Vol. 8, No. 1, February 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Event-Based Detection of Provocative Political Discourse on Indonesian Twitter: A Comparative Study of SVM and IndoBERT

Evril Fadrekha Cahyani[1], Ali Nur Ikhsan[2], Deuis Nur Astrida[3]

[1]Information Systems Study Program, Universitas Amikom Purwokerto, Indonesia

[2]Department of Informatics, Universitas Amikom Purwokerto, Indonesia

[3]Department of Information Technology, Universitas Amikom Purwokerto, Indonesia

Received:
November 15, 2025
Revised:
January 12, 2026
Accepted:
January 28, 2026
Published:
February 24, 2026

Corresponding Author:
Author Name*:
Ali Nur Ikhsan
Email*:
alinurikhsan@amikompurwokerto.ac.id

DOI:
10.63158/journalisi.v8i1.1409

Abstract. Political polarization on Indonesian social media intensified during the August 2025 House of Representatives (DPR) demonstrations, where provocative and sarcastic tweets helped amplify institutional criticism and widen public conflict. This study examines event-based automatic detection of provocative political discourse by comparing a feature-based Support Vector Machine (SVM) classifier with a transformer-based IndoBERT model on a large-scale Indonesian Twitter (X) corpus collected from 15 August to 15 September 2025. Tweets were preprocessed and labeled using a rule-based proxy lexicon to distinguish provocative from neutral content, then both models were trained and evaluated under the same experimental setting. Results show that SVM is highly effective for recognizing explicit provocation expressed through repetitive and lexically salient slogans, whereas IndoBERT provides more stable detection of implicit and context-dependent provocation, including irony and sarcasm that are common in Indonesian political talk online. In addition, temporal exploration indicates sharp spikes in tweet volume that align with key offline protest moments, suggesting a close coupling between street-level mobilization and digital discourse dynamics. Overall, the findings support the use of contextual NLP models within event-centered social media analysis to strengthen scalable monitoring of polarization and to inform early-warning approaches for escalating conflict in Indonesia's digital public sphere.

Keywords: Provocative discourse; Political polarization; IndoBERT; Support Vector Machine; Event-based Twitter analysis

1. INTRODUCTION

The rapid advancement of information technology has reconfigured the role of conventional media and elevated digital platforms into a central public sphere for political discourse. In this environment, citizens are no longer positioned merely as audiences; they actively access, produce, and circulate political information at scale [1], [2]. In Indonesia, X (formerly Twitter) is particularly influential because it accelerates agenda formation, shapes public opinion, and can even steer what becomes salient in mainstream media coverage [3], [4]. Yet the very openness that enables broad participation also lowers the friction for harmful communication—allowing hate speech, provocation, and polarization to spread quickly, repeatedly, and often strategically [5]. The core problem, therefore, is not simply that political talk has moved online, but that the digital public sphere increasingly rewards inflammatory engagement, making provocative discourse more visible and more impactful than deliberative debate.

This risk became especially visible during the August 2025 DPR controversy, when online discussions surged and were dominated by provocative, sarcastic, and derisive rhetoric directed at the institution's perceived performance. Beyond spontaneous public frustration, the intensity, repetition, and framing patterns observed in such moments can signal coordinated political communication—where particular narratives are amplified to manufacture consensus, intensify distrust, or sustain outrage cycles [6], [7]. In other words, the challenge is twofold: (1) identifying provocative political speech as a linguistic phenomenon (often implicit, sarcastic, or context-dependent), and (2) understanding how amplification dynamics—potentially including buzzer activity—shape what becomes prominent in the conversation.

Addressing this challenge requires an automated system capable of detecting provocative tweets accurately and consistently in large-scale streams. Machine learning and Natural Language Processing (NLP) provide practical foundations for such detection because they enable the classification of massive text corpora with systematic criteria [8], [9]. However, model choice matters because provocation in political discourse is rarely expressed only through explicit slurs or direct threats; it often appears as insinuation, ridicule, coded language, or sarcasm. Classical approaches such as Support Vector Machine (SVM) remain attractive due to their efficiency and strong baseline performance

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

in text classification, but they can struggle to capture semantic nuance when meaning depends on context, irony, or pragmatic cues [10]. Transformer-based architectures, by contrast, can represent contextual relationships between words and better infer implicit meaning—making models such as IndoBERT promising for Indonesian-language political discourse, where sarcasm and indirectness are common rhetorical strategies [11], [12].

Despite growing attention to harmful and provocative discourse, an important gap remains in how Indonesian Twitter studies operationalize provocation within specific political episodes and the localized dynamics that accompany them. Prior work by Rahman et al. [13], Miqdad [14], and Saputra & Sibaroni [15] has provided valuable foundations, but their emphases leave unresolved issues for event-driven political provocation. Rahman et al. [13] showed that SVM can effectively classify hate speech, yet their analysis did not engage with amplification dynamics that may be buzzer-driven within public policy debates. Miqdad [14] examined the role of political buzzers in spreading provocative narratives, but did not integrate machine-learning detection to model these patterns at scale. Saputra & Sibaroni [15] demonstrated that BERT-based models outperform SVM in capturing semantic context, yet their approach was not applied to local political cases shaped by mass mobilization, institutional conflict, and rapidly shifting frames. Consequently, the literature still lacks an integrated empirical framework that simultaneously (a) anchors analysis in a real-world political event, (b) compares feature-based and contextual NLP models for provocation detection, and (c) incorporates interaction/temporal signals relevant to buzzer amplification and polarization.

