JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 1269~1282
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.5138

Analysis of Public Sentiment Towards Retired Military Officers' Pressure to Impeach the Vice President Through X Using Decision Trees

Dimas Andrean[1*], Rakhmat Kurniawan[2]
[1,2]Computer Science Program, State Islamic University of North Sumatra

Article Info

Article history:

Received 12 07, 2025
Revised 12 16, 2025
Accepted 12 27, 2025

Keywords:

Sentiment Analysis,
Impeachment of Vice President,
Retired TNI Officer,
Social Media X,
CART Decision Tree

ABSTRACT

The advancement of information technology has strengthened the analysis of social media data, particularly in understanding public opinion on national political issues. This study examines public sentiment on the X (Twitter) platform regarding the issues of Gibran's Impeachment and the Urging by Retired Military Officers using the Decision Tree CART algorithm. Data were collected through a crawling process, resulting in 1,020 tweets for the Impeachment issue and 89 tweets for the Urging issue. After preprocessing, the dataset was labeled using a Lexicon-Based method that classifies text into positive, negative, and neutral sentiments. The evaluation results show that for the Impeachment issue, the model achieved an accuracy of 97.05%–99.51%, with the highest performance found in the Neutral class (F1-Score 98.46%). For the Urging issue, the model obtained an overall accuracy of 88.89%, with the highest performance also in the Neutral class (F1-Score 94.12%). Model performance decreased in the Positive and Negative classes due to data imbalance. Overall, the findings indicate that Decision Tree CART is effective for sentiment classification on small to medium datasets and reveal that public sentiment toward both issues is predominantly Neutral.

Corresponding Author:

Dimas Andrean
Computer Science Program,
State Islamic University of North Sumatra
North Sumatra, Indonesia
Email: dimasandrean112237@gmail.com

1. Introduction

Advances in information technology have driven progress in data processing and analysis, particularly social media data, which now directly influences social interaction patterns and political dynamics [1]. Current advances in information technology have had a major impact on how data is managed and stored [2]. Social media, as part of information technology, has become an active public discussion space, where people voice their opinions on various political issues and policies [3]. In Indonesia, platforms such as X (formerly Twitter) play an important role in shaping public perception and can even exert pressure on political decisions [4]. Opinions that develop on social media can create polarization or widespread support for certain issues in a short period of time.

In recent years, social media use in Indonesia has experienced significant growth, especially in the context of conveying public opinion on social and political issues. Twitter, now known as the X platform, is one of the most popular social media platforms in Indonesia and has become the main digital space for the public to express their views, criticism, and support for various national political events [5]. The open, fast, and real-time nature of X allows every user to participate in public discussions, including on sensitive issues such as government policies, elections, and statements by national figures [5]. This condition is clearly seen in a case that is currently in the public spotlight, namely the insistence of a number of retired TNI officers on the MPR to impeach the Vice President. This issue has caused a huge stir on social media, especially on the X platform, with thousands of posts, replies, and threads showing the public's involvement in the political discourse. The public's responses have been very diverse, with some supporting, some rejecting, and some remaining neutral. However, these interactions as a whole provide a concrete picture of how social media has become a digital reflection of public opinion.

This phenomenon shows that social media is not merely a space for sharing information, but has transformed into an alternative political space that influences discourse and even the potential direction of state policy. However, the abundance of opinions that appear on social media are spontaneous, unstructured, and scattered across a large volume of data, making them difficult to understand directly. Therefore, even though many public statements are recorded in the digital space, understanding the direction and strength of public opinion becomes unclear if it is not analyzed scientifically and systematically. This is where the importance of conducting in-depth studies based on social media data lies, in order to gain an objective and comprehensive understanding of public perception [6].

The main problem in analyzing public opinion on social media, especially on complex and sensitive issues such as the pressure from retired TNI officers on the MPR to impeach the Vice President, lies not only in the large volume of unstructured data, but also in the socio-political implications it raises. Unmanaged public opinion can trigger polarization, spread misinformation, and even influence the general public's perception of government stability.

Content that is full of emotional, provocative, and even sarcastic language is often amplified through virality on platforms such as X, which can cause unrest, division of opinion, and potentially disrupt social cohesion. If this data is not analyzed systematically, the voice of the people will only become “digital noise” that masks real aspirations, when in fact it can be used as an indicator of the health of democracy and public trust in state institutions. Therefore, data mining analysis methods are needed that are not only capable of classifying sentiment into positive, negative, and neutral categories, but also provide insights that can be used as a basis for public policy-making, social conflict mitigation, and maintaining a digital space that is healthier and more productive for society.

One appropriate data mining approach to address this issue is to use the Decision Tree algorithm. Decision Tree is a method in data mining that can form a classification model based on labeled historical data [8]. This algorithm is considered suitable because it has advantages in terms of ease of interpretation, clear visual structure (in the form of a decision tree), and its ability to handle text data that has been converted into numerical features, such as keywords or word frequency [9].

Previous studies have applied Decision Tree algorithms to social media sentiment analysis, but in different contexts. Hardyatman & Hasan (2025) examined public perceptions of plans to increase VAT by 12% using 1,000 tweets, focusing on practical fiscal policy issues. In contrast, this study examines a strategic and sensitive political issue, namely the discourse on the impeachment of the Vice President, which has a higher level of opinion polarization because it involves the dimensions of political legitimacy and power interests[10]. In addition, this study uses a larger and longitudinal dataset to obtain a more comprehensive mapping of opinions. Meanwhile, Octa N. et al. (2024) examined public sentiment towards marketplaces in Indonesia with a Decision Tree accuracy of 70.27%, in the relatively stable and politically uncontroversial realm of the digital economy. From this comparison, it appears that there is a gap in research in the form of the absence of a Decision Tree model optimized for the context of strategic political opinion in Indonesia [11]. Therefore, this study aims to develop a Decision Tree-based public opinion classification model to improve the accuracy of sentiment analysis on national political issues, particularly those related to the pressure exerted by retired TNI officers to impeach the Vice President. The objective of this study is to utilize the Decision Tree algorithm to classify public opinion regarding the retired TNI's demand to the MPR for the impeachment of the Vice President, thereby obtaining an objective picture of public perception and scientific contribution in the study of digital political opinion.

