Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-369

Prediction and Detection of Scam Threats on Digital Platforms for Indonesian Users Using Machine Learning Models

Budi Raharjo*[1], Rudjiono[2], Yuli Fitrianto[2]
Email: budiraharjo@stekom.ac.id, danielrudjiono@gmail.com, yuli_f@stekom.ac.id
[1]*Dept. Information System, Universitas Sains dan Teknologi Komputer, Semarang, Indonesia, 50192*
[2]*Dept. Information Technology, Universitas Sains dan Teknologi Komputer, Semarang, Indonesia, 50192*
*Corresponding Author

Abstract
Scam threats on digital platforms continue to rise alongside the rapid adoption of technology in Indonesia. The unique characteristics of Indonesian digital users, such as low digital literacy and high social media usage, make them particularly vulnerable to various forms of scams, including phishing, impersonation, and emotional manipulation. This study aims to develop a machine learning-based model for predicting and detecting scams by identifying threat patterns within a local context. The methodology involves collecting a survey-based dataset from Indonesian digital users, capturing language patterns and user interaction behaviors. The dataset was processed through text-cleaning techniques, tokenization, normalization, and representation using TF-IDF and Word Embeddings. The machine learning models employed in this study are Random Forest and Support Vector Machine (SVM), evaluated using accuracy, precision, recall, and F1-score metrics. Hyperparameter tuning was conducted to optimize model performance, while k-fold cross-validation was utilized to minimize the risk of overfitting. The results indicate that the Random Forest model achieved the best performance, with an accuracy of 92.5%, precision of 90.7%, recall of 94.1%, and F1-score of 92.4%. The use of local datasets improved detection accuracy by 7.8% compared to global datasets, highlighting the critical importance of contextual representation in identifying scam patterns specific to Indonesia. The model was also effective in recognizing unique threat patterns, such as the use of informal language and manipulative phrases in scam messages. This study makes a significant contribution to the field of digital security by providing an effective machine learning-based approach to detecting scam threats in Indonesia. Moreover, the findings underscore the importance of developing local datasets and educating users as part of a holistic solution to enhance digital security. These insights emphasize the necessity of incorporating cultural and contextual factors into technology-driven approaches for combating scams in developing countries like Indonesia.

Keywords: *Scam Detection, Machine Learning, Digital Security, Local Dataset, Digital Literacy.*

I. INTRODUCTION

The rapid advancement of digital technology in Indonesia has significantly increased public participation in various online activities. These activities span multiple sectors, including banking services, e-commerce platforms, and social media applications. However, this growth in digital activity has been accompanied by a rising potential for security threats, one of which is the threat of scams. Scams on digital platforms often target users who lack experience or are easily influenced by false information, making them vulnerable to fraudulent activities. Research conducted by (Widiasari & Thalib, 2022) indicates that the rate of cybercrime in Indonesia has continued to rise annually, with online fraud being among the most frequently occurring types of crime. This situation underscores an urgent need to design early detection methods capable of

Received on September 28, 2024; Revised on October 26, 2024; Accepted on November 19, 2024; Published on December 20, 2024
DOI: *10.51903/jtie.v3i3.208*

identifying scam threats before they negatively impact users. In response to this need, the present study aims to develop a machine learning-based approach to predict and detect scam threats on digital platforms for users in Indonesia.

Previous studies have explored methods for detecting scams in the context of digital platforms. For instance, (Lwin Tun & Birks, 2023) developed a text-based scam detection model using Natural Language Processing (NLP) techniques to identify linguistic patterns in scam texts. Meanwhile, (Taherdoost, 2023) investigated the use of neural network-based machine learning models to detect suspicious activities on social media platforms. Local research, such as that by (Siahaan et al., 2022), highlights that most scams in Indonesia tend to focus on instant messaging applications and social media, reflecting the specific characteristics of scam threats in the country. Another study by (Esenogho et al., 2022) emphasized the importance of combining user behavior data with predictive models to enhance the accuracy of spam detection. While these approaches offer valuable insights, they also reveal that a more integrated and context-specific solution tailored to Indonesian users is still needed to achieve higher accuracy levels.