This study addresses that gap through an event-based design centered on the DPR demonstrations, using a large-scale dataset collected from 15 August to 15 September 2025. Rather than treating provocation as a generic category detached from context, the study positions provocative discourse as an event-contingent phenomenon shaped by real-time contestation, sarcasm, and narrative competition in Indonesia's digital politics. Methodologically, it contributes a comparative evaluation of SVM and IndoBERT, pairing feature-based classification with contextual embeddings to test how each approach performs when faced with implicit meaning, rhetorical provocation, and sarcasm in Indonesian political communication. Substantively, it extends detection beyond text-only categorization by examining temporal patterns and user interactions as signals that may reflect amplification behavior more realistically in the digital public sphere [6], [7].

Based on this framework, the study aims to detect and classify provocative tweets related to the August 2025 DPR issue using SVM and IndoBERT, and to compare the performance of both approaches in capturing provocative linguistic patterns and semantic context in Indonesia's online political discourse [12], [16]. By integrating event grounding, comparative NLP modeling, and attention to amplification dynamics, the study offers a more context-sensitive and empirically actionable approach to understanding—and detecting—provocative political discourse on Indonesian Twitter.

2. METHODS

This study adopts a quantitative experimental design employing machine learning based text mining techniques to detect provocative tweets related to the August 2025 DPR issue. This approach facilitates the analysis of linguistic and emotional patterns in social media texts through empirically optimized classification algorithms [13]. The experiment implements two models, Support Vector Machine (SVM) and IndoBERT, to compare the performance of a feature-based classifier with that of a contextual-embedding model in identifying provocative Indonesian-language tweets [15]. Figure 1 shows how this research flow diagram.

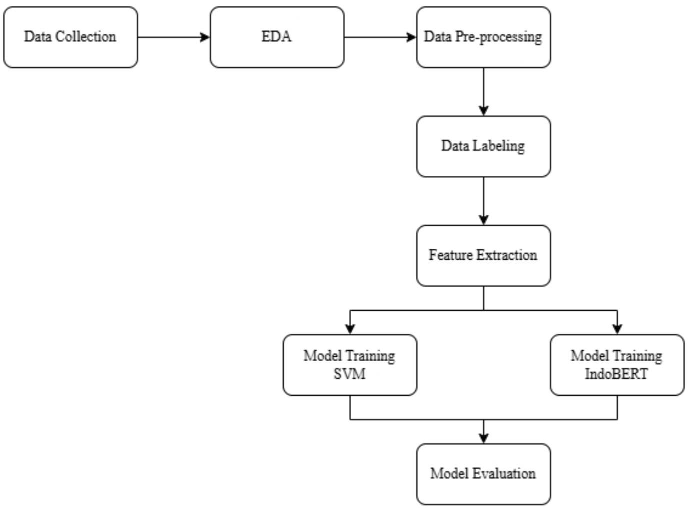

Figure 1. Research flow diagram

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

2.1 Data Collection

Data were collected using the Twitter API and the n8n automation platform to scrape tweets based on the query "*dpr*" OR *"bubarkan dpr"* with an Indonesian-language filter (lang:id) during the period 15 August–15 September 2025. The resulting dataset comprises several thousand public tweets relevant to the demonstrations and online discourse surrounding the DPR. Each entry includes key attributes such as upload date, username, tweet content, and metadata (retweets, replies, and likes). Tweets containing media (images or videos) were removed to maintain the analytical focus on textual content [17].

Table 1. Dataset size at each processing stage

Processing Stage	Tweets	Description
Raw data	27,496	Tweets collected directly from Twitter API using keyword queries related to DPR
Cleaned data	26,435	After removing duplicates, empty tweets, and media-only tweets
Labeled - Neutral	17,409	Tweets without provocative keywords based on rule-based labeling
Labeled - Provocative	9,502	Tweets containing provocative or conflict-inducing expressions

2.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to examine tweet volume dynamics, user engagement patterns, and topic distributions prior to preprocessing and model development. Temporal analysis of tweet frequencies constitutes a fundamental step in social media research, as it enables the identification of bursty communication patterns and peak activity periods associated with salient offline political events. In particular, trend analysis of tweet volumes can reveal pronounced surges surrounding politically significant moments, thereby situating online discourse within real-world political timelines. Furthermore, EDA of issue-focused Twitter corpora facilitates the identification of salient keywords and content clusters, which subsequently inform preprocessing strategies and feature extraction for downstream modeling [18], [19].

2.3 Data Preprocessing

Data preprocessing was applied to normalize linguistic variation and reduce noise typical of Indonesian Twitter discourse prior to feature extraction and model training. This design follows standardized practices established in Indonesian NLP benchmarks such as IndoNLU and IndoLEM, which emphasize consistent normalization, token handling, and morphology-aware processing to ensure robust downstream evaluation across diverse Indonesian language tasks [20], [21]. Empirical studies on Indonesian social-media and sentiment datasets further demonstrate that operations such as case folding, normalization, stopword removal, and stemming mitigate orthographic and morphological variability (e.g., informal spellings, affixation, and word elongation), thereby improving the reliability of text representations for classification models [22], [23]. Accordingly, the standardized preprocessing pipeline summarized in Table 2 was adopted to generate a refined textual field for both TF-IDF-based SVM and transformer-based IndoBERT.