The research method employed in this study is a quantitative approach, which is designed to test existing theories by examining measurable data and identifying patterns through systematic and objective observation. Quantitative research emphasizes the use of numerical data and statistical analysis to explain relationships between variables in a structured and replicable manner. In this context, the method is considered appropriate because it allows the researcher to assess public sentiment in a comprehensive, data-driven way, minimizing subjective interpretation and enhancing the reliability of the findings.

The primary focus of this study is to analyze public sentiment toward the pressure exerted by retirees of the Indonesian National Armed Forces (Tentara Nasional Indonesia/TNI) on the People’s Consultative Assembly (Majelis Permusyawaratan Rakyat/MPR) in relation to the impeachment of the Vice President.

This issue is examined as a public and political phenomenon that has generated significant attention and debate within Indonesian society. By concentrating on sentiment, the study seeks to capture how the public perceives, supports, criticizes, or responds to the actions and demands of TNI retirees within the broader political discourse.

The data used in this research are derived from opinions disseminated on social media platform X (formerly Twitter), which serves as a space for public expression and real-time political discussion. Social media data are particularly relevant because they reflect spontaneous public reactions and diverse viewpoints from various segments of society. Through quantitative analysis of this opinion data, the study aims to identify dominant sentiment trends and patterns, thereby providing an empirical understanding of public responses to the issue under investigation without altering the original meaning or reference framework of the research.

2. Research Method

2.1. Application of Decision Tree

This study uses the Decision Tree algorithm as a sentiment classification method to group public opinion into positive, negative, and neutral categories. Several stages are carried out gradually, starting from data collection, text pre-processing, feature extraction, data labeling, classification modeling, to model evaluation to obtain valid and reliable analysis results. The following are the stages of the Flowchart application used in this study :

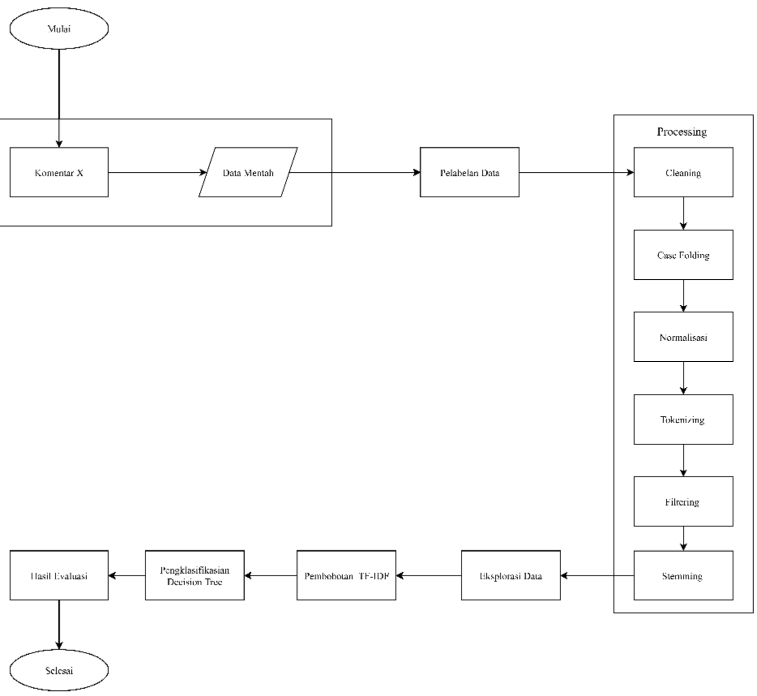


Figure 1. Flowchart Decision Tree

Based on Figure 1, the design flow for each process of the system to be built can be explained as follows.

1. Data Collection

Initial data is taken from comments (in this case referred to as "Comment X"). At this stage, all comments received are collected in their original form without any modification. The data obtained is still raw data that is heterogeneous, unstructured, and often contains irrelevant elements, duplications, or inconsistencies. The collection is carried out thoroughly so that no information is missed, considering that each comment has the potential to provide an overview of the academic conditions and patterns of student interaction. This raw data is then stored in a database as the basis for the cleaning, normalization, and transformation processes in the next stage, so that the system can produce accurate and reliable information.

2. Data Labeling

Once the data has been collected, it is labeled to categorize it, for example, positive, negative, or neutral comments. At this stage, each comment is analyzed to determine the class that best fits the meaning and context of the comment. The labeling process can be done manually by human evaluators or using a semi-automatic approach with the help of an initial model, but it still requires verification to ensure consistent results. This stage is crucial because the quality of the labels provided will affect the model's ability to recognize language patterns, tone, and comment content trends. With proper labeling, the system obtains a structured and ready-to-use dataset for the training process, enabling the model to learn more accurately and produce reliable predictions during the implementation stage.

3. Data Preprocessing

   The labeled data then goes through several preprocessing stages to prepare the data for further analysis [12]:

   a. Cleaning: Removing irrelevant elements such as punctuation marks, numbers, and special characters.
   b. Case Folding: Converting all letters to lowercase for consistency.
   c. Tokenizing: Breaking text into tokens (words) based on spaces.
   d. Filtering: Removing unimportant words using a stopword list.
   e. Stemming: Converting inflected words to their base form according to applicable rules [13].

4. Data Exploration

After preprocessing, data exploration was conducted with sentiment visualization using WordCloud to display the words that appeared most frequently in each sentiment category. At this stage, the system began to present an initial overview of word patterns based on the sentiment labels that had been assigned previously. WordCloud was used as an exploratory tool to help researchers see language trends, dominant terms, and characteristics that appeared in positive, negative, and neutral comments. Through this visualization, researchers can understand the narrative structure underlying each sentiment group, identify words that contribute significantly to sentiment assessment, and find early indications of the issues or topics most frequently discussed by users. The results of this exploration then become an initial reference before entering a more in-depth analysis stage or developing a machine learning-based model.

