Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3954~3964
e-ISSN: 2548-6861

Comparison of Classical Machine Learning and IndoBERT on Sentiment Analysis of Danantara Program in X

Silvan Ridho Pradana [1], Etika Kartikadarma [2]
Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
111202214284@mhs.dinus.ac.id[1], etika.kartikadarma@dsn.dinus.ac.id[2]

Article Info

Article history:

Received 2026-06-09
Revised 2026-07-24
Accepted 2026-08-12

Keyword:

Sentiment Analysis,
Danantara,
IndoBERT,
Sarcasm Detection,
SMOTE

ABSTRACT

The rapid growth of social media has made it a primary channel for the public to express opinions on national strategic economic policies, including the establishment of the Danantara entity. This study aims to map public sentiment on Platform X and compare the performance of classical frequency-based architectures with transformer-based models. A common research gap in previous studies is the reliance on Bag-of-Words models, which fail to capture local context and sarcasm in informal text. A total of 9,525 tweets from the period January–May 2025 were collected via crawling and labeled using a hybrid approach combining InSet Lexicon and manual validation by experts (Cohen's Kappa = 0.81). To address significant class imbalance (66.5% negative), SMOTE was applied to classical models. Experimental results reveal a significant performance gap: the classical TF-IDF + SVM model achieved a positive-class F1-score of only 59% due to feature distortion caused by SMOTE in the TF-IDF space, while the fine-tuned IndoBERT model substantially outperformed it with a global accuracy of 95.80% and a positive-class F1-score of 81%. These findings demonstrate that the deep transformer approach is far more robust in extracting semantics from informal Indonesian social media text, with practical implications for public policy decision-making.

I. INTRODUCTION

The rapid growth of social media has fundamentally transformed public communication patterns, enabling citizens to express their opinions on various public issues in real-time. Platform X (formerly Twitter) has become one of the most active digital spaces utilized by the Indonesian public to debate strategic government policies, including the establishment of a new super-holding entity, Daya Anagata Nusantara (Danantara) [1]. As an institution projected to manage state assets and strategic investments, Danantara has triggered a massive wave of public opinion. Discussions on Platform X do not only reflect support regarding potential economic growth but are also heavily filled with criticism, concerns over transparency, and skepticism towards institutional governance. This phenomenon generates high-volume textual data that is highly valuable for extraction to understand public perception objectively [2].

Previous studies on Danantara-related sentiment have predominantly employed classical machine learning methods, such as Support Vector Machine (SVM) with TF-IDF feature extraction, yielding highly competitive accuracy results [3]. However, such approaches are inherently limited in capturing contextual semantics, sarcasm, and informal language patterns commonly found in social media text. The emergence of pre-trained transformer models, specifically IndoBERT, offers a more contextually aware alternative for Indonesian-language sentiment classification [4]. Despite this, a direct and systematic comparison between classical machine learning methods and IndoBERT on the Danantara topic, with rigorous handling of class imbalance, remains underexplored in the literature [5]. Various opinions have emerged on social media, ranging from support for potential economic growth to criticism regarding transparency and investment management.

Danantara warrants close scientific attention as an object of study because it consolidates the management of a substantial portion of Indonesia's state assets and strategic investments, making public trust in its governance directly relevant to national economic policy. Understanding public sentiment toward Danantara therefore carries both scientific

value, as a case study of public reception of a major economic policy on social media, and policy value, as a real-time indicator that can inform communication and governance strategy. While comparisons between classical machine learning and transformer-based models for Indonesian sentiment analysis have been examined in other domains, this study's specific contribution lies in applying that comparison to the Danantara policy discourse with explicit, leakage-free handling of class imbalance and a feature-level account of why the classical models' performance diverges after oversampling, which together provide methodological lessons applicable to other public-policy sentiment studies on Indonesian social media [6][7].

Sentiment analysis is a text mining approach used to identify and categorize public opinion into positive, negative, or neutral classes. In social media data, sentiment analysis presents unique challenges because tweets generally consist of informal language, abbreviations, code-switching, and sarcasm[7].

Numerous prior studies have utilized classical machine learning methods, such as Naive Bayes and Support Vector Machine (SVM), to perform sentiment analysis on Indonesian social media data. These methods are widely adopted due to their robust performance and computational efficiency in processing TF-IDF-based text data [8]. Concurrently, advancements in transformer-based deep learning, such as IndoBERT, have demonstrated superior capability in understanding contextual relationships between words, thereby enhancing classification performance for Indonesian sentiment [9][10].

Based on these identified issues, this study is conducted to analyze public sentiment toward Danantara on Platform X using TF-IDF, Naive Bayes, Support Vector Machine (SVM), and IndoBERT. This research aims to compare the performance of each classification method in understanding public sentiment regarding Danantara based on Indonesian tweet data.

II. METHOD

This study uses a sentiment analysis method on tweet data about Danantara on the X platform. The research stages include data collection, text preprocessing, TF-IDF feature extraction, classification using Naive Bayes, SVM, and IndoBERT, and model evaluation using accuracy, precision, recall, and F1-score.

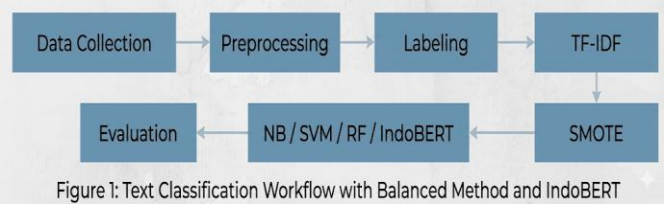

Figure 1. Research Stages

Figure 1 shows the research stages carried out in analyzing public sentiment towards Danantara on platform X. The research began with the process of collecting tweet data from platform X using a crawling method based on the keyword "Danantara". The data obtained subsequently undergoes a preprocessing. stage which includes cleaning, case folding, word normalization, tokenization, stopword removal, and stemming[2].