Table 2. Data preprocessing steps

Step	Description
Case folding	Converting all text to lowercase.
Cleaning	Removing URLs, numbers, punctuation, emojis, and non-alphabetic symbols.
Tokenization	Splitting text into individual word units for subsequent processing.
Stopwords removal	Eliminating common words with minimal semantic contribution (e.g., *yang, dan, di*) using the indonesian stopword list from the nltk library.
Stemming	Reducing words to their base form using the Sastrawi library. Example: *bubarkan* → *bubar*.
Slang normalization	Converting non-standard or informal words into their formal equivalents (e.g., *ga* → *tidak*, dprrr → dpr).

2.4 Data Labeling

Tweet labeling was conducted automatically using a rule-based proxy approach that detects the presence of provocative words or phrases based on a lexicon of keywords compiled from observations of DPR-related discourse and prior studies. Tweets containing terms with negative, sarcastic, or conflict-inducing connotations were assigned the label 1 (provocative), while the remainder were labeled 0 (neutral). This approach was selected to enable efficient labeling of a large dataset without manual annotation [24]. Keywords such as *bubar DPR* and *tunjangan DPR* were incorporated into a rule-based function to generate binary labels.

2.5 Feature Extraction and Data Representation

This stage transforms the cleaned text into numerical representations suitable for machine learning algorithms. For the SVM model, TF-IDF (with 1–2 n-grams and max_features = 5000) was used as the vector representation. This method emphasizes words that hold significant weight in distinguishing between provocative and non-provocative classes [25]. For the IndoBERT model, the built-in tokenizer was used to convert text into subword-based tokens while capturing bidirectional contextual relationships among words [26].

2.6 Model Construction and Training

The SVM model was implemented using the scikit-learn library with a linear kernel. The model was trained using TF-IDF features and optimized through an 80:20 train–test split, where 80% of the data was used for training and 20% for performance evaluation. Class imbalance was addressed by applying class_weight = 'balanced' to compensate for disparities between provocative and neutral labels [27]. The 80:20 ratio was chosen because it provides sufficient training data to optimize model learning while maintaining a representative test set for independent validation, aligning with standard practices in text classification. TF-IDF vectorization was fitted exclusively on the training data and then applied to the test set to prevent information leakage.

IndoBERT was fine-tuned using the HuggingFace Transformers library based on the indobenchmark/indobert-base-p1 model. The model was trained for three epochs with a per-device batch size of 16 for both training and evaluation. Training was conducted using mixed-precision (FP16) optimization, with 500 warm-up steps and a weight decay of 0.01 to improve training stability and generalization. Fine-tuning was performed on Google Colab using an NVIDIA Tesla T4 GPU with 16 GB VRAM. The model was configured for binary classification (provocative vs. neutral) and trained using the same 80:20

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

stratified train–test split as the SVM model to ensure a consistent comparative evaluation [28].

2.7 Model Evaluation

Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score, complemented by a confusion matrix to provide a detailed breakdown of classification results. These metrics were used to assess each model's effectiveness in distinguishing between provocative and neutral tweets [29].

3. RESULTS AND DISCUSSION

3.1. Performance Evaluation

This section presents the empirical outcomes of preprocessing, exploratory analysis, labeling, and supervised model evaluation for the August 2025 DPR Twitter corpus. From the initial 27,496 raw records, a total of 26,435 tweets were retained after cleaning, deduplication, and normalization. The final dataset was then divided into two automatically assigned classes—neutral (64.7%) and provocative (35.3%)—based on a rule-based labeling scheme using DPR-related provocative keywords. The resulting class balance is shown in Figure 2, which indicates that although neutral tweets dominated the conversation, more than one-third of the corpus contained language patterns associated with provocation, signaling a substantial presence of conflict-oriented discourse within the event-driven public sphere.

