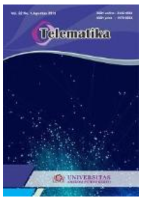
Telematika – Vol. 18, No. 2, August (2025) pp. 190-202 e-ISSN 2442-4528 | p-ISSN 1979-925X



Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/

Telematika

Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS AMIKOM PURWOKERTO

Comparative Analysis of Support Vector Machine and IndoBERT Algorithms in Stance Detection on Political Issues in Social Media X: A Case Study of BPI Danantara

Dhammananda Taniputra[1], Beny Beny[2], Lidwina Demai Lucia[3], Rachel Velysha[4]

[1,2,3,4] Department of Informatics, Faculty of Computer Science, Dinamika Bangsa University, Jambi, Indonesia

ARTICLE INFO

History of the article:
Received April 15, 2025
Revised May 20, 2025
Accepted August 27, 2025

Keywords:
Stance Detection
IndoBERT
SVM
Social Media
BPI Danantara

Correspondence:
E-mail:
dhammanandatan28@gmail.com

ABSTRACT

Stance detection is an NLP task aimed at identifying and classifying a writer's attitude toward a topic as supportive, opposing, or neutral based on text analysis, providing deeper insights into public opinion and supporting data-driven decision-making. This study focuses on Indonesian society's stance toward the National Investment Management Agency (BPI Danantara), which has received positive responses for its economic potential as well as negative reactions due to concerns over governance and corruption risks. In this research, a machine learning approach using the Support Vector Machine algorithm and a deep learning approach using the IndoBERT model were applied to detect pro, contra, and neutral stances in posts from the X social media platform. A total of 6,805 tweets were collected through scraping and manually labeled by three annotators. The dataset was then processed through cleaning, undersampling, and modeling, and evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics. Experiments were conducted across various scenarios, including binary and three-class classification as well as balanced and imbalanced datasets, to assess the effectiveness of each model. The results indicate that IndoBERT consistently outperforms SVM across all scenarios, particularly in capturing nuanced stances in Indonesian text. However, statistical evaluation using the paired t-test and the Wilcoxon signed-rank test reveals that the performance differences between the two models are generally not statistically significant, except in the three-class classification scenario with undersampling, where IndoBERT shows a significant advantage in handling balanced multi-class stance detection. These findings demonstrate the advantage of Transformer-based approaches for complex stance detection tasks and highlight their potential for developing automated public opinion monitoring systems. Nevertheless, this study has limitations, including the relatively small dataset, the focus on a single social media platform, and the methods applied. Future research could explore larger and more diverse datasets, incorporate multiple social media platforms, and employ other Transformer-based models to enhance generalization and improve stance detection accuracy.

1. INTRODUCTION

BPI Danantara is a national investment management institution inaugurated by President Prabowo Subianto on February 24, 2025. It was established based on Law No. 1 of 2025 and Government Regulation No. 10 of 2025 (Maulana et al., 2025). The institution oversees seven state-owned enterprises (BUMN) and the Indonesia Investment Authority (INA) with the aim of reducing the national savings–investment gap and promoting inclusive, globally competitive economic growth. Its mandate includes strengthening BUMN investment through an investment holding (managing dividends and assets) as well as an operational holding (overseeing business activities). Around USD 20 billion has been allocated for strategic projects such as renewable energy, downstream industries, and food security (Simanjuntak & Widyadhana, 2025). Despite its significant potential as a sovereign wealth fund, BPI Danantara still faces risks such as mismanagement and corruption. International cases, such as Malaysia's 1MDB scandal, highlight the importance of strong governance, transparency, and strict oversight. Therefore, public perception of BPI Danantara becomes a crucial aspect to consider (Ayunia et al., 2025).

The establishment of the Badan Pengelola Investasi Daya Anagata Nusantara (BPI Danantara) has generated a wide range of public responses, from support to opposition (Hidayat et al., 2025). This issue has continued to grow on the X platform (formerly Twitter), which provides an open space for users to express their opinions on various matters, including political policies. In today's digital era, social media platforms such as X play an important role as public spaces where society can actively discuss trending and widely debated topics (Anggraheni et al., 2022). The high level of user activity produces a large volume of data (big data) containing public opinion, including discussions related to BPI Danantara. In Indonesia, the number of X users reaches approximately 48.45 million around 4.23% of the global user base generating a continuous stream of opinion data every day. This abundance of user-generated content can be utilized for stance detection tasks to understand public perceptions toward specific topics (Setiawati et al., 2024).

Stance detection is a task within Natural Language Processing (NLP) that aims to identify an individual's attitude or position toward a particular topic based on the text they write, whether it is pro (supportive), contra (opposing), or neutral (Alturayeif et al., 2023). Although it is often compared to sentiment analysis, the two focus on different aspects. Stance detection evaluates the writer's position toward a predefined issue, whereas sentiment analysis focuses on identifying the emotional tone of the text—positive, negative, or neutral—toward the statement itself (Wankhade et al., 2022). Stance detection can be performed using various approaches, including traditional machine learning methods as well as deep learning models.

The Support Vector Machine (SVM) algorithm and IndoBERT represent two widely used approaches in Natural Language Processing (NLP), particularly for text classification tasks such as sentiment analysis, hoax detection, topic classification, hate speech detection, and other related applications. SVM operates by identifying the optimal hyperplane that best separates data points from different classes (Cotfas et al., 2021). Meanwhile, IndoBERT is a Transformer-based model and a variant of BERT that has been specifically pretrained on large-scale Indonesian language corpora, enabling it to capture the linguistic characteristics of Indonesian text and perform effectively on various Indonesian NLP tasks (Vincent et al., 2025). With these capabilities, both SVM and IndoBERT can be applied to stance detection, a task that aims to determine a writer's position or attitude toward a specific issue. In the context of this study, stance detection is used to assess public perception of BPI Danantara whether supportive (pro) or opposing (contra) based on user posts on the X (Twitter) platform. As no prior Indonesian stance detection studies have specifically employed IndoBERT, this research evaluates IndoBERT's performance with SVM as a baseline for comparison.