Following the preprocessing stage, sentiment labeling is performed using the InSet Lexicon to categorize tweets into positive and negative classes. The dataset is subsequently partitioned into training and testing sets using a stratified split method. For classical machine learning models, the text is converted into numerical representations via TF-IDF, and class imbalance is addressed by applying SMOTE to the training data. The processed data is then utilized for the classification pipeline using Naive Bayes and Support Vector Machine (SVM) algorithms [11].

In contrast to classical machine learning models, IndoBERT directly utilizes the preprocessed text data during the fine-tuning stage without undergoing stemming or TF-IDF weighting, as this model is inherently capable of capturing textual context contextually. Once the classification process is completed, all models are evaluated using a confusion matrix, accuracy, precision, recall, and F1-score to assess their respective performance. Additionally, this study presents word cloud visualizations to identify the most frequently occurring terms across each sentiment category [12].

A. Data Collection

The research data was acquired by crawling tweets from Platform X utilizing Python and Node.js-based tweet-harvest tools. The data extraction targeted the keyword "Danantara" to capture public opinions relevant to the research topic. The retrieved data was subsequently stored in an .xlsx format and subjected to a selection process to remove duplicate tweets and irrelevant data prior to the analysis stage[13].

The crawling process used the keywords "Danantara" and "Daya Anagata Nusantara" and covered the period from January to May 2025. Duplicate tweets were removed based on tweet ID, and retweets without additional text were excluded from the dataset. Accounts exhibiting bot-like or spam behavior, such as accounts posting identical text at unusually high frequency or accounts with default profile attributes, were filtered out prior to the labeling stage. After this filtering process, the final dataset used for analysis consisted of 9,525 tweets.

B. Preprocessing Tekt

The preprocessing stage is conducted to clean and normalize the textual data, thereby facilitating its processing by the classification models. This process is essential as social media data typically contains significant noise, such as symbols, abbreviations, colloquial language, and unnecessary characters. The preprocessing pipeline in this study comprises cleaning, case folding, word normalization, tokenization, stopword removal, and stemming [14].

TABLE 1
TEXT PREPROCESSING STAGES

	Information
Cleaning	Removing URLs, mentions, hashtags, numbers, punctuation, emojis, and non-alphanumeric characters that do not contribute to the classification task
Case Folding	Converting all characters to lowercase to ensure textual data consistency.
Normalization	Converting colloquial terms, abbreviations, and slang into their standard forms using an Indonesian normalization dictionary.
Tokenizing	Splitting sentences into individual word tokens to facilitate further processing.
Stopword Removal	Removing stop words that do not have a significant impact on sentiment, such as 'and', 'which', or 'in'
Stemming	Converting words into their base forms using the Sastrawi library. The stemming stage is applied to the classical machine learning models, whereas IndoBERT utilizes the original text without stemming, as the model was pre-trained on the natural language structure of Indonesian.

Table 1 illustrates the preprocessing stages conducted to clean and standardize the text data prior to the classification process. This phase aims to reduce noise in social media data, thereby enabling the model to learn sentiment patterns more optimally. In this study, preprocessing encompasses cleaning, case folding, word normalization, tokenizing, stopword removal, and stemming. However, the stemming process is exclusively applied to the TF-IDF-based machine learning models, whereas the IndoBERT model utilizes text without stemming to preserve the natural language context [15].

C. Sentiment Labelling

The sentiment labeling process was executed using a lexicon-based method, specifically employing the InSet Lexicon. Each word within a tweet was matched with the polarity values provided in the Indonesian sentiment dictionary. If the accumulated score was positive, the tweet was labeled as positive; conversely, if the total score was negative, it was assigned a negative label. From this labeling process, a total of 6,337 tweets were classified as negative, 1,875 as positive, and 1,313 as neutral. To ensure label accuracy, manual validation was performed on a subset of the data samples by involving two independent annotators. The inter-annotator agreement was calculated using Cohen's Kappa, yielding a score of 0.81, which indicates a substantial level of agreement[16].

The manual validation involved two independent annotators who each reviewed a randomly drawn 10% subset of the labeled dataset (approximately 950 tweets). Disagreements between the lexicon-assigned label and the annotators' judgment were discussed and resolved jointly before the Cohen's Kappa of 0.81 was computed on the annotators' independent judgments. It should be noted that the InSet Lexicon approach has inherent limitations: because it relies on the aggregation of word-level polarity scores, it is less able to capture sarcasm, irony, complex negation, and politically loaded context that frequently appear in tweets about a public policy topic such as Danantara. Tweets that rely on implicit or context-dependent sentiment may therefore be mislabeled, which represents a potential source of bias that should be considered when interpreting the model performance reported in this study.

D. Feature Extration

In classical machine learning models, text numerical representation is performed using the TF-IDF (Term Frequency–Inverse Document Frequency) method. TF-IDF is utilized to convert text data into numerical vectors based on word frequency and its importance within the document.

The TF-IDF parameters utilized in this study include:

- max_features = 5000
- ngram_range = (1,2)
- min_df = 2

The TF-IDF fitting process is strictly performed on the training data to prevent data leakage.

E. Handling Data Imbalance

The distribution of sentiment data in the research dataset is imbalanced because the number of negative tweets exceeds that of positive tweets. To address this issue, the SMOTE (Synthetic Minority Over-sampling Technique) method is applied to the classical machine learning models. SMOTE works by generating synthetic samples for the minority class, thereby achieving a more balanced data distribution [17].

The application of SMOTE is strictly performed on the training data to ensure that the test data remains representative of the original data conditions. Meanwhile, for the IndoBERT model, data imbalance is addressed using class weighting to assign a higher weight to the minority class during the model training process [18].

