IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3376~3388
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3376-3388

Detection of illegal drug activities in Indonesia via social media X

Adi Hanif Sedar[1], Lukman Yudokusumo[1], Indra Budi[1], Amanah Ramadiah[1], Aris Budi Santoso[1], Prabu Kresna Putra[2]
[1]Department of Computer Science, Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia
[2]Research Center for Data Science and Information, National Research and Innovation Agency (BRIN), Tangerang, Indonesia

Article Info

Article history:

Received Jan 27, 2025
Revised May 17, 2026
Accepted May 24, 2026

Keywords:

Illicit drugs
Public health
Social media
Text classification
X

ABSTRACT

In today's digital world, social media platforms such as X (formerly Twitter) are widely used for communication but also for illegal activities, such as selling illicit drugs. This study examines how posts from X related to illegal drug sales in Indonesia can be analyzed using machine learning techniques, with a focus on incorporating features related to personally identifiable information (PII) and web URLs. The main questions are how to use classification methods to detect these posts and which drugs are most often mentioned. A total of 11,777 posts from X were collected using web scraping. After cleaning and processing the data, five machine learning models were used. The support vector machine (SVM) model showed the best results with a weighted F1-score of 97.08%. The findings show that abortion pills are the most frequently mentioned drugs, followed by sexual enhancement and anesthetic drugs. This study also used visual tools to show the most common hashtags, posts distribution over time, and types of drugs discussed. This study underlines the importance of monitoring social media to protect public health and highlights how incorporating PII features can improve the detection of illegal activities. Future research should look at posts in other languages and from other platforms, considering PII features for enhanced analysis.

Corresponding Author:

Adi Hanif Sedar
Department of Computer Science, Faculty of Computer Science, Universitas Indonesia
Jakarta, Indonesia
Email: adi.hanif@ui.ac.id

1. INTRODUCTION

In today's digital era, social media is a big part of our daily lives, it lets people share information, ideas, and opinions [1]. X (formerly Twitter), one of the most popular social media platforms, is used by millions of people every day for different reasons, like talking to friends or spreading news [2]. But social media's ease of use and the anonymity it offers have brought up new challenges, especially when it comes to illegal activities like selling fake or illegal drugs [3]. Selling illegal drugs is a big problem worldwide, it's bad for public health, safety, and the economy, the World Health Organization (WHO) [4] says illegal drugs are those sold without approval from health authorities. This includes fake drugs or selling drugs without a prescription that shouldn't be sold over the counter [5].

X has become a favored space for drug vendors to post illegal advertisements, reaching a wide audience at low cost [6]–[8]. Previous research found that approximately 13 in 100 social media posts advertise illicit drugs [9]. The anonymity and broad reach of social media provide an efficient means for illegal sellers to promote their products. Indonesia has specific regulations governing the sale of drugs, which are outlined in several laws. The primary legislation includes Health Law No. 36 of 2009 (*Undang-Undang*

Kesehatan No. 36 Tahun 2009) and the Narcotics Law No. 35 of 2009 (*Undang-Undang Narkotika No. 35 Tahun 2009*). These laws restrict the sale of certain medications and narcotics without proper authorization. For instance, drugs like abortion pills, sedatives, and anesthetics are regulated under these laws to prevent misuse and illegal distribution. For example, abortion in Indonesia is highly restricted, yet common, with an estimated 42.5 abortions per 1,000 women aged 15-49 in Java, Indonesia [10]. Unsafe abortions contribute significantly to maternal mortality, accounting for up to 30% of maternal deaths [11]. Online marketplaces have become a source for obtaining misoprostol, an abortion medication; however, many sellers provide incomplete information about side effects and complications, and some packages contain an insufficient quantity of the drug for a complete abortion [12]. Enforcement remains challenging, especially on digital platforms where sellers can easily conceal their identities.

Because of the harmful effects of illegal drugs sales, it's important to find efficient ways to spot and stop these activities on social media [8], [13]. In Indonesia, the sale of illegal drugs through social media platforms has become a major concern [14]. It provides an easy and efficient means for illegal sellers to promote their products. The freedom of communication and the ability to conceal identities make these sellers feel secure operating on the platform [15]. Indonesia's large population of over 175 million active internet users (2021–2024), combined with the growing adoption of smartphones, has created a strong foundation for the rapid growth of social media and online commerce [16]. Moreover, Indonesia, being a country with high social media usage, serves as a potential market for illegal drug sellers to reach more consumers. With the lack of effective supervision and law enforcement, the activity of selling illegal drugs can lead to serious health consequences for the public.

Advanced technology like data mining, machine learning, and natural language processing (NLP) can be used to identify and analyze communication patterns related to illegal activity sales on X [13]. By automating text analysis and using text classification techniques, researchers and law enforcement can quickly identify and respond to illegal activities on social media [17]. While these methodologies have shown promise, there is a notable gap in the consideration of personally identifiable information (PII) detection within the feature sets used for classification. PII, such as phone numbers, email addresses, and web URLs, often serve as direct indicators of illicit intent [5], especially in the context of illegal drug sales, where sellers include contact information for potential buyers. Previous studies [6], [13], [17] have primarily focused on textual and linguistic features without extensively incorporating PII and web URLs detection into their models. This may limit the effectiveness of classification models in accurately identifying illicit activities.

This study addresses this gap through two research questions: i) how classification techniques and text mining can be applied to identify posts related to the sale of illegal drugs on X, and ii) which types of drugs are most frequently promoted on X in Indonesia. To answer these questions, the main contributions of this work are threefold: i) systematically incorporate PII features like phone numbers, email addresses, and web URLs into illicit drug detection on Indonesian X, demonstrating through feature importance analysis that these features rank among the top five most influential signals; ii) construct and analyze a domain-specific dataset Indonesian posts covering abortion pills, sexual enhancement, and anesthetic drugs, contributing empirical evidence of online illicit pharmaceutical activity in this regional context; and iii) offer practical insights into deployment considerations, including error analysis and ethical implications relevant to law enforcement and public health agencies. The findings can assist health departments and law enforcement in identifying and addressing illicit drug sales on social media, while informing the development of stronger regulatory and monitoring frameworks.

