Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Analyzing Public Sentiment on the Proposal to Return Regional Head Elections to DPRD on Platform X Using the C4.5 Algorithm

Ade Novia Maulana[1], Wan Mohd Yusoff bin Wan Yaacob[2], Fatima Felawati[3]

[1,3]Department of Information Systems, UIN Sulthan Thaha Saifuddin Jambi, Jambi, Indonesia
[2]Department of Information and Communication Technology, Politeknik Tuanku Syed Sirajuddin, Perlis, Malaysia

Received:
October 4, 2025
Revised:
March 10, 2026
Accepted:
March 26, 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Ade Novia Maulana
Email*:
ade@uinjambi.ac.id

DOI:
10.63158/journalisi.v8i2.1483

Abstract. This study examines public sentiment among X users toward the proposal to return regional head elections (Pilkada) to an indirect electoral mechanism through the Regional People’s Representative Council (DPRD), using a decision-tree classifier based on the C4.5 approach. A dataset of 4,127 tweets collected via X API v2 between December 2024 and January 2026 was analyzed using a seven-stage text preprocessing pipeline. Sentiment labels were generated through a hybrid lexicon-based approach, followed by manual verification of 500 stratified tweets by two independent annotators, yielding substantial inter-annotator agreement (Cohen’s Kappa = 0.78). TF-IDF was used for feature extraction, and the dataset was divided using an 80:20 stratified train-test split. The classifier achieved 81% accuracy, 82% precision, 79% recall, and an F1-score of 80%, outperforming Naive Bayes (74%) and Support Vector Machine (79%) baselines on the same dataset. The sentiment distribution showed that 45% of tweets were negative, 32% were positive, and 23% were neutral, indicating a predominantly critical response among X users toward the proposal. These findings describe discourse on X during the study period and should not be interpreted as representative of broader public opinion. Overall, the study highlights the potential of machine learning methods for analyzing Indonesian political discourse on social media.

Keywords: C4.5 algorithm, sentiment analysis, political communication, social media X, TF-IDF, machine learning classification

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

1. INTRODUCTION

The implementation of direct regional head elections (Pilkada) in Indonesia since 2005 represented a significant advance in the country's democratization after the reform era. This system replaced the previous mechanism where DPRD elected regional heads, giving citizens direct sovereignty in choosing their leaders [1]. On 12 December 2024, however, a controversial proposal emerged to revert Pilkada to indirect elections through DPRD, sparking intense national debate that continued into 2026.

The proposal gained prominence during Golkar Party's 60th anniversary at Sentul International Convention Center on 12 December 2024, when President Prabowo Subianto directly proposed returning Pilkada to DPRD selection, arguing that indirect elections could substantially reduce state expenditures and candidate political costs [2]. Golkar Party Chairman Bahlil Lahadalia strongly supported the proposal, followed by endorsements from multiple coalition parties. By 2026, Koalisi Merah Putih members including Democratic Party expressed support, while PDIP firmly opposed. Indonesia Corruption Watch notes that Pilkada 2024 cost approximately Rp 37 trillion in regional budget grants, compared to Rp 71.3 trillion for the 2024 national elections and Rp 71 trillion for the free nutritious meal program in 2025 [3].

Proponents argue that indirect elections offer greater efficiency, cost savings, and political stability [4]. Critics counter that the proposal threatens democratic principles, eliminates independent candidates, increases corruption potential, and weakens people's sovereignty [5]. Historical precedent from 2014 shows similar attempts were met with widespread public opposition, prompting President Susilo Bambang Yudhoyono to issue emergency regulations (Perppu) restoring direct elections. Recent polling by LSI Denny JA in January 2026 indicates that 71 percent of Prabowo-Gibran voters oppose indirect Pilkada, preferring direct elections, reflecting significant disconnect between elite party support and grassroots sentiment [6].

Social media platform X, with over 23,76 million active Indonesian users, serves as a critical arena for political discourse and real-time public opinion measurement [7]. Understanding sentiment patterns on the indirect Pilkada discourse is essential for policymakers, researchers, and stakeholders to gauge public acceptance of such fundamental democratic changes. Sentiment analysis using machine learning provides systematic tools to quantify opinions expressed in social media texts [8].

Previous studies demonstrate the effectiveness of the C4.5 algorithm in sentiment classification found C4.5 achieved 71, 96% accuracy compared to Naïve Bayes at 66,11 % [9]. The C4.5 algorithm outperformed Naïve Bayes with 90% versus 85% accuracy [10]. The C4.5 algorithm is an effective algorithm for sentiment classification, with evidence of high accuracy in a number of studies and advantages in model transparency.

This research addresses a critical gap in understanding public sentiment toward proposed changes in Indonesia's electoral system. While previous studies analyzed sentiment on various political topics [11], [12], [13], limited research has examined public opinion on the DPRD-based indirect Pilkada proposal using rigorous machine learning. Prior Indonesian political sentiment studies have addressed presidential elections, party preferences, and general electoral discourse [11], [12], [13]. To the best of the authors' knowledge, no prior study has applied a decision-tree classifier with documented inter-annotator validation specifically to the DPRD indirect Pilkada debate on platform X, and this gap motivates the present work.