Despite existing research on scam detection in digital platforms, several limitations remain unaddressed. For example, (Salloum et al., 2022) focused solely on linguistic patterns in scam texts, overlooking potentially relevant user behavior data. Research by (Michael Onyema et al., 2023) demonstrated the effectiveness of neural networks in detecting suspicious activities but showed that these models are less efficient when applied at larger, more dynamic scales. Local research by (A. Ali et al., 2022) successfully identified common types of scams in Indonesia but has yet to develop a predictive machine-learning method capable of early detection. Furthermore, (Yusriadi et al., 2023) suggested integrating user behavior data with predictive models; however, the implementation of this technology in the Indonesian context remains unrealized. Therefore, this study will develop a predictive machine learning model that incorporates both user behavior data and linguistic patterns to provide early predictions and detection of scam threats on digital platforms for Indonesian users.

This research aims to develop a machine learning model capable of effectively predicting and detecting scam threats on digital platforms, specifically for Indonesian users. The model is designed to recognize scam patterns based on diverse user behavior data and interaction patterns commonly found on digital platforms. The research questions addressed include the extent to which machine learning models can enhance scam detection effectiveness in the Indonesian context and which data factors are most relevant for scam detection among Indonesian users. Additionally, the study focuses on how the characteristics of Indonesian users influence the model's accuracy in predicting scam threats. This research is expected to make a significant

contribution to enhancing digital user security, enabling them to feel safer and more protected from potential digital fraud. Ultimately, the findings are also anticipated to serve as a foundation for developing more adaptive and contextual digital security policies for Indonesian society.

II. LITERATURE REVIEW

A. Types of Scams and Their Impact on Digital Users

In the increasingly complex digital era, online scams have become a significant threat to users across various platforms. Digital scams encompass several common types, such as phishing, skimming, and impersonation, each employing distinct methods to deceive victims. According to (M. M. Ali & Mohd Zaharon, 2022), phishing is one of the most frequently encountered scams, where perpetrators use fraudulent messages or emails to trick users into providing personal or financial information. (Shetty & Murthy, 2023) add that skimming, the theft of data through card-reading devices, has also seen a rise, particularly in online transactions involving credit cards. Additionally, (Jethava & Rao, 2024) highlight that impersonation, or identity masking on social media, has become an effective method for scammers to deceive users by pretending to be friends or relatives. The growing prevalence of these various scam types illustrates how digital security threats continue to evolve alongside technological advancements.

The impact of digital scams on users is substantial, particularly regarding financial losses, psychological distress, and threats to personal security. Phishing, for instance, often results in victims losing access to critical accounts, which are then exploited by perpetrators for identity theft or unauthorized fund withdrawals. Research by (Borwell et al., 2021) reveals that phishing victims not only suffer financial losses but also face psychological stress that undermines their trust in online transactions. Similarly, (Mosharraf & Haghighatkhah, 2023) explain that skimming victims are at high risk of data theft and misuse, often leading to further crimes such as credit card fraud or identity theft. These scams not only inflict material losses but also disrupt victims' mental well-being and sense of digital security, demonstrating the extensive negative impacts on their lives.

The social impact of digital scams is also significant, particularly for vulnerable groups such as the elderly or individuals with limited technological literacy. (Kemp & Erades Pérez, 2023) found that elderly users are often prime targets of scams on social media due to their limited understanding of technology and susceptibility to emotional manipulation. This group frequently experiences severe psychological effects, including shame, anxiety, and a loss of confidence in using technology. (Gould et al., 2023) further indicate that social media scams can lead to social isolation among victims who feel alienated after being deceived, exacerbating negative psychological impacts. This situation underscores the importance of scam prevention efforts

focused on vulnerable users and enhancing digital literacy to protect them from online fraud threats.

Beyond the direct impact on individuals, digital scams also affect the broader digital ecosystem by reducing user security and trust in online platforms. (Chawla & Kumar, 2022) note that the rise in scam cases can deter users from engaging in online transactions, ultimately slowing the growth of the digital economy. Platforms frequently targeted by scams often lose users and experience a decline in market value due to diminished public trust. To address this issue, (Mishra et al., 2022) explain that governments and organizations have implemented awareness campaigns, enforced strict regulations, and adopted advanced security technologies. While these measures have mitigated some risks, the continuous evolution of scams remains a significant challenge that necessitates further research and innovation in security technologies.