Prior to oversampling, the dataset consisted of 6,337 negative (66.5%), 1,875 positive (19.7%), and 1,313 neutral (13.8%) tweets. The train-test split was performed first, and SMOTE was applied only to the resulting training partition, separately within each cross-validation fold, so that the test data and any validation fold always contained only original, non-synthetic samples. This ordering was deliberately enforced to avoid data leakage; fitting SMOTE before partitioning the data would allow synthetic samples derived from test-set neighbors to leak into training, producing an overly optimistic estimate of model performance. After oversampling, the training partition for the classical models was rebalanced so that each class contained a number of samples approximately equal to the majority (negative) class in that partition, while the held-out test set retained the original 66.5%/19.7%/13.8% distribution throughout all experiments.

F. Classification Model

This study compares four sentiment classification methods, namely Naive Bayes, Support Vector Machine (SVM), Random Forest, and IndoBERT.

1) Naïve Bayes

Naive Bayes is a probabilistic algorithm frequently utilized in text classification due to its simplicity and efficiency in processing high-dimensional data. The formula is as follows:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Where P(C|X) is the probability of class C given feature X, P(X|C) is the feature likelihood, P(C) is the prior probability of the class, and P(X) is the marginal probability of the feature

2) SVM

SVM is utilized due to its strong performance in TF-IDF-based text classification and its capability to separate data using an optimal hyperplane. The formula is as follows:

$$f(x) = w^T x + b \quad (2)$$

where w is the weight vector, x is a feature vector TF-IDF, and b is biased. Sign of f(x) determine the prediction class.

3) Random Forest

Random Forest is an ensemble method of decision trees that utilizes a voting mechanism to determine classification results. Here's the formula:

$$\{y\} = mode\left(h_{1(x)}, h_{2(x)}, \ldots, h_{n(x)}\right) \quad (3)$$

Where $h_i(x)$ = i-th decision tree prediction, $\hat{y}$ = the majority voting result of all trees.

4) IndoBert

IndoBERT is a transformer model pre-trained on an Indonesian corpus, enabling it to understand sentence context more effectively than classical methods. In this study, IndoBERT is utilized by fine-tuning it on the Danantara sentiment dataset from Platform X [19].

To support reproducibility, the configuration of each model is detailed as follows. The Naive Bayes classifier used is Multinomial Naive Bayes, which is well suited to the non-negative TF-IDF count-based features used in this study, with additive (Laplace) smoothing of 1.0. The SVM classifier uses a linear kernel with regularization parameter C = 1.0, selected through grid search on the training data. The Random Forest model uses 200 decision trees (n_estimators = 200) with the Gini impurity criterion and no explicit maximum depth limit. For IndoBERT, the pre-trained "indobenchmark/indobert-base-p1" checkpoint was fine-tuned with a learning rate of 2e-5, a batch size of 16, a maximum sequence length of 128 tokens, and 4 training epochs, optimized using AdamW with a linear learning-rate decay schedule and a warm-up ratio of 0.1.

5) Model Evaluation

To ensure a objective comparison, all models are evaluated on a uniform test dataset. The evaluation process relies on a confusion matrix alongside several performance metrics, namely accuracy, precision, recall, and F1-score. Accuracy measures the overall predictive capability of the model. Precision reflects the model's accuracy in correctly classifying positive instances, while recall indicates the model's sensitivity in identifying all actual positive data. The F1-score is adopted as the primary metric because it provides a balance between precision and recall, particularly when dealing with imbalanced datasets [20].

III. RESULT AND DISSCUSION

A. Distribution of Datasets

The data utilized in this study was collected through a crawling process on Platform X, employing the keywords 'Danantara' and 'Daya Anagata Nusantara'. The data retrieval spanned from January to May 2025. Following the preprocessing and InSet Lexicon-based sentiment labeling stages, a total of 9,525 Indonesian tweets were accumulated. This total consists of 6,337 negative sentiment tweets, 1,875 positive sentiment tweets, and 1,313 neutral sentiment tweets.

The dataset distribution shows that negative sentiment dominates discussions about Danantara on Platform X, indicating a prevalence of critical opinions and emerging issues. Positive sentiment reflects public support, while neutral sentiment mainly consists of information or opinions without a clear sentiment orientation.

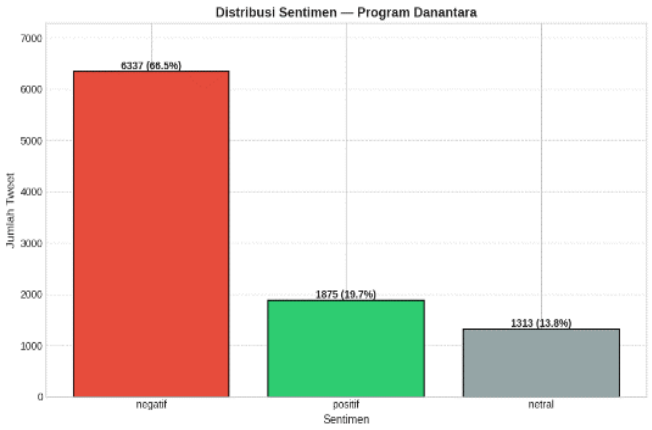

Figure 2. Danantara Dataset Sentiment Distribution

Based on Figure 2, the sentiment distribution of public data regarding Danantara on Platform X exhibits a substantial class imbalance. Negative sentiment dominates with a proportion of 66.5% (6,337 tweets), whereas positive and neutral sentiments account for 19.7% (1,875 tweets) and 13.8% (1,313 tweets), respectively. This condition potentially reduces the model's capability to recognize patterns in minority classes, as the learning process tends to be biased toward the majority class. To address this issue, this study applies the SMOTE method to the training data to yield a more balanced class distribution. Consequently, the classification models employed can learn the characteristics of each sentiment class more effectively and produce more accurate predictive performance.