This study makes three contributions. First, it applies a decision-tree classifier approximating C4.5 to classify X user sentiment toward the DPRD indirect Pilkada proposal, with documented preprocessing and hybrid labeling procedures including inter-annotator agreement (Cohen's Kappa = 0.78). Second, it benchmarks this classifier against Naive Bayes and SVM on the same dataset and feature space. Third, it situates the quantitative findings within the political communication literature on Indonesian electoral reform. The primary objective is to evaluate the C4.5 approximation for Indonesian political sentiment classification and to provide an empirically grounded account of X user opinion on this specific policy proposal.

2. METHODS

This research employs a quantitative approach using machine learning for sentiment analysis. The methodology follows the Knowledge Discovery in Database (KDD) framework: data collection, preprocessing, feature extraction, model development, evaluation, and visualization [14].

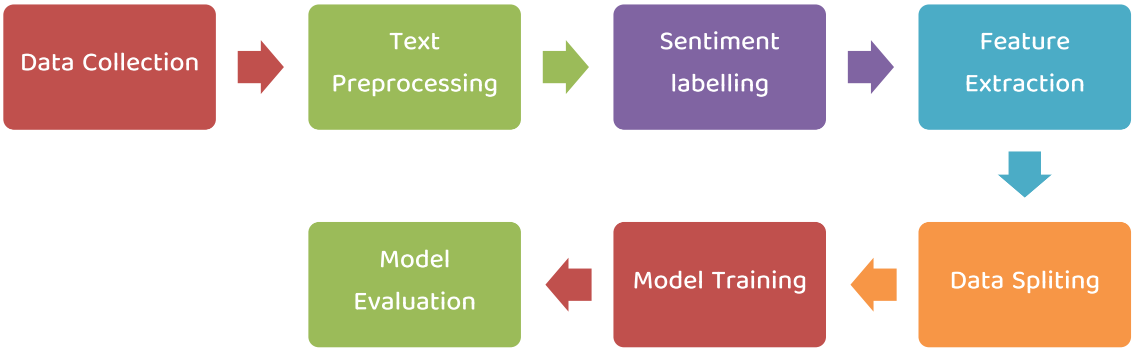

Figure 1. Research Procedure

2.1. Data Collection

Data were collected from platform X using X API v2 during 12 December 2024 to January 2026. Keywords included #PilkadaTidakLangsung, #PilkadaDPRD, #PilkadaLewatDPRD, "Pilkada dipilih DPRD", "kepala daerah dipilih DPRD", and related terms. Data extraction captured tweet text, timestamp, user location, retweet count, and like count [15]. Regarding data inclusion and exclusion: original tweets and quoted tweets were retained as they contain original opinion content; pure retweets (prefixed with RT @) and reply threads were excluded to avoid content duplication and to ensure each record reflects an independently authored expression. After removing duplicates, spam, and irrelevant posts under these criteria, 4,127 unique tweets were obtained for analysis and stored in CSV format.

2.2. Text Preprocessing

Text preprocessing was conducted to clean raw social media data and improve classification accuracy [16]. In this study, the preprocessing stage followed a seven-step pipeline designed to transform unstructured text into a more consistent and analyzable form. This process is essential in sentiment classification because social media texts often contain noise, informal expressions, and inconsistent formatting that can reduce model performance if left untreated. The first step was case folding, in which all text was converted to lowercase. This was done to ensure consistency across the dataset and to reduce vocabulary duplication caused by differences in letter casing. For example, words such as "Pilkada," "pilkada," and "PILKADA" were treated as the same term after normalization.

The second step involved URL and mention removal. Hyperlinks and @username mentions were eliminated because they generally do not contribute meaningful sentiment information and may introduce unnecessary noise into the dataset. Removing these elements helped the model focus more directly on the textual content that reflects user opinions. The third step was noise removal, which included deleting punctuation marks, numbers, and special symbols that did not carry semantic value for sentiment analysis. Since social media posts often contain excessive non-alphabetic characters, this stage was necessary to simplify the text and improve feature quality. Next, the cleaned text underwent tokenization, where each sentence or post was split into individual word tokens. Tokenization enabled the text to be processed at the word level, which is required for subsequent steps such as stopword removal, stemming, and feature extraction.

The fifth stage was stopword removal, in which common Indonesian words with little or no contribution to sentiment were removed. These included highly frequent function words that do not meaningfully distinguish between positive, negative, or neutral expressions. Eliminating stopwords helped reduce dimensionality and improve the relevance of the retained terms. The sixth step was stemming, which reduced words to their root forms using the Sastrawi library. This process helped group different variations of a word under a single base form, thereby reducing redundancy in the dataset and improving consistency in feature representation. The final step was lemmatization, which normalized words into their dictionary base forms to reduce inflectional variation. While similar to stemming, lemmatization aimed to produce linguistically valid root forms, thereby improving the semantic consistency of the processed text. Together, these seven preprocessing stages enhanced the quality of the textual data and prepared it for more effective sentiment classification.