5. TF-IDF Weighting

At this stage, each term in the training data is weighted using the TF IDF (Term Frequency Inverse Document Frequency) method to assess the importance of each word in the context of the document. At this stage, the system calculates how often a word appears in a comment compared to how common that word is in the entire collection of comments. This weighting process aims to highlight words that have high informational value, while reducing the influence of words that are too common and do not provide specific meaning to the analysis. The TF IDF calculation results in a more structured numerical representation, allowing the model to learn language patterns more effectively. This stage also ensures that each comment has a feature vector that reflects the actual characteristics of the text, before it is finally used in the sentiment classification model training process.

6. Classification with the Decision Tree Algorithm

The weighted data is then classified using the Decision Tree algorithm to produce a classification model based on predetermined categories. At this stage, the model begins to learn the relationship patterns between features that have been represented through TF IDF weights and the target sentiment categories. Decision Tree was chosen for its ability to clearly map the decision-making process through a branching structure, so that each classification decision can be traced back based on specific conditions. During the training process, the algorithm builds a decision tree by searching for the most informative feature separation, so that the model can accurately group comments into positive, negative, or neutral classes. This stage is an important foundation for producing a stable, interpretable model that can be used in the prediction process on new data.

7. Evaluation Results

   The resulting model is evaluated using a Confusion Matrix to calculate:

   a. True Positive (TP): Correct predictions for the positive class.
   b. True Negative (TN): Correct predictions for the negative class.
   c. False Positive (FP): Incorrect predictions for the positive class.
   d. False Negative (FN): Incorrect predictions for the negative class. From the Confusion Matrix, accuracy, precision, recall, and F1-score can be calculated to assess the model's performance.

3. Result and Discussion

The results and discussion in this study are presented through several stages that are arranged systematically.

3.1. Data Collection

The dataset in this study uses secondary data sources, namely public data obtained directly from the internet through the social media platform X (formerly Twitter). The data used consists of public opinions or responses to the Indonesian Armed Forces retirees' push for the impeachment of the Vice President. Data collection was carried out using the crawling (scraping) technique using the Python programming language, which was run in the Google Colab environment with the help of the Tweet-Harvest tool that utilizes Twitter API token authentication.

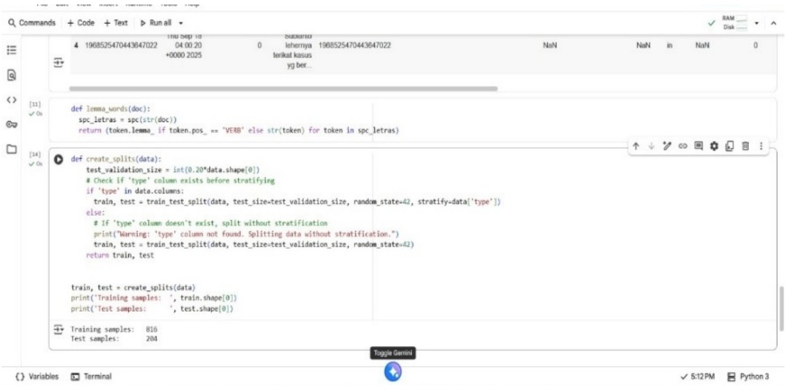


Figure 2. Data Crawling Process

After crawling data with the keywords, we obtained 1,020 tweets from the two hashtags above and 89 tweet data.

3.2. Data Processing

At this stage, data preprocessing will be carried out. This process is necessary because the data still contains many unnecessary elements (noise) in the sentiment analysis process. Therefore, the data must first be cleaned of all distracting elements. The obtained dataset is stored and will be processed in a file with the .csv format. The following is a system flow of the data preprocessing process, which includes cleaning, case folding, normalization, tokenizing, stopword removal, and stemming.

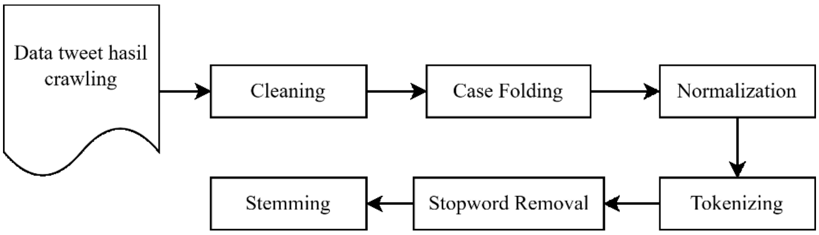


Figure 3. Data Preprocessing Process Block Diagram

1. Cleaning

   The first step is cleaning, which aims to clean up the comment data that has been obtained. The elements that are removed are mentions, hashtags, user names, links, email addresses, numbers, punctuation marks, and emoticons. Examples of data that has undergone the Cleaning stage of data preprocessing are presented in Tables 1 and 2.

Table 1. Data on the Results of Cleaning Gibran's Impeachment

NO	Tweet	Cleaning
1	Roy Suryo said he was asked by the TNI Retirees Forum to be an expert witness for Gibran's impeachment, including matters related to the Fufufafa account and analysis related to the constitution and alleged corruption.	Roy Suryo was appointed as a telecommunications expert by the Indonesian Armed Forces Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka.
2	Roy Suryo's statement on September 10, 2025, as reported by various media outlets, mentioned that he was asked by retired TNI members to serve as an expert to support Gibran's impeachment, including	According to Roy Suryo's statement on September 10, 2025, he was asked by Indonesian Armed Forces retirees to be an expert for Gibran's impeachment, including

NO	Tweet	Cleaning
	matters related to the Fufufafa account. Gibran previously denied this.	matters related to the Fufufafa account. Gibran denied this.

