. The word cloud visually represents the words employed within the dataset. Words that are used more frequently are displayed in a larger size within the word cloud.

Overall, it can be concluded that Naïve Bayes shows a positive performance and responds well to increasing the amount of training data. This makes Naïve Bayes worth considering for application to similar classification cases in the future. High community support is an important asset for implementing a single-use plastic ban policy in the future [52]. However, efforts to increase support from the government should be continued through outreach and education to change negative and neutral opinions.

Table 6: Word count per sentiment

Positive		Negative		Neutral	
Word	Count	Word	Count	Word	Count
plastik (plastic)	251	plastik (plastic)	44	plastik (plastic)	24
kantong (bag)	182	kantong (bag)	38	kantong (bag)	18
pakai (use)	169	pakai (use)	33	pakai (use)	18
penggunaan (utilize)	64	sampah (trash)	9	penggunaan (utilize)	8
sampah (trash)	52	belanja (shop)	6	bebas (free)	4
belanja (shop)	36	kertas (paper)	6	belanja (shop)	3
mengurangi (reduce)	35	bahan (material)	5	Indonesia	3
lingkungan (environment)	29	galon (gallon)	5	Jakarta	3
pasar (market)	23	penggunaan (utilize)	5	larang (ban)	3
bawa (carry)	20	kain (fabric)	4	larangan (prohibition)	3
galon (gallon)	19	ulang (recycle)	4	ramah (friendly)	3
botol (bottle)	17	botol (bottle)	3	daerah (region)	2
kemasan (packaging)	15	diciptakan (created)	3	lingkungan (environment)	2
masyarakat (community)	14	dipakai (used)	3	LLHPB*	2
bebas (free)	13	dipake (used)	3	tingkat (level)	2

* Lembaga Lingkungan Hidup dan Penanggulangan Bencana (Environment and Disaster Management Unit)

(a) (b) (c)

Fig. 7: Wordcloud: Positive (a), Negative (b), Neutral (c)

4. Conclusions

This study analyzes public sentiment on the X Platform regarding the policy of single-use plastic ban in Indonesia by utilizing machine learning algorithms. The Naïve Bayes algorithm is proven to be the most effective, with the highest accuracy of 89.73% at 80:20 ratio. This algorithm can improve its performance as the amount of training data increases without overfitting. The results of the analysis show that the majority of the public has a positive attitude (70.6%), supporting this policy. Only 19.6% have a negative attitude and 9.8% have a neutral attitude.

The results of this study can be used as an effort to gain public support for environmental policies. This allows for a faster and more accurate understanding of the dynamics of public opinion, which is valuable for policymakers in formulating and implementing effective and supportive plastic waste management strategies. There needs to be socialization and discussion to reduce negative opinions from several groups. In addition, neutral opinions must be changed to support policies through massive education and information. Thus, the government can formulate effective and environmentally friendly policies that are supported by the majority of the community.

Future analysis can be realized by combining data from various social media platforms, conducting offline surveys or direct interviews, and creating additional machine learning algorithms to better understand the Indonesian language and cultural context. Several options, such as Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN), can be further incorporated to validate the current findings due to higher predictive power. The k-fold cross-validation method will also be utilized to improve model robustness.

Acknowledgments

This study has no financial support. Soen Steven acknowledges the Postdoctoral Scheme at Research Center for Sustainable Production System and Life Cycle Assessment, National Research and Innovation Agency (BRIN), Indonesia 2024-2025. Also, all authors have read and agreed to this article with contributions as follows: **ZW**: Conceptualization, Writing—original draft, Writing—review and editing, Methodology, Data

Collection, Simulation, Formal analysis, Validation, Visualization. **SS**: Writing—review and editing, Formal analysis, Validation, Visualization, Critical revising. **RYHS, MM, GO, MLDW,**: Formal analysis, Writing—review and editing, Validation. **NSL, ABA, MS, VL, AY, DD, AAS**: Formal analysis, Validation **AARS, ESAS**: Conceptualization, Formal analysis, Writing—review and editing.