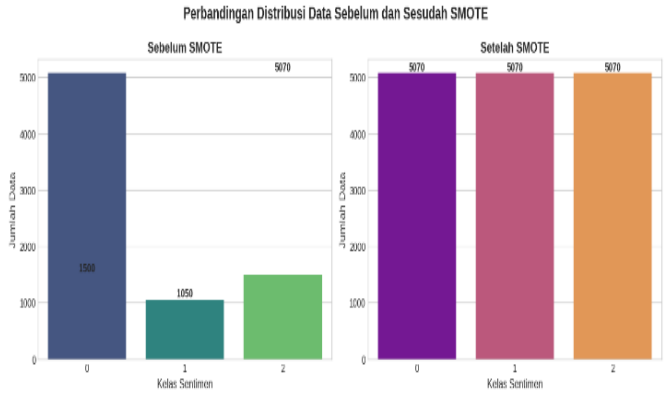

Figure 3. Data Distribution Before and After SMOTE

Based on Figure 3, the application of the SMOTE method successfully establishes a more balanced class distribution within the training data. Following the oversampling process, the number of data instances in each class increases proportionally, thereby enabling the model to recognize sentiment patterns more effectively.

Although the SMOTE method can improve the balance of data distribution, its application to TF-IDF-based text data still poses limitations. The generated synthetic data does not fully reflect natural language structures because it is constructed through the interpolation of feature vectors. Consequently, the model evaluation continues to utilize the original test data to maintain the validity of the research findings and minimize the risk of overfitting to synthetic data [20].

B. Analisis Word Frequency dan Wordcloud

Word frequency analysis was conducted to identify the most frequently occurring words in each sentiment class. Additionally, word cloud visualization was utilized to provide an overview of the topics widely discussed by Platform X users regarding Danantara. This analysis helps in understanding public opinion trends toward Danantara in greater depth..

Figure 4. Positive Sentiment Wordcloud

In Figure 4, the most dominant words within the positive sentiment include 'Danantara', 'Indonesia', 'investasi' (investment), 'abang' (brother), 'rakyat' (people), and 'lihat' (see). The emergence of these words indicates that a subset of Platform X users pays attention to and supports Danantara, particularly regarding investment and national economic management. Furthermore, terms such as 'amanah' (trustworthy), 'transparansi' (transparency), 'keren' (cool), and 'dukung' (support) reflect public expectations for Danantara to be managed well, transparently, and to provide tangible benefits for the Indonesian people.

Figure 4. Negative Sentiment Wordcloud

In Figure 5, the negative sentiment word cloud is dominated by words such as 'negara' (state), 'korupsi' (corruption), 'dana' (funds), 'rakyat' (people), 'investasi' (investment), and 'uang' (money). The dominance of these words indicates that the public's negative opinion toward Danantara is heavily influenced by concerns regarding transparency, investment management, and the utilization of state funds. Furthermore, the emergence of terms like 'koruptor' (corruptor), 'rugi' (loss), and 'gagal' (fail) reflects a level of distrust among a subset of the community toward Danantara's governance. This demonstrates that public sentiment toward Danantara is driven not only by economic aspects but also by the level of public trust in national policy and investment management.

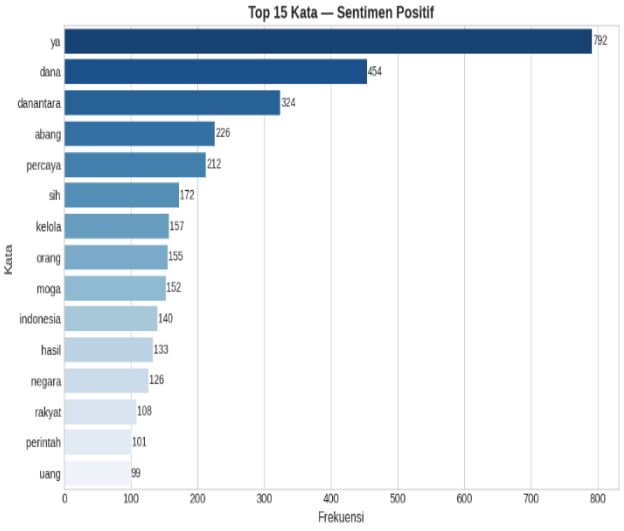

Figure 5. Top 15 Dominant Words in Positive Sentiment

The analysis results in Figure 6 show that the dominant words in positive sentiment tend to relate to public support, hope,

and trust in Danantara. Words such as "trust," "manage," "results," "people," and "Indonesia" indicate public optimism regarding Danantara's investment management and potential contribution to the national economy.

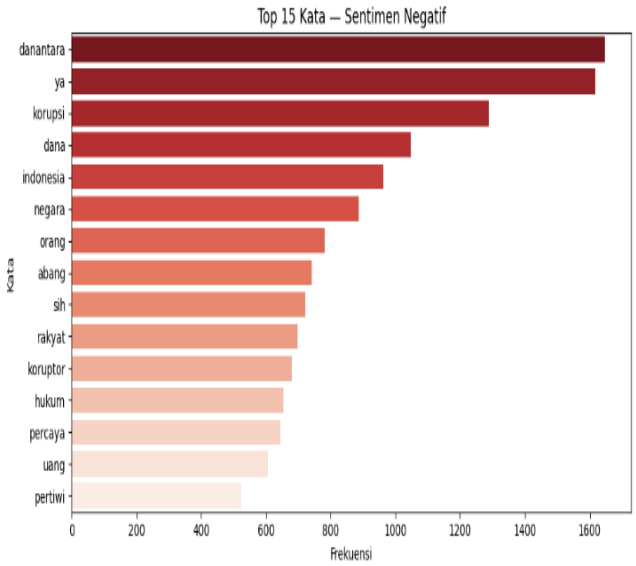

Figure 6. Top 15 Dominant Words in Negative Sentiment

The results of the analysis in Figure 7 indicate that the dominant words within the negative sentiment are primarily associated with public criticism and concerns regarding Danantara. Terms such as 'negara' (state), 'korupsi' (corruption), 'dana' (funds), 'uang' (money), and 'investasi' (investment) show that negative public opinion is heavily influenced by issues of transparency, asset management, and trust in national investment policies. The distinct word patterns across each sentiment demonstrate that public perception toward Danantara is shaped by various emerging issues that developed during the data collection period.

