Jurnal Krisnadana (Jurnal Komputer, Sistem Kendali & Jaringan)
Vol 5, Issue 3, May 2026, pp. 598-605
https://doi.org/10.58982/ha2x2j34

E-ISSN: 2808-3563
P-ISSN: 2808-3520

Sentiment Analysis of Public Opinion on Danantara Via Social Media X

Welda [1*], Aniek Suryanti Kusuma [2], I Putu Robin Laksamana Putra [3]

[1*] Informatics Dapartment, Institut Bisnis and Technology Indonesia, Denpasar, Bali, Indonesia
[2*,3] Magister of Informatics, Institut Bisnis and Technology Indonesia, Denpasar, Bali, Indonesia
welda@instiki.ac.id

ARTICLE INFO

Article history:
Received 3 June 2026
Revised 15 July 2026
Accepted 15 July 2026

Keywords:
Sentiment Analysis;
DANANTARA;
Social Media;
Naïve Bayes;
Web Scraping;

ABSTRACT

Social media has become a primary channel for the public to express opinions on various issues, including public services. One topic that has been widely discussed on platform X is Daya Anagata Nusantara (DANANTARA), a government-owned strategic investment management agency. Public sentiment toward DANANTARA is divided into three categories—positive, negative, and neutral—which may influence broader public perception. This study aims to analyze public sentiment toward DANANTARA using data obtained from platform X. The methods employed include web scraping for automated data collection and the Naïve Bayes algorithm for classifying sentiment into three labels: positive, negative, and neutral. Platform X was selected because it has over 19.5 million active users in Indonesia, making it a representative data source. From the 930 data points analyzed, it was found that public sentiment toward DANANTARA is predominantly neutral. This research is expected to provide an objective overview of public perception regarding DANANTARA.

© 2026 Authors. Licensed under CC BY-SA 4.0

1. Introduction

In the current digital era, social media has become an inseparable tool from the daily lives of modern society. Social media typically serves as a platform for the public to express their opinions and views on various aspects, ranging from social life to public services. One public service issue that has become a hot topic and is currently being widely discussed in the digital world, particularly on application X, is Daya Anagata Nusantara (DANANTARA). DANANTARA is a strategic investment management body that optimizes government investments to support economic growth and development (Daya Anagata Nusantara - Danantara, n.d.). The opinions of X application users are divided into three sentiment categories: neutral, positive, and negative. The opinions of X users toward DANANTARA will indirectly influence the sentiment or perception of the broader public, including those outside the X application, regarding DANANTARA.

To understand the public sentiment or perception toward DANANTARA on application X, the author concludes that an in-depth analysis of X user posts related to DANANTARA is necessary. In this analysis, the author will employ two methods: first, the web scraping method, which aims to retrieve post data from the X application (website); and second, for classification, the author selects a data mining algorithm, namely Naïve Bayes, to classify

posts related to DANANTARA, where the classification results will be labeled into three categories: Positive, Negative, and Neutral.

Application X, formerly known as Twitter, is a social media platform used by netizens or social media users as a space to express both positive and negative views on currently trending topics. Currently, there are 19.5 million active X users recorded in Indonesia [1]. Therefore, application X serves as the data source for this research. Sentiment analysis is one method used to examine public sentiment and opinion. One commonly used sentiment analysis method is classification using Natural Language Processing, one of which is Naïve Bayes. Using the Naïve Bayes method, sentiment grouping can be performed, with the intended groupings being positive, negative, and neutral sentiment. In conducting classification, data collection is certainly required for subsequent processing. This data collection will be carried out through web scraping on the X website application. The rationale for this approach includes the limitations of humans in extracting large amounts of data from a website within a short time, as well as the need for a sufficiently large dataset for model training to avoid inaccurate and imprecise training results.