2. METHOD

In this study, data were collected using web scraping, targeting X's search results to find posts related to the keywords illegal drugs. The features extracted include user ID, posts text for classification purposes, and the posts date, which were then extracted using Python. After obtaining the collected data, text pre-processing (filtering, case folding, tokenization, stop word removal, and data stemming) was performed to clean it. The data was then manually processed and classified. Next, the data was divided into testing and training sets. Construct a machine learning model, then conducted using by comparing the models. Through these methods, text classification was assessed to see which posts lead to the sale of illicit drugs. Finally, the data was created and visualized using the Python library for creating graphs matplotlib. The overall flow process of this study is illustrated in Figure 1.

2.1. Data collection

The dataset used in this research was from X, collected using web scraping, targeting X's search results. Web scraping can be designed to collect conversational data from social media platforms in JavaScript object notation (JSON) format, which can then be filtered and formatted for specific use cases [18]. Using Python scripts with scraping libraries, we targeted data involving

specific keywords in Indonesian, such as "*jual obat*" (sell drugs), "*obat aborsi*" (abortion pills), "*penggugur kandungan*" (pregnancy termination), "*obat tanpa resep*" (drugs without prescription), "*pil penenang*" (sedative pills), "*obat bius*" (anesthetic drugs), and also slang or code words used in the narcotics world to evade detection, including "putaw" for heroin. This approach, designed to cover the broad spectrum of illegal drugs and drug sales online, ensures a thorough data collection process by including additional terms used by dealers and users to keep their activities underground. This scraping process was conducted in May 2024. During the data collection stages, several challenges were encountered, such as many posts containing repetitive or duplicated, required additional steps to remove these to avoid skewing the results. Before filtering duplicated results, the dataset for this study consisted of 13,196 posts collected over a one-year period (April 2023–2024). Out of the initial 13,196 posts collected, 1,419 were identified as duplicates or irrelevant and removed during pre-processing, resulting in a final dataset of 11,777 posts. The data collected is then stored in the JSON.

After the data was collected the data was labeled manually. To ensure the quality of manual classification, the labeling process was conducted by two annotators following specific guidelines. These guidelines included criteria such as mentioning the brand name or the type of drug (e.g., anesthetic) or its effects, ensuring consistent and accurate categorization of the posts. This collaborative approach aimed to minimize bias and enhance the reliability of the labeling process. To measure inter-annotator agreement, Cohen's Kappa (κ) was computed on the 11,777 labeled posts, yielding κ =0.87, which indicates almost perfect agreement [19]. Disagreements were resolved through discussion before training. The labeling criteria for intent classification are described in Table 1. Table 2 presents examples of posts classified as intent to sell and not intent to sell. Examples of posts categorized by drug type are shown in Table 3.

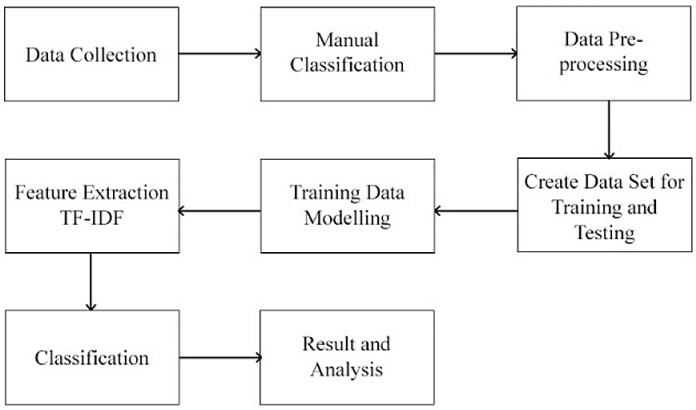

Figure 1. Flow process of the proposed research

Table 1. Intent classification

Label	Description
Yes	Texts that indicate the user is offering to sell illegal drugs. This could include specific offers, pricing information, details about drug types or quantities.
No	Texts that mention illegal drugs but do not demonstrate a clear intent to sell. This could include discussions about news, illegal drug effects, legal issues, or personal experiences that are not related to transactions.

Table 2. Example of intent classification posts from X

Label	Posts from X
Yes	*"Jual obat aborsi penggugur kandungan misoprostol Gastrul ampuh untuk tuntaskan masalah terlambat haid usia 1 2 3 4 5 6 bulan !!#obataborsi"* "Selling abortion drugs, pregnancy terminators, misoprostol Gastrul, effective for resolving late period issues for 1 to 6 months !!#abortionpill"
No	*"Kayanya konotasi Narkoba itu negatif. Tapi memang ada obat flu yg didalamnya ada golongan narkotika dan itu harus sesuai resep dokter. Kayanya yg ada kandungan zat narkotika gak segampang itu diedarkan dipasar tanpa resep. Cmiiw"* "It seems like the connotation of drugs is negative. But there are indeed flu medicines that contain narcotics and must be prescribed by a doctor. It seems that substances containing narcotics cannot be easily distributed in the market without a prescription. Correct me if I'm wrong."

Table 3. Example of posts with type of drug classification

Label	Post from X
Abortion pills	*"Jual obat cytotec aborsi penggugur kandungan asli WA"* "Selling original abortion drug cytotec WhatsApp"
Anesthetic drugs	*"Jual obat bius cair liquid sex ampuh langsung pingsang WA"* "Selling liquid anesthetic drug sex liquid effective, WhatsApp"
Sexual enhancement drugs	*"Jual obat viagra asli"* "Selling original viagra drug"

2.2. Data processing

X data is hard to work with because the way people write is inconsistent, and there aren't clear categories, making analysis and classification challenging [20]. The goal of data pre-processing is to cleanse and standardize the data to aid subsequent processes. Various data cleaning techniques ensure that extracted information is accurate and devoid of any unnecessary noise or missing information due to scraping. In this study, the author uses several steps for data pre-processing. The initial phase involves selecting and curating the collected X data by filtering out duplicated posts and unwanted phrases not related to drugs or posts from advertisers, resulting in only 11,777 posts being used. Once the data is filtered, all the letters in the posts are transformed into lowercase to make the data uniform and easier to process in the subsequent steps. The tokenization process breaks down each post into individual words, or "tokens", while simultaneously removing punctuation, using the natural language toolkit (NLTK)'s word tokenize function. Specific words that do not contribute to the overall meaning of the posts, known as stop words, are then removed. Additionally, URLs, hashtags, and mentions are also eliminated using a set of Indonesian and X-specific stop words. The final step in the pre-processing pipeline transforms all words into their root forms or stems using the Sastrawi stemmer for the Indonesian language, simplifying the data and reducing its complexity for subsequent processing. Each processed posts are then stored in a list, with its root words and a version of the posts with the processed content. The processed data is finally saved in a JSON file for further use. Table 4 displays the transformation of the data before and after the pre-processing phase for the collected posts.