In a previous study on Indonesian-language stance detection, researchers analyzed user comments on the X (formerly Twitter) platform related to the 2024 Indonesian General Election. The system developed in that study retrieved tweet URLs, performed web crawling to collect comment threads, processed the text, and classified each comment's stance into three categories: Favor, Against, and None. The results showed that the Support Vector Machine (SVM) algorithm achieved the best performance, obtaining a macro F1-score of 77.3%, outperforming five other machine learning algorithms Multinomial Naive Bayes (MNB), Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbor (KNN). Based on these findings, SVM was selected as the primary model for the web-based application developed in the study. The application is capable of performing stance detection on comments from the X platform and was designed to help users avoid the echo chamber effect (Sulistyo & Pamungkas, 2025).

In addition, a related study investigated public stance toward distance education in Saudi Arabia during the COVID-19 pandemic using a collection of Arabic tweets. The research developed a stance classification system and evaluated various algorithms, including Random Forest, SVM, AdaBoost, Multinomial Naive Bayes, CNN, and LSTM, using precision, recall, F1-measure, and AUC as evaluation metrics. The results demonstrated that SVM with TF-IDF features achieved the best performance, obtaining an F-measure of 0.859 and an AUC of 0.951, outperforming all other classical machine learning and deep learning models evaluated in the study. Based on these findings, the SVM model was chosen for the system to analyze public stance patterns toward distance education policies (Alqurashi, 2022).

Another study examined stance detection in Arabic tweets related to Hezbollah, aiming to identify whether a tweet expresses support, opposition, or neutrality toward the organization as part of extremist discourse analysis. The study compared several traditional machine learning algorithms, including SVM (with linear, RBF, polynomial, and sigmoid kernels), Multinomial Naïve Bayes, and Weighted K-NN. Experimental results showed that the combination of TF-IDF features and the SVM with an RBF kernel delivered the best performance, outperforming Naïve Bayes, K-NN, and FastText. Without applying SMOTE, the SVM-RBF model achieved a macro F-score of 72.43%, and after SMOTE was applied, its performance improved to 78.62%, representing the highest score among all models evaluated in the study (Alkhraiji & Azmi, 2025).

However, there is also research on stance detection in English-language tweets that analyzed users' attitudes toward online learning during the COVID-19 pandemic. In this study, all traditional machine learning models such as SVM, Random Forest, KNN, Naïve Bayes, Decision Tree, and Logistic Regression achieved performance levels below 84% accuracy. In contrast, BERT attained an accuracy of 84.3% and outperformed BiLSTM, Attention-BiLSTM, and NBSVM, with an accuracy of 84.3%, a precision of 83.6%, and an F1-score of 84.9%. These findings, along with results from prior stance detection studies, provide the basis for selecting SVM and IndoBERT (a variant of BERT for Indonesian) in this research to compare the performance of both approaches in the stance detection task (Hamad et al., 2022).

2. RESEARCH METHODS

This section explains the procedures carried out in this study to perform stance detection using two approaches, namely the Support Vector Machine (SVM) algorithm as a machine learning method and the IndoBERT model (Indonesian Bidirectional Encoder Representations from Transformers) as a transformer-based architecture. The overall research methodology, including the stages of data processing, model implementation, and evaluation, is illustrated in Figure 1.

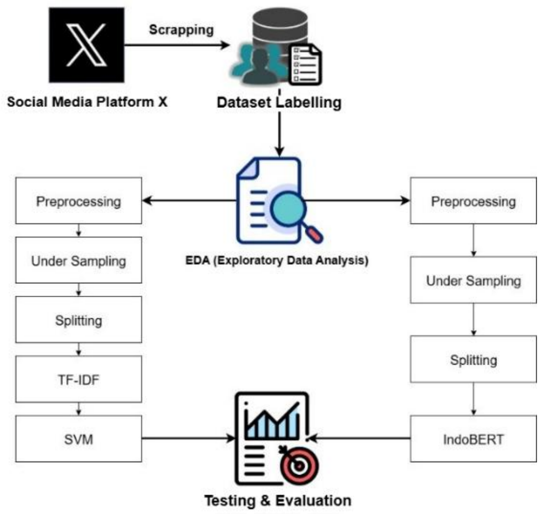


Figure 1. Overview of the Research Methodology

2.1. Data Collection

The research data were collected through crawling using Tweet Harvester v2.6.1 with authentication via an auth token to ensure stable access. This process filtered Indonesian-language tweets containing the keyword "Danantara" within the period from January 1 to June 1, 2025, which were then stored in .csv format and cleaned from duplicates, including retweets. During the crawling process, the researchers employed a specific search query as follows:

Search_keyword = 'danantara since:2025-01-01 until:2025-06-01 lang:id"

2.2. Data Labeling

After collecting the tweet data containing the keyword "Danantara," a manual labeling process was conducted by three annotators to classify user stances into three categories:

(1) Pro category includes tweets that explicitly or implicitly express support, appreciation, or positive views toward "Danantara.".
(2) Contra category includes tweets that express rejection, criticism, or negative views toward "Danantara," which may take the form of disappointment, cynicism, or direct opposition.
(3) Neutral category includes tweets that do not display a clear stance, are purely informative, or do not take a position on the issue.

In cases where disagreements occurred among annotators, a discussion session was held to reach a consensus. If consensus could not be reached immediately, the annotators reviewed the tweet context together and applied predefined labeling guidelines to ensure consistent and reliable annotation.

2.3. EDA (Exploratory Data Analysis)

Figure 2 shows the label distribution in the stance detection task on BPI Danantara, consisting of pro (3,524 tweets), contra (2,221 tweets), and neutral (1,060 tweets). The imbalance, with pro as the majority class may bias the model. therefore, undersampling was applied to reduce the size of the dominant class, which helps decrease computational time and prevents overfitting often observed in oversampling methods. Unlike random oversampling that duplicates identical samples or SMOTE-based techniques that may generate synthetic samples containing noise, undersampling avoids these issues by simply removing redundant majority instances (Wongvorachan & He, 2023).