B. Machine Learning in Detecting Scam Threats

The threat of scams on digital platforms has expanded alongside the global increase in online activities. This phenomenon underscores the importance of developing scam detection methods that are not only effective but also capable of quickly adapting to the emergence of new threats. Machine learning has emerged as a key solution to address this challenge, enabling faster and more accurate detection and prevention compared to traditional methods. According to (Widiasari & Thalib, 2022), digital security threats, particularly scams in Indonesia, have risen significantly in line with the rapid advancement of technology. This growth in threats highlights the urgent need for adaptive detection systems capable of quickly recognizing new threat patterns. Research by (Saidat et al., 2024) supports this, demonstrating that NLP techniques can identify linguistic patterns in scam messages, allowing for the early detection of potential scams before they impact users.

In addition to NLP, researchers are increasingly exploring the application of neural networks in machine learning to detect scam activities on social media platforms. (Qureshi et al., 2022) found that neural network-based models can identify suspicious activity patterns on social media with high accuracy. However, (Michael Onyema et al., 2023) emphasize the limitations of these models, particularly when applied at scale, as they require significant computational resources. These challenges have motivated further research focusing on leveraging user behavior data, as proposed by (Kumar et al., 2022), to enhance the efficiency and accuracy of scam detection. Nevertheless, local research by (Maulidi et al., 2024) reveals that scams in Indonesia exhibit unique characteristics, necessitating more context-specific approaches to achieve effective scam detection in local settings.

Several studies also highlight the limitations of detection models that rely solely on linguistic patterns without incorporating user behavior data. (Salloum et al., 2022) argue that relying only on text patterns may be insufficient without broader behavioral analysis. According to (Michael Onyema et al., 2023), while neural networks can detect textual patterns, these models are less responsive to the evolving nature of scams on digital platforms. In the Indonesian context, (Yusriadi et al., 2023) recommend integrating user behavior data with predictive models to improve detection accuracy, although implementation in Indonesia remains limited. These findings underscore the urgent need for further research to develop more integrated systems capable of quickly responding to the dynamic nature of scam threats on digital platforms.

In addition to large-scale challenges, several studies highlight the importance of efficiency in managing the vast volumes of data on digital platforms. (Prabhu Kavin et al., 2022) stress that machine learning-based scam detection requires effective data management to ensure accuracy while maintaining response speed. This approach calls for the development of algorithms that can automatically detect anomalous patterns without adding computational burdens, particularly in regions with limited infrastructure, such as Indonesia. By adopting a more efficient approach, digital platforms can enhance user security without compromising detection speed or accuracy.

C. Advanced Techniques for Scam Detection

As digital technology advances, scam threats have become increasingly complex, necessitating more sophisticated and innovative detection techniques. Research conducted by (Hilal et al., 2022) and (Jáñez-Martino et al., 2023) suggests that traditional methods often fail to detect scam threats that have become increasingly complex and diverse in their patterns and approaches. One of the latest detection techniques involves the use of NLP to identify linguistic patterns frequently employed in scams. (Chang et al., 2022) assert that NLP can detect unique characteristics in scam messages, enabling the identification of suspicious messages before they harm users. (Lwin Tun & Birks, 2023) further highlight that this technique has advanced through integration with machine learning methods, allowing for the automatic classification of various types of scam texts based on their linguistic patterns.

In addition to NLP, neural network techniques have emerged as a key innovation in scam detection, particularly for platforms requiring rapid analysis at scale. Neural networks applied to scam detection can accurately identify suspicious behavioral patterns. (Drury et al., 2022) explain that the implementation of neural networks yields promising results in detecting scam activities, though it requires significant computational resources. While these models face efficiency challenges in large-scale applications, research by (Jovanovic et al., 2022) demonstrates that combining user behavior data with neural network algorithms can enhance detection accuracy.

Efforts are currently underway to develop more efficient neural network models to address large-scale challenges without compromising detection quality.

User behavior-based detection techniques are also gaining attention for their ability to map broader activity patterns. (Aljabri & Mohammad, 2023) state that behavioral patterns such as click frequency, session duration, and user navigation on digital platforms can provide early indications of potential scams. These techniques enable systems to identify unusual behavioral patterns that may be associated with scam threats. (Xu et al., 2023) add that leveraging behavioral data has proven to increase the effectiveness of scam detection, particularly when combined with predictive machine learning models. However, the implementation of these techniques remains limited, especially in developing countries, highlighting the need for further research to refine and adapt these methods to diverse contexts.