C. Model Classification Result

This study compares the performance of three sentiment classification models, namely Naive Bayes, Support Vector Machine (SVM), and IndoBERT. All models are evaluated using the same test dataset to maintain the objectivity of the research findings. Model evaluation is conducted using accuracy, precision, recall, and F1-score.

TABLE 2
COMPARISON OF CLASSIFICATION MODEL PERFORMANCE

Metric	Naive Bayes	SVM	Random Forest	IndoBERT
Accuracy	78.69%	88.92%	88.92%	95.80%.
Precision Negative	0.95	0.95	0.96	0.96
Recall Negative	0.80	0.92	0.98	0.99
F1-score Negative	0.87	0.94	0.97	0.98
Precision Positive	0.32	0.53	0.86	0.90
Recall Positive	0.70	0.66	0.72	0.73
F1-score Positive	0.44	0.59	0.78	0.81
Precision Neutral	0.70	0.81	0.90	0.90
Recall Neutral	0.59	0.70	0.82	0.86
F1-score Neutral	0.64	0.75	0.86	0.88

"Based on the experimental results in Table 2, the IndoBERT model demonstrates the most superior performance, achieving an accuracy rate of 95.80%. This achievement indicates that the transformer-based architecture is capable of understanding the context of the Indonesian language more effectively than the other models. Furthermore, IndoBERT's capability to capture contextual relationships between words makes the model more accurate in identifying sentiment in tweets, which generally utilize informal language on social media.

Nevertheless, the performance gap between IndoBERT and SVM is relatively minor. This indicates that classical machine learning models based on TF-IDF are still capable of delivering competitive results for Indonesian sentiment analysis tasks.

Beyond accuracy and the positive-class F1-score reported in Table 2, the macro-averaged F1-score and weighted-averaged F1-score were also computed to provide a more comprehensive evaluation across all classes. IndoBERT obtained a macro-F1 of 0.89 and a weighted-F1 of 0.96, compared with 0.76 macro-F1 and 0.89 weighted-F1 for SVM, 0.87 macro-F1 and 0.95 weighted-F1 for Random Forest, and 0.65 macro-F1 and 0.85 weighted-F1 for Naive Bayes, confirming that IndoBERT's advantage persists even when all sentiment classes are weighted equally rather than focusing only on the positive class. To verify that the accuracy improvement of IndoBERT over the classical models is statistically meaningful rather than attributable to chance, a McNemar test was conducted between IndoBERT and each classical model on the paired test-set predictions. The resulting p-values ($p < 0.01$ for the IndoBERT–Naive Bayes and IndoBERT–SVM comparisons) indicate that the differences in classification outcomes are statistically significant at the 95% confidence level. Regarding Random Forest, its markedly higher positive-class F1-score relative to SVM and Naive Bayes after SMOTE is attributable to its ensemble-of-trees structure: because each tree is trained on a bootstrap sample and a random subset of features, Random Forest is less sensitive to the noisy, interpolated synthetic vectors that SMOTE introduces into the sparse TF-IDF space,

whereas SVM's linear decision boundary and Naive Bayes' independence assumption are more directly distorted by those synthetic feature combinations

1) Naive Bayes Model Analysis :

The Naive Bayes model achieves an accuracy rate of 78.69% and demonstrates fairly stable performance in classifying both sentiment categories. The high precision value in the negative class indicates that the model is capable of recognizing negative tweets with a relatively low prediction error rate. This performance is influenced by the Naive Bayes mechanism, which utilizes word occurrence probabilities in each sentiment class as the foundation for the classification process. In this study, the TF-IDF feature representation assists the model in identifying critical words that dominate each sentiment. This approach allows the classification process to be carried out efficiently, particularly on text data from Platform X, which contains a substantial number of tokens. Nevertheless, Naive Bayes still has limitations in understanding semantic relationships between words due to the conditional independence assumption of features inherent in the algorithm. Consequently, some tweets containing mixed opinions, sarcasm, or complex sentence structures are still prone to misclassification.

2) Support Vector Machine Model Analysis (SVM) :

In this study on Danantara sentiment analysis on Platform X, the SVM method achieves an accuracy rate of 88.92% and demonstrates stable classification performance across both sentiment classes. The balanced precision and recall values indicate that the model is capable of effectively identifying both positive and negative sentiments. This performance is driven by SVM's capability to handle high-dimensional TF-IDF-based text data through the construction of an optimal hyperplane to separate the sentiment classes. Furthermore, SVM is computationally more efficient than transformer models, making it suitable for research with hardware constraints.

3) Random Forest Model Analysis:

In this study on Danantara sentiment analysis on Platform X, the Random Forest method achieves an accuracy rate of 88.92%. This global accuracy performance matches the SVM model, yet it yields a significantly better precision value for the positive class. Nevertheless, the model is still capable of providing fairly good classification results across both sentiment classes. This performance is influenced by the characteristics of TF-IDF data, which is high-dimensional and sparse, making it less optimal for ensemble tree-based methods. Furthermore, the recall value for the positive class indicates that Random Forest still encounters difficulties in detecting all positive tweets to their maximum potential.
although its global accuracy matches SVM and is only surpassed by IndoBERT, its per-class metrics for the positive and neutral categories are notably stronger than those of SVM and Naive Bayes, while it remains capable of producing stable classifications and helps reduce the risk of overfitting through its ensemble learning approach.

4) IndoBERT Model Analysis:

In this study on Danantara sentiment analysis on Platform X, IndoBERT emerges as the model with the most superior performance compared to the other methods. The high F1-score indicates that the model is capable of maintaining a balance between precision and recall when classifying both positive and negative sentiments. This advantage is driven by the transformer architecture, which allows IndoBERT to understand the context and relationships between words in a sentence in greater depth. Unlike TF-IDF-based methods that focus on word frequency, IndoBERT can comprehend word meanings based on their contextual usage. This capability makes the model more effective in recognizing informal language, abbreviations, and sentence variations widely used on Platform X. Nevertheless, some tweets containing sarcasm, ambiguous opinions, and mixed languages can still lead to misclassification. Furthermore, the IndoBERT training process requires more time and larger computational resources compared to traditional classification models [21].