2.3. Sentiment Labelling

Sentiment labelling used a hybrid approach combining lexicon-based automatic labelling with manual verification [17] . A comprehensive Indonesian sentiment lexicon containing positive and negative word lists with polarity scores was utilized [18]. Tweets received sentiment scores based on cumulative word polarity [19]. Positive scores indicated positive sentiment, negative scores negative sentiment, and near-zero scores neutral sentiment. Manual verification was performed by two independent annotators on a stratified random sample of 500 tweets (12% of the dataset), selected to represent proportional sentiment class distribution. Annotator 1 is a researcher with expertise in Indonesian political discourse; Annotator 2 is a computer science researcher. Both annotators labeled each tweet independently without access to the lexicon-generated labels. Inter-annotator agreement was measured using Cohen's Kappa ($\kappa = 0.78$), indicating substantial agreement and confirming labeling reliability. Disagreements ($n = 43$) were resolved through adjudication by a third expert. Context-dependent expressions, sarcasm, and colloquial Pilkada-related terms were documented and used to update the domain-specific lexicon. It is important to note that the lexicon-based approach serves as an initialization mechanism, with human-verified annotations constituting the ground-truth labels used for model training and evaluation.

2.4. Feature Extraction Using TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) converted pre-processed text into numerical feature vectors. TF-IDF assigns word weights based on document frequency (TF) inversely proportional to corpus frequency (IDF), highlighting distinctive terms while reducing common word importance. The following configuration was used: max_features = 5,000, min_df = 2, max_df = 0.95, sublinear_tf = True, analyzer = 'word', ngram_range = (1, 2). Unigrams and bigrams were included to capture negation patterns and common political phrases characteristic of Indonesian Pilkada discourse [20]. The Formulas**:**

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

$$\text{TF-IDF}(t, d) = \frac{ft,\ d}{Nd} \quad (2)$$

$$\text{IDF}(t) = \log \frac{N}{Ft} \quad (3)$$

Where, TF (Term Frequency) : The frequency of words that appear in a particular document. IDF (Inverse Document Frequency) : How rarely the word appears in each document in the corpus. N : Number of documents in the corpus.

2.5. Data Splitting

In the data splitting stage, the dataset is divided into training data (80%) and testing data (20%). The training data is used to build the classification model, while the testing data evaluates the model's performance. Previous studies have shown that an 80:20 ratio yields the highest accuracy when tested using a confusion matrix for three sentiments (negative, positive, and neutral) [21].

2.6. C4.5 Algorithm Implementation

This study uses a decision-tree classifier approximating C4.5 characteristics through entropy-based splitting. The classifier constructs decision trees through iterative dataset splitting, calculating information gain for each attribute and selecting the highest-gain attribute as the splitting criterion. The process continues recursively until reaching stopping conditions such as pure leaf nodes or minimum sample thresholds. As noted, the implementation relies on scikit-learn's DecisionTreeClassifier with entropy criterion, which approximates C4.5 in its information-theoretic splitting behavior but does not implement gain ratio or post-pruning as in Quinlan's original formulation; this distinction is acknowledged as a study limitation [22]. The entropy formula:

$$\text{Entropy (S)} = \sum_{t=1}^{n} -pi*\log2pi \tag{4}$$

Where:

S : Represents a collection of text documents

N : The number of sentiment classes

pi : Proportion of documents belonging to the sentiment class

$$\text{Gain (S, A)} = \text{Entropy (S)} = \sum_{i=1}^{n} \frac{|Si|}{|S|} *\text{Entropy (Si)} \tag{5}$$

Where

A : Refers to a feature or attribute extracted from text

S_i : Represents subsets of documents formed after splitting by attribute A

$| S_i |/|S|$: The proportion of documents in subset S_i

The dataset was split 80% training (3,302 tweets) and 20% testing (825 tweets) using stratified random sampling to maintain proportional sentiment class representation in both subsets. A single stratified 80:20 split was adopted to replicate the evaluation design used in the most directly comparable prior Indonesian political NLP studies on datasets of similar size, enabling consistent external benchmarking; the trade-off against cross-validation is acknowledged as a limitation. The decision-tree classifier was implemented using scikit-learn's DecisionTreeClassifier with the following hyperparameters: criterion = 'entropy', max_depth = None (unpruned), min_samples_split = 2, random_state = 42. For baseline comparison, Naive Bayes (MultinomialNB, default settings) and Support Vector Machine (LinearSVC, C = 1.0) were trained and evaluated on the identical 825-tweet test set produced from the same stratified split, using the same TF-IDF feature matrix. This ensures that performance differences between classifiers reflect model behavior rather than differences in data or feature configuration.