Table 4. Before and after pre-processing

Label	Posts from X
Before	*"Jual obat cytotec aborsi penggugur kandungan asli WA"* "Selling original cytotec abortion drug pregnancy terminator WA"
After	*jual obat cytotec aborsi gugur kandung asli* "Selling original cytotec abortion drug pregnancy terminator"

2.3. Feature extraction

The term frequency-inverse document frequency (TF-IDF) algorithm involves two calculations: TF and IDF. The TF-IDF score highlights which feature holds the greatest significance across the entire document collection [21]. After the weights are assigned, the level of similarity is determined by arranging the words into a vector and utilizing cosine similarity [22]. The processed JSON file is loaded, and the content of each post is collected into separate lists. The CountVectorizer and TfidfVectorizer from sklearn are used to build a vocabulary and encode the documents into a bag of words (BoW) and TF-IDF matrices.

2.4. Classification and evaluation

Once the data was processed and verified, researchers proceeded to build a machine learning model. To perform future classification on the data, five machine learning algorithms were tested: naïve Bayes, logistic regression, random forest, gradient boosting, and support vector machine (SVM). The selection of these models was informed by their strong performance in similar studies. Previous research has shown that SVM is particularly effective in text classification tasks due to its ability to handle high-dimensional feature spaces and perform well even with small datasets [6], [7]. Additionally, other algorithms were used as baseline models to evaluate the relative performance of classifiers. While transformer-based models (e.g., bidirectional encoder representations from transformers (BERT)) and deep learning architectures (convolutional neural network (CNN)-long short-term memory (LSTM) hybrids) have shown strong performance in text classification [7], this study deliberately focuses on classical machine learning approaches for two reasons: interpretability of feature contributions, which is essential for understanding the role of PII features and supporting human moderator workflows [23], [24] and lower computational overhead, enabling potential deployment in real-time monitoring systems [25]. Benchmarking against transformer-based approaches is left for future work. To ensure model performance, from 11,777 posts, for training the classification models, 80% of the data (9,421 posts) was used for training, while the remaining

20% (2,356 posts) was reserved for testing. The size of the training data was chosen to ensure a model training process while maintaining sufficient data for unbiased testing and evaluation.

This research compared models and ranked the models based on a specified metric, in this case, accuracy, thus providing a clear indication of which model performed best for our specific task. Then compare the results in the spider chart for metrics recall, precision, and F1-score. Precision and recall are crucial metrics to evaluate when dealing with balanced data. However, for unbalanced data, the F1-score becomes the most significant metric to consider [26]. This evaluation provides insights into the strengths and weaknesses of each model in the context of detecting intent to sell illegal drug activities on X. To enhance the classification of posts indicating intent to sell illegal drugs, regular expression (regex) patterns were developed to identify and extract PII elements such as phone numbers, email addresses, and web URLs from the post texts. Regular expressions are a powerful tool for pattern matching within text, allowing for the efficient detection of specific patterns corresponding to different types of PII. Coldea *et al.* [27] developed a regex-based approach for flagging sensitive data in large health datasets, improving data quality reporting. Table 5 summarizes the regex patterns implemented.

By applying these regex patterns to the posts data, the occurrences of phone numbers, email addresses, and web URLs were systematically identified and counted. These elements were then encoded as additional binary features in the dataset, indicating the presence or absence of each type of PII in a post. By including PII and web URLs features alongside traditional textual features (e.g., TF-IDF vectors), the impact of PII detection on the performance of the classification models was assessed.

Table 5. Implemented regex pattern

Name	Regex pattern	Purpose
phone_patterN	\b(?:\+?(\d{1,3})?[-.●]?(\d{1,4})?[-.●]?(\d{1,4})?[-.●]?(\d{1,9}))\b	Detects phone numbers, including international formats with -, ., or ● as separators
email_patterN	\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b	Detects email addresses
url_pattern	\b(?:http[s]?://)?(?:www\.)?[a-zA-Z0-9.-]+\.[a-zA-Z]{2,6}\b	Detects website URLs (with or without http(s):// and www. prefix)

3. RESULTS AND DISCUSSION

A total of 11,777 posts were processed and used to train and evaluate five classifiers. SVM achieved the highest overall F1-score of 97.08%, outperforming naïve Bayes (94.85%), logistic regression (96.65%), random forest (96.83%), and gradient boosting (96.49%). SVM was therefore selected as the best-performing classifier for further analysis. Figure 2 presents the comparison of text classification performance across all models using a spider chart. Table 6 summarizes the classification performance metrics across all models.

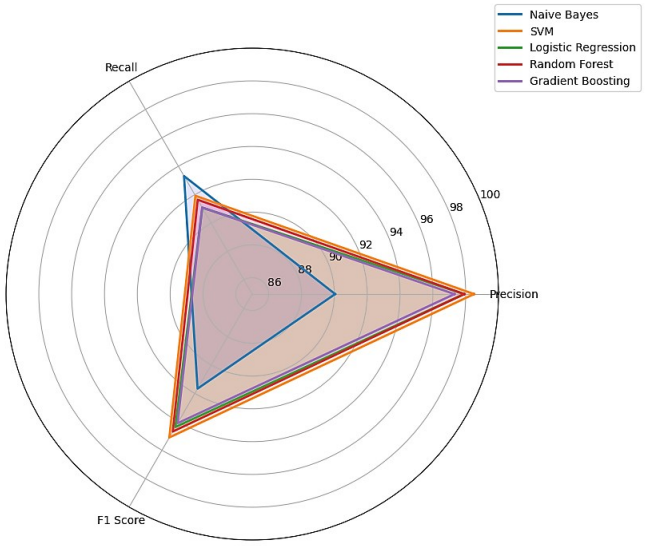

Figure 2. Comparison of text classification performance

Table 6. Classification performance on intent to sell (positive) class on the validation set