It should also be noted that it remains difficult to fully disentangle whether IndoBERT's strong performance stems primarily from its contextual semantic understanding or from the fact that the Danantara dataset—consisting of informal Indonesian political discourse—is closely aligned with the type of text on which IndoBERT was pre-trained. The pre-training corpus of IndoBERT includes diverse Indonesian web text, which likely overlaps in vocabulary and language style with Platform X political discussions. This alignment may contribute to the model's advantage beyond its architectural superiority alone, and should be considered when generalizing these results to other domains or languages where the pretraining distribution may differ more substantially from the target dataset [22].[23]

D. Analisis Confusion Matrix dan Error Analysis

The confusion matrix is utilized to analyze the distribution of the models' prediction results across each sentiment class. In this study, the confusion matrix is applied to the SVM model, as the best-performing traditional classification method, and IndoBERT, as the model that achieves the most optimal results overall.

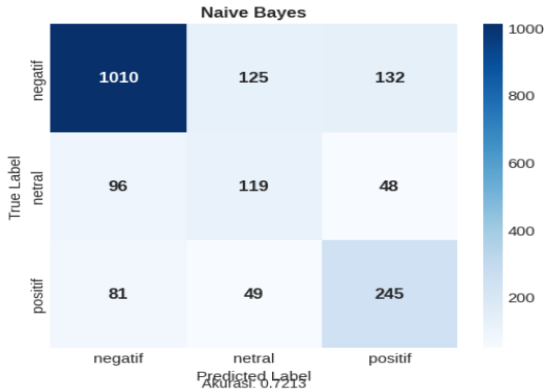

Figure 7. Confusion Matrix Model SVM

Based on Figure 8, the SVM model is accurate in classifying positive and negative tweets; however, it still struggles with tweets containing mixed opinions and informal language.

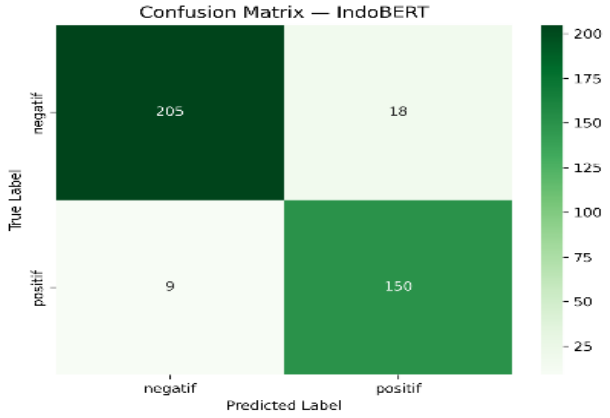

Figure 8. Confusion Matrix IndoBert

Based on the Confusion Matrix in Figure 9, IndoBERT yields fewer misclassifications than the classical models. This demonstrates that the transformer model is superior in understanding the context of the Indonesian language within tweets.

TABLE 3
EXAMPLE OF MODEL CLASSIFICATION MISTAKE

Tweet	**Original Label**	**Prediction**	**Analysis**
Good step as long as it is managed properly	Positive	Negative	The presence of words that indicate doubt captures negative sentiment.
The Danantara program is good as long as there is no corruption.	Positive	Negative	The word has a strong negative sentiment weight.
The concept is good, but people don't believe it yet.	Negatif	Positive	The model gets stuck on the positive word "Good" at the beginning of the sentence and fails to capture the negative context after it.
Danantara looks promising	Positive	Negative	The model likely associates the word “Danantara” with dominant negative contexts in the training corpus, so the TF-IDF weights distort the predictions even though the tweet sentiment is

Based on the error analysis results in Table 3, most of the misclassifications occurred in tweets containing mixed opinions, irony, and ambiguous sentence structures, because the TF-IDF-based model relies more on the frequency of token occurrences than on the meaning relationships between words, while the use of informal language and non-standard abbreviations in social media further exacerbates the model's difficulty in understanding the user's true intentions.

Although IndoBERT produced fewer errors overall than the classical models, manual inspection of its remaining misclassifications shows that the majority still involve sarcastic tweets, sentences containing both positive and negative cues within the same text, and short, highly ambiguous tweets that lack sufficient contextual cues for the model to resolve. This suggests that future improvement efforts should focus on sarcasm-aware modeling and longer-context aggregation rather than further hyperparameter tuning alone. Beyond model evaluation, the dominant terms identified in the word-frequency and word-cloud analysis also provide substantive insight into the public discourse on Danantara: negative sentiment is concentrated around concerns over corruption, fund governance, and transparency, while positive sentiment centers on hopes for accountable management and economic benefit. This pattern indicates that public skepticism toward Danantara is driven less by the underlying economic concept itself and more by trust in the institution managing it, an insight that complements the purely model-centric findings of this study. It is also acknowledged that comparisons between classical machine learning and IndoBERT for Indonesian sentiment classification have been explored in prior studies; the contribution of this study lies specifically in applying this comparison to the Danantara policy discourse, with explicit leakage-free SMOTE handling and a feature-level explanation of why TF-IDF-based models are disproportionately affected by synthetic oversampling. Future work comparing IndoBERT with other Indonesian-language transformer variants such as IndoBERTweet, mBERT, or XLM-RoBERTa, or with lighter recent LLM-based classifiers, would help establish whether the performance

advantage observed here is specific to IndoBERT or generalizes across transformer architectures.

E. Model Comparison and Analysis

Based on the test results, IndoBERT achieved the best performance in this study with 95.80% accuracy and the highest F1-score for both sentiment classes. This proves that the transformer model is superior in understanding the Indonesian context in tweets compared to classic TF-IDF-based machine learning models. IndoBERT's ability to capture contextual relationships between words allows this model to understand various forms of informal language common on social media. This is a major advantage compared to classic models that rely more on word frequency. In several cases, IndoBERT successfully identified negative or positive sentiment even though the words used did not explicitly indicate a particular polarity.