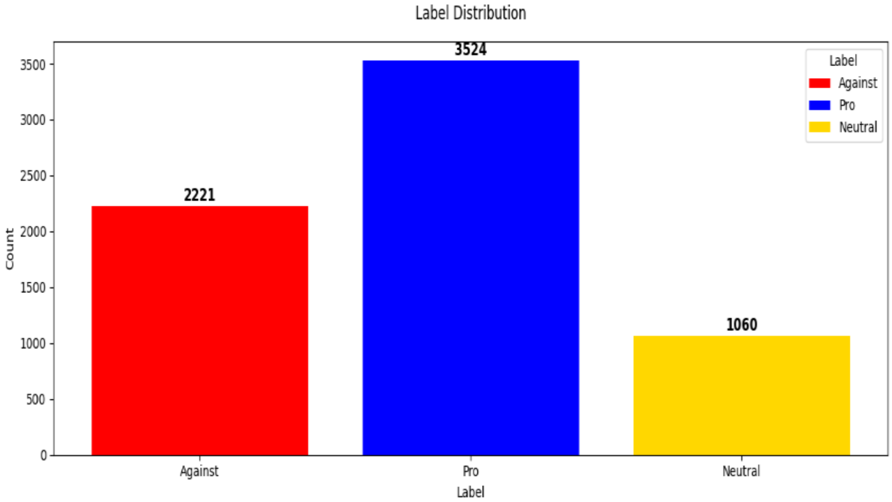


Figure 2. Label Distribution

Figure 3 presents a word cloud of the stance detection dataset on BPI Danantara, with “Danantara” as the most dominant term, highlighting its role as the central focus of discourse. Frequently appearing words such as “Presiden,” “Prabowo,” and “BUMN” reflect the issue’s link to government policies, while terms like “masa depan,” “investasi,” and “ekonomi” are associated with supportive contexts. In contrast, words such as “korupsi,” “tidak,” and “gak” appear in critical or opposing statements, illustrating the diversity of stances ranging from support to opposition and neutrality.

Figure 3. Word Cloud

Figure 4 presents the frequency of the most frequently occurring words in the dataset used for the stance classification task on the BPI Danantara issue. The word “Danantara” has the highest frequency, appearing 4,610 times, which indicates that this entity is the primary focus in the analyzed public discourse. In addition, words such as “dan,” “yang,” and “di” also appear very frequently. However, these words are categorized as stopwords, which are functional words in Indonesian that do not provide significant semantic contribution to stance analysis. The dominance of stopwords in this frequency distribution highlights the importance of applying stopword removal during the preprocessing stage. This step is essential to reduce noise so that the textual features learned by the model better represent the underlying semantics. Moreover, stopword removal is particularly important for Support Vector Machine (SVM) models, as the presence of excessive stopwords can introduce irrelevant features, inflate dimensionality, and ultimately reduce classification performance. Therefore, eliminating stopwords ensures that the SVM model focuses on more informative and discriminative terms (Isnan et al., 2023).

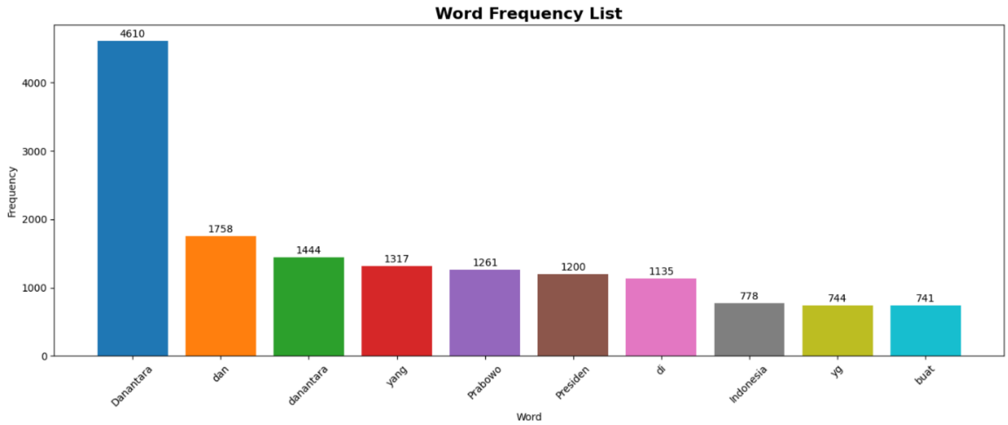


Figure 4. Word Frequency List

Figure 5 presents the frequency of various text patterns in the analyzed set of tweets. The most frequently observed pattern is punctuation, with 6,506 occurrences. Mentions (@) appear 3,891 times, reflecting the common practice of users referring to other accounts in conversations. URLs are recorded 2,167 times, indicating the prevalence of tweets containing links to external sources. Hashtags (#) occur 1,204 times, highlighting their role in grouping topics or signaling discussion themes. Non-ASCII characters appear 181 times, while HTML entities such as "&amp" are found 136 times. Emojis occur 59 times, representing a form of visual expression despite their relatively low frequency. Extra whitespace is observed only once, and no HTML elements are present, suggesting that the text is relatively free from markup. Overall, this chart provides an initial overview of textual characteristics relevant to preprocessing. These text patterns must be carefully addressed during the preprocessing stage, as uncleaned mentions, URLs, emojis, punctuation, or HTML entities can introduce noise, reduce feature quality, and negatively affect the performance of both SVM and IndoBERT models. Proper preprocessing ensures that the models focus on meaningful linguistic content rather than irrelevant or distracting patterns (Chai, 2023).