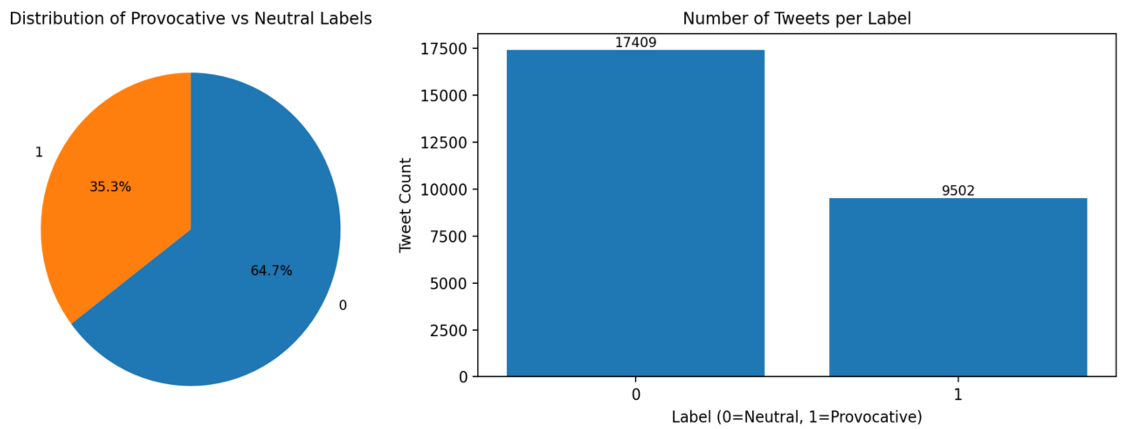

Figure 2. Distribution of tweet labels in the cleaned dataset

Exploratory temporal analysis demonstrates that tweet activity was not evenly distributed across the observation window. As illustrated in Figure 3, discussion volume fluctuated sharply, with the most pronounced spike occurring on 31 August 2025. This surge coincided with heightened offline mobilization and a large-scale demonstration responding to DPR policy proposals, suggesting that online political discourse closely tracked major real-world political triggers rather than evolving gradually. In practical terms, this temporal clustering implies that provocative content detection systems must remain reliable under bursty conditions—precisely when attention, emotional intensity, and amplification behavior tend to increase. Complementing this temporal insight, dominant lexical patterns also reflect issue concentration: terms such as *dpr*, *rakyat*, *demo*, and *anggota dpr* were among the most frequently occurring tokens, pointing to themes of institutional critique, representation, and dissatisfaction. These dominant terms are summarized visually in Figure 4, where the word cloud highlights the rhetorical center of the discourse. In addition, engagement exploration revealed a strong positive association between retweets and likes, indicating that messages gaining visibility through retweets tended to accumulate endorsement signals as well—an amplification pattern consistent with viral circulation dynamics during contentious political moments.

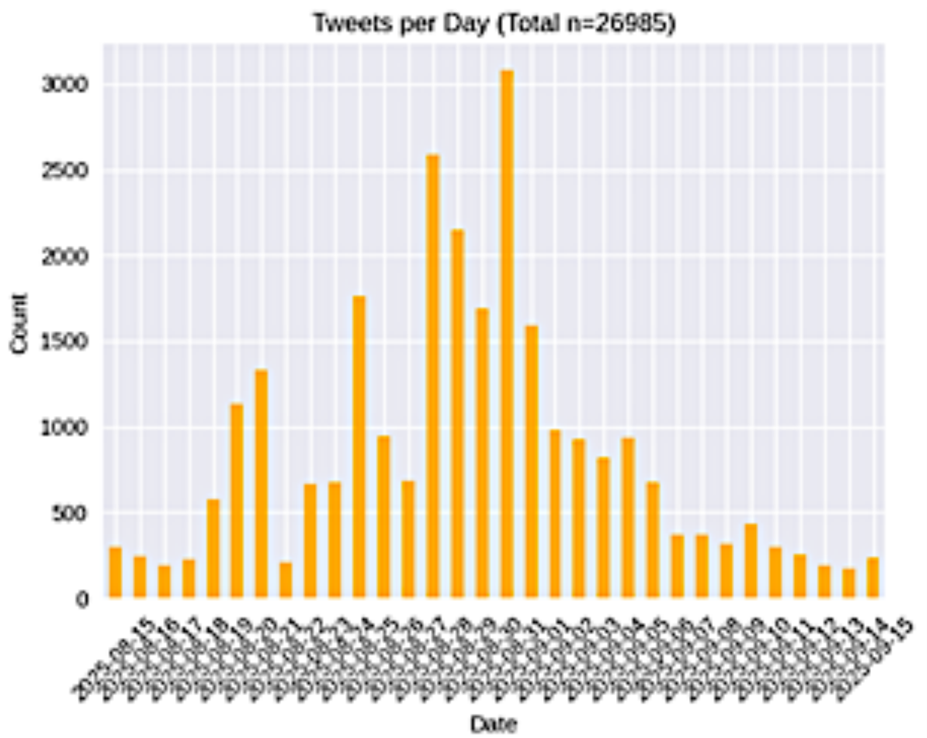

Figure 2. Daily tweet distribution related to the August 2025 DPR issue

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

Figure 3. Word cloud of dominant terms in DPR-related tweets

Data preprocessing played a critical role in improving corpus quality for modeling. The pipeline removed duplicates, URLs, symbols, and informal variants, reducing textual noise and producing a more standardized representation for downstream classification. After normalization, the rule-based labeling procedure identified 9,502 tweets as provocative, primarily triggered by conflict-laden expressions (e.g., *bubar DPR*, *tunjangan DPR*) and similar rhetorical markers of confrontation. While this labeling approach is intentionally simple and scalable for event-level datasets, the distribution in Figure 2 and the activity spikes in Figure 3 underscore that provocation in this corpus is not merely incidental; it is temporally concentrated and thematically anchored in the institutional controversy itself.

Using the labeled dataset, both SVM and IndoBERT were trained and evaluated under identical experimental conditions using an 80:20 train–test split. The SVM model employed TF–IDF n-gram representations and delivered very strong results, reflecting its effectiveness in capturing explicit lexical signals and recurrent provocative phrases. IndoBERT, fine-tuned on the same dataset, achieved slightly higher and more consistent performance across both classes, suggesting greater robustness for handling indirect, context-dependent, or rhetorically nuanced expressions. A direct comparison of the two models is reported in Table 3, which summarizes accuracy, precision, recall, F1-score, AUC, and 5-fold cross-validation. Overall, both models achieved near-ceiling performance; however, IndoBERT showed the strongest overall stability, particularly reflected in its higher cross-validation score and marginally improved overall metrics.