2.7. Evaluation Metrics

The model's performance was evaluated using standard classification metrics: accuracy, precision, recall, and F1-score, calculated from the confusion matrix on test data [23].

1) Accuracy Formula

Accuracy measures how well the model correctly classifies data, with higher values indicating better model performance. However, accuracy is less effective on imbalanced datasets as it does not reflect the model's ability to handle minority classes. The formula as shown in Equation 6.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Where TP (True Positive): Text correctly classified as positive sentiment, TN (True Negative): Text correctly classified as negative sentiment, FP (False Positive): Text incorrectly classified as positive, and FN (False Negative): Text incorrectly classified as negative.

2) Precision

Precision measures the proportion of correctly predicted positive cases out of all predicted positive cases. This metric evaluates how accurately the model classifies positive data. In sentiment classification, precision measures how many tweets classified as positive truly have positive sentiment. The formula as shown in Equation 7.

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

$$\text{Precision}=\frac{TP}{TP+FP} \tag{7}$$

Where TP (True Positive): Text correctly classified as positive sentiment and FP (False Positive): Text incorrectly classified as positive.

3) Recall

Recall evaluates the model's ability to identify actual positive data. This metric is crucial to determine how much positive data is correctly identified, even if the model produces incorrect negative predictions. The formula as shown in Equation 8.

$$\text{Recall}=\frac{TP}{TP+FN} \tag{8}$$

Where TP (True Positive): Text correctly classified as positive sentiment and FN (False Negative): Text incorrectly classified as negative

4) F1-Score

F1-Score Formula is a metric that describes the balance between precision and recall. In classification tasks, the F1-score demonstrates how effectively the model combines precision and recall, providing an overview of its ability to classify opinions on Twitter. The formula as shown in Equation 9.

$$F1=2\ \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \tag{9}$$

Where, Precision: Measures how many texts predicted as positive sentiment are truly positive. Recall : Measures how many texts that truly have positive sentiment are correctly identified.

2.8. Visualization

Result Visualization is the final stage of the research, where model evaluation outcomes are presented using graphs, tables, or diagrams. This visualization aims to simplify the understanding and interpretation of the analysis results [24] [25].

3. RESULTS AND DISCUSSION

3.1. Data Collection

The 4,127 collected tweets underwent sentiment labeling, producing the distribution shown in Table 1. The dataset reveals a predominance of negative sentiment (45%), followed by positive (32%) and neutral (23%). This skewed distribution is consistent with patterns documented in studies of social media reactions to top-down policy proposals that affect established democratic rights [14], [15]. The relatively high proportion of negative sentiment suggests that, within the X user community, the proposal was broadly perceived as a regression in democratic participation. This aligns with the historical precedent of 2014, when a similar indirect election bill triggered widespread protests, ultimately prompting President Susilo Bambang Yudhoyono to issue an emergency regulation restoring direct elections.

Table 1. Sentiment Distribution in Dataset

No	Sentiment Class	Numbers of Tweets	Percentage (%)
1	Positive	1321	32
2	Negative	1857	45
3	Neutral	949	23
4	Total	4127	100

The dataset was divided 80% training and 20% testing as detailed in Table 2. This maintains proportional sentiment representation in both subsets for robust training and unbiased evaluation.

Table 2. Training and Testing Data Distribution

No	Data Type	Positive	Negative	Neutral	Total
1	Training (80%)	1057	1486	759	3302
2	Testing (20%)	264	371	190	825

3.2. C4.5 Algorithm Classification

The C4.5 algorithm demonstrated strong performance classifying sentiments toward the indirect Pilkada discourse. Table 3 presents detailed performance metrics for each

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

sentiment class and overall model performance, while Figure 3 visualizes the comparative performance across different metrics.

Table 3. C4.5 Algorithm Classification Performance

No	Class	Precision	Recall	F1-Score
1	Positive	0,85	0,80	0,82
2	Negative	0,81	0,84	0,82
3	Neutral	0,78	0,75	0,76
4	Accuracy			0,81
5	Weighted Avg	0,82	0,79	0,80

Results indicate C4.5 achieved 81% overall accuracy with weighted average precision of 82%, recall of 79%, and F1-score of 80%. Positive sentiment classification showed highest precision (0.85), suggesting the model effectively identified positive tweets with minimal false positives. Negative sentiment achieved highest recall (0.84), indicating strong capability to capture actual negative sentiments. Neutral sentiment showed slightly lower performance (precision 0.78, recall 0.75), possibly due to ambiguous expressions or context-dependent language requiring deeper semantic understanding.