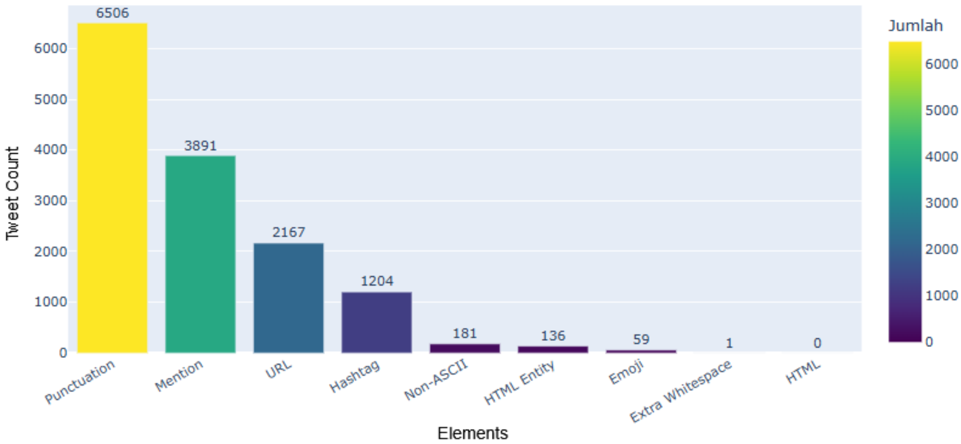


Figure 5. Word Frequency List

2.4. Implementation

This section provides a detailed explanation of the implementation stages in the research on stance detection related to the BPI Danantara issue on the social media platform X, using two methods: Support Vector Machine (SVM) as a machine learning algorithm and IndoBERT as a transformer-based model. The discussion includes the workflow of each model, the research scenarios applied, and the implementation of the code starting from dataset loading, data preprocessing, feature extraction, model training, and performance evaluation. Both methods are then applied to the stance detection task with BPI Danantara as the research object to compare the effectiveness of traditional machine learning approaches and modern transformer models in understanding public opinions on social media.

Support Vector Machine (SVM) is a classification algorithm that identifies an optimal hyperplane to separate data classes while maximizing the margin, which improves model generalization (Ke et al., 2024). SVM can apply kernel functions such as the Radial Basis Function (RBF) kernel to map data into a higher-dimensional feature space and enable non-linear separation (Rochim et al., 2021). The RBF kernel is effective in learning complex and non-linear decision boundaries, allowing separation of data that cannot be linearly divided in the original feature space (Gopi et al., 2023). In text classification, SVM is commonly combined with TF–IDF feature extraction to transform text into numerical representations (Umar & Abdullahi, 2024). TF–IDF assigns higher weights to informative terms while reducing the influence of common words, thereby enhancing SVM performance in text classification tasks (Al-Qablan et al., 2023).

IndoBERT is a deep learning model based on the Transformer architecture and is a variant of BERT specifically designed for the Indonesian language. This model is bidirectional, meaning that in understanding a word, IndoBERT simultaneously considers the context from both directions (left and right) (9). IndoBERT was trained on a total of 220 million words derived from three main sources: Indonesian Wikipedia (74 million words), news articles from Kompas, Tempo, and Liputan6 (55 million words), and an Indonesian web corpus (90 million words) (Purnomo et al., 2024). Architecturally, IndoBERT Base adopts the standard BERT-Base (uncased) configuration consisting of 12 encoder layers, resulting in 12 hidden layers, 12 attention heads, a hidden size of 768, and a feed-forward layer size of 3,072, with a total of approximately 110 million parameters (Syazal & Yulianti, 2025). Before being processed by the model, each input text in IndoBERT is first converted into tokens, and each token is then represented in the form of embeddings, which are a combination of token embeddings, segment embeddings, and position embeddings. These embedding representations serve as the primary input to the model, enabling IndoBERT to understand the structure, position, and contextual meaning of each token in a sentence more accurately (Nabiilah et al., 2024).

The implementation stages of the SVM model for stance classification in this study consisted of the following steps :

(1) Load Dataset, loading a dataset consisting of 6,801 tweets labeled as pro, contra, and neutral.

(2) Text Cleaning, cleaning tweets through removal and normalization processes, including lowercasing all text and removing noise such as URLs, mentions, hashtags, numbers, punctuation marks, emojis, and other unwanted symbols.

(3) Case Folding, converting all characters into lowercase form.

(4) Tokenization, splitting text into tokens in the form of words.

(5) Stopword Removal, removing common words (stopwords).

(6) Stemming, reducing words to their root form using the Sastrawi library.

(7) Undersampling, applied to balance class distribution in the imbalanced dataset.

(8) Dataset Splitting, dividing the dataset into 70% for training, 15% for validation, and 15% for testing the SVM model.

(9) TF-IDF Feature Extraction, applied as the feature extraction method for the SVM model.

(10) The SVM model was trained using the Radial Basis Function (RBF) kernel with a hyperparameter value of C = 1.0 and gamma set to ‘scale’.”

(11) Evaluation was conducted using standard performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. In addition, statistical significance evaluation was performed to examine performance differences between models by applying the paired t-test and the Wilcoxon signed-rank test at a predefined significance level.

In addition, the implementation stages of the IndoBERT model for the stance classification task in this study consisted of the following steps:

(1) Load Dataset, loading a dataset consisting of 6,801 tweets labeled as pro, contra, and neutral.

(2) Text Cleaning, cleaning tweets through removal and normalization processes, including lowercasing all text and removing noise such as URLs, mentions, and hashtags, along with other unwanted symbols, without applying stopword removal or stemming since these processes can degrade IndoBERT’s performance.

(3) Tokenization with IndoBERT Tokenizer, performed using AutoTokenizer.

(4) Transformation into Tensor Format, where the tokenized IndoBERT data are wrapped using the CustomTorchDataset to ensure compatibility with the DataLoader and to maintain tracking of the original data during training and evaluation. In this step, the maximum sequence length is set to 256 tokens.

(5) Undersampling, applied to balance the class distribution in the imbalanced dataset.

(6) Dataset Splitting, dividing the dataset into 70% for training, 15% for validation, and 15% for testing the IndoBERT model.

(7) Data Loading, constructing three DataLoaders (train, validation, test) with a batch size of 32 to process the data incrementally and efficiently during training and evaluation.

(8) Load IndoBERT Model, loading the pretrained model *indobenchmark/indobert-base-p1*.

(9) IndoBERT modeling was conducted by performing fine-tuning on the pre-trained IndoBERT model using a learning rate of 2e-5 and a maximum of 10 training epochs. Early stopping was applied with a patience value of 3 to prevent overfitting. A dropout rate of 0.1 was used to improve model generalization, and the AdamW optimizer was employed during training. The best-performing model was selected based on the highest validation accuracy achieved in a single epoch, using the validation dataset as the model selection criterion. The selected model was then used for evaluation on the test dataset to obtain the final performance results.

(10) Evaluation was conducted using standard performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. In addition, statistical significance evaluation was performed to examine performance differences between models by applying the paired t-test and the Wilcoxon signed-rank test at a predefined significance level.