Table 3. Model performance comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC	Validation
SVM	0.995	0.994	0.996	0.995	0.998	0.992
IndoBERT	0.998	0.998	0.998	0.998	0.999	0.997

Table 3 also indicates that both approaches produced excellent class-wise balance (precision and recall are highly symmetric), implying that the models were not merely overpredicting the majority class. This is especially important given the dataset composition in Figure 2, where neutral tweets are more prevalent. The ROC-based evaluation further supports this conclusion: the ROC curve for the SVM classifier in Figure 5 demonstrates near-perfect separability, consistent with the high AUC value reported in Table 3. Meanwhile, error analysis via confusion matrices confirms that misclassifications were rare. In particular, the confusion matrix for IndoBERT shown in Figure 6 indicates fewer errors in borderline or ambiguous cases compared to the feature-based baseline, aligning with the expectation that transformer embeddings better encode semantics beyond surface-level keyword cues.

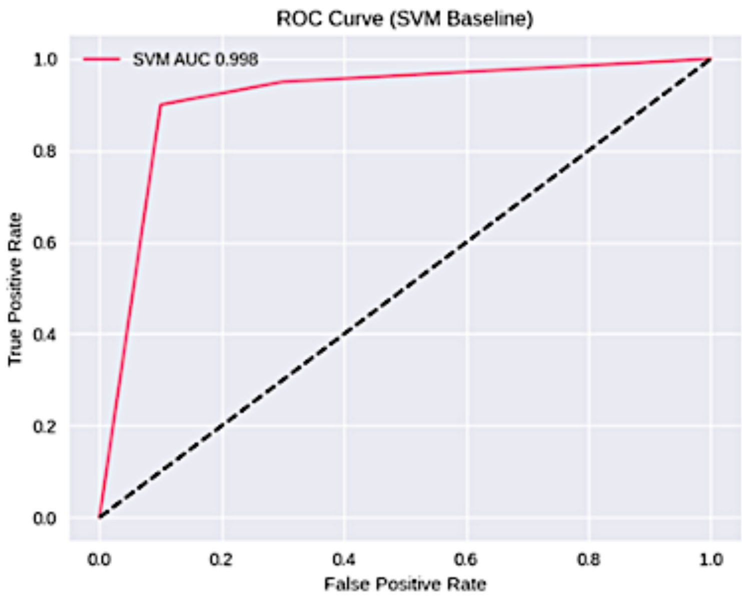

Figure 5. ROC curve of the SVM classifier

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

Taken together, the results demonstrate that both feature-based and transformer-based methods are highly effective for large-scale detection of provocative political discourse in Indonesian Twitter data. The quantitative evidence in Table 3, supported by separability in Figure 5 and error patterns in Figure 6, suggests that SVM remains an efficient high-performing baseline when provocation is expressed explicitly, whereas IndoBERT provides additional robustness when provocation emerges indirectly through sarcasm, implication, or contextual framing—conditions that are common in event-driven political communication.

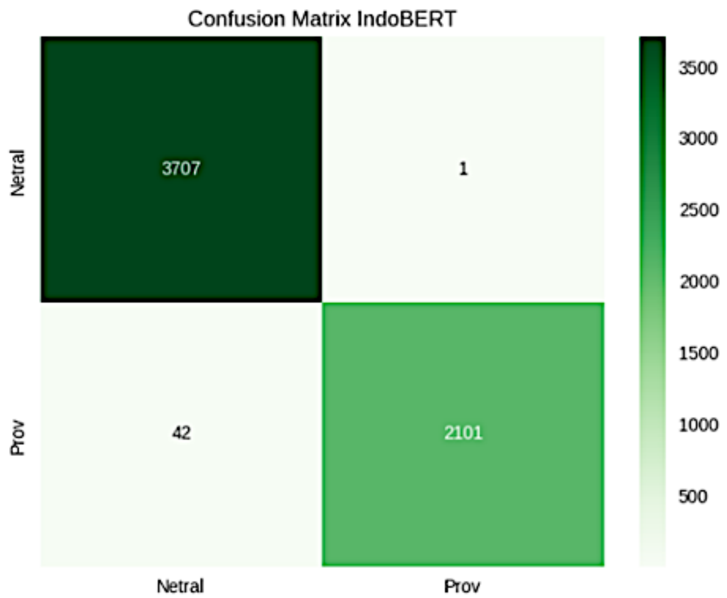

Figure 6. Confusion matrix of the IndoBERT model

3.2. Discussion

The results in Table 3 show that both SVM and IndoBERT achieve near-ceiling performance in detecting provocative discourse in the August 2025 DPR Twitter corpus. Rather than indicating redundancy between the two approaches, this convergence highlights an important property of the dataset itself: much of the provocation during this event was systematically patterned and strongly separable from neutral commentary. The high AUC reported for SVM in Table 3, together with the near-perfect separability visible in the ROC curve (Figure 5), suggests that the provocative class contained lexical and stylistic markers that were consistently present across many tweets. This is further supported by the label distribution in Figure 2, where the provocative class constitutes a substantial minority (35.3%) rather than a sparse outlier category. In event-driven controversies, this kind of separability is meaningful: it implies that provocative discourse is not merely anecdotal, but a recurring communicative mode that can be detected with high reliability when operationalized clearly.