There are several previous studies that have addressed similar themes and employed similar methods, namely web scraping for data collection and Naïve Bayes for classification. Among these is a study titled "Sentiment Analysis of Online Loans on Twitter Social Media Using the Naïve Bayes Method." The results of that study showed that out of 2,912 data points used for classification, negative sentiment accounted for 68.61% with 1,998 data points, and positive sentiment accounted for 31.39% with 914 data points. The model accuracy obtained was 80% [2]. Another study is titled "Public Sentiment Analysis of the Personal Data Protection Law on Application X Using the Support Vector Machine Method." The results of that study obtained 372 tweets on PDP Law posts using crawling techniques. The analysis process yielded 131 tweets categorized as positive, 182 tweets as neutral, and 59 tweets as negative, with the Support Vector Machine method used for modeling, producing an accuracy of 75.89% [3]. The third study is titled "Sentiment Analysis of Dana Application Reviews Using the Random Forest Method." The results of this study showed that a total of 1,354 data points were obtained through the scraping method from the Google Play Store. These data were processed using text preprocessing methods to simplify and clean the dataset, followed by modeling with Random Forest. Model testing was conducted using the confusion matrix method, yielding a model precision of 84% and a recall of 84% [4]. The next study is titled "Public Sentiment Analysis of Covid-19 Vaccination on Twitter Social Media Using the Support Vector Machine (SVM) Algorithm." This study used text mining techniques to clean and simplify the dataset. Data crawling techniques were used to collect tweet data regarding Covid-19 vaccination. The results showed that 4,708 tweet data points were collected, and the accuracy of the Support Vector Machine algorithm used for analysis was 89%. The study also produced a web-based application that can be used for crawling and analysis using the Support Vector Machine algorithm [5]. The fourth study is titled "Sentiment Analysis of Jd.Id Online Store Customers Using the Naïve Bayes Classifier Method Based on Emotion Icon Conversion," conducted by Sari [6]. The analysis was performed using the Naïve Bayes data mining method, and the results showed that the Naïve Bayes method used for sentiment analysis achieved an accuracy of 98%. The fifth study served as a reference for the selection of the data scraping method, titled "Web Scraping of Herbal Compounds in Indonesia Using Selenium Python." In this study, data were collected on herbal compounds in Indonesia through the KNAPSAC website link:

OPEN ACCESS

http://www.knapsackfamily.com/knapsac_k_core/top.php (KNAPSAC). Data scraping was performed using the Selenium library in Python. The scraping results successfully collected 1,341 data points [7].

Based on the problems described, the author concludes to design and build a system for web scraping and data mining analysis, namely "Sentiment Analysis of Daya Anagata Nusantara (Danantara) on Social Media X Using the Naïve Bayes Method." Based on the background described previously, the problem formulation of this study includes several key points. First, this study aims to identify the sentiment of X application users toward Daya Anagata Nusantara (DANANTARA). Second, this study will explain the process of collecting public opinion data regarding DANANTARA through the web scraping method on application X. Third, this study will also discuss the application of the Naïve Bayes algorithm in classifying X application users' sentiment into positive, negative, and neutral categories toward DANANTARA. With this problem formulation, it is expected that the research will provide in-depth insights into public perception of the policy under investigation.

2. Method

This study employs sentiment analysis with the Naive Bayes algorithm to classify public perception of Daya Anagata Nusantara (DANANTARA) on the social media platform X. Naïve Bayes is a classification method that uses probability and statistical methods proposed by the British scientist Thomas Bayes, which specifically predicts future probabilities based on previous experience [8]. This method is frequently used in text analysis, including sentiment analysis, due to its capability to handle large datasets, as well as its accuracy and speed in the classification process. The Naive Bayes algorithm is highly effective for large datasets and is widely used in text classification tasks due to its simplicity and efficiency [9]. The research flow is systematically presented in Figure 1.

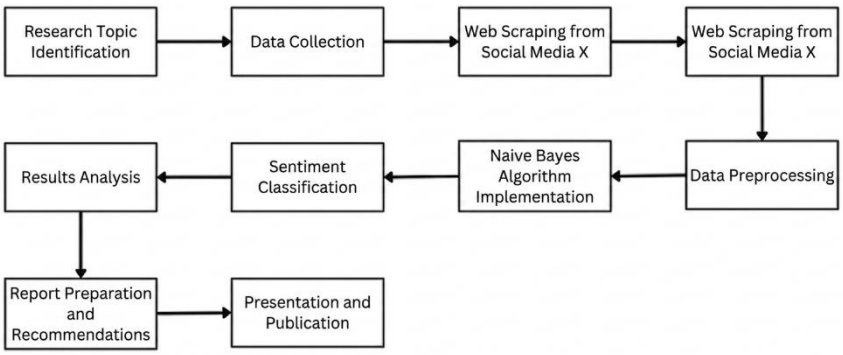

Figure 1. Research Process Flow

Primary data were collected through web scraping techniques using the Python programming language with Selenium and BeautifulSoup libraries, then stored in an SQL database. Sentiment labeling was carried out automatically based on predetermined keywords.

The text preprocessing stage was applied to clean the data and reduce ambiguity, encompassing four processes: case folding to convert all letters to lowercase, tokenization to

separate sentences into individual words, stop word removal to eliminate unimportant words, and stemming to transform affixed words into their root forms [10].