Model	Precision (%)	Recall (%)	F1-score (intent) (%)	F1-score (weighted) (%)
Naïve Bayes	90.05	93.31	91.66	94.85
SVM	98.51	91.92	95.10	97.08
Logistic regression	97.90	91.09	94.37	96.65
Random forest	97.92	91.64	94.68	96.83
Gradient boosting	97.32	91.09	94.10	96.49

Following the comparison of all models, a closer examination of the best-performing classifier was conducted. Figure 3 presents the confusion matrix of the SVM classifier on the testing data (2,356 posts), showing 1,633 true negatives, 14 false positives, 39 false negatives, and 670 true positives. Overall, the model correctly classified 2,303 out of 2,356 posts, with only 53 misclassifications. Figure 3 presents the confusion matrix.

An error analysis of the misclassified posts in Figure 3 presents the confusion matrix revealed distinct patterns for each error type. The 39 false negatives were dominated by short, conversational posts that mentioned drug names without explicit commercial cues, such as brief replies referencing "*obat bius*" (anesthetic drugs) or hashtag-only posts without supporting context. Many of these posts lacked phone numbers, URLs, or pricing information, which made the selling intent difficult to detect from textual features alone. The 14 false positives, in contrast, were typically promotional-style posts containing strong PII signals such as phone numbers and URLs, indicating that the model occasionally over-relied on these features when textual context was ambiguous. The higher number of false negatives compared to false positives also explains the model's lower recall (94.50%) relative to its precision (97.95%) on the testing set (2,356 posts). Table 7 presents representative examples of misclassified posts from each error type.

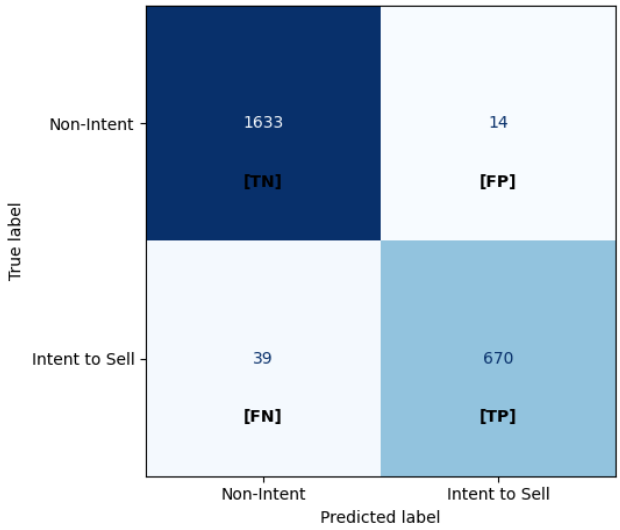

Figure 3. Confusion matrix of SVM classifier on testing data

Table 1. Examples of misclassified posts by SVM

Type	Posts from X	True label	Predicted
False negative	"*@user obat bius kak cepet kok*" (Anesthetic drugs bro, works fast)	Intent to sell	Non-intent
False negative	*"#penggugurkandungan #misoprostol #cytotecoriginal #aborsi [URL]"*	Intent to sell	Non-intent
False positive	"Art of the day: 'cytotec misoprostol'. View at: [URL]"	Non-intent	Intent to sell
False positive	"*Tuntas Boyolali 2 bln lebih [URL]*" (Complete Boyolali 2+ months)	Non-intent	Intent to sell

These findings suggest that incorporating contextual cues beyond keywords and PII for example, leveraging conversational context or slang-aware embeddings, could further improve classification accuracy. After the classification and evaluation process, matplotlib and panda's libraries were used to visualize the distribution and trends of the posts. The distribution of posts classified as intent to sell vs. non-intent is shown in Figure 4. Figure 4 represents the text classification of posts that intend to sell illicit drugs using a

visual representation. 3,392 posts showed intent to sell and 8,385 posts are neutral or indicated no intent to sell. The distribution of posts over time that express an intent to sell illegal drugs is presented in Figure 5.