At the same time, the two models arrive at strong performance through different representational mechanisms, and those differences matter for interpreting Indonesian political communication on X. The strong SVM performance indicates that provocation around the DPR demonstrations was frequently expressed through repetitive, slogan-like, and lexically salient constructions that are easily captured by TF–IDF n-grams. This aligns with the broader logic of protest communication on social media, where short, punchy phrases and institutional references often function as rallying points. Expressions such as bubar DPR or grievance-framed mentions like tunjangan DPR produce distinctive token patterns; when such phrases recur across thousands of tweets, a linear classifier can separate classes effectively. The temporal spike in Figure 3 reinforces this interpretation: on 31 August 2025, the conversation surged dramatically, and such surges typically encourage message standardization—users repeat the same hashtags, slogans, and key phrases to maximize visibility and collective resonance. In other words, the dataset characteristics visible in Figure 3 provide a contextual explanation for why TF–IDF + SVM performs so well: when discourse becomes highly "chant-like," surface lexical cues become unusually predictive.

IndoBERT's performance advantage in Table 3—although numerically small—still carries interpretive significance because it reflects greater robustness in handling rhetorical nuance. Indonesian political criticism on X often relies on insinuation, satire, sarcasm, and indirect accusation, especially when discussing powerful institutions such as the DPR. These rhetorical strategies can be difficult to capture with frequency-based features because the same words may appear in both neutral and provocative contexts depending on framing. IndoBERT's contextual embeddings help address this by modeling how words function in relation to surrounding tokens, enabling the model to infer pragmatic intent even when explicit trigger words are absent. The pattern of fewer misclassifications shown in the IndoBERT confusion matrix (Figure 6) supports this claim: errors are rare overall, but IndoBERT appears to reduce borderline mistakes where provocation is implied rather than directly stated. Put differently, while SVM excels when provocation is "spoken out loud," IndoBERT is better positioned when provocation is "smuggled in" through tone, irony, or contextual framing—features that are common in Indonesian digital culture.

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

Taken together, the findings suggest that the DPR controversy produced a hybrid discursive environment: overt mobilization rhetoric coexisted with subtler forms of critique. This hybridity is visible when the EDA outputs are read alongside the model results. The dominant tokens in Figure 4 confirm the discourse focus on institutions (dpr, anggota dpr) and collective identity (rakyat), while the sharp activity clustering in Figure 3 suggests that heightened attention periods likely intensified both emotional expression and the repetition of mobilizing frames. In such settings, slogan-driven messages spread easily because they are simple to replicate, but sarcasm and insinuation also flourish because they offer socially legible ways to ridicule authority without always using explicit hostility. The dual success of SVM and IndoBERT therefore reflects not only algorithmic competence, but also the linguistic reality of event-driven Indonesian political talk: it blends direct calls and complaints with performative, indirect rhetoric.

Beyond modeling, the ability to automatically detect provocative political discourse has practical implications for political monitoring and democratic governance in Indonesia. During high-tension periods like the DPR demonstrations, rapid increases in provocative content can serve as early indicators of polarization, coordinated agitation, or escalating mobilization. The dataset dynamics in Figure 3 illustrate why: online attention can rise sharply around offline triggers, and the informational environment can shift within days—or even hours—from routine debate to emotionally intense contention. Tools that can reliably flag provocative discourse at scale can therefore help researchers and civil society map escalation trajectories, identify critical time windows, and analyze how narratives spread. Importantly, the strong engagement association noted in the results (retweets and likes moving together) implies that visibility and endorsement can reinforce each other; in this context, detecting provocation is not just about content moderation, but also about understanding which rhetorical forms are most likely to be amplified.

IndoBERT is particularly relevant for such monitoring because it offers a pathway beyond simple keyword filtering. When political messaging is shaped by humor, satire, and indirect critique, keyword lists may miss large portions of provocative framing, especially if actors deliberately avoid explicit trigger terms to evade detection. IndoBERT's contextual sensitivity allows the detection system to better capture emotionally loaded insinuations

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

and rhetorical attacks that operate at the level of implied meaning. This matters for the broader concern raised in the introduction: coordinated amplification and buzzer-like dynamics may not always rely on explicit hate speech; they can also exploit sarcasm, ridicule, and delegitimizing frames to intensify distrust [6], [7]. While this study does not directly attribute content to coordinated networks, combining accurate provocation detection with temporal and interaction signals creates an analytical foundation for examining amplification patterns more systematically in future work.

Nevertheless, the strong classification scores in Table 3 should be interpreted with methodological caution because they are partly shaped by the labeling strategy. Labels were generated through a rule-based lexicon of provocative keywords and expressions. This means that both models were trained on an operational definition of "provocation" that is anchored in lexical triggers. In such a setup, SVM is naturally advantaged because TF–IDF features align closely with the labeling mechanism: if labels are assigned based on the presence of specific phrases, frequency-based representations will learn those phrase boundaries efficiently. IndoBERT's higher stability suggests better generalization within this framework, but it is still learning from labels that originate from a predefined lexicon. The near-perfect separability shown by Figure 5 and the minimal error counts in Figure 6 therefore reflect strong internal consistency between labeling rules and model learning objectives, not necessarily the full complexity of provocation in natural political language.