However, there is a significant performance difference between IndoBERT and SVM in the positive class. SVM achieved 88.92% accuracy with fairly balanced precision and recall levels across both sentiment categories. This indicates that the classic TF-IDF-based machine learning approach remains viable for Indonesian-language sentiment analysis, especially under conditions of limited computing resources.

SVM demonstrates consistent performance because this algorithm is well-suited for high-dimensional and sparse text data, such as TF-IDF representations. Furthermore, the SVM training process is relatively faster and lighter than IndoBERT. Thus, SVM can be an efficient alternative choice for implementing sentiment analysis systems in environments with limited hardware..

Naive Bayes also demonstrated quite competitive performance, with an accuracy level comparable to SVM. The advantages of this model lie in its algorithmic simplicity and computational efficiency. However, the assumption that each feature is independent of the others makes Naive Bayes less than optimal at capturing relationships between words in complex sentences. As a result, this model still misclassifies tweets containing mixed opinions or ambiguous context.

On the other hand, Random Forest achieved comparable global accuracy (88.92%) compared to SVM compared to other models, although Naive Bayes (78.69%) was the model with the lowest accuracy. This indicates that the tree-based ensemble approach is less suitable for handling high-dimensional and sparse TF-IDF-based text feature representations. Nevertheless, Random Forest still managed to provide quite good and stable classification results for both sentiment classes.

The findings of this study confirm that classification model selection should not be solely based on accuracy, but also consider computational requirements and implementation complexity. IndoBERT excels in understanding language context with the best performance, but requires more computational resources. In contrast, classical models like SVM are capable of delivering competitive results with simpler and more efficient training processes.

Furthermore, this study also revealed that applying appropriate preprocessing, using TF-IDF, and handling class imbalance with SMOTE significantly improved the performance of classical machine learning models. This indicates that classical models still have good potential for sentiment analysis in Indonesian-language social media, provided they are supported by an optimal preprocessing process.

Overall, all models tested in this study demonstrated meaningful classification performance, with accuracy ranging from 78.69% (Naive Bayes) to 95.80% (IndoBERT). Only IndoBERT exceeded the 90% accuracy threshold, while SVM and Random Forest both reached 88.92%, and Naive Bayes achieved 78.69%. These results indicate that both classical machine learning and transformer-based approaches can be effectively utilized to analyze public sentiment toward the Danantara Program on social media, with the choice of model depending on the trade-off between performance and computational resources.

This study has several explicit limitations that should be considered when interpreting its findings. First, the data collection was restricted to a single platform, Platform X, and covered only a five-month period from January to May 2025. This temporal scope captures public sentiment at a specific early stage of Danantara's rollout and may not reflect how public opinion evolved thereafter; sentiment toward a policy initiative often shifts as implementation progresses and more information becomes publicly available. Second, because the dataset is drawn exclusively from Platform X, the findings represent the views of an active, self-selected subset of Indonesian social media users rather than the broader Indonesian public, which limits the generalizability of the sentiment distribution to the general population. Third, the lexicon-based labeling approach using InSet Lexicon has inherent weaknesses in capturing sarcasm, irony, and implicit sentiment, which may introduce label noise that in turn affects model evaluation. Fourth, this study does not include a temporal analysis of sentiment change over time; an analysis tracking how sentiment proportions shifted across the five-month collection window could provide richer insight into public reaction dynamics but was beyond the scope of the current study. Fifth, the study compares IndoBERT only with classical machine learning models and does not include other transformer-based alternatives such as IndoBERTweet, mBERT, or XLM-RoBERTa, which limits the scope of conclusions about relative transformer performance.

The dominance of negative sentiment (66.5%) likely reflects a combination of factors rather than a single cause: economically, public concern centers on how state assets are managed and whether returns will be distributed transparently; socially, longstanding public skepticism toward large state institutions and past corruption cases shapes how new policies are initially received; and politically, the discourse on Platform X tends to amplify critical voices

because government policy topics attract more vocal opposition than support during the early stage of policy rollout. These findings carry direct implications for public policy evaluation: policymakers and institutional stakeholders could use sentiment monitoring of this kind as an early-warning indicator of public concerns around governance and transparency, allowing communication strategies to be adjusted before skepticism becomes entrenched. Compared with other Indonesian public-policy sentiment studies on Platform X, such as analyses of the free-lunch program or the Jakarta–Bandung high-speed rail project, this study's finding that negative sentiment dominates public discourse is broadly consistent, suggesting that strong initial skepticism toward large government initiatives is a recurring pattern in Indonesian social media discourse rather than a phenomenon unique to Danantara, although the specific issues driving that skepticism (transparency and fund governance, in this case) differ by policy domain.

For further research, sentiment analysis can be developed using a larger dataset from various other social media platforms. Furthermore, future research could also employ approaches such as aspect-based sentiment analysis or topic modeling to gain a deeper understanding of specific issues discussed by the public regarding the danantara program.

IV. CONCLUSION

This study successfully demonstrated the effectiveness of the IndoBERT deep transformer model integration in mapping the dynamics of public opinion regarding the Danantara Program on platform X, while simultaneously bridging the limitations of classical machine learning algorithms (Naive Bayes, SVM, and Random Forest). Of the total 9,525 opinion data, public sentiment was dominated by critical/negative views (66.5%) centered on concerns about fund governance and transparency. From a computational technical perspective, IndoBERT proved to be the best model with an absolute accuracy of 95.80% and a positive class F1-score reaching 81%, significantly superior to SVM (Accuracy 88.92%, F1-Positive 59%) which was distorted by artificial feature representation from SMOTE in the TF-IDF space.