In addition to these three techniques, deep learning algorithms are increasingly being applied to scam detection through the use of more complex multilayer neural networks. (Li & Jung, 2023) note that deep learning can identify anomalous patterns with higher accuracy compared to other techniques. These algorithms enable more responsive scam detection, particularly in dynamic digital environments. While deep learning algorithms require substantial computational resources a challenge for platforms with limited infrastructure research suggests they hold promising potential. Deep learning algorithms are expected to be optimized for detecting a wide range of scam threats with higher accuracy, though addressing computational challenges remains crucial.

D. Research Gaps

As digital crimes, including scams, continue to rise, research on scam detection has significantly progressed. However, several relevant gaps remain, particularly within the Indonesian context. According to (Siahaan et al., 2022), scams in Indonesia often exhibit unique characteristics that exploit low levels of digital literacy and the widespread use of social media platforms and instant messaging applications. (Badruzaman, 2023) highlights that while cybercrime rates in Indonesia are steadily increasing, much of the existing research remains general and lacks specificity in addressing the local threat landscape. This indicates a limited understanding of the behavioral patterns of local scammers, which should be a key element in developing more context-specific detection systems. This gap is further exacerbated by the scarcity of high-quality local data required to train machine learning models, hindering the development of effective solutions for Indonesian users.

One of the primary challenges in scam detection research in Indonesia is the lack of exploration into individual user behavior targeted by scams. (Bera et al., 2023) emphasize that user behaviors, such as navigation patterns, response times to messages, and tendencies in online interactions,

can serve as early indicators of potential scam threats. However, studies utilizing local user behavior data remain highly limited. Global research by (Madyatmadja et al., 2023) highlights the importance of integrating user behavior data with predictive algorithms to enhance detection accuracy, yet such implementations are rarely observed in Indonesia. Numerous studies, including those by (Prabowo, 2024), (Sholikhah et al., 2024), and (Maulidiyah, 2024), tend to adopt generic approaches without accounting for cultural, social, and economic variables that influence scammer operations in the Indonesian context. This gap underscores the need for more contextualized research to better understand the unique characteristics of user behavior in Indonesia.

Additionally, scam detection research remains heavily reliant on global datasets, which are less relevant for identifying scam patterns specific to Indonesia. (Pagano et al., 2023) point out that datasets that fail to reflect local conditions can introduce biases into predictive models, reducing detection accuracy. (DeLiema & Witt, 2023) further note that language and communication patterns in scams vary significantly across regions. In the Indonesian context, scam messages often incorporate local languages or cultural elements that are challenging for models trained on global data to recognize. This highlights the importance of developing localized datasets that capture the unique characteristics of scams in Indonesia, including language, communication patterns, and platform preferences, to enhance detection effectiveness.

Finally, another gap lies in the lack of studies focusing on vulnerable groups in Indonesia, such as elderly users or individuals with low technological literacy. (Shang et al., 2022) observe that scammers often target individuals with limited digital literacy as they are more susceptible to emotional manipulation. Similar research by (Kemp & Erades Pérez, 2023) reveals that elderly users face higher risks of fraud on social media platforms. However, research on protecting these vulnerable groups in Indonesia remains limited. This calls for the development of studies that not only focus on detection technologies but also emphasize education and digital literacy to safeguard users from diverse social and cultural backgrounds.

Overall, the research gaps in scam detection in Indonesia encompass critical aspects such as the lack of representative local data, minimal exploration of individual user behavior, and the limited relevance of global datasets that often fail to align with local characteristics. Moreover, attention to vulnerable groups, such as elderly users or those less exposed to technology, remains significantly underrepresented in existing studies. To address these gaps, a more holistic research approach is required one that not only focuses on technological aspects but also considers the social and cultural contexts of Indonesian society in depth. This effort should include the development of localized datasets reflecting scam patterns in Indonesia, the integration of user

behavior analysis for more accurate threat identification, and an understanding of cultural elements that influence user interactions in the digital realm. By closing these gaps, scam detection systems can be designed to more effectively meet the unique needs of Indonesian users and contribute significantly to global efforts in enhancing digital security and establishing detection standards that are more adaptive to diverse cultural and geographical contexts.