Table 2. Data on the Results of Cleaning Gibran's Pressure

NO	Tweet	Cleaning
1	@NgkongRoses, the Indonesian House of Representatives (DPR_RI) and the People's Consultative Assembly (MPR) play a very important role, especially since they have received a letter from the TNI Retirees Forum regarding the demand for Gibran's impeachment..	The Indonesian House of Representatives and the People's Consultative Assembly play a very important role, especially since they have received a letter from the TNI Retirees Forum regarding the demand for Gibran's impeachment, but the House of Representatives.
2	The demand from the TNI Retirees Forum for Gibran's replacement as Vice President must be taken seriously by President Prabowo. It must also be reviewed from a constitutional perspective. (Member)	The demand from the TNI Retirees Forum for Gibran's replacement as Vice President must be taken seriously by President Prabowo. It must also be reviewed from a constitutional perspective.

2. Case Folding

The second step is case folding, which aims to convert all letters in the text to lowercase to standardize their form. Examples of data that have undergone the Case Folding preprocessing stage are presented in Tables 3 and 4.

Table 3. Data on the Results of Gibran's Impeachment Case

NO	Cleaning	Case Folding
1	Roy Suryo said he was appointed by the TNI Retirees Forum as a telecommunications expert to support the impeachment of Vice President Gibran, including preparing data on Fufufafa accounts and analyzing constitutional issues and allegations of corruption.	Roy Suryo appointed by the TNI Retirees Forum as a telecommunications expert to support the impeachment process
2	Roy Suryo was officially appointed by the TNI Retirees Forum.	Roy Suryo officially appointed by the TNI Retirees Forum

Table 4. Data on Gibran's Case Folding Results

NO	Cleaning	Case Folding
1	The Indonesian House of Representatives and the People's Consultative Assembly played a very important role, especially since they had already received a letter from the TNI Retirees Forum regarding the demand for Gibran's impeachment, but the House of Representatives did not have the courage to follow up on it.	The Indonesian House of Representatives and the People's Consultative Assembly played a very important role, especially since they had already received a letter from the TNI Retirees Forum regarding the demand for Gibran's impeachment, but the House of Representatives did not have the courage to follow up on it.
2	The demand from the TNI Retirees Forum for the replacement of Vice President Gibran	The TNI Retirees Forum demanded his replacement.

3. Spelling Normalization

The third stage is normalization, which aims to normalize words in the text into standard Indonesian according to the KBBI (Big Indonesian Dictionary). This normalization stage uses the Indonesian kamuskatabaku.xlsx dictionary file accessed from Github. The following examples of data that have undergone the Normalization stage of data preprocessing are presented in Tables 5 and 6.

Table 5. Data Results of Gibran's Impeachment Normalization

NO	Case Folding	Spelling Normalization
1	Roy Suryo was appointed as a telecommunications expert by the Indonesian	Roy Suryo was appointed as a telecommunications expert by the Indonesian

NO	Case Folding	Spelling Normalization
2	Armed Forces (TNI) retirees forum to support the impeachment of Vice President Gibran Rakabuming Raka. Yes, the news is true based on Roy Suryo's statement on September 10, 2025, as reported by other media outlets. He was appointed as an expert by retired TNI personnel to	Armed Forces (TNI) retirees forum to support the impeachment of Vice President Gibran Rakabuming Raka. He ya berita itu benar berdasarkan pernyataan Yes, the news is true based on Roy Suryo's statement on September 10, 2025, as reported by other media outlets. He was appointed as an expert by retired TNI personnel to support

Table 6. Data on the Results of Normalization of Gibran's Pressure

NO	Case Folding	Spelling Normalization
1	The Indonesian House of Representatives and the People's Consultative Assembly play a very important role, especially since they have received a letter from the TNI retirees forum regarding the demand for impeachment.	The Indonesian House of Representatives and the People's Consultative Assembly play a very important role, especially since they have received a letter from the TNI retirees forum regarding the demand for Gibran's impeachment, but the House of Representatives does not have the courage to follow up on it.
2	The TNI retirees' forum's demand for the replacement of Vice President Gibran must be	he TNI retirees' forum's demand for the replacement of Vice President Gibran must be taken seriously by

4. Tokenizing

The fourth stage is tokenizing, which aims to separate words in a sentence or tweet. The following examples of data that have undergone this stage of data preprocessing are presented in Tables 7 and 8.

Table 7. Data Results of Tokenizing Gibran's Impeachment

NO	Spelling Normalization	Tokenizing
1	Roy Suryo was appointed as a telecommunications expert by the Indonesian Armed Forces Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka. He prepared data on the Fufufafa account allegedly linked to Gibran, along with other analyses related to constitutional issues and corruption.	[Roy, Suryo, appointed, as, expert, telematics, by, forum, retired, TNI, to, support, impeachment, vice, president, Gibran, Rakabuming, Raka, he, prepared, data, account, fufufafa, which, allegedly, related, Gibran, along with, other, analysis, related, issues, constitution, and, corruption]
2	Roy Suryo was officially appointed by the Indonesian Armed Forces Retirees Forum.	[Roy, Suryo, officially appointed by the TNI veterans forum]