References

1) A.Dey, C.V. Dhumal, P. Sengupta, A. Kumar, N.K. Pramanik, and T. Alam, "Challenges and possible solutions to mitigate the problems of single-use plastics used for packaging food items: a review," J. Food Sci. Technol., 58 (9) 3251–3269 (2021). doi:10.1007/s13197-020-04885-6.
2) C.Liu, T. Thang Nguyen, and Y. Ishimura, "Current situation and key challenges on the use of single-use plastic in hanoi," Waste Manag., 121 422–431 (2021). doi:10.1016/j.wasman.2020.12.033.
3) L.R.Lima, R.F. Gutierrez, and S.A. Cruz, "Challenges in the context of single-use plastics and bioplastics in brazil: a legislative review," Waste Manag. Res. J. a Sustain. Circ. Econ., 40 (7) 998–1006 (2022). doi:10.1177/0734242X211055548.
4) Syafrudin,Mochamad Arief Budihardjo, Nany Yuliastuti, and Bimastyaji Surya Ramadan, "Assessment of greenhouse gases emission from integrated solid waste management in semarang city, central java, indonesia," Evergreen, 8 (1) 23–35 (2021). doi:10.5109/4372257.
5) I.Febijanto, S. Steven, N. Nadirah, H. Bahua, A. Shoiful, D.P. Dewanti, I.P.A. Kristyawan, K.A. Haris, M. Yuliani, M. Hanif, M.H. Robbani, N.R. Yusuf, Prihartanto, P. Alfatri, R.A. Pratama, W. Purwanta, Wiharja, R. Nugroho, and S.K. Ramadhan, "Municipal solid waste (msw) reduction through incineration for electricity purposes and its environmental performance: a case study in bantargebang, west java, indonesia," Evergreen, 11 (1) 32–45 (2024). doi:10.5109/7172186.
6) A.Zhumadilova, and S. Zhigitova, "Features of modern areas of solid waste disposal," Evergreen, 10 (2) 640–648 (2023). doi:10.5109/6792809.
7) S.Steven, N.T.U. Culsum, I.C. Sophiana, I. Febijanto, E. Syamsudin, N. Ghazali, N. Nadirah, E.S.A. Soekotjo, and I.M. Hidayatullah, "Potential of corn cob sustainable valorization to fuel-grade bioethanol: a simulation study using superpro designer®," Evergreen, 10 (4) 2287–2298 (2023). doi:10.5109/7160904.
8) L.Di Paolo, S. Abbate, E. Celani, D. Di Battista, and G. Candeloro, "Carbon footprint of single-use plastic items and their substitution," Sustainability, 14 (24) 16563 (2022). doi:10.3390/su142416563.
9) M.Shen, W. Huang, M. Chen, B. Song, G. Zeng, and Y. Zhang, "(Micro)plastic crisis: un-ignorable contribution to global greenhouse gas emissions and climate change," J. Clean. Prod., 254 120138 (2020). doi:10.1016/j.jclepro.2020.120138.
10) A.Berisha, and L. Osmanaj, "Kosovo scenario for mitigation of greenhouse gas emissions from municipal waste management," Evergreen, 8 (3) 509–516 (2021). doi:10.5109/4491636.
11) D.A.Trirahayu, A.Z. Abidin, R.P. Putra, A.S. Hidayat, E. Safitri, and M.I. Perdana, "Process simulation and design considerations for biodiesel production from rubber seed oil," Fuels, 3 (4) 563–579 (2022). doi:10.3390/fuels3040034.
12) D.A.Trirahayu, A.Z. Abidin, R.P. Putra, F.D. Putri, A.S. Hidayat, and M.I. Perdana, "Process assessment of integrated hydrogen production from by-products of cottonseed oil-based biodiesel as a circular economy approach," Hydrogen, 4 (2) 272–286 (2023). doi:10.3390/hydrogen4020019.
13) E.Cowan, A.M. Booth, A. Misund, K. Klun, A. Rotter, and R. Tiller, "Single-use plastic bans: exploring stakeholder perspectives on best practices for reducing plastic pollution," Environments, 8 (8) 81 (2021). doi:10.3390/environments8080081.
14) Shabbiruddin,N. Kanwar, V.K. Jadoun, J. N. S., A. Afthanorhan, N. Fatema, H. Malik, and M.A. Hossaini, "Industry—challenge to pro-environmental manufacturing of goods replacing single-use plastic by indian industry: a study toward failing ban on single-use plastic access," IEEE Access, 11 77336–77346 (2023). doi:10.1109/ACCESS.2023.3296097.
15) S.Ali, W. Ahmed, Y.A. Solangi, I.S. Chaudhry, and N. Zarei, "Strategic analysis of single-use plastic ban policy for environmental sustainability: the case of pakistan," Clean Technol. Environ. Policy, 24 (3) 843–849 (2022). doi:10.1007/s10098-020-02011-w.
16) M.Pearson, E. Khare, and O. Tantawi, "Addressing the single-use plastic proliferation problem," MIT Sci. Policy Rev., 3 85–93 (2022). doi:10.38105/spr.ro3uhoc2lz.
17) H.Kamaruddin, Maskun, F. Patittingi, H. Assidiq, S.N. Bachril, and N.H. Al Mukarramah, "Legal aspect of plastic waste management in indonesia and malaysia: addressing marine plastic debris," Sustainability, 14 (12) 6985 (2022). doi:10.3390/su14126985.
18) Y.Wang, and R. Karasik, "Plastic Pollution Policy Country Profile: Indonesia," Duke University, 2022.
19) A.R.Bizhanova, A.S. Koshkinbayeva, G.A. Zhunisova, G.Z. Osmanova, D. Belkhozhayeva, and D.S. Baisymakova, "Regulatory issues of depollution in kazakhstan," Evergreen, 9 (4) 903–908 (2022).