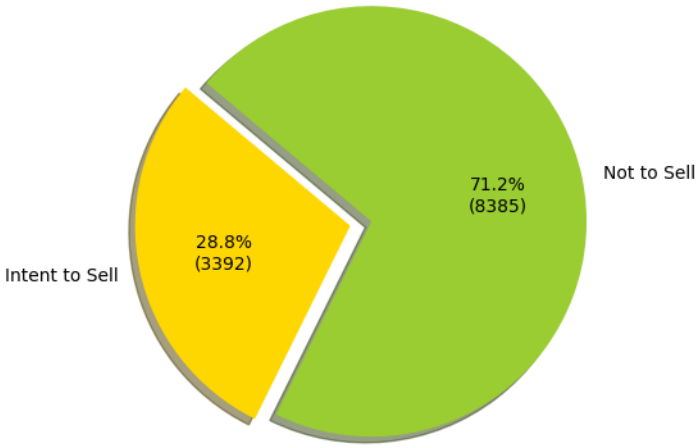

Figure 4. Intent to sell classification

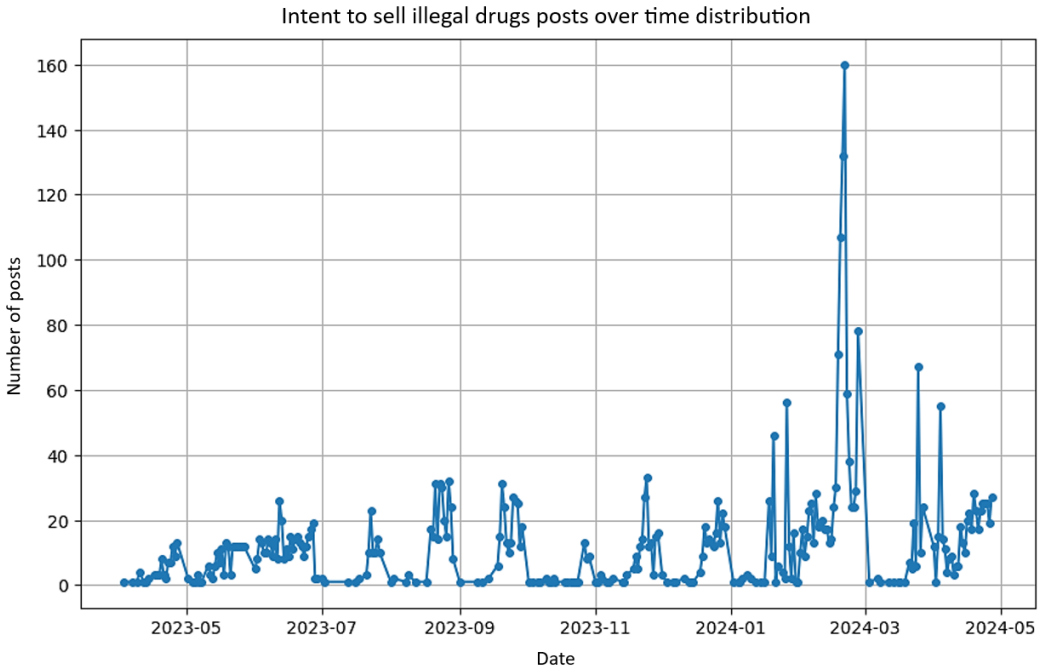

Figure 5. Distributions of post over time

Figure 5 shows the distribution of intent-to-sell posts from May 2023 to May 2024. Activity remained low and steady until late November 2023, followed by a sharp spike in December 2023 (peaking at 160 posts), with smaller fluctuations through May 2024. This spike may indicate heightened illicit sales activity during that period. The number of posts grouped by topic is presented in Figure 6. Figure 6 illustrates the number of posts grouped by the type of drug mentioned in the posts. The chart compares three categories: abortion pills, sexual enhancement drugs, and anesthetic drugs. The data shows that abortion pills are the majority, with a total of 3,064 posts. This is significantly higher than the other categories, indicating a predominant focus on this type of drug in the dataset. Sexual enhancement drugs are mentioned in 139 posts, while anesthetic drugs appear in 189 posts. The stark difference in the number of posts between abortion pills and the other two categories suggests that discussions related to abortion pills are far more prevalent on

social media. This could reflect current events, public interest, or ongoing discussions in society related to this drug. The count of posts containing PII is presented in Figure 7.

Figure 7 illustrates the count of posts that contain PII, categorized into phone numbers, email addresses, and web URLs. The data reveals a significant presence of PII in posts, with phone numbers appearing in 4,024 posts and web URLs in 4,822 posts. It is important to note that some posts may contain both phone numbers and web URLs, leading to overlapping counts. The total number of posts containing PII is not the sum of these two categories, as some posts included both types of PII. Notably, no posts were found to contain email addresses. This highlights the extent to which users may be sharing sensitive information on social media.