Consequently, the findings should be understood primarily as evidence that (1) event-driven political provocation during August–September 2025 contained highly learnable linguistic patterns, and (2) contextual models provide additional robustness when provocation is implicit. To strengthen external validity, future research should incorporate manual or semi-supervised annotation with clearer pragmatic guidelines (e.g., distinguishing sarcasm, delegitimization, incitement, and hate speech as separable categories) and test generalization across different political events. Such extensions would allow evaluation of whether the models can detect evolving rhetorical strategies beyond predefined lexical cues, particularly in environments where provocation is intentionally obfuscated.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

4. CONCLUSION

This study developed an event-based comparative framework to detect provocative political discourse on Indonesian Twitter (X) during the August–September 2025 DPR demonstrations. Using the same corpus and evaluation split, it compared SVM (TF–IDF) and IndoBERT, showing that both models can identify provocative discourse effectively at scale: SVM performs strongly when provocation is expressed through explicit, repetitive phrases, while IndoBERT is more robust for implicit, context-dependent rhetoric such as sarcasm and indirect attacks. Substantively, the results suggest that Indonesian digital political discourse during the DPR controversy combined slogan-driven mobilization with ironic and emotionally loaded critique, reinforcing the need for contextual NLP methods rather than keyword-only monitoring. However, the findings should be interpreted cautiously because labels were generated via a rule-based lexicon, the dataset covers only one event window (15 August–15 September 2025), and the analysis excludes network-level signals relevant to buzzer coordination and amplification. Future research should incorporate manual or semi-supervised annotation, expand to multiple political episodes for stronger generalizability, and integrate NLP with Social Network Analysis to examine how provocative narratives spread and whether amplification is coordinated. Overall, this study offers a scalable baseline for monitoring political provocation in Indonesian social media while outlining clear pathways to improve validity and explanatory power.

REFERENCES

[1] S. Gayatri and I. B. C. Satwika, "The Role of Social Media as a Medium for Political Information (Peran Media Sosial sebagai Media Sarana Informasi Politik)," *Anubhava: J. Ilmu Komun. Hindu*, vol. 2, no. 1, pp. 273–282, May 2022, doi: 10.25078/anubhava.v2i1.1050.

[2] L. Judijanto, R. Maulinda, S. Zulaika, I. Tjahyadi, and S. Suroso, "The Influence of Information Sources and Social Interaction on Social Media toward the Formation of Public Political Opinion in Indonesia (Pengaruh Sumber Informasi dan Interaksi Sosial di Media Sosial terhadap Pembentukan Opini Politik Masyarakat di Indonesia)," *Sanskara Ilmu Sos. Humaniora*, vol. 1, no. 01, pp. 21–31, Dec. 2023, doi: 10.58812/sish.v1i01.303.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

[3] A. Arpandi, "Online Media in Increasing Public Political Participation in General Elections (Media Online dalam Meningkatkan Partisipasi Politik Masyarakat pada Pemilihan Umum (Pemilu))," *Edu Soc.: J. Pendidik. Ilmu Sos. Pengabdi. Masy.*, vol. 3, no. 1, pp. 843–855, Aug. 2023, doi: 10.56832/edu.v3i1.293.

[4] A. A. Santoso, "The Nomination of Ganjar Pranowo as Presidential Candidate: Topic Analysis Study on Reverse Agenda Setting Related to the Ganjar Pranowo Case (Penetapan Ganjar Pranowo Sebagai Calon Presiden: Studi Analisis Topik pada Reverse Agenda Setting Terkait Kasus Ganjar Pranowo)," *J. Ilm. Ilmu Pendidik.*, vol. 7, no. 4, pp. 3805–3812, Apr. 2024, doi: 10.54371/jiip.v7i4.4140.

[5] A. C. Rosdiana and A. A. Suryaningtyas, "Identity Politics in Ganjar Pranowo's Political Campaign Ahead of the 2024 Presidential Election (Politik Identitas dalam Kampanye Politik Ganjar Pranowo Menjelang Pilpres 2024)," *J. Audiens*, vol. 5, no. 1, pp. 77–90, Mar. 2024, doi: 10.18196/jas.v5i1.336.

[6] A. Bilbao-Jayo and A. Almeida, "Improving Political Discourse Analysis on Twitter With Context Analysis," *IEEE Access*, vol. 9, pp. 104846–104863, 2021, doi: 10.1109/ACCESS.2021.3099093.

[7] M. Dynel, "Do We Know Whether to Laugh or Cry? User Responses to @Ukraine's Dark-humour Meme," *J. Creative Commun.*, vol. 19, no. 3, pp. 243–258, Nov. 2024, doi: 10.1177/09732586241239908.

[8] P. Kanungo and H. Singh, "A Feature Extraction based Improved Sentiment Analysis on Apache Spark for Real-time Twitter Data," *Scalable Comput.: Pract. Exp.*, vol. 24, no. 4, pp. 847–855, Nov. 2023, doi: 10.12694/scpe.v24i4.2343.

[9] H. Setyawan, L. M. Azizah, and A. Y. Pradani, "Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine," *Emerg. Inf. Sci. Technol.*, vol. 4, no. 1, pp. 1–7, May 2023, doi: 10.18196/eist.v4i1.18681.

[10] A. R. W. Sait and M. K. Ishak, "Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification," *Comput. Syst. Sci. Eng.*, vol. 44, no. 3, pp. 2553–2567, 2023, doi: 10.32604/csse.2023.029603.