doi:10.5109/6622877.
20) C.A.G.Gutierrez, A. Whittaker, K.M. Patenio, J. Gehman, L.M. Lefsrud, D. Barbosa, and E. Stroulia, “Analyzing and visualizing Twitter conversations,” in: Proc. 31st Annu. Int. Conf. Comput. Sci. Softw. Eng., IBM Corp., USA, 2021: pp. 4–13.
21) M.Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: approaches, challenges and trends,” Knowledge-Based Syst., 226 107134 (2021). doi:10.1016/j.knosys.2021.107134.
22) A.Mitra, “Sentiment analysis using machine learning approaches (lexicon based on movie review dataset),” J. Ubiquitous Comput. Commun. Technol., 2 (3) 145–152 (2020). doi:10.36548/jucct.2020.3.004.
23) P.Monika, C. Kulkarni, N. Harish Kumar, S. Shruthi, and V. Vani, “Machine learning approaches for sentiment analysis,” Int. J. Health Sci. (Qassim)., 1286–1300 (2022). doi:10.53730/ijhs.v6nS4.6119.
24) J.M.Helm, A.M. Swiergosz, H.S. Haeberle, J.M. Karnuta, J.L. Schaffer, V.E. Krebs, A.I. Spitzer, and P.N. Ramkumar, “Machine learning and artificial intelligence: definitions, applications, and future directions,” Curr. Rev. Musculoskelet. Med., 13 (1) 69–76 (2020). doi:10.1007/s12178-020-09600-8.
25) D.Hsu, “Overview of Machine Learning,” 2022.
26) S.Uddin, and H. Lu, “Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data,” PLoS ONE, 19 (4) e0301541 (2024). doi:10.1371/journal.pone.0301541.
27) R.Nayak, S. Dinesh, and S.Thirunavukkarasu, “A Novel Method Improvement of Rapid Miner for the Data Mining Applications,” 2019.
28) M.Bjaoui, H. Sakly, M. Said, N. Kraiem, and M.S. Bouhlel, “Depth insight for data scientist with RapidMiner « an innovative tool for AI and big data towards medical applications»,” in: Proc. 2nd Int. Conf. Digit. Tools Uses Congr., ACM, New York, NY, USA, 2020: pp. 1–6. doi:10.1145/3423603.3424059.
29) G.D.Israel, “Determining Sample Size,” University of Florida, 2003.
30) W.G.Cochran, “The Estimation of Sample Size,” in: Sampl. Tech. Third Ed., John Wiley & Sons, Inc., New York, NY, USA, 1977: pp. 86–102.
31) H.T.Y.Achsan, H. Suhartanto, W.C. Wibowo, D.A. Dewi, and K. Ismed, “Automatic extraction of indonesian stopwords,” Int. J. Adv. Comput. Sci. Appl., 14 (2) 166–171 (2023). doi:10.14569/IJACSA.2023.0140221.
32) A.Jalilifard, V.F. Caridá, A.F. Mansano, R.S. Cristo, and F.P.C. da Fonseca, “Advances in Computing and Network Communications,” Springer Singapore, Singapore, 2021. doi:10.1007/978-981-33-6977-1.
33) Z.Zhang, Z. Wu, and Z. Shi, “An improved algorithm of TFIDF combined with Naive Bayes,” in: 7th Int. Conf. Multimed. Image Process., ACM, New York, NY, USA, 2022: pp. 167–171. doi:10.1145/3517077.3517104.
34) U.Naseem, I. Razzak, and P.W. Eklund, “A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter,” Multimed. Tools Appl., 80 (28–29) 35239–35266 (2021). doi:10.1007/s11042-020-10082-6.
35) C.Villavicencio, J.J. Macrohon, X.A. Inbaraj, J.-H. Jeng, and J.-G. Hsieh, “Twitter sentiment analysis towards covid-19 vaccines in the philippines using naïve bayes,” Information, 12 (5) 204 (2021). doi:10.3390/info12050204.
36) S.Sazzed, and S. Jayarathna, “SSentiA: a self-supervised sentiment analyzer for classification from unlabeled data,” Mach. Learn. with Appl., 4 100026 (2021). doi:10.1016/j.mlwa.2021.100026.
37) I.Wickramasinghe, and H. Kalutarage, “Naive bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation,” Soft Comput., 25 (3) 2277–2293 (2021). doi:10.1007/s00500-020-05297-6.
38) W.W.Cholil, F. Panjaitan, F. Ferdiansyah, A. Arista, R. Astriratma, and T. Rahayu, “Comparison of Machine Learning Methods in Sentiment Analysis PeduliLindungi Applications,” in: 2022 Int. Conf. Informatics, Multimedia, Cyber Inf. Syst., IEEE, 2022: pp. 276–280. doi:10.1109/ICIMCIS56303.2022.10017669.
39) E.Y.Boateng, J. Otoo, and D.A. Abaye, “Basic tenets of classification algorithms k-nearest-neighbor, support vector machine, random forest and neural network: a review,” J. Data Anal. Inf. Process., 08 (04) 341–357 (2020). doi:10.4236/jdaip.2020.84020.
40) A.Jovic, K. Brkic, and N. Bogunovic, “An overview of free software tools for general data mining,” in: 2014 37th Int. Conv. Inf. Commun. Technol. Electron. Microelectron., IEEE, 2014: pp. 1112–1117. doi:10.1109/MIPRO.2014.6859735.
41) J.Han, M. Kamber, and J. Pei, “Data Mining: Concepts and Techniques, 3rd edition,” Elsevier, 2012. doi:10.1016/C2009-0-61819-5.
42) BPSIndonesia, “Statistical Yearbook of Indonesia,” 2023.
43) Kepios,“Digital 2024: indonesia,” Meltwater Inc., (2024). https://datareportal.com/reports/digital-2024-indonesia.
44) B.Auxier, and M. Anderson, “Social media use in 2021,” Pew Res. Cent., (2021). https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.
45) B.Auxier, and M. Anderson, “Social media and news fact sheet,” Pew Res. Cent., (2023).