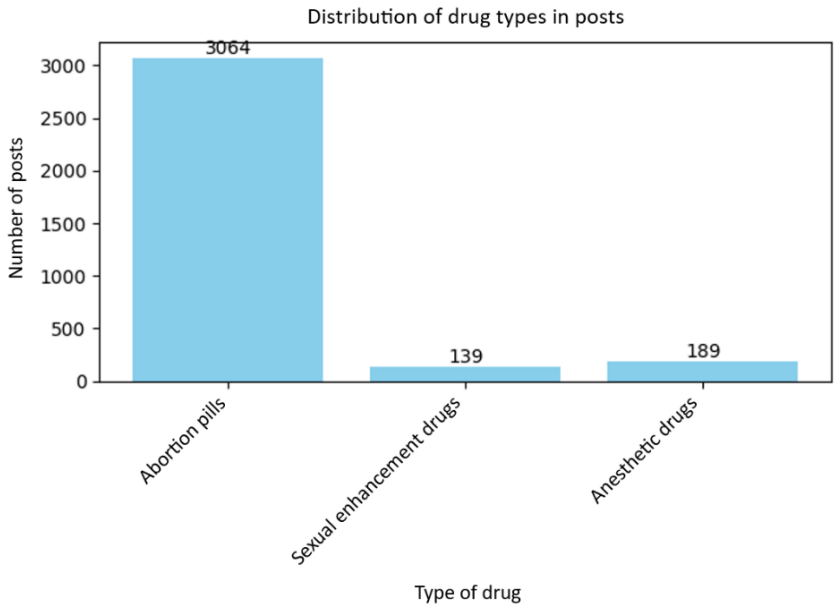

Figure 1. Number of posts grouped by topic

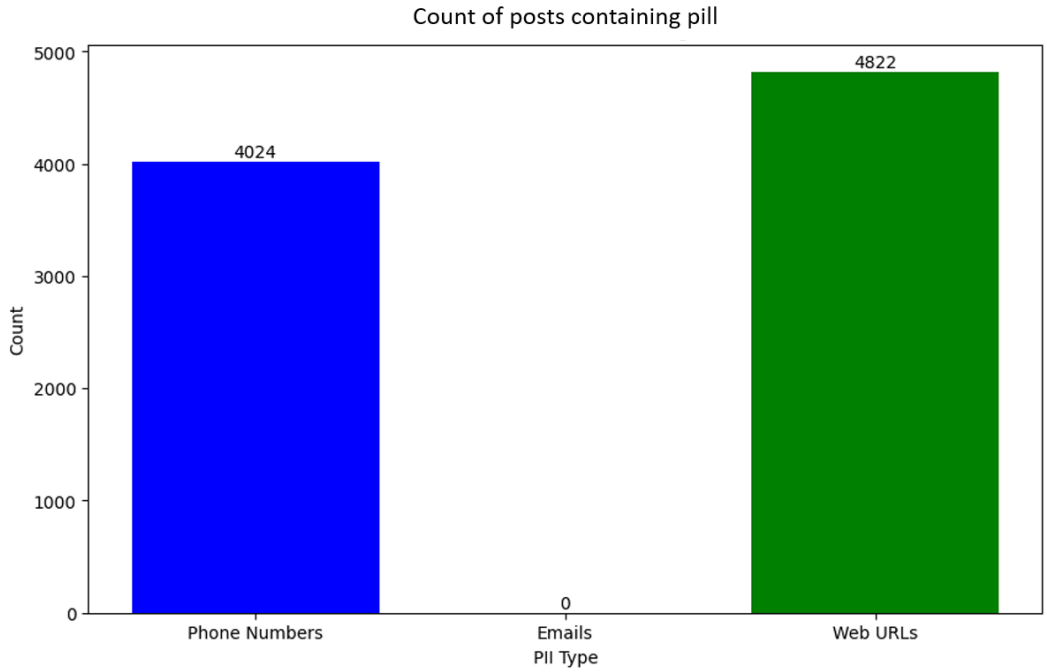

Figure 2. Count of posts containing PII

To evaluate the significance of the PII features in our classification task, we performed feature importance analysis using the random forest classifier, which provides inherent measures of feature importance based on the Gini impurity reduction. The feature importance analysis identified the features that did not contribute to the classifier's performance, enabling improvements in accuracy [23], [24]. The results indicated that PII features, particularly the presence of phone numbers and web URLs, were among the top contributors to the model's ability to accurately classify posts with intent to sell illegal drugs. Table 8 shows the feature importance score.

Table 8. Feature importance score

Rank	Feature	Importance score
1	Word: *aborsi* (abortion)	0.0816
2	Word: cytotec	0.0706
3	Word: *penggugur* (abortion inducer)	0.0492
4	Presence of phone number	0.0480
5	Presence of URL	0.0403
6	Word: misoprostol	0.0242
7	Word: *tuntas* (complete)	0.0221
8	Word: *asli* (original)	0.0149

The feature importance scores reveal that the presence of a phone number and URL are on top five significant features, indicating that posts containing phone numbers are strong indicators of intent to sell illegal drugs. Similarly, the presence of a web URL is also highly influential. Traditional textual features like specific keywords (e.g., "*aborsi* (abortion)", "cytotec (abortion pill)", "*penggugur* (abortion inducer)") remain important, but the incorporation of PII features enhances the model's ability to detect illicit intent.

While the presence of phone numbers and other PII significantly boosts the accuracy of detecting illicit drug sale posts, it does not automatically lead to easy apprehension of offenders. Critical challenge in using phone numbers and PII for investigating narcotics crimes in Indonesia is due to weaknesses in SIM card registration policies and data sharing between telecom providers and law enforcement [28]. People engaging in drug transactions on social media also generally report low concern for law enforcement exposure, but still employ risk management strategies, such as using self-deleting messages and code language [29]. These factors make it difficult for authorities to link phone numbers to actual individuals or dismantle broader networks, despite the availability of PII. The word cloud of most used hashtags is presented in Figure 8.

Figure 8. Word cloud of the most used hashtags

Figure 8 shows a word cloud of the most used hashtags in the posts we collected. The bigger the word, the more often it appeared. The biggest words, like "*obat aborsi*," "cytotec," and "*penggugur kandungan*," are the most common hashtags. This shows that posts are mostly about abortion pills, especially the "cytotec" brand. Other common hashtags are "jualobatcytotec" and "misoprostol," which also relate to specific drugs and their sale. The results from this highlight the importance and feasibility of utilizing social media data, specifically X data, as a valuable source of text classification analysis. Such analysis can be crucial in guiding decision-making processes in various sectors, like policy formulation.

The performance of our SVM model, achieving an overall F1-score of 97.08%, is consistent with findings from related studies on drug-related text classification in social media. Alruwaili *et al.* [30] reported an accuracy of 96% when using SVM to detect drug-selling behaviors on Arabic posts, which closely aligns with our results and suggests that SVM is robust across different languages for this task domain. Similarly, Kanan *et al.* [7] developed an intelligent health care system based on low-resource language processing and demonstrated the effectiveness of machine learning models for drug abuse detection in social media. In contrast, Tassone *et al.* [20] found that SVM achieved only 59.33% accuracy on a broader drug use detection task using English posts, highlighting that task-specific keyword targeting and domain-focused preprocessing, as employed in this study, may substantially improve classification performance.

This study also contributes a novel methodological advancement by incorporating PII features (phone numbers and web URLs) into the classification pipeline. While previous studies [8], [13], [17] primarily relied on textual and linguistic features for detecting drug-related content, our feature importance analysis demonstrates that PII elements rank among the top five most influential features. This finding suggests that sellers of illegal drugs on Indonesian X frequently include contact information to facilitate transactions, making PII detection a valuable complement to traditional NLP features. Moreover, the PII-enhanced framework is generalizable: phone numbers and URLs are language-agnostic signals, allowing this approach to be adapted to other regions and languages with minimal modification to the regex patterns. Future work could explore combining our PII-enhanced feature engineering with transformer-based models to further improve detection accuracy.

Beyond technical performance, the deployment of this detection model requires careful ethical considerations. As shown in the error analysis, false positives may incorrectly flag posts discussing drugs in non-commercial contexts such as personal experiences or harm-reduction discourse. The sensitive nature of certain categories, particularly abortion pills, also intersects with reproductive health and personal safety. Therefore, the model is recommended for use as a decision-support tool for human moderators rather than as a standalone enforcement mechanism, with transparent oversight and compliance with data protection regulations.