Table 8. Data from Tokenizing Gibran's Urgency

NO	Spelling Normalization	Tokenizing
1	The Indonesian House of Representatives and the People's Consultative Assembly played a very important role, especially since they had received a letter from the TNI retirees' forum regarding the demand for Gibran's impeachment, but the House of Representatives did not have the courage to follow up on it.	[‘very’, ‘important’, ‘role’, ‘that’, ‘DPR’, ‘RI’, ‘and’, ‘MPR’, ‘especially’, ‘they’, ‘have’, ‘received’, ‘letter’, ‘forum’, ‘retired’, ‘TNI’, ‘regarding’, 'demands', ‘impeachment’, ‘Gibran’, ‘but’, ‘DPR’, ‘does’, ‘not’, ‘have’, ‘the courage’, ‘to’, ‘follow up on it’]
2	The pressure from the TNI retirees' forum demanding the replacement of Vice President Gibran must be taken seriously by President Prabowo and reviewed from a constitutional perspective. Members of the PDIP faction in the House of Representatives have stated this.	[‘demand’, ‘forum’, ‘retired’, ‘military’, ‘officers’, ‘requesting’, ‘replacement’, ‘vice’, ‘president’, ‘gibran’, ‘must’, ‘be’, ‘taken’, ‘seriously’, ‘by’, ‘president’, ‘prabowo’, ‘and’, ‘reviewed’, 'also', ‘from’, ‘the’, ‘constitutional’, ‘aspect’, ‘members’, ‘of’, ‘the’, ‘House’, ‘of’, ‘Representatives’, ‘from’, ‘the’, ‘PDI-P’,

NO	Spelling Normalization	Tokenizing
		‘faction’, ‘stated’, ‘this’, ‘point’]

5. Stopword Removal

The next step is stopword removal, which aims to remove words in the document that have no meaning, such as pronouns, conjunctions, and so on. The following examples of data that have undergone this stage of data preprocessing are presented in Tables 9 and 10.

Table 9. Data Results of Stopword Removal for Gibran's Impeachment

NO	Tokenizing	Stopword Removal
1	[Roy Suryo was appointed as a telecommunications expert by the Indonesian Armed Forces (TNI) Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka. He prepared data from the Fufufafa account, which is suspected of being linked to Gibran, along with other analyses related to constitutional issues and corruption.]	[Roy Suryo was appointed as a telecommunications expert by the Indonesian Armed Forces (TNI) Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka. He prepared data from the fufufafa account, which is suspected of being linked to Gibran, along with other analyses related to constitutional issues and corruption.]
2	[Yes, the news is true, based on a statement by Roy Suryo on September 10, 2025, as reported by other media outlets. He was appointed as an expert by retired TNI officers to support Gibran's impeachment, including data from the fufufafa account, but Gibran has denied this].	[news, true, statement, Roy Suryo, September 10, 2025, reported, media, appointed, expert, retired TNI officer, supporting, impeachment, Gibran, data, account, fufufafa, Gibran, denied]

Table 10. Data Results of Stopword Removal Gibran's Urgency

NO	Tokenizing	Stopword Removal
1	[‘which’, ‘plays’, ‘a’, ‘very’, ‘important’, ‘role’, ‘in’, ‘the’, ‘DPR’, ‘RI’, ‘and’, ‘MPR’, ‘especially’, ‘since’, ‘they’, ‘have’, ‘already’, ‘received’, ‘a’, ‘letter’, ‘from’, ‘the’, ‘forum’, ‘of’, ‘retired’, ‘TNI’, ‘officers’, ‘regarding’, 'demands', ‘impeachment’, ‘Gibran’, ‘but’, ‘DPR’, ‘does not’, ‘have’, ‘the courage’, ‘to’, ‘follow up on it’]	[‘play a role’, ‘DPR’, ‘RI’, ‘MPR’, ‘receive’, ‘letter’, ‘forum’, ‘retired’, ‘TNI’, ‘demand’, ‘impeachment’, ‘Gibran’, ‘DPR’, ‘have’, ‘the courage’, ‘to follow up on it’]
2	[‘pressure’, ‘forum’, ‘retired’, ‘military’, ‘who’, ‘are’, ‘requesting’, ‘replacement’, ‘vice’, ‘president’, ‘gibran’, ‘must’, ‘be’, ‘responded’, ‘to’, ‘seriously’, ‘by’, ‘president’, ‘prabowo’, ‘and’, ‘reviewed’, 'also', ‘from’, ‘the’, ‘constitutional’, ‘aspect’, ‘members’, ‘of’, ‘the’, ‘House’, ‘of’, ‘Representatives’, ‘from’, ‘the’, ‘PDI-P’, ‘faction’, ‘declare’, ‘that’, ‘the’, ‘matter’]	[‘urge’, ‘forum’, ‘retired’, ‘military’, ‘request’, ‘replacement’, ‘vice’, ‘president’, ‘gibran’, ‘responded’, ‘seriously’, ‘president’, ‘prabowo’, ‘reviewed’, 'aspect', ‘constitution’, ‘member’, ‘DPR’, ‘faction’, ‘PDIP’, ‘declare’]

6. Stemming

The final stage of the preprocessing process is stemming, which aims to remove all affixes in words found in the document, such as prefixes, suffixes, and word pluralization. Examples of data that have undergone this stage of data preprocessing are presented in Tables 11 and 12.

Table 11. Data from the Stemming Results of Gibran's Impeachment

NO	Tokenizing	Stopword Removal
1	Roy Suryo was appointed as a telecommunications expert by the Indonesian Armed Forces Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka. He prepared data on the Fufufafa account, which is suspected of being linked to Gibran,	Roy Suryo was appointed as a telecommunications expert by the TNI Retirees Forum to support the impeachment of Vice President Gibran Rakabuming Raka. He prepared data on

NO	Tokenizing	Stopword Removal
	along with other analyses related to constitutional issues and corruption.	the Fufufafa account allegedly linked to Gibran, along with analysis on constitutional issues and corruption.
2	Yes, the news is true based on Roy Suryo's statement on September 10, 2025, as reported by other media outlets. He was appointed as an expert by retired TNI officers to support Gibran's impeachment, including data from the fufufafa account, but Gibran has denied this.	The news is true. Roy Suryo's statement on September 10, 2025, was reported by the media. He was appointed as an expert by retired TNI officers to support Gibran's impeachment, including data from the fufufafa account. Gibran has denied this.