https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/.

46) S.Chen, G.I. Webb, L. Liu, and X. Ma, "A novel selective naïve bayes algorithm," Knowledge-Based Syst., 192 105361 (2020). doi:10.1016/j.knosys.2019.105361.
47) S.Uddin, I. Haque, H. Lu, M.A. Moni, and E. Gide, "Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction," Sci. Rep., 12 (1) 6256 (2022). doi:10.1038/s41598-022-10358-x.
48) B.Charbuty, and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," J. Appl. Sci. Technol. Trends, 2 (01) 20–28 (2021). doi:10.38094/jastt20165.
49) M.Ismail, N. Hassan, and S. Saleh Bafjaish, "Journal of soft computing and data mining comparative analysis of naive bayesian techniques in health-related for classification task," J. Soft Comput. Data Min., 1 (2) 1–10 (2020). doi:10.30880/jscdm.2020.01.02.001.
50) K.Maswadi, N.A. Ghani, S. Hamid, and M.B. Rasheed, "Human activity classification using decision tree and naïve bayes classifiers," Multimed. Tools Appl., 80 (14) 21709–21726 (2021). doi:10.1007/s11042-020-10447-x.
51) A.J.Wilson, B.S. Lakeland, T.J. Wilson, and T. Naylor, "A naive bayes classifier for identifying class ii ysos," Mon. Not. R. Astron. Soc., 521 (1) 354–388 (2023). doi:10.1093/mnras/stad301.
52) S.Dwiki, "Development of environmental policy in indonesia regarding mining industry in comparison with the united states and australia : the lesson that can be learned," Evergreen, 5 (2) 50–57 (2018). doi:10.5109/1936217.