In this study, four experimental scenarios were designed to evaluate the performance of two classification approaches for the stance detection task on the BPI Danantara topic, namely Support Vector Machine (SVM) with TF-IDF feature representation and IndoBERT as a pretrained transformer-based language model for Indonesian text. Each scenario considers the number of labels, the degree of class imbalance, and variations in the dataset resulting from the preprocessing stage. Detailed information regarding the models used and the label distributions for each scenario is presented in Table 1.

Table 1. Overview of Experimental Scenarios and Class Distributions

Scenario	Model	Label Distribution After Preprocessing
Binary classification without undersampling	SVM (TF-IDF)	Contra: 2194, Pro: 3072
	IndoBERT	Contra: 2023, Pro: 3112
Three-class classification without undersampling	SVM (TF-IDF)	Contra: 2192, Pro: 3072, Neutral: 983
	IndoBERT	Contra: 2203, Pro: 3111, Neutral: 1009
Binary classification with undersampling	SVM (TF-IDF)	Contra: 2194, Pro: 2194
	IndoBERT	Contra: 2023, Pro: 2023
Three-class classification with undersampling	SVM (TF-IDF)	Contra: 983, Pro: 983, Neutral: 983
	IndoBERT	Contra: 1009, Pro: 1009, Neutral: 1009

3. RESULTS AND DISCUSSION

After the training process was conducted for each model, including SVM and IndoBERT, in accordance with the research methodology applied to each scenario, the experimental results are presented in this section. The experiments compare a TF-IDF–based SVM model and IndoBERT across four scenarios, namely binary and three-class classification, each with and without the application of undersampling. The models are evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics to assess their effectiveness under both balanced and imbalanced data conditions.

3.1. Evaluation Results of SVM and IndoBERT in the Binary Classification Scenario without Undersampling

Table 2 shows that IndoBERT slightly outperforms SVM in binary stance detection without undersampling. IndoBERT achieves higher accuracy (93.73%) and F1-score (94.64%) than SVM (93.42% accuracy and 93.32% F1-score), as well as better precision and recall. These results indicate that IndoBERT provides more robust and consistent performance across all evaluation metrics.

Table 2. Summary of Performance Metric Comparison between SVM and IndoBERT in Binary Classification without Undersampling

Metric	SVM (Test)	IndoBERT (Test)
Accuracy	93.42%	93.73%
F1-Score	93.32%	94.64%
Precision	93.03%	94.84%
Recall	94.01%	94.43%

Table 3 shows that although IndoBERT generally achieved higher accuracy and F1-score than SVM across most folds in binary classification without undersampling, the improvements were not consistent in all folds. Statistical testing using paired t-test (p = 0.7386) and Wilcoxon signed-rank test (p = 0.5566) indicated no statistically significant performance difference between the two models, despite IndoBERT's higher average results.

Table 3. Test Fold Results from 10-Fold Cross-Validation of SVM and IndoBERT for Binary Classification without Undersampling

Fold	SVM				IndoBERT			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
1	0.9367	0.9400	0.9367	0.9370	0.9750	0.9787	0.9787	0.9787
2	0.9114	0.9120	0.9114	0.9116	0.9375	0.9200	0.9787	0.9485
3	0.9494	0.9548	0.9494	0.9497	0.9500	0.9778	0.9362	0.9565
4	0.8987	0.9105	0.8987	0.8994	0.8750	0.8776	0.9149	0.8958
5	0.9620	0.9652	0.9620	0.9622	0.9500	0.9778	0.9362	0.9565
6	0.9620	0.9625	0.9620	0.9621	0.9250	0.9362	0.9362	0.9362
7	0.9241	0.9357	0.9241	0.9246	0.9500	0.9778	0.9362	0.9565
8	0.9494	0.9510	0.9494	0.9495	0.9125	0.9333	0.9130	0.9231
9	0.9114	0.9198	0.9114	0.9120	0.9494	0.9773	0.9348	0.9556
10	0.9367	0.9453	0.9367	0.9372	0.9494	0.9375	0.9783	0.9574

3.2. Evaluation Results of SVM and IndoBERT in the Three-Label Classification Scenario without Undersampling

Table 4 illustrates that in the three-label stance classification scenario without undersampling, IndoBERT demonstrates superior overall performance on the test set. IndoBERT achieves higher accuracy (83.35%) and a markedly higher F1-score (78.26%) than SVM, which records an accuracy of 81.98% and an F1-score of 70.72%. The substantial improvement in F1-score indicates that IndoBERT attains a better balance between precision and recall in the multi-class setting.

Table 4. Summary of Performance Metric Comparison between SVM and IndoBERT in the Three-Label Classification Scenario without Undersampling

Metric	SVM (Test)	IndoBERT (Test)
Accuracy	81.98%	**83.35%**
F1-Score	70.72%	**78.26%**
Precision	**82.35%**	78.88%
Recall	70.52%	**77.77%**

Although SVM slightly surpasses IndoBERT in precision, its considerably lower recall reflects a weaker ability to correctly identify instances across all stance categories. In contrast, IndoBERT shows a much higher recall, indicating stronger capability in capturing true positive instances for each class. Overall, IndoBERT demonstrates more balanced and robust performance in handling the complexity of three-class stance classification on the test set.

Table 5 shows that although IndoBERT achieved higher accuracy and F1-score than SVM in several folds of the three-label classification without undersampling, the performance differences varied across folds. Statistical testing using paired t-test (p = 0.3720) and Wilcoxon signed-rank test (p = 0.3750) indicates no statistically significant performance difference between the two models, despite IndoBERT showing higher average results.