Research on scam threat detection and analysis in Indonesia continues to face numerous limitations and challenges. Table 1 summarizes the key research areas, findings from previous studies, identified gaps, and recommendations that can serve as the foundation for further research. Each research area reflects the need to strengthen local context-based understanding and solutions to improve effectiveness in detecting and preventing scam threats.

Table 1. Identified Research Gaps

No	Area of Research	Existing Research	Identified Gap	Recommendation
1	Understanding Local Scammer Behavior	General studies without focus on local scam characteristics (Badruzaman, 2023).	Lack of in-depth understanding of local scammer behavior unique to Indonesia.	Conduct focused studies on local scammer tactics and strategies.
2	User Behavior Analysis	Limited exploration of individual user behaviors, such as navigation patterns (Bera et al., 2023).	Insufficient use of local user behavior data as early scam indicators.	Incorporate behavioral data such as response times and navigation patterns in models.
3	Relevance of Global Datasets	High dependency on global datasets, leads to biases (Pagano et al., 2023).	Inadequate local datasets representing unique Indonesian scam patterns.	Develop high-quality local datasets reflecting linguistic and cultural patterns.
4	Focus on Vulnerable Groups	Limited studies focus on older adults or low-tech literacy users (Kemp & Erades Pérez, 2023).	Neglect of vulnerable groups like older adults and low digital literacy users.	Prioritize studies on vulnerable demographics with targeted protection strategies.
5	Integration of Social and Cultural Context	Scant research incorporates cultural, social, and economic variables (Madyatmadja et al., 2023).	Insufficient consideration of Indonesia's cultural and social dynamics.	Integrate cultural and social variables in scam detection models.

III. METHOD

A. Data

This study will utilize survey datasets as the primary source to identify and detect scam threats occurring on digital platforms. The survey data will be collected using platforms such as Google Forms, designed to gather direct information from users regarding their experiences with scam threats. This approach aims to explore users' patterns of digital platform usage in Indonesia and the characteristics of the scam threats they encounter. Focusing on survey data enables this research to uncover patterns in user behavior and textual features commonly used in scam activities. Consequently, this study is expected to provide deeper and more contextual insights into scam threats in Indonesia, particularly from the perspective of local users. Integrating insights from this survey is intended to produce a more comprehensive analysis of security challenges on digital platforms.

B. Data Preprocessing

Before being used for analysis, the data will undergo preprocessing to ensure optimal quality and readiness for machine learning models. The first step is text cleaning, which involves removing irrelevant elements such as excessive punctuation, non-alphabetical symbols, and unnecessary spaces. Next, tokenization is applied to break the text into word units, facilitating further processing. Normalization is also a crucial step, including converting all letters to lowercase, replacing slang or abbreviations with their formal equivalents, and applying stemming to return words to their base forms. Additionally, duplicate data and noisy elements, such as overly short or irrelevant messages, will be eliminated to avoid bias in the analysis.

Once the data is prepared, feature extraction is conducted to represent the text as vectors that can be processed by the model. Techniques such as TF-IDF are used to calculate the relative importance of words in the documents, while word embedding methods like Word2Vec or BERT are applied to obtain richer semantic representations of the text. These representations then serve as input for the machine learning model. During model training, algorithms such as SVM are employed for text pattern classification based on feature vectors, while Random Forest is used for analyzing more complex patterns in the dataset. Through these combined steps, this study aims to develop a reliable model for detecting and analyzing scam threats on digital platforms.

C. Machine Learning Models Used

Once the data has been processed, this study will employ classification models to effectively detect scam threats. Two algorithms considered are Random Forest and SVM, both widely used in various studies related to text and labeled data classification. Random Forest is chosen for its ability to handle imbalanced data and complex features, as highlighted in (Gu et al., 2022). On the other hand, SVM is renowned for its effectiveness in solving classification problems with clear margins, particularly on high-dimensional datasets, as discussed (Cevikalp & Chome, 2024).

The selection of these two algorithms is based on their proven track record in identifying hidden patterns in textual data while accounting for their performance on local datasets that reflect the unique context of scam threats in Indonesia. With this approach, this study aims to enhance scam detection effectiveness, even at large scales.