Table 12 .Data from Gibran's Stemming Results

NO	Tokenizing	Stopword Removal
1	The Indonesian House of Representatives and the People's Consultative Assembly played a very important role, especially since they had already received a letter from the TNI retirees' forum regarding the demand for Gibran's impeachment, but the House did not have the courage to follow up on it.	The Indonesian House of Representatives and the People's Consultative Assembly received a letter from the TNI Retirees Forum urging the impeachment of Gibran. The House of Representatives had the courage to follow up on this.
2	The demand by the TNI retirees forum for the replacement of Vice President Gibran must be taken seriously by President Prabowo and reviewed from a constitutional perspective. Members of the DPR from the PDIP faction have stated this	The demand by the TNI retirees forum for the replacement of Vice President Gibran has been taken seriously by President Prabowo and reviewed from a constitutional perspective. Members of the DPR from the PDIP faction have stated this.

3.3. CART Decision Tree Classification

After the word weighting stage is complete, the test data will be classified using the CART Decision Tree method. The following is an example of classification calculations for test data using training data that has undergone preprocessing and labeling. The CART Decision Tree method divides based on the feature with the largest Gini reduction. The Gini calculation is performed using the following formula.

$$Gini = 1 - \sum(p_i)^2 \tag{1}$$

Explanation:
Gini = impurity value of a node.
Σ = summation symbol.
p_i = proportion of data belonging to class i in the dataset.
$(p_i)^2$ = square of the proportion of class i.

$$Gini_{split} = \sum\left(\frac{|D_i|}{|D|}.Gini(D_i)\right) \tag{2}$$

Explanation:
D = entire dataset before splitting.
D_i = i-th subset of the dataset after splitting.
|D| = number of data points in the initial dataset.
$|D_i|$ = number of data points in the i-th subset.
$|D_i|$ / |D| = proportion of data points in the i-th subset compared to the total data.
Gini(D_i) = Gini value in the i-th subset.

$$Gain = Gini_{parent} - Gini_{split} \tag{3}$$

Explanation:
Gain = the value of the increase in split quality.

Gini_{parent} = the Gini value before the split (parent node).
Gini_{split} = the Gini value after the split.

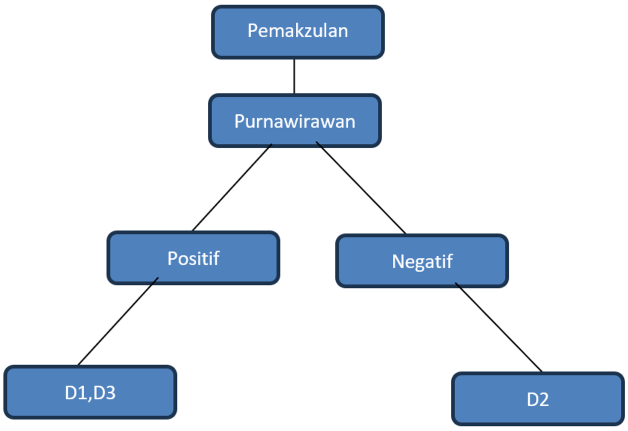


Figure 4. Simple CART Decision Tree

Showing the CART method decision tree that classifies documents based on the most influential words. The word “impeachment” was chosen as the root because it had the highest gain, making it the most effective at separating documents into Positive and Negative. Documents containing that word were then tested at the next node, “retired,” which was able to further differentiate the documents. At the leaf node, the final labels are determined: the left branch is Positive (D1 and D3), while the right branch is Negative (D2). Thus, CART classifies documents gradually from the root to the leaves based on distinguishing words, resulting in the following predictions: D1 → Positive, D2 → Negative, D3 → Positive.

Tabel 1 Hasil Uji Untuk Klasifikasi Decision Tree

Document	Tweet	Initial Label	Predicted Label
D1	Roy Suryo appointed as telecommunications expert by the Indonesian Armed Forces Retirees Forum in support of the impeachment of Vice President Gibran	Positif	Positif
D2	Indonesian Armed Forces Retirees Forum prepares options for the forced impeachment of Gibran, President Prabowo, the Indonesian House of Representatives, and the Indonesian Regional Representative Council	Negatif	Negatif
D3	Fact check: tens of thousands of Indonesian Armed Forces retirees stage a demonstration demanding the impeachment of Vice President Gibran, the Indonesian House of Representatives	Positif	Negatif

3.4. Evaluating Results

After completing the testing stage of the Decision Tree algorithm, a set of results is produced as the output of the model. These results take the form of labels assigned to the test data, which are generated based on the patterns learned by the model during the training process. In the training phase, the algorithm constructs a decision tree by identifying important features and decision rules that best separate the data into predefined sentiment categories. Once this learning process is complete, the trained model is then applied to previously unseen test data to predict their sentiment labels.

The classification results of the test data, expressed in the form of sentiment classes produced by the program, are subsequently compared with the actual or true class labels contained in the dataset. This comparison is essential to evaluate the performance of the Decision Tree model. Through this process, several standard evaluation metrics can be calculated, including accuracy, precision, recall, and f1-score. Accuracy measures the overall correctness of the model, while precision and recall provide more detailed insights into how well the model identifies specific sentiment classes. The f1-score, as a harmonic mean of precision and recall, offers a balanced measure of the model’s performance.

However, because the dataset used in this study exhibits an imbalance between positive and negative sentiment data, special attention is given to this condition during evaluation. Negative sentiment data are significantly more dominant than positive sentiment data, which can affect the interpretation of overall performance metrics. Therefore, the calculation and analysis of precision, recall, and f1-score are focused primarily on the negative sentiment class, as it represents the majority of the data and plays a crucial role in reflecting the model's effectiveness. To clearly illustrate the classification performance, a confusion matrix resulting from the sentiment analysis using the Decision Tree algorithm is presented, showing the distribution of correct and incorrect predictions across sentiment classes.