Table 5. Test Fold Results from 10-Fold Cross-Validation of SVM and IndoBERT for Three-Label Classification without Undersampling

Fold	SVM				IndoBERT			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
1	0.8085	0.8503	0.8085	0.7889	0.8315	0.7866	0.7776	0.7812
2	0.8404	0.8555	0.8404	0.8194	0.7894	0.7496	0.7009	0.7134
3	0.8297	0.8507	0.8297	0.8205	0.8421	0.7911	0.7847	0.7862
4	0.8510	0.8841	0.8510	0.8335	0.8105	0.7743	0.7272	0.7422
5	0.8085	0.8027	0.8085	0.7839	0.8631	0.8209	0.7959	0.8059
6	0.7978	0.8554	0.7978	0.7624	0.8736	0.8381	0.8363	0.8371
7	0.8297	0.8229	0.8297	0.8084	0.8315	0.7841	0.7928	0.7880
8	0.7978	0.8073	0.7978	0.7704	0.8631	0.8698	0.8071	0.8289
9	0.7849	0.8316	0.7849	0.7687	0.8210	0.7698	0.7789	0.7729
10	0.8494	0.8882	0.8494	0.8301	0.8085	0.7610	0.7747	0.7658

3.3. Evaluation Results of SVM and IndoBERT in the Binary Classification Scenario with Undersampling

Table 6 shows that undersampling slightly improved the performance of both SVM and IndoBERT in the binary stance classification scenario. IndoBERT achieved the highest performance across all evaluation metrics, with an accuracy of 94.55%, an F1-score of 94.52%, precision of 94.55%, and recall of 94.49%. Meanwhile, SVM also showed competitive results, attaining an accuracy of 93.93%, an F1-score of 93.92%, precision of 94.25%, and recall of 93.92%. The consistently higher precision and recall values obtained by IndoBERT suggest a more balanced ability to minimize false positives while effectively identifying true positive instances. Overall, these findings demonstrate that undersampling positively contributed to model performance in the binary classification setting, with IndoBERT remaining the more robust and consistently superior model on the test set.

Table 6. Summary of Performance Metric Comparison between SVM and IndoBERT in the Binary Classification Scenario with Undersampling

Metric	SVM (Test)	IndoBERT (Test)
Accuracy	93.93%	**94.55%**
F1-Score	93.92%	**94.52%**
Precision	94.25%	**94.55%**
Recall	93.92%	**94.49%**

Table 7 shows that in the binary classification setting with undersampling, both models achieved strong performance across folds. IndoBERT attained higher accuracy, recall, and F1-score in several folds, including some with perfect results, while SVM showed more consistent performance across different splits. Statistical testing using paired t-test (p = 0.7404) and Wilcoxon signed-rank test (p = 0.6250) indicates no statistically significant performance difference between the two models, despite IndoBERT outperforming in several individual folds.

Table 7. Test Fold Results from 10-Fold Cross-Validation of SVM and IndoBERT for Binary Classification with Undersampling

Fold	SVM				IndoBERT			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
1	0.9697	0.9714	0.9697	0.9697	0.9701	0.9429	1.0000	0.9706
2	0.8333	0.8490	0.8333	0.8314	0.9242	0.9118	0.9394	0.9254
3	0.9242	0.9342	0.9242	0.9238	0.9697	0.9429	1.0000	0.9706
4	0.9394	0.9459	0.9394	0.9392	0.8788	0.9032	0.8485	0.8750
5	0.9242	0.9278	0.9242	0.9241	1.0000	1.0000	1.0000	1.0000
6	0.9394	0.9459	0.9394	0.9392	0.9545	0.9412	0.9697	0.9552
7	0.9848	0.9853	0.9848	0.9848	0.9394	0.9677	0.9091	0.9375
8	0.9848	0.9853	0.9848	0.9848	0.9394	0.9394	0.9394	0.9394
9	0.9394	0.9410	0.9394	0.9393	0.8788	0.9032	0.8485	0.8750

Fold	SVM				IndoBERT			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
10	0.9538	0.9577	0.9538	0.9537	1.0000	1.0000	1.0000	1.0000

3.4. Evaluation Results of SVM and IndoBERT in the Three-Label Classification Scenario with Undersampling

Table 8 shows that IndoBERT outperforms SVM by a substantial margin in the three-label stance classification scenario with undersampling. IndoBERT significantly outperformed SVM across all evaluation metrics, achieving an accuracy of 79.56%, an F1-score of 79.34%, precision of 79.46%, and recall of 79.56%. In contrast, SVM obtained considerably lower results, with an accuracy of 72.17%, an F1-score of 72.73%, precision of 74.04%, and recall of 72.18%. The consistent improvements in both precision and recall indicate IndoBERT's superior capability in capturing true stance instances while reducing misclassifications across all three classes. Although IndoBERT consistently outperformed SVM, the results indicate a decrease in performance for both models compared to the three-label classification without undersampling, suggesting that undersampling reduces the ability of both models to detect pro, contra, and neutral stances.

Table 8. Summary of Performance Metrics Comparison between SVM and IndoBERT in the Three-Label Classification Scenario with Undersampling

Metric	SVM (Test)	IndoBERT (Test)
Accuracy	72.17%	**79.56%**
F1-Score	72.73%	**79.34%**
Precision	74.04%	**79.46%**
Recall	72.18%	**79.56%**

Table 9 shows that in the three-label classification setting with undersampling, IndoBERT achieved higher accuracy and F1-score than SVM across most folds, with particularly notable improvements in several folds. While SVM produced competitive results in a few folds, its performance was generally lower and more variable. Statistical testing using paired t-test ($p = 0.0426$) and Wilcoxon signed-rank test ($p = 0.0488$) indicates a statistically significant performance difference between the two models. These results demonstrate that IndoBERT benefits more effectively from undersampling and provides significantly better performance than SVM in this experimental setting.

Table 9. Test Fold Results from 10-Fold Cross-Validation of SVM and IndoBERT for Three-Label Classification with Undersampling

Fold	SVM				IndoBERT			
	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
1	0.9697	0.9714	0.9697	0.9697	0.9701	0.9429	1.0000	0.9706
2	0.8333	0.8490	0.8333	0.8314	0.9242	0.9118	0.9394	0.9254
3	0.9242	0.9342	0.9242	0.9238	0.9697	0.9429	1.0000	0.9706
4	0.9394	0.9459	0.9394	0.9392	0.8788	0.9032	0.8485	0.8750
5	0.9242	0.9278	0.9242	0.9241	1.0000	1.0000	1.0000	1.0000
6	0.9394	0.9459	0.9394	0.9392	0.9545	0.9412	0.9697	0.9552
7	0.9848	0.9853	0.9848	0.9848	0.9394	0.9677	0.9091	0.9375
8	0.9848	0.9853	0.9848	0.9848	0.9394	0.9394	0.9394	0.9394
9	0.9394	0.9410	0.9394	0.9393	0.8788	0.9032	0.8485	0.8750
10	0.9538	0.9577	0.9538	0.9537	1.0000	1.0000	1.0000	1.0000