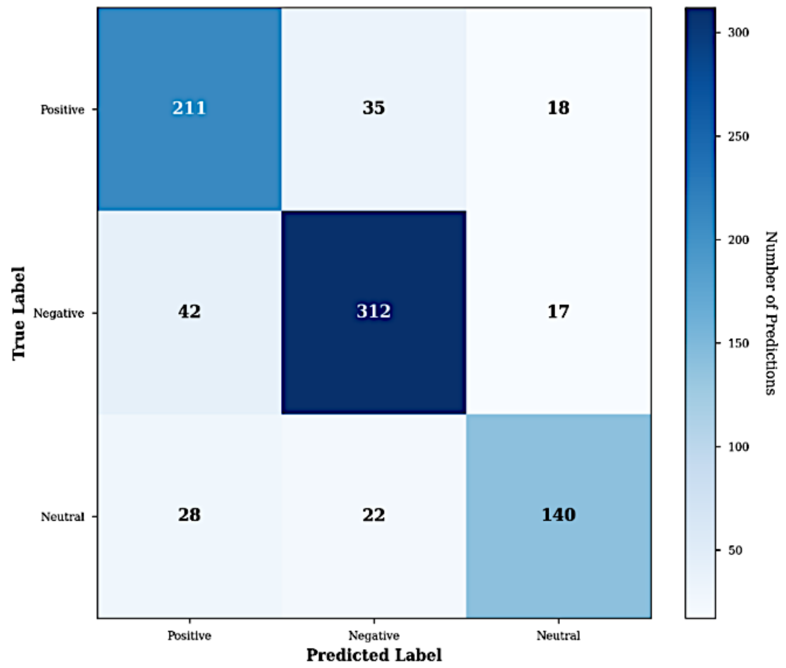

Figure 2. Confusion Matrix

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

The confusion matrix presented in Figure 2 provides detailed insight into the model's classification behavior across sentiment classes. The diagonal elements represent correct classifications, while off-diagonal elements indicate misclassifications. The matrix reveals that the model performs best on negative sentiment classification (312 out of 371 correctly classified), followed by positive sentiment (211 out of 264), and neutral sentiment (140 out of 190).

These results compare favorably with previous sentiment analysis studies, as illustrated in Figure 3. Hidayat et al [9] reported C4.5 accuracy of 71.96%, while this study achieved 81%, suggesting improved preprocessing and feature extraction techniques. The performance aligns with Rahmayanti et al [10] 90% accuracy on different datasets, validating the decision-tree classifier's effectiveness for Indonesian political sentiment analysis. The relatively balanced performance across sentiment classes indicates reasonable model generalization despite moderate class imbalance (45% negative, 32% positive, 23% neutral). These performance figures should be interpreted with appropriate caution, given that training labels were initialized through a semi-automatic lexicon-based approach; however, the manual verification and inter-annotator agreement (κ = 0.78) provide reasonable confidence in label quality. For the internal baseline comparison, all three classifiers (decision-tree approximation, Naive Bayes, SVM) were evaluated on the identical 825-tweet test set from the same stratified 80:20 split, using the same TF-IDF feature matrix, ensuring that performance differences reflect model behavior rather than data or feature differences. On this basis, Naive Bayes achieved 74% accuracy and SVM achieved 79%, confirming that the decision-tree classifier (81%) outperforms both baselines. It should be noted that comparisons with prior published studies in Figure 3 involve different datasets and are presented for contextual reference only, not direct equivalence. The relatively lower neutral class recall across all three models reflects a well-documented challenge in bag-of-words political sentiment classification: neutral tweets in this corpus tend to contain factual reporting, hedged language, or irony without clear lexical polarity markers, making their boundary with mildly negative or positive expressions difficult to distinguish. Alfina et al. [16] and Koto et al. [17] document the same neutral-class performance deficit in Indonesian political NLP tasks, attributing it to the high proportion of context-dependent and figurative language in social media political discourse. Among X users in this dataset, 45% of tweets toward the DPRD indirect election proposal expressed negative sentiment, a distribution that is consistent with

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

patterns observed in social media responses to top-down policy proposals affecting established democratic rights. This finding aligns with political communication research on Indonesian electoral discourse, which shows that proposals perceived as elite-driven tend to generate concentrated critical expression on social media. It is emphasized that these findings describe discourse patterns among X users in the observed period and should not be interpreted as a representative measure of the general Indonesian public's views. The neutral class (23%), comprising informational tweets and ambivalent commentary, reflects the tendency of users in polarized political topics to engage with content without committing to an evaluative stance.