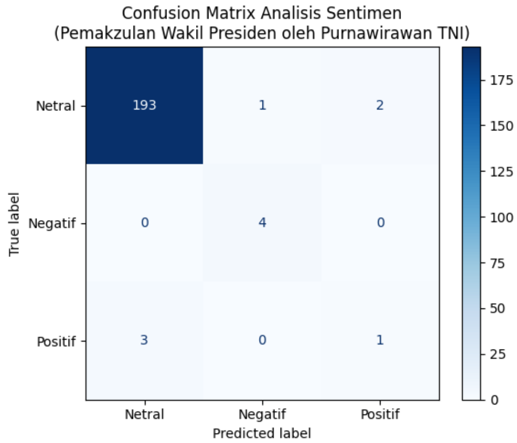


Figure 5. Confusion Matrix: Gibran's Impeachment

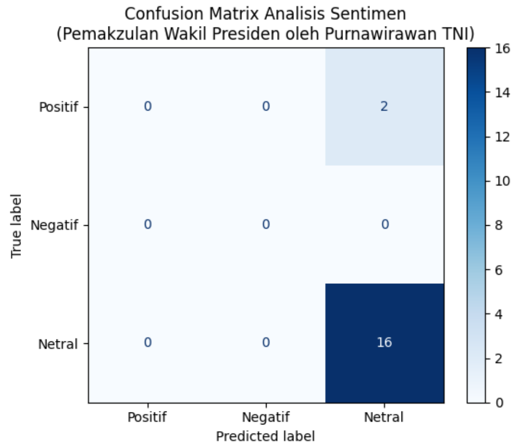
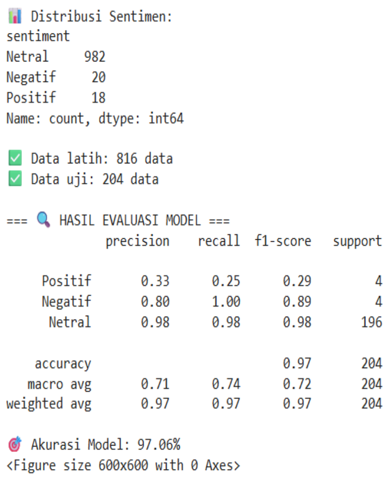


Figure 6. Confusion Matrix of Gibran's Pressure

Figures 4.10 and 4.11 show the classification results for accuracy, precision, recall, and F1-score.



```
📊 Distribusi Sentimen:
sentiment
Netral     982
Negatif     20
Positif     18
Name: count, dtype: int64

✅ Data latih: 816 data
✅ Data uji: 204 data

=== 🔍 HASIL EVALUASI MODEL ===
              precision    recall  f1-score   support

     Positif       0.33      0.25      0.29         4
     Negatif       0.80      1.00      0.89         4
      Netral       0.98      0.98      0.98       196

    accuracy                           0.97       204
   macro avg       0.71      0.74      0.72       204
weighted avg       0.97      0.97      0.97       204

🎯 Akurasi Model: 97.06%
<Figure size 600x600 with 0 Axes>
```

Figure 7. Model Evaluation Results

The image above shows the results of evaluating the sentiment classification model using data consisting of three categories, namely Neutral, Negative, and Positive. The data distribution is very unbalanced, with the Neutral class dominating with 982 data points, while the Negative class has only 20 and the Positive class has 18. The data was then divided into 816 training data points and 204 test data points. The evaluation results show that the model performs very well on the Neutral class with precision, recall, and f1-score of 0.98, respectively, due to the large amount of data. However, the performance on minority classes such as Positive and Negative is lower, especially the Positive class, which only has an f1-score of 0.29 due to the very small amount of data. Although the overall accuracy reached 97%, this value does not fully reflect the actual performance because it is influenced by the dominance of the Neutral class. This can be seen from the macro average, which only reached an f1-score of 0.72, which provides a fairer picture of the performance of each class. Overall, the model is very good at recognizing Neutral sentiment, but is still not optimal in classifying Positive and Negative sentiment.

4. Conclusion

This study conducted sentiment analysis on the X (Twitter) application related to the issues of impeachment and pressure on Gibran using the CART Decision Tree method. Data was collected through a crawling process, resulting in 1020 tweets on the issue of impeachment and 89 tweets on the issue of pressure. After preprocessing and removing duplicates, the data was labeled using the Lexicon-Based method based on polarity scores ranging from −1 to +1, which were then classified into positive, negative, and neutral sentiments. The CART Decision Tree model was trained and tested on both issues. For the impeachment issue, the model's accuracy ranged from 97.05% to 99.51%, with the best performance in the Neutral class and the lowest in the Positive class. For the Gibran pressure issue, the accuracy reached 88.89% with the highest results in the Neutral class. Overall, CART was able to classify sentiment well, especially in the Neutral class, which was dominant, while performance in the Positive and Negative classes tended to be low due to data imbalance. These results indicate that public sentiment on both issues was mostly Neutral, signifying a response that tended to be informative and unbiased.

References

[1]. O. Manullang, C. Prianto, and N. H. Harani, “Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest,” *J. Ilm. Inform.*, vol. 11, no. 02, pp. 159–169, 2023.

[2]. R. V. Alif, M. R. Alhafiz, and M. Fakhriza, “Perancangan Sistem Informasi Dokumen P44 (Hasil Putusan) Berbasis Web Studi Kasus Kejaksaan Negeri Deli Serdang,” *J. Informatics Busisnes*, vol. 1, no. 4, pp. 233–238, 2024.

[3]. A. Heryanto and R. Pramudita, “Opini Media Sosial Facebook Terhadap Produk Hijab Menggunakan Metode Text Mining,” *Inf. Syst. Educ. Prof. J. Inf. Syst.*, vol. 4, no. 2, pp. 168–177, 2020.

[4]. S. Suhendra and F. S. Pratiwi, “Peran komunikasi digital dalam pembentukan opini publik: Studi kasus media sosial,” in *Iapa Proceedings Conference*, 2024, pp. 293–315.

[5]. M. K. Gibran, M. I. Rifki, A. H. Hasugian, A. T. A. A. Siahaan, A. Sahputra, and R. Ong, “Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes,” *Instal J. Komput.*, vol. 16, no. 03, pp. 210–220, 2024.