Visualization of dominant keywords was performed using word clouds to identify the most frequently occurring words in each sentiment class [11]. The visualization results for positive, negative, and neutral sentiments are displayed in Figure 2.

Figure 2. Word Cloud of Positive, Negative, and Neutral Sentiments

Sentiment classification was carried out using the Multinomial Naive Bayes algorithm with Laplace Smoothing to handle words that do not appear in the training data. The calculation began with determining the prior probability for each sentiment class, then calculating the likelihood of each word for each class. The likelihood calculation results for each word in the test sentence example are presented in Table 1.

Table 1. Log Likelihood of Each Word for Sentiment Classes

Word	log(P (word|positive))	log(P (word|neutral))	log(P (word|negative))
step	-8.943	-9.197	-8.386
pertama	-7.147	-7.002	-8.386
expan	-8.943	-9.197	-8.386
tkdn	-5.832	-6.938	-7.702
jadi	-6.025	-7.123	-6.693
skala	-8.943	-8.510	-8.386
asean	-7.763	-8.820	-8.386
kedua	-7.147	-8.510	-7.702
masukkan	-8.943	-9.197	-8.386
danantara	-5.424	-7.530	-7.702
ke	-4.332	-4.191	-5.398
global	-6.136	-7.002	-7.702
supply	-7.763	-7.530	-8.386
chain	-7.471	-7.649	-8.386

The likelihood values were then accumulated with the prior probability on a logarithmic scale to avoid floating-point underflow. The log probability accumulation was calculated using the formula:

$$Total\_\log = Log(C) + \Sigma\, Log(P|Word) \quad (1)$$

The accumulation calculation results for each sentiment class are summarized in Table 2.

Table 2. Log Probability Accumulation for Each Sentiment Class

Sentiment Class	Log Probability Value
Positive	-101,829
Neutral	-101,135
Negative	-113,876

The sentiment class was determined based on the highest log probability value. Based on Table 2, the highest value was obtained for the neutral class (-101.135), therefore the test sentence was classified as neutral sentiment.

Overall model performance evaluation was conducted using accuracy, precision, recall, and f1-score metrics. The evaluation results for all test data are displayed in Table 3.

Table 3. Classification Model Evaluation Results

Class	Precision	Recall	F1-Score	Support
Positive	0,85	0,79	0,82	336
Neutral	0,81	0,88	0,84	453
Negative	0.78	0,71	0,74	141
Accuracy			**0,82**	**930**

Based on Table 3, the Naive Bayes model built demonstrates good performance with an overall accuracy of 82%, with the neutral class having the highest f1-score (0.84), indicating that the model is most effective in identifying neutral sentiment.

3. Result and Discussions (Heading 4) (Palatinotype, bold, 11 pt)

The implementation of the sentiment analysis system was carried out through several stages, from data collection, text preprocessing, classification, model testing, to the interpretation of results.

Data Collection and Dataset

Data were automatically collected from the social media platform X using the Python programming language with Selenium and BeautifulSoup libraries. The search keyword used was "DANANTARA" with a time frame from January 1, 2024 to December 12, 2025. An initial sentiment label was assigned in a simple manner by matching tweet content against a predefined list of positive and negative keywords. From 1,000 raw data points collected, a filtering process was conducted to remove duplicates and irrelevant content, resulting in a final dataset of 930 tweets. This dataset consists of two main columns, namely tweet content and initial sentiment label.

Figure 3. Final Dataset from Scraping Results

Data Preprocessing

The text went through several preprocessing stages, including: case folding (converting all characters to lowercase), tokenization (splitting sentences into words), stopword removal (removing common uninformative words using NLTK and a custom stopword list), and stemming using the Sastrawi library to convert words to their base forms. The sentiment labels were then converted into numerical codes: 1 (positive), 0 (negative), and 2 (neutral). To provide a visual overview of the most dominant words in each sentiment category, word clouds were generated.

Figure 4. Word Clouds by Sentiment Category: Negative, Positive, Neutral

Naïve Bayes Classification and Model Testing

The classification model was built using the Multinomial Naïve Bayes algorithm from the Scikit-learn library. The dataset was randomly split into training and testing sets, then the text features were extracted using Count Vectorization. The complete prediction results were saved in JSON format. Model performance evaluation using a confusion matrix and classification report yielded an accuracy of 79.9%, indicating that the model possesses fairly good predictive capability.