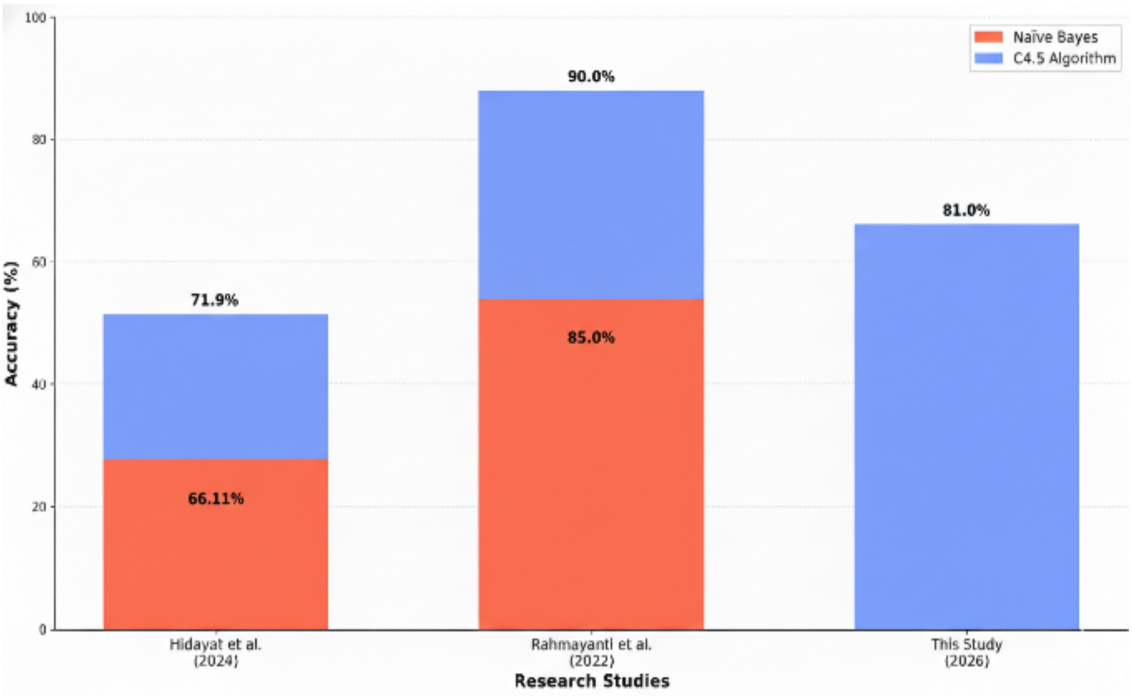

Figure 3. Accuracy Comparison with Previous Study

3.3. Discussion

This study demonstrates that a decision-tree classifier approximating C4.5 can be effectively applied to political sentiment analysis in the context of Indonesian social media. By combining a hybrid lexicon-based labeling strategy, manual validation by independent annotators, and TF-IDF feature extraction, the study offers a practical and transparent analytical pipeline for examining time-sensitive political discourse. This approach is especially valuable in situations where fully manual annotation is difficult to implement due to time and resource constraints. The inclusion of inter-annotator verification strengthens the reliability of the labeling process and addresses a common limitation of purely lexicon-based methods, which often lack sufficient quality control when dealing with informal and context-dependent social media language.

The sentiment distribution identified in this study shows that negative sentiment was the most dominant category (45%), indicating substantial concern among X users regarding the proposal to return Pilkada to an indirect electoral mechanism through the DPRD. Negative tweets in the dataset were largely associated with themes of democratic backsliding, the elimination of independent candidates, the potential expansion of corrupt political bargaining, the weakening of popular sovereignty, and repeated references to the 2014 indirect-election precedent. This pattern suggests that opposition on X was not merely emotional or reactive, but often framed in explicitly democratic and institutional terms. In contrast, positive sentiment (32%) was mainly linked to arguments about cost efficiency, the possibility of greater political stability through reduced campaign tensions, and the perceived benefit of strengthening party system consolidation. Meanwhile, neutral sentiment (23%) consisted primarily of informational posts, news-sharing activity, and tweets that lacked a clearly evaluative stance. Taken together, these results indicate that discourse on X during the observed period was more critical than supportive, suggesting that the proposal encountered a notable degree of legitimacy resistance within this platform's public conversation.

This interpretation gains additional context when read alongside survey evidence referenced in the study, including LSI Denny JA polling from January 2026, which reportedly found that 71% of Prabowo–Gibran voters opposed indirect elections. Although polling data and social media discourse capture different dimensions of public response, the similarity in direction between these two sources strengthens the argument that resistance to the proposal was not limited to isolated online reactions. At the same time, this study does not claim that X discourse is equivalent to broader public opinion. Social media users are self-selecting, platform demographics are uneven, and public posting behavior may amplify more polarized or politically engaged voices. For that reason, the findings should be interpreted as a description of platform-specific discourse dynamics, not as a direct measure of nationwide sentiment. Even so, the results remain meaningful for policymakers and political communicators because online

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

discourse increasingly shapes issue framing, agenda visibility, and the perceived legitimacy of institutional reform proposals.