[6]. R. N. Muhammad and B. Tanggahma, “Pengaruh media sosial pada persepsi publik terhadap sistem peradilan: analisis sentimen di Twitter,” *UNES Law Rev.*, vol. 7, no. 1, pp. 507–516, 2024.

[7]. R. A. Kurniawan, M. S. Hasibuan, P. Piramida, and R. S. Ramadhan, “Penerapan Algoritma K-Means Untuk Clustering Tempat Makan Di Batubara,” *J. Comput. Sci. Informatics Eng.*, vol. 1, no. 1, pp. 10–18, 2022.

[8]. E. Fauziningrum and E. I. Sulistyaningsih, “Penerapan Data Mining Metode Decision Tree Untuk Mengukur Penguasaan Bahasa Inggris Maritim (Studi Kasus Di Universitas Maritim Amni),” *J. SAINS DAN Teknol. Marit.*, vol. 22, no. 1, pp. 41–50, 2021.

[9]. Y. Akbar and T. Sugiharto, “Analisis Sentimen Pengguna Twitter di Indonesia Terhadap

ChatGPT Menggunakan Algoritma C4. 5 dan Naïve Bayes," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023.
[10]. I. D. Hardyatman and F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Rencana Kenaikan PPN 12% Di Indonesia Pada Media Sosial X Menggunakan Metode Decision Tree," *J. Inf. Syst. Res.*, vol. 6, no. 2, 2025.
[11]. N. Agustina and U. A. Yogyakarta, "ALGORITMA DECISION TREE UNTUK ANALISIS SENTIMEN PUBLIC TERHADAP ALGORITMA DECISION TREE UNTUK ANALISIS SENTIMEN PUBLIC TERHADAP MARKETPLACE DI," no. June 2023, 2024, doi: 10.53580/naratif.v5i1.186.
[12]. S. K. Wardani, Y. A. Sari, and I. Indriati, "Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5582–5590, 2021.
[13]. D. Rifaldi and A. Fadlil, "Teknik preprocessing pada text mining menggunakan data tweet 'mental health,'" *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023.
[14]. E. Novalia and A. Voutama, "Black Box Testing dengan Teknik Equivalence Partitions Pada Aplikasi Android M-Magazine Mading Sekolah," *Syntax J. Inform.*, vol. 11, pp. 23–35, Jun. 2022, doi: 10.35706/syji.v11i01.6413.
[15]. E. Oktafanda, N. W. Al-Hafiz, A. Latif, and F. Santosa, "Analysis and Design of Monolithic System Architecture Migration to Microservices at PT. MALINDO Conceptual Approach", *JTOS*, vol. 8, no. 1, pp. 54 - 63, May 2025.
[16]. R. Hardiartama, A. A. Arifiyanti, and Ana Wati3S. F., "Application of Ensemble Machine Learning Methods for Aspect-Based Sentiment Analysis on User Reviews of the Wondr by BNI App", *JTOS*, vol. 8, no. 1, pp. 97 - 111, Jun. 2025.
[17]. Mirza Khazim Nugraha, Nur Cahyo Wibowo, and Iqbal Ramadhani Mukhlis, "Design and Development of the Boarding House Management Information System (SIMKO) Using Laravel with Agile Methodology", *JTOS*, vol. 8, no. 1, pp. 112 - 122, Jun. 2025.
[18]. M. A. Sifaul Anam, A. Faroqi, A. Faroqi, and T. Lathif Mardi Suryanto, "Analysis of MyPertamina Application Acceptance Using a Modified Technology Acceptance Model (TAM)", *JTOS*, vol. 8, no. 1, pp. 133 - 141, Jun. 2025.
[19]. D. Handoko Putra and N. Rahditiantoro, "Implementation of Forward Chaining in a Web-Based Expert System for Passport Services", *JTOS*, vol. 8, no. 1, pp. 217 - 277, Jun. 2025.
[20]. R. A. P. Alisia, Anita Wulansari, and Rafika Rahmawati, "Analysis of Factors Influencing the Acceptance of the Indodax Application Using the UTAUT 2 Model", *JTOS*, vol. 8, no. 1, pp. 317 - 327, Jun. 2025.
[21]. H. Nopriandi, N. W. Al-Hafiz, and S. Chairani, "Analysis and Modeling of the Internal Quality Audit Information System Islamic University of Kuantan Singingi", *JTOS*, vol. 8, no. 1, pp. 398 - 408, Jun. 2025.
[22]. M. Yusfahmi, Febri Haswan, Nofri Wandi Al-Hafiz, Elgamar Syam, Helpi Nopriandi, Jasri, Aprizal, Harianja, Erlinda, Sri Chairani, Gunardi Hamzah, & Morine Delya Octa. (2025). SOSIALISASI DAN PENERAPAN APLIKASI BERBASIS TEKNOLOGI INFORMASI UNTUK MENDUKUNG TRANSFORMASI DIGITAL BUMDes TEBING TINGGI. BHAKTI NAGORI (Jurnal Pengabdian Kepada Masyarakat), 5(2), 712 - 719. https://doi.org/10.36378/bhakti_nagori.v5i2.4910
[23]. R. Nazli, A. Amrizal, H. Hendra, and S. Syukriadi, "Modeling User Interface Design E-Business Applications for Marketing Umkm Products in Payakumbuh City Using Pieces

Framework", *JTOS*, vol. 7, no. 2, pp. 55 - 66, Nov. 2024.

[24]. M. Syamsul Azis and H. Basri, "Building A Powerfull File Specification Application For Database System Design", *JTOS*, vol. 7, no. 2, pp. 103 - 109, Dec. 2024.

[25]. Nofri Wandi Al-Hafiz, Helpi Nopriandi, and Harianja, "Design of Rainfall Intensity Measuring Instrument Using IoT-Based Microcontroller", *JTOS*, vol. 7, no. 2, pp. 202 - 211, Dec. 2024.