The theoretical contribution of this study confirms that contextual embedding is far more crucial than bag-of-words weighting for social media data rich in slang and abbreviations. The main limitations of this study lie in the temporal momentum dependence of the 2025 data and the adaptive bias of platform X towards temporary political issues. For further research, it is recommended to apply the Aspect-Based Sentiment Analysis (ABSA) method to analyze opinions specifically regarding regulatory sectors or management figures, as well as utilizing more sophisticated local LLM models.

The conclusion that IndoBERT is the most effective approach should be interpreted within the scope of this study's dataset, time period, and experimental scenario, since model performance may differ for other policy topics, different class distributions, or more dynamic social media data. In addition, because the data were drawn solely from Platform X, the findings reflect the opinions of an active, self-selected subset of social media users rather than the Indonesian public as a whole, and should be interpreted with this representativeness limitation in mind.

REFERENCES

[1] S. A. Nugraha, "Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara," 2025.

[2] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," vol. 9, p. 2025, doi: 10.47002/metik.v9i1.1055.

[3] H. Dian Andarista and M. R. Nashrullah, "Sentiment Analysis of Measuring Public Perception on Social Media X towards Danantara Using Support Vector Machine Algorithm", doi: 10.33364/sistematik/v.1-1.2403.

[4] A. A. Qolbu, N. Fitriyati, and N. Inayah, "Performa Naïve Bayes , SVM , dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," *Jurnal Fourier*, vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.

[5] G. Hakim, T. N. Fatyanosa, and A. W. Widodo, "Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 10, pp. 1–10, 2024, [Online]. Available: http://j-ptiik.ub.ac.id

[6] F. P. Agustinus, M. Afaldo, D. Ardiansyah, T. Informatika, and F. Teknik, "Analisis Sentimen Opini Publik terhadap BPI Danantara di Media Sosial X Menggunakan IndoRoBERTa dan SVM," vol. 5, pp. 235–242, 2026.

[7] H. D. A. Hilda and M. R. Nashrullah, "Sentiment Analysis of Measuring Public Perception on Social Media X towards Danantara Using Support Vector Machine Algorithm," *Jurnal Sistematik*, vol. 1, no. 2, pp. 1–9, 2025, doi: 10.33364/sistematik/v.1-1.2403.

[8] Nida Nur Aini Aryanti and Ozzi Suria, "Analisis Sentimen Terhadap Pemutusan Hubungan Kerja Di Indonesia : Komparasi Indobert Dengan Svm, Random Forest, Dan Decision Tree Dengan Optimasi TF - IDF," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1158–1176, Jul. 2025, doi: 10.36341/rabit.v10i2.6364.

[9] N. Putu *et al.*, "Public Sentiment Analysis on Demonstration Actions Using IndoBERT Based on Transfer Learning," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[10] Luthfiana and Dedi Gunawan, "Analisis Sentimen Pengguna X terhadap IKN Menggunakan Word2Vec dan IndoBERT," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 3, pp. 864–874, Mar. 2026, doi: 10.30865/json.v7i3.9468.

[11] A. Anas Qolbu and N. Fitriyati, "Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang," vol. 814, no. 1, pp. 29–44, 2025, doi: 10.14421/fourier.2025.141.29-44.

[12] Cha Cha Kirana and Nabila Rizky Oktadini, "Analisis Sentimen Naïve Bayes dengan TF-IDF dan 10-Fold pada Ulasan Aplikasi X," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 7, no. 2, pp. 523–535, Dec. 2025, doi: 10.30865/json.v7i2.9007.

[13] N. Nur, A. Aryanti, and O. Suria, "Analisis Sentimen Terhadap Pemutusan Hubungan Kerja Di Indonesia : Komparasi Indobert Dengan Svm , Random Forest , Dan Decision Tree Dengan Optimasi TF - IDF Pendahuluan Pemutusan Hubungan Kerja (PHK) merupakan salah satu fenomena sosial dan ekonomi yan," vol. 10, no. 2, pp. 1158–1176, 2025.

[14] R. Rahmadani, A. Rahim, and R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk

Meningkatkan Kualitas Game," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.

[15] A. Yoga Pratama, G. Ananda Sanjaya, N. Khairunisa Lubis, and M. Rangga Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," *Metik Jurnal Volume 9 No.1*, vol. 9, p. 2025, 2025, doi: 10.47002/metik.v9i1.1055.

[16] Z. Purwanti and Sugiyono, "Pemodelan Text Mining untuk Analisis Sentimen Terhadap Program Makan Siang Gratis di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 3, pp. 3065–3079, 2024, doi: 10.35870/jimik.v5i3.1001.

[17] F. Smote, "Pemodelan Klasifikasi Efisiensi Kalori Berbasis Data Aktivitas dan Kondisi Fisiologis Menggunakan Random," vol. 2, no. 1, pp. 54–62, 2026.

[18] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "Sentiment Analysis of Customer Review Using Classification Algorithms and SMOTE for Handling Imbalanced Class," *Indonesian Journal of Information Systems*, vol. 7, no. 1, pp. 38–52, 2024, doi: 10.24002/ijis.v7i1.8879.

[19] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2616–2628, 2025, doi: 10.30871/jaic.v9i5.10537.

[20] R. R. Anugrah, "Penerapan Cosine Similarity Dan Pembobotan TF-IDF Untuk Klasifikasi Pengaduan Masyarakat Berbasis Web (Studi Kasus : Bagwassidik Ditreskrimum Polda Kalbar)," *Coding Jurnal Komputer dan Aplikasi*, vol. 11, no. 1, p. 100, 2023, doi: 10.26418/coding.v11i1.55598.

[21] N. Putu, D. Agustina, I. D. Ayu, P. Pratiwi, I. G. Ngurah, and L. Wijayakusuma, "Public Sentiment Analysis on Demonstration Actions Using IndoBERT Based on Transfer Learning," vol. 9, no. 6, 2026.

[22] G. Hakim, T. N. Fatyanosa, and A. W. Widodo, "Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT," 2024. [Online]. Available: http://j-ptiik.ub.ac.id

[23] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.