From a modeling perspective, the decision-tree approximation produced better performance than both Naive Bayes and SVM on the same dataset, supporting earlier findings by Hidayat et al. [9] and Rahmayanti et al. [10], who also reported competitive results for decision-tree methods in Indonesian text-classification tasks. This suggests that interpretable classifiers still have practical value in political sentiment analysis, particularly when the goal is not only prediction but also methodological transparency. However, the relatively weaker performance in the neutral class points to an important limitation. Neutral expressions on social media are often subtle, context-dependent, or mixed in tone, making them difficult to capture through bag-of-words-style representations alone. This challenge is consistent with observations from Koto et al. [17] and Alfina et al. [16] in related Indonesian NLP studies. A logical extension of this work would therefore be the use of contextual embedding models such as IndoBERT, which are better suited to capturing semantic nuance, contextual ambiguity, and implicit evaluative language. Future research could also improve robustness by expanding the annotation sample, testing additional ensemble methods, and comparing platform-specific discourse patterns across multiple social media environments.

This study demonstrates the applicability of a decision-tree classifier approximating C4.5 for political sentiment analysis on Indonesian social media, contributing a documented pipeline that combines hybrid labeling, inter-annotator validation, and TF-IDF feature extraction. The methodology proves practical for time-sensitive political discourse where fully human-annotated datasets are not feasible, while the verification procedure provides a quality baseline that lexicon-only approaches lack. The negative sentiment predominance among X users suggests that the indirect Pilkada proposal faces legitimacy resistance within this platform's discourse, a pattern that policymakers and democratic researchers may find relevant when assessing the public communication dimensions of electoral reform. These findings should be read alongside existing survey data rather than as standalone indicators of public opinion. The decision-tree approximation outperformed both Naive Bayes and SVM baselines, a result consistent with Hidayat et al. [9] and Rahmayanti et al. [10] who find decision-tree classifiers competitive in

Indonesian text classification tasks. The lower neutral-class performance is a shared limitation across all three models in this study, and bag-of-words representations in general, which is consistent with findings from Koto et al. [17] and Alfina et al. [16] in related Indonesian NLP benchmarks. Addressing this limitation through contextual embeddings such as IndoBERT is recommended as a direct extension of this work.

4. CONCLUSION

This study applied a decision-tree classifier approximating C4.5 to classify public sentiment among X users toward the proposal to return regional head elections (Pilkada) to indirect elections through the DPRD, using 4,127 tweets collected between December 2024 and January 2026. The pipeline incorporated hybrid lexicon-based labeling with inter-annotator verification (Cohen's Kappa = 0.78) and TF-IDF feature extraction, yielding 81% accuracy and outperforming Naive Bayes (74%) and SVM (79%) baselines on the same dataset. Sentiment distribution revealed 45% negative, 32% positive, and 23% neutral among X users in this period, reflecting a critical stance toward the proposal within this platform's discourse; these findings describe X user opinion specifically and should not be generalized to the broader Indonesian electorate. Future work may apply transformer-based models such as IndoBERT across multiple platforms to extend coverage and improve contextual classification depth.

REFERENCES

[1] R. Lubis, N. R. Yunus, R. Herlina, and A. P. Shapiulayyevna, "Political law means that the election of regional heads is democratically elected: Politik hukum memaknai pemilihan kepala daerah dipilih secara demokratis," *DJL*, vol. 1, no. 1, pp. 65–79, Mar. 2025, doi: 10.64344/djl.v1i1.17.

[2] A. Gyarani, J. Rifqi, and A. Putranto, "Peranan komunikasi publik pada partai pengusung Pilkada 2024 Kabupaten Bogor (studi terhadap pasangan No. 1 Rudy Susmanto dan Ade Ruhandi)," *JIMIK*, vol. 6, no. 1, pp. 706–718, Jan. 2025, doi: 10.35870/jimik.v6i1.1296.

[3] Indonesia Corruption Watch, "Demokrasi tidak boleh ditarik mundur: Tolak wacana pengembalian Pilkada oleh DPRD!," Dec. 30, 2025. [Online]. Available: https://antikorupsi.org

[4] R. Riyanti, “Maintaining direct regional head elections, solutions or challenges?,” *JOG Untirta*, vol. 7, no. 3, Sep. 2022, doi: 10.31506/jog.v7i3.16325.

[5] T. Hidayat and L. Fitrianingrum, “Election of regional head based on hierarchy: Strengthening argumentation of the discussion of abolishing direct elections in Indonesia and proposed models,” *CLS*, vol. 1, no. 2, pp. 140–152, Sep. 2022, doi: 10.36448/cls.v1i2.27.

[6] Tempo, “Survei: Mayoritas pemilih Prabowo tolak Pilkada lewat DPRD,” Dec. 30, 2025. [Online]. Available: https://www.tempo.co

[7] F. Panjaitan, W. Ce, H. Oktafiandi, G. Kanugrahan, Y. Ramdhani, and V. H. C. Putra, “Evaluation of machine learning models for sentiment analysis in the South Sumatra governor election using data balancing techniques,” *JournalISI*, vol. 7, no. 1, pp. 461–478, Mar. 2025, doi: 10.51519/journalisi.v7i1.1019.