4. CONCLUSION

The study examined posts related to illegal drug sales. Utilizing machine learning techniques, the researchers analyzed 11,777 posts. Using naïve Bayes, logistic regression, random forest, gradient boosting, and SVM classifiers to identify posts indicating intent to sell illegal drugs. The results showed that the SVM model performed best, achieving an F1-score of 97.08%. These results suggest that the SVM algorithm is more suitable for future tasks on this dataset. The feature importance analysis demonstrated that PII elements were critical in enhancing the accuracy of the classification models. These findings suggest that integrating PII detection into the analysis can greatly enhance the effectiveness of monitoring and detecting illicit drug sales activities online. This study analyzes that posts mentioning abortion pills were the most common, followed by those about sexual enhancement drugs and anesthetic drugs. There was a notable increase in posts related to illegal drug sales in December 2023, suggesting more activity during that period. The study included visualizations to show the data, such as the distribution of posts over time, the number of posts by drug type, and a word cloud of the most used hashtags. These visualizations showed trends in the data, with abortion pills being the most discussed, especially the "cytotec". This study has some limitations. The data might not fully reflect the public because it focused on specific posts and topics. Only posts in the Indonesian language were included, which might miss relevant data in other languages. Future research should look at a wider range of topics and include data from other social media platforms, forums, and comment sections. This would give a better understanding of public opinion. Studying data in different languages and regions could provide a more global view. Adding data from news reports and law enforcement could help confirm the findings and give more insights into illegal drug sales on social media. The results of this study can be applied in several real-world contexts, particularly for law enforcement and public health agencies. For instance, the classification model could be integrated into a monitoring system that flags potential illegal drug sale posts in real-time, allowing for faster response and intervention. Public health campaigns can also be tailored to address the rise in discussions about certain drugs, such as abortion pills, providing targeted information and resources to combat their misuse.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Adi Hanif Sedar	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Lukman Yudokusumo		✓		✓		✓		✓	✓	✓	✓			
Indra Budi	✓	✓	✓	✓	✓		✓			✓		✓	✓	
Amanah Ramadiah	✓	✓			✓		✓			✓		✓		
Aris Budi Santoso	✓	✓		✓	✓					✓	✓	✓	✓	✓
Prabu Kresna Putra	✓	✓		✓	✓					✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [AHS], upon reasonable request.

REFERENCES

[1] K. Tammisalo, M. Danielsbacka, A. O. Tanskanen, and B. Arpino, "Social media contact with family members and happiness in younger and older adults," *Computers in Human Behavior*, vol. 153, Apr. 2024, doi: 10.1016/j.chb.2023.108103.

[2] M. Arazzi, M. Ferretti, S. Nicolazzo, and A. Nocera, "The role of social media on the evolution of companies: a Twitter analysis of streaming service providers," *Online Social Networks and Media*, vol. 36, Jul. 2023, doi: 10.1016/j.osnem.2023.100251.

[3] J. B. Hertig, S. M. James, C. J. Hummel, and M. J. Rubin, "Evaluation of pharmacists' awareness of illegal online pharmacies and perceived impact on safe access to medicines," *Medicine Access @ Point of Care*, vol. 5, Jan. 2021, doi: 10.1177/23992026211005642.

[4] World Health Organization, *WHO global surveillance and monitoring system for substandard and falsified medical products*, Geneva, Switzerland: World Health Organization, 2017.

[5] R. V. D. Sanden, C. Wilkins, M. Rychert, and M. J. Barratt, "'Choice' of social media platform or encrypted messaging app to buy and sell illegal drugs," *International Journal of Drug Policy*, vol. 108, Oct. 2022, doi: 10.1016/j.drugpo.2022.103819.

[6] M. K. Alhumayani and H. N. Alhazmi, "Detecting reported side effects of COVID-19 vaccines from Arabic Twitter (X) data," *IEEE Access*, vol. 12, pp. 55367–55388, 2024, doi: 10.1109/ACCESS.2024.3389655.

[7] T. Kanan *et al.*, "An intelligent health care system for detecting drug abuse in social media platforms based on low-resource language," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 691–703, 2024, doi: 10.1109/TASLP.2023.3294699.

[8] F. Zhao *et al.*, "Computational approaches to detect illicit drug ads and find vendor communities within social media platforms," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 19, no. 1, pp. 180–191, Jan. 2022, doi: 10.1109/TCBB.2020.2978476.

[9] A. Fuller, M. Vasek, E. Mariconti, and S. D. Johnson, "Understanding and preventing the advertisement and sale of illicit drugs to young people through social media: a multidisciplinary scoping review," *Drug and Alcohol Review*, vol. 43, no. 1, pp. 56–74, Jan. 2024, doi: 10.1111/dar.13716.

[10] M. M. Giorgio *et al.*, "Estimating the incidence of induced abortion in Java, Indonesia, 2018," *International Perspectives on Sexual and Reproductive Health*, vol. 46, 2020, doi: 10.1363/46e0220.

[11] P. W. Saraswati, "Saving more lives on time: strategic policy implementation and financial inclusion for safe abortion in Indonesia during COVID-19 and beyond," *Frontiers in Global Women's Health*, vol. 3, Sep. 2022, doi: 10.3389/fgwh.2022.901842.

[12] A. M. Moore *et al.*, "Attempts to purchase misoprostol online in Indonesia: a mystery client study design," *BMC Women's Health*, vol. 22, no. 1, Aug. 2022, doi: 10.1186/s12905-022-01889-6.

[13] S. Yuenyong, N. Hnoohom, K. Wongpatikaseree, and T. P. N. Ayutthaya, "Classification of Tweets related to illegal activities in Thai language," in *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Nov. 2018, pp. 1–6, doi: 10.1109/iSAI-NLP.2018.8692858.