3.5. Comparison of ROC-AUC between SVM and IndoBERT in the Binary Classification Scenario

Table 10 presents a comparison of ROC-AUC values between the SVM and IndoBERT models in the binary classification scenario (pro and contra) for stance detection on the BPI Danantara issue, both with and without the application of undersampling. The results indicate that IndoBERT (without undersampling) achieved the highest AUC score of 0.9763, slightly outperforming SVM (without undersampling), which obtained an AUC of 0.9660. When undersampling was applied, both models experienced a decrease in AUC, with SVM reaching 0.9654 and IndoBERT 0.9744. Despite this performance decline due to undersampling, IndoBERT consistently demonstrated better performance than SVM under both conditions.

Table 10. Comparison of ROC-AUC between SVM and IndoBERT in the Binary Classification Scenario

No.	Model	AUC (ROC)
1	SVM (Without Undersampling)	0.9660
2	IndoBERT (Without Undersampling)	0.9763
3	SVM (With Undersampling)	0.9654
4	IndoBERT (With Undersampling)	0.9744

3.6. Comparison of ROC-AUC between SVM and IndoBERT in the Three-Label Classification Scenario

Table 11 presents the ROC–AUC comparison of SVM and IndoBERT in the three-label classification scenario (pro, contra, neutral). IndoBERT achieved the highest performance with an AUC of 0.9374 without undersampling, outperforming SVM, which obtained an AUC of 0.9184. When undersampling was applied, the AUC values of both models decreased to 0.8848 for SVM and 0.9226 for IndoBERT. Despite this decline, IndoBERT consistently outperformed SVM. The larger performance drop compared to the binary classification scenario indicates the increased difficulty of accurately classifying the neutral class in the three-label setting.

Table 11. Comparison of ROC-AUC between SVM and IndoBERT in the Binary Classification Scenario

No.	Model	AUC (ROC)
1	SVM (Without Undersampling)	0.9184
2	IndoBERT (Without Undersampling)	0.9374
3	SVM (With Undersampling)	0.8848
4	IndoBERT (With Undersampling)	0.9226

3.7. Overview of the Comparative ROC-AUC between the SVM and IndoBERT Methods

Figure 6 illustrates the comparison of AUC (ROC) values between SVM and IndoBERT in binary and three-label classification, both with and without undersampling. In the binary scenario, both models achieved excellent results with AUC values exceeding 0.96, where IndoBERT slightly outperformed SVM by recording 0.9763 without undersampling and 0.9744 with undersampling, compared to SVM's 0.9660 and 0.9654. In the three-label classification, the decline was more pronounced, particularly for SVM, which dropped to 0.9184 without undersampling and 0.8848 with undersampling. IndoBERT, however, remained more stable with AUC values of 0.9374 and 0.9226 under the same conditions. These findings confirm that IndoBERT consistently achieves higher AUC values than SVM across all scenarios and demonstrates stronger resilience to performance degradation when undersampling is applied.

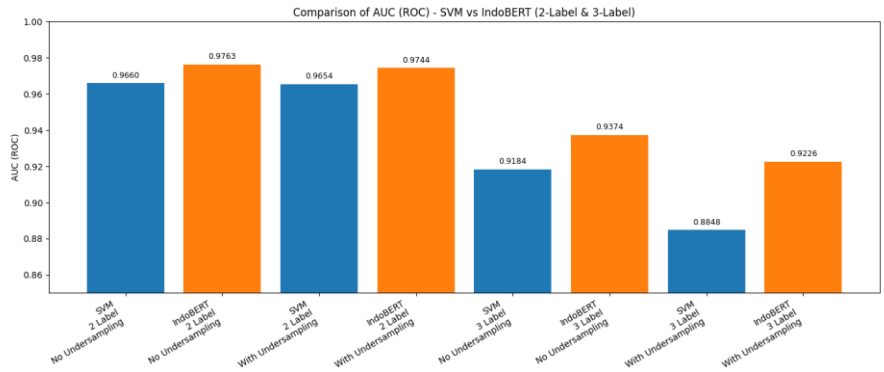


Figure 6. Overall Overview of the Comparative ROC-AUC between the SVM and IndoBERT Methods

4. CONCLUSIONS AND RECOMMENDATIONS

This study compares the performance of the Support Vector Machine (SVM) and IndoBERT models in a stance detection task on the BPI Danantara issue using data from the X social media platform. Overall, IndoBERT consistently outperformed SVM across all scenarios. In binary classification, IndoBERT achieved 93.73% accuracy and 94.64% F1-score without undersampling, and improved to 94.55% accuracy and 94.52% F1-score with undersampling, compared to SVM's 93.42%–

93.93% accuracy and 93.32%–93.92% F1-score. In three-label classification, IndoBERT also showed clear superiority, achieving 83.35% accuracy and 78.26% F1-score without undersampling, and 79.56% accuracy and 79.34% F1-score with undersampling, substantially higher than SVM's 81.98% and 72.17% accuracy, and 70.72% and 72.73% F1-score, respectively. However, based on statistical tests using the paired t-test and the Wilcoxon signed-rank test, the performance differences between the two models in most scenarios were not statistically significant. A statistically significant difference was observed only in the three-label classification scenario with undersampling, indicating that IndoBERT is significantly more effective in handling multi-class stance detection on balanced data. These findings suggest that IndoBERT has an advantage in modeling the complex contextual characteristics of the Indonesian language, although further validation using larger datasets and additional experiments is required to strengthen the generalizability of the results.

REFERENCES

Al-Qablan, T. A., Mohd Noor, M. H., Al-Betar, M. A., & Khader, A. T. (2023). A survey on sentiment analysis and its applications. In *Neural Computing and Applications* (Vol. 0123456789). Springer London. https://doi.org/10.1007/s00521-023-08941-y

Alkhraiji, A. K., & Azmi, A. M. (2025). Stance Detection in Arabic Tweets: A Machine Learning Framework for Identifying Extremist Discourse. *Mathematics*, *13*(18), 2965. https://doi.org/https://doi.org/10.3390/math13182965

Alqurashi, T. (2022). *Stance Analysis of Distance Education in the Kingdom of Saudi Arabia during the COVID-19 Pandemic Using Arabic*.