[8] Moch. R. K. Muhaemin, Fitriyani, and L. S. Darfiansa, “Analysis of public sentiment regarding the 2024 Jakarta election on platform X using deep learning,” *IJOICT*, vol. 11, no. 1, pp. 13–25, Jun. 2025, doi: 10.21108/ijoict.v11i1.9176.

[9] R. Hidayat, A. B. Hakim, and R. Nugraha, “Perbandingan metode Naïve Bayes dan decision tree C4.5 untuk analisis sentimen produk Es Teh Indonesia di media sosial Twitter,” *SISKOM-KB*, vol. 7, no. 2, pp. 88–98, Mar. 2024, doi: 10.47970/siskom-kb.v7i1.537.

[10] A. Rahmayanti, L. Rusdiana, and S. Suratno, “Perbandingan metode algoritma C4.5 dan Naïve Bayes untuk memprediksi kelulusan mahasiswa,” *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, Aug. 2022, doi: 10.21580/wjit.2022.4.1.9654.

[11] A. S. Hanin and M. Maryam, “Sentiment analysis of Twitter towards the free lunch program using the C4.5 algorithm,” *Int. J. Adv. Data Inf. Syst.*, vol. 6, no. 1, pp. 31–45, Apr. 2025, doi: 10.59395/ijadis.v6i1.1357.

[12] S. Puad, G. Garno, and A. Susilo Yuda Irawan, “Analisis sentimen masyarakat pada Twitter terhadap pemilihan umum 2024 menggunakan algoritma Naïve Bayes,” *JATI*, vol. 7, no. 3, pp. 1560–1566, Oct. 2023, doi: 10.36040/jati.v7i3.6920.

[13] T. N. Wijaya, R. Indriati, and M. N. Muzaki, “Analisis sentimen opini publik tentang undang-undang Cipta Kerja pada Twitter,” *JJEEE*, vol. 3, no. 2, pp. 78–83, Jul. 2021, doi: 10.37905/jjeee.v3i2.10885.

[14] N. Nabiilah, S. Rohimah, and S. K. P. Loka, “Analyzing customer sentiment towards marketplace reviews using classification algorithms,” *Int. J. Adv. Technol. Inf. Sci.*, vol. 2, no. 1, pp. 23–30, Feb. 2025, doi: 10.57152/ijatis.v2i1.1774.

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

[15] V. C. Storey and D. E. O'Leary, "Text analysis of evolving emotions and sentiments in COVID-19 Twitter communication," *Cogn. Comput.*, vol. 16, no. 4, pp. 1834–1857, Jul. 2024, doi: 10.1007/s12559-022-10025-3.

[16] S. Khairunnisa, A. Adiwijaya, and S. A. Faraby, "Pengaruh text preprocessing terhadap analisis sentimen komentar masyarakat pada media sosial Twitter (studi kasus pandemi COVID-19)," *MIB*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.

[17] Y. Qi and Z. Shabrina, "Sentiment analysis using Twitter data: A comparative application of lexicon- and machine-learning-based approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 31, Feb. 2023, doi: 10.1007/s13278-023-01030-x.

[18] M. Hayaty and A. H. Pratama, "Performance of lexical resource and manual labeling on long short-term memory model for text classification," *J. Ilm. Tek. Elektro Komput. Dan Inform.*, vol. 9, no. 1, pp. 74–84, Feb. 2023, doi: 10.26555/jiteki.v9i1.25375.

[19] M. Raees and S. Fazilat, "Lexicon-based sentiment analysis on text polarities with evaluation of classification models," *arXiv*, 2024, doi: 10.48550/ARXIV.2409.12840.

[20] M. Das, S. K., and P. J. A. Alphonse, "A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset," *arXiv*, 2023, doi: 10.48550/ARXIV.2308.04037.

[21] I. Y. Kairupan, A. Angdresey, and H. Arif, "An extreme gradient boosting approach for classification and sentiment analysis," *AJTM*, vol. 16, no. 3, pp. 211–225, 2023, doi: 10.12695/ajtm.2023.16.3.5.

[22] Y. Diao and Q. Zhang, "Optimization of management mode of small- and medium-sized enterprises based on decision tree model," *J. Math.*, vol. 2021, pp. 1–9, Dec. 2021, doi: 10.1155/2021/2815086.

[23] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.

[24] K. Md. S. Hasan, S. Saha, M. S. Mehrab, Nipa, T. Islam, and A. Salam, "Data analysis using visualization techniques," in *Proc. 2025 Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Chittagong, Bangladesh, Feb. 2025, pp. 1–6, doi: 10.1109/ECCE64574.2025.11013513.

[25] A. Lavanya, S. Sindhuja, L. Gaurav, and W. Ali, "A comprehensive review of data visualization tools: Features, strengths, and weaknesses," *Int. J. Comput. Eng. Res. Trends*, vol. 10, no. 1, pp. 10–20, Jan. 2023, doi: 10.22362/ijcert/2023/v10/i01/v10i0102.