[11] H. Jayadianti *et al.*, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM J. Ilm.*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[12] D. I. Putri *et al.*, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," *J. Appl. Inform. Comput.*, vol. 8, no. 1, pp. 7–12, Jul. 2024, doi: 10.30871/jaic.v8i1.7440.

[13] O. H. Rahman, G. Abdillah, and A. Komarudin, "Classification of Hate Speech on Twitter Social Media Using Support Vector Machine (Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine)," *J. RESTI*, vol. 5, no. 1, pp. 17–23, Feb. 2021, doi: 10.29207/resti.v5i1.2700.

[14] M. Miqdad, "Literature Review: Political Buzzers and Opinion Development on Social Media in Indonesia (Literature Review: Buzzer Politik dan Pengembangan Opini di Media Sosial di Indonesia)," *NeoRespublica: J. Ilmu Pemerintah.*, vol. 5, no. 2, pp. 689–698, Mar. 2024, doi: 10.52423/neores.v5i2.231.

[15] R. A. Saputra and Y. Sibaroni, "Multilabel Hate Speech Classification in Indonesian Political Discourse on X using Combined Deep Learning Models with Considering Sentence Length," *J. Ilmu Komput. Inform.*, vol. 18, no. 1, pp. 113–125, Feb. 2025, doi: 10.21609/jiki.v18i1.1440.

[16] N. N. A. Aryanti and O. Suria, "Sentiment Analysis of Layoffs in Indonesia: Comparison of IndoBERT with SVM, Random Forest, and Decision Tree with TF-IDF Optimization (Analisis Sentimen terhadap Pemutusan Hubungan Kerja di Indonesia: Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree dengan Optimasi TF-IDF)," *Rabit: J. Teknol. Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1158–1176, Jul. 2025, doi: 10.36341/rabit.v10i2.6364.

[17] I. Riadi, A. Fadlil, and U. A. Dahlan Yogyakarta, "Identifying Hate Speech in Tweets with Sentiment Analysis on Indonesian Twitter Utilizing Support Vector Machine Algorithm," 2023.

[18] S. Kumar *et al.*, "An anatomical comparison of fake-news and trusted-news sharing pattern on Twitter," *Comput. Math. Organ. Theory*, vol. 27, no. 2, pp. 109–133, Jun. 2021, doi: 10.1007/s10588-019-09305-5.

[19] T. Lynn *et al.*, "An Exploratory Data Analysis of the #Crowdfunding Network on Twitter," *J. Open Innov.: Technol. Mark. Complex.*, vol. 6, no. 3, p. 80, Sep. 2020, doi: 10.3390/joitmc6030080.

[20] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *Proc. AACL-IJCNLP 2020*, Suzhou, China, 2020, pp. 843–857.

[21] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proc. COLING 2020*, Barcelona, Spain, 2020, pp. 757–770.

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

[22] A. W. Pradana and M. Hayaty, "The Effect of Stemming and Removal of Stopwords on the Accuracy of Sentiment Analysis on Indonesian-language Texts," *Kinetik*, vol. 4, no. 4, pp. 375–380, Oct. 2019, doi: 10.22219/kinetik.v4i4.912.

[23] N. H. Jeremy, "The Impact of Text Preprocessing in Sarcasm Detection on Indonesian Social Media Contents," *J. EMACS (Eng. Math. Comput. Sci.)*, vol. 7, no. 2, pp. 183–189, 2025, doi: 10.21512/emacsjournal.v6.

[24] E. W. Pamungkas *et al.*, "Enhancing hate speech detection in Indonesian using abusive words lexicon," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 33, no. 1, p. 450, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp450-462.

[25] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," *J-KOMA J. Comput. Sci. Appl.*, 2025, doi: 10.21009/j.

[26] R. Santosa, A. B. Nusantara, and S. Imron, "Comparative Analysis of SVM and IndoBERT for Intent Classification in Indonesian Overtime Chatbots," *J. Syst. Comput. Eng. (JSCE)*, vol. 6, no. 3, pp. 258–270, Aug. 2025, doi: 10.61628/jsce.v6i3.2058.

[27] R. Romindo, J. J. Pangaribuan, and O. P. Barus, "Implementation of TF-IDF and Support Vector Machine Algorithms for Detecting Cyberbullying Comments on TikTok Social Media (Implementasi Algoritma TF-IDF dan Support Vector Machine terhadap Analisis Pendeteksi Komentar Cyberbullying di Media Sosial TikTok)," *J. Device*, vol. 13, no. 1, pp. 124–134, 2023.

[28] F. Baharuddin and M. F. Naufal, "Fine-Tuning IndoBERT for Indonesian Exam Question Classification Based on Bloom's Taxonomy," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 2, pp. 253–263, Oct. 2023, doi: 10.20473/jisebi.9.2.253-263.

[29] A. R. Hanum *et al.*, "Performance Analysis of BERT Text Classification Algorithm in Detecting Hoax News (Analisis Kinerja Algoritma Klasifikasi Teks BERT dalam Mendeteksi Berita Hoaks)," *J. Teknol. Inf. Ilmu Komput.*, vol. 11, no. 3, pp. 537–546, 2024, doi: 10.25126/jtiik2024118093.