Alturayeif, N., Luqman, H., & Ahmed, M. (2023). A systematic review of machine learning techniques for stance detection and its applications. *Neural Computing and Applications*, *35*, 5113–5144. https://doi.org/https://doi.org/10.1007/s00521-023-08285-7

Anggraheni, P., Setyowati, N. T., & Harry, H. (2022). *SOCIAL MEDIA AND POLITICAL PARTICIPATION IN INDONESIA : RESTRICTIONS ACCESS AT ANNOUNCEMENT RESULTS OF 2019 PRESIDENTIAL ELECTION*. *2*(July). https://doi.org/10.56353/aspiration.v2i1.23

Ayunia, C., Pellokila, Z., Sugiharti, D. K., & Kusumawardhani, N. (2025). *Legal Implications of The Danantara Superholding Model on Public Accountability in State-Owned Bank Asset Management in Indonesia*. *5*(5), 3721–3730.

Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, *29*(3), 509–553. https://doi.org/10.1017/S1351324922000213

Cotfas, L.-A., Delcea, C., Gherai, R., & Roxin, I. (2021). Unmasking People's Opinions behind Mask-Wearing during COVID-19 Pandemic—A Twitter Stance Analysis. *Symmetry*, *13*(11). https://doi.org/https://doi.org/10.3390/sym13111995

Gopi, A. P., Jyothi, R. N. S., Narayana, V. L., & Et, A. (2023). Classification of tweets data based on polarity using improved RBF kernel of SVM. *International Journal of Information Technology*, *15*, 965–980. https://doi.org/10.1007/s41870-019-00409-4

Hamad, O., Hamdi, A., Hamdi, S., & Shaban, K. (2022). StEduCov: An Explored and Benchmarked Dataset on Stance Detection in Tweets towards Online Education during COVID-19 Pandemic. *Big Data and Cognitive Computing*, *6*(3), 88. https://doi.org/https://doi.org/10.3390/bdcc6030088

Hidayat, A., Alhakim, B. A., Litamahuputty, J. V., & Patty, P. (2025). *Effectiveness Of Danantara Policy In Creating Golden Indonesia 2045*. *4*(1), 69–74.

Isnan, M., Natanael, G., & Pardamean, B. (2023). ScienceDirect ScienceDirect Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model. *Procedia Computer Science*, *227*, 168–175. https://doi.org/10.1016/j.procs.2023.10.514

Ke, T., Ge, X., Yin, F., Zhang, L., Zheng, Y., Zhang, C., Li, J., Wang, B., & Wang, W. (2024). A general maximal margin hyper-sphere SVM for multi-class classification. *Expert Systems with Applications*, *237*(Part C). https://doi.org/https://doi.org/10.1016/j.eswa.2023.121647

Maulana, A., Durya, N. P. M. A., Pratiwi, R. D., & Andika, A. D. (2025). Reaksi Pasar Saham Terhadap Peresmian Danantara pada Himpunan Bank Negara. *Permana : Jurnal Perpajakan, Manajemen, Dan Akuntansi*, *17*(3), 1506–1519. https://doi.org/https://doi.org/10.24905/permana.v17i3.1033

Nabiilah, G. Z., Alam, I. N., Purwanto, E. S., & Hidayat, M. F. (2024). *Indonesian multilabel classification* *using IndoBERT embedding and MBERT classification*. *14*(1), 1071–1078. https://doi.org/10.11591/ijece.v14i1.pp1071-1078

Purnomo, T. D., Sutopo, J., & History, A. (2024). *COMPARISON OF PRE-TRAINED BERT-BASED TRANSFORMER MODELS FOR REGIONAL*. *3*(3), 11–21.

Rochim, A. F., Widyaningrum, K., & Eridani, D. (2021). *Performance Comparison of Support Vector Machine Kernel Functions in Classifying COVID-19 Sentiment*. 224–228. https://doi.org/10.1109/ISRITI54043.2021.9702845

Setiawati, P. A., Imad, S., & Polisi, T. (2024). Analisis Sentimen Pengangguran di Indonesia Selama dan Setelah COVID-19 pada X (Twitter) Menggunakan Naïve Bayes dan Support Vector Machine. *Jurnal Sistem Informasi Dan Informatika*, *6*(2), 662–675. https://doi.org/10.51519/journalisi.v6i2.713

Simanjuntak, J. M., & Widyadhana, K. N. (2025). *Danantara , the SOE Superholding , and the Pillar of Indonesia ' s Economic Future ?* *6*(4), 1163–1182.

Sulistyo, A. O., & Pamungkas, E. W. (2025). *Pemanfaatan stance detection untuk mitigasi dampak echo chamber effect dalam media sosial x berbasis web*. *10*(4), 3211–3220.

Syazal, M. R., & Yulianti, E. (2025). *Classification of Economic Activities in Indonesia Using IndoBERT Language Model*. *2*, 155–165.

Umar, M., & Abdullahi, A. (2024). Special Issue on Computing & Advances in Information Technology. *SLUJST (Sri Lanka University Journal of Science & Technology)*, *4*(1). https://doi.org/10.56471/SLUJST.V4I.266

Vincent, J., Vincent, J., Santoso, M. I., Pohan, S., Hasani, M. F., Maulina, A., Vincent, J., Ivan, M., Pohan, S., Fikri, M., & Maulina, A. (2025). ScienceDirect ScienceDirect ScienceDirect Extractive Indonesian News Text Summarization using Extractive Indonesian News MBERT , Text Summarization using and RoBERTa Extractive Indonesian News Text Summarization using and. *Procedia Computer Science*, *269*, 1087–1096. https://doi.org/10.1016/j.procs.2025.09.050

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*, 5731–5780. https://link.springer.com/article/10.1007/S10462-022-10144-1

Wongvorachan, T., & He, S. (2023). *A Comparison of Undersampling , Oversampling , and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining*. https://doi.org/https://doi.org/10.3390/info14010054