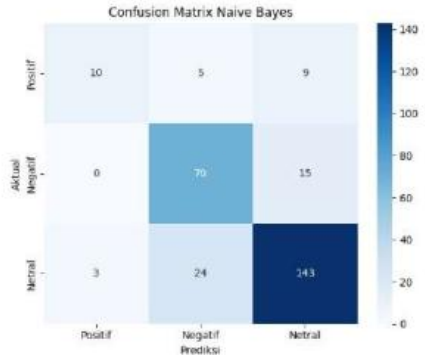

Figure 5. Confusion Matrix from Model Testing

Classification Results

The application of the model to the entire dataset revealed that the distribution of public sentiment towards DANANTARA was dominated by neutral sentiment at 61.40%, followed by positive sentiment at 32.47%, and negative sentiment at only 6.13%. These results indicate that public discourse on platform X is predominantly informational or neutral in nature, rather than constituting strong expressions of support or rejection.

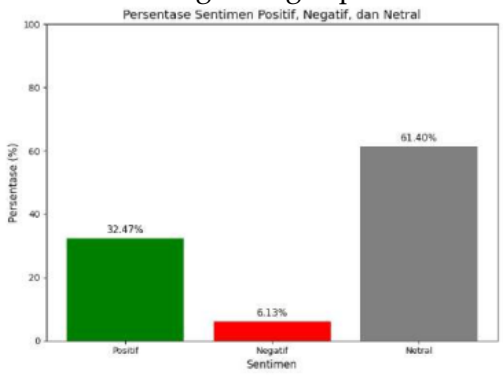

Figure 6. Distribution of Public Sentiment towards DANANTARA

4. Conclusion

The This research successfully conducted a sentiment analysis of Daya Anagata Nusantara (DANANTARA) on social media X using the Naive Bayes method. Data were collected through web scraping from platform X, yielding a total of nine hundred and thirty data points which served as the dataset for the classification process. The classification results revealed that public sentiment towards DANANTARA tends to be neutral, with the proportion of positive sentiment at 32.47%, negative sentiment at 6.13%, and neutral sentiment at 61.40%. The Naïve Bayes model was tested on the entire dataset of nine hundred and thirty data points and achieved an accuracy of 79.9%, indicating that seven hundred and sixty-three data points were correctly predicted. This achievement demonstrates that the model exhibits good performance in carrying out the sentiment prediction process.

References

[1] A. Sentimen and D. A. N. Dana, "Twitter Dengan Metode Naive Bayes Classifier," vol. 8, no. 5, pp. 9858–9864., 2024.

[2] M. I. Ghozali, W. H. Sugiharto, and A. F. Iskandar, "Analisis Sentimen Pinjaman Online Di Media Sosial Twitter Menggunakan Metode Naive Bayes," *KLIK Kaji. Ilm. Inform. Dan Komput.*, vol. 3, no. 6, pp. 1340–1348, 2023.

[3] R. Abdul *et al.*, "Analisis Sentimen Masyarakat Terhadap Undang-Undang Perlindungan Data Pribadi Pada Aplikasi X Dengan Metode Support Vector Machine," *J. Method.*, vol. 10, no. 1, pp. 6–11, 2024.

[4] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," vol. 6, no. 9, pp. 4305–4313, 2022.

[5] H. Syah and A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector

Machine (SVM)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, pp. 59–67, 2022.

[6] S. V. Sari, "Analisis Sentimen Pelanggan Toko Online Jd . Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," vol. 10, no. 2, pp. 681–686, 2019.

[7] N. Fadhilla, R. Afrianti, and A. Badawi, "Web Scraping Senyawa Herbal di Indonesia Menggunakan Selenium Python," vol. 4307, no. 4, pp. 1362–1366, 2024.

[8] V. D. Lestari, J. Fredricka, and F. I. Nizar, "Implementasi Metode Bayes untuk Klasifikasi Penyakit Pasien pada Puskesmas Anggut Atas Kota Bengkulu," *Digit. Transform. Technol.*, vol. 5, no. 1, pp. 409–419, 2025.

[9] B. Rahmatullah, S. A. Saputra, P. Budiono, and D. P. Wigandi, "Sentiment Analysis of Free Nutritious Meals Using Naive Bayes Algorithm," *J. Inf. Technol.*, vol. 5, no. 1, pp. 259–264, 2025.

[10] L. Hermawan and M. B. Ismiati, "Pembelajaran Text Preprocessing Berbasis Simulator Untuk Mata Kuliah Information Retrieval," *J. Transform.*, vol. 17, no. 2, pp. 188–199, 2022.

[11] J. A. Wibowo, V. C. Mawardi, and T. Sutrisno, "Visualisasi Word Cloud Hasil Analisis Sentimen Berbasis," *J. Serina Sains, Tek. dan Kedokt.*, vol. 2, no. 1, pp. 61–70, 2024.