[14] Fathurrohman and G. Bichler, "Explaining the positional importance of actors involved in trafficking methamphetamine into Indonesia," *Global Crime*, vol. 22, no. 2, pp. 93–122, Apr. 2021, doi: 10.1080/17440572.2020.1819249.
[15] S. A. Bakken, "Drug dealers gone digital: using signalling theory to analyse criminal online personas and trust," *Global Crime*, vol. 22, no. 1, pp. 51–73, Jan. 2021, doi: 10.1080/17440572.2020.1806826.
[16] A. S. Muzakki and S. Suraji, "Legal protection of consumers in e-commerce through social media in Indonesia in the industrial era 4.0," *International Journal of Law, Crime and Justice*, vol. 1, no. 3, pp. 297–302, Sep. 2024, doi: 10.62951/ijlcj.v1i3.183.
[17] D. Petrou, V. M. Gil, F. Castillo, C. Tunc, and R. Bryce, "Twitter account analysis for drug involvement detection," in *2023 3rd Intelligent Cybersecurity Conference (ICSC)*, 2023, pp. 9–16, doi: 10.1109/ICSC60084.2023.10349992.
[18] M. Khder, "Web scraping or web crawling: state of art, techniques, approaches and application," *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 145–168, Dec. 2021, doi: 10.15849/IJASCA.211128.11.
[19] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment," *IEEE Access*, vol. 9, pp. 78368–78381, 2021, doi: 10.1109/ACCESS.2021.3084050.
[20] J. Tassone, P. Yan, M. Simpson, C. Mendhe, V. Mago, and S. Choudhury, "Utilizing deep learning and graph mining to identify drug use on Twitter data," *BMC Medical Informatics and Decision Making*, vol. 20, no. S11, 2020, doi: 10.1186/s12911-020-01335-3.
[21] N. S. M. Nafis and S. Awang, "An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification," *IEEE Access*, vol. 9, pp. 52177–52192, 2021, doi: 10.1109/ACCESS.2021.3069001.
[22] E. A. Sukma, A. N. Hidayanto, A. I. Pandesenda, A. N. Yahya, P. Widharto, and U. Rahardja, "Sentiment analysis of the new Indonesian government policy (omnibus law) on social media Twitter," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Nov. 2020, pp. 153–158, doi: 10.1109/ICIMCIS51567.2020.9354287.
[23] H. AlSagri and M. Ykhlef, "Quantifying feature importance for detecting depression using random forest," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, 2020, doi: 10.14569/IJACSA.2020.0110577.
[24] A. M. Musolf, E. R. Holzinger, J. D. Malley, and J. E. B. Wilson, "What makes a good prediction? feature importance and beginning to open the black box of machine learning in genetics," *Human Genetics*, vol. 141, no. 9, pp. 1515–1528, Sep. 2022, doi: 10.1007/s00439-021-02402-z.
[25] A. H. Nasution, A. Onan, Y. Murakami, W. Monika, and A. Hanafiah, "Benchmarking open-source large language models for sentiment and emotion classification in Indonesian Tweets," *IEEE Access*, vol. 13, pp. 94009–94025, 2025, doi: 10.1109/ACCESS.2025.3574629.
[26] S. Riyanto, I. S. Sitanggang, T. Djatna, and T. D. Atikah, "Comparative analysis using various performance metrics in imbalanced data for multi-class text classification," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.01406116.
[27] A.-I. Coldea, M. A. Elmessary, D. S. Thayer, I. Scanlon, and H. Davies, "Sensitive data flagging within data quality reports: R and regex integration for effective text flagging in large datasets," *International Journal of Population Data Science*, vol. 9, no. 5, Sep. 2024, doi: 10.23889/ijpds.v9i5.2730.
[28] I. Pambudi and A. Redi, "Criminal law politics and reform to strengthen law enforcement roles in Indonesia in combatting narcotics crimes," *Asian Journal of Social and Humanities*, vol. 3, no. 3, pp. 660–672, Dec. 2024, doi: 10.59888/ajosh.v3i3.467.
[29] R. V. D. Sanden, C. Wilkins, and M. Rychert, "'I straight up criminalized myself on messenger': law enforcement risk management among people who buy and sell drugs on social media," *Drugs: Education, Prevention and Policy*, vol. 31, no. 3, pp. 378–390, May 2024, doi: 10.1080/09687637.2023.2224497.
[30] W. Alruwaili, B. Protano, T. Sirigiriraju, and H. Alhoori, "Using Arabic Tweets to understand drug selling behaviors," *Procedia Computer Science*, vol. 164, pp. 559–564, 2019, doi: 10.1016/j.procs.2019.12.220.

BIOGRAPHIES OF AUTHORS

Adi Hanif Sedar is a master's degree student in Information Technology at Universitas Indonesia. He received his bachelor's degree in Information Systems from Universitas Bina Nusantara. Beyond his academic work, he has practical experience in system analysis and software development and currently works at PT. Telkom Indonesia as a UX researcher and then switched to a software engineer. He also builds software-as-a-service solutions, such as automated faceless video creation and a survey tool, both features integrated with next-gen artificial intelligence. He can be contacted at email: adi.hanif@ui.ac.id.

Lukman Yudokusumo is a master's degree student in Information Technology at Universitas Indonesia. He received his bachelor's degree in Electrical and Electronics Engineering from Telkom University. He can be contacted at email: lukman.yudokusumo@ui.ac.id.

Indra Budi is a distinguished researcher and academic whose pioneering work in computer science and artificial intelligence spans diverse domains such as natural language processing, machine learning, and data mining. With a keen focus on enhancing human-computer interaction, his ResearchGate profile highlights a prolific array of publications and collaborations that underscore his dedication to innovative approaches. He is also a respected faculty member at UI's Department of Computer Science, where his insights and expertise significantly contribute to shaping the evolving landscape of AI research and application. He can be contacted at email: indra@cs.ui.ac.id.

Amanah Ramadiah a proficient data science lead at FNA, is driving suptech solutions for Asia-Pacific central banks and co-developing Central Bank Digital Currency simulations. Also, an adjunct faculty at Universitas Indonesia, her interdisciplinary expertise encompasses computer science, finance, and economics, focusing on systemic risk and digital currencies. A visiting scholar at the Bank of England and IMF, she holds a Ph.D. and M.Sc. from University College London, and a B.Sc. from Universitas Indonesia. She can be contacted at email: amanah.ramadiah06@ui.ac.id.

Aris Budi Santoso is a skilled software engineer with a master's degree in Information Technology from the Universitas Indonesia. Currently works as an app developer at the Ministry of Finance, Indonesia, he excels in full-stack development, web technologies, and software architecture. With a meticulous problem-solving approach, he delivers high-quality solutions and actively contributes to innovative projects. Proficient in JavaScript and Python, he is a proactive collaborator who remains committed to staying updated with industry trends. He can be contacted at email: aris.budi@ui.ac.id.

Prabu Kresna Putra an accomplished researcher in computer science and technology, he works in the National Research and Innovation Agency's Center for Geospatial Research. With a master's degree in Information Technology from the Universitas Indonesia, his specialized research spans data mining, social media analytics, and spatial analytics. Through collaborative engagement in a variety of projects, he consistently brings innovation to the forefront, addressing practical challenges. Proficient in software development and artificial intelligence, his impactful advancements resonate across professional platforms, solidifying his dedication to excellence as a prominent researcher. He can be contacted at email: prab003@brin.go.id.