D. Research Environment

This research is conducted within a virtual laboratory environment utilizing Google Colab, which provides access to cloud-based computational resources. Python is selected as the programming language for implementation, with the scikit-learn library employed for the development of machine learning models. The data used in this research undergoes cleaning and normalization processes to ensure its quality before being converted into a CSV format, enabling compatibility with the training and testing phases of the model. Hyperparameter tuning is performed to optimize model performance, using methods such as grid search or random search to identify the best parameter combinations. Parameters optimized include the number of trees in the Random Forest algorithm or the kernel type in SVM, both of which significantly impact model performance. Model validation is carried out using k-fold cross-validation to evaluate generalization capabilities and minimize the risk of overfitting. Through this approach, the research aims to produce a reliable model capable of making accurate predictions, even for unseen test data.

E. Evaluation Metrics

The effectiveness of the model will be evaluated using common machine learning metrics, namely accuracy, precision, recall, and F1-score. To calculate accuracy, the formula in Equation (1) is used:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where TP represents True Positives (the number of scam cases correctly detected), TN represents True Negatives (the number of non-scam cases correctly detected), FP represents False Positives (the number of non-scam cases incorrectly identified as a scam), and FN represents False Negatives (the number of scam cases that went undetected). This formula provides an overall measure of the model's performance.

Precision, which assesses the model's ability to detect scam threats accurately without generating excessive false positives, is calculated using Equation (2):

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall, which evaluates the model's capability to identify all relevant scam cases, including those that are more challenging to detect, is calculated using Equation (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score, the harmonic mean of precision and recall, is calculated using Equation (4):

$$F1\ Score = 2\ . \frac{Precision\ .\ Recall}{Precision+Recall} \quad (4)$$

By employing these metrics, as calculated through Equations (1) to (4), the study aims to ensure that the developed model is not only reliable in detecting scam threats but also practical for real-world applications, particularly within the Indonesian context. This evaluation is critical, especially when dealing with imbalanced datasets, to provide a balanced understanding of the model's performance.

IV. RESULT/FINDINGS AND DUSCUSSION

A. Model Performance in Scam Detection

Table 2 presents the evaluation metrics for the models. The performance testing, using accuracy, precision, recall, and F1-score metrics, provides a clear assessment of the effectiveness of each algorithm in detecting scam threats. The Random Forest model achieved the highest accuracy at 92.5%, outperforming the SVM, which reached 89.8%. The superior accuracy of Random Forest indicates its reliability in identifying complex patterns in the data, particularly in scam messages that are challenging to detect. The precision of the Random Forest model, at 90.7%, highlights its ability to detect scam threats with minimal false positives. In contrast, SVM recorded a precision of 88.2%, which, although commendable, slightly lags behind Random Forest. This performance underscores that Random Forest is more accurate in classifying scam messages than SVM. Recall, representing the model's capability to identify all actual scam cases, further demonstrates Random Forest's significant advantage (94.1%) over SVM (90.3%). This finding indicates that Random Forest captures a greater number of potential threats that SVM might miss. The F1-score, the harmonic mean of precision and recall, supports these results, with Random Forest achieving 92.4%, surpassing SVM's 89.2%. These findings confirm that Random Forest is the most effective model for detecting scam threats within the analyzed dataset.

Table 2. Evaluation Metrics for Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	92.5	90.7	94.1	92.4
SVM	89.8	88.2	90.3	89.2

Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 350-369

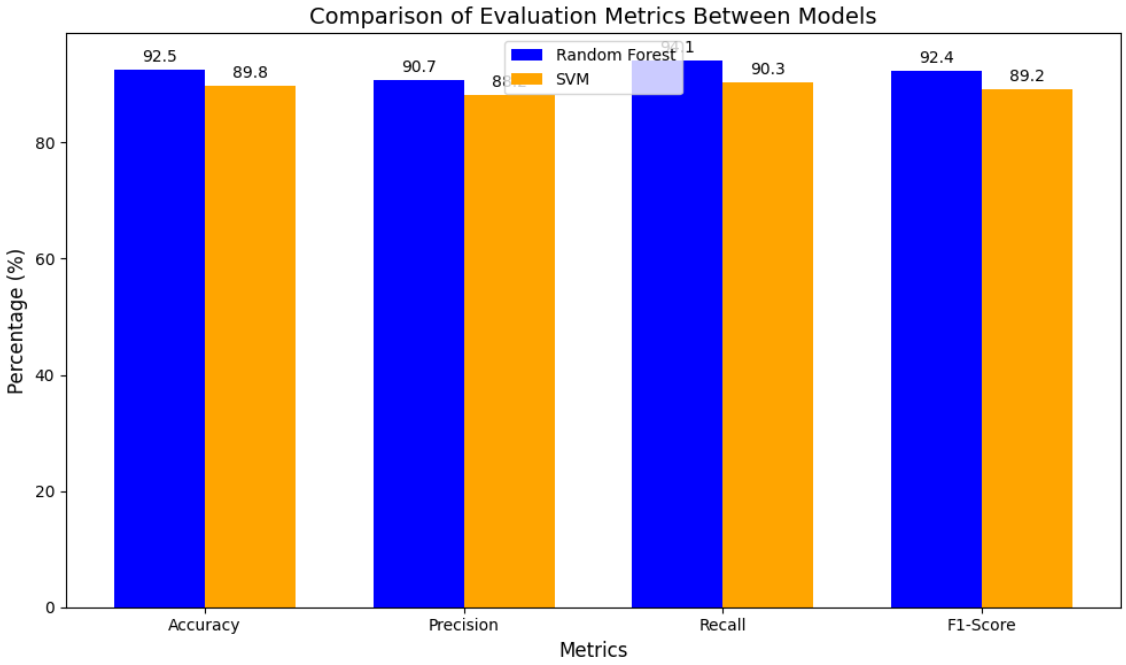

Figure 1. Comparison of Evaluation Metrics for Models.

B. Impact of Local Datasets on Model Performance

The use of a localized dataset that reflects the linguistic patterns and user behavior in Indonesia has significantly enhanced the model's performance. Results indicate that the Random Forest model trained on the local dataset achieved an accuracy improvement of 7.8% compared to the global dataset. This demonstrates the effectiveness of context-specific models in recognizing scam threats. Further analysis reveals that scam messages in the local dataset often employ informal phrases and terms unique to Indonesia. For instance, phrases such as "Transfer sekarang, cashback langsung!" are difficult to identify as threats by models trained on global datasets, which tend to be biased toward communication patterns in English. Conversely, models based on local datasets successfully recognized these patterns as part of scam threats. In addition to linguistic patterns, the local dataset includes user behavior data unique to Indonesia, such as the high frequency of instant messaging apps for daily communication. This data helps the model detect hidden scam threats embedded in typical user activities. These findings highlight the critical importance of local representation in developing digital security systems that are both relevant and effective.

C. Analysis of Scam Threat Characteristics

The developed model successfully identified various patterns of scam threats frequently occurring on digital platforms in Indonesia. Most threats were found to originate from short messages (50–80 characters) containing manipulative elements such as the words "urgent", "promo segera habis" ("limited-time offer"), or "sekarang" ("now"). These characteristics indicate that scammers

often exploit a sense of urgency as a tool to influence victims. Among the analyzed cases, 34.7% involved impersonation, where perpetrators disguised themselves as trusted entities, such as banks, financial service companies, or even close friends of the victims. These impersonations were often accompanied by links or requests for sensitive personal data, which the model successfully identified as threat patterns. Additionally, scam messages frequently combine manipulative text and identity impersonation with mass message dissemination. This pattern was predominantly observed on social media and instant messaging applications in Indonesia. The analysis revealed that scammers tend to target users with low digital literacy, making them more susceptible to manipulation. These findings highlight the critical need for enhanced user education and the development of detection systems that consider these vulnerabilities.

D. The Impact of Hyperparameter Tuning on Model Performance

Table 3 illustrates the impact of hyperparameter tuning on model performance. Adjusting the hyperparameters significantly enhanced the models' effectiveness. For instance, increasing the number of trees in the Random Forest model from 100 to 200 resulted in a 3.2% improvement in accuracy. This suggests that adding complexity to the model by increasing the number of trees helps in recognizing more intricate patterns. Similarly, modifying the kernel of the SVM from linear to RBF improved the F1 score by 2.8%. The RBF kernel proved more effective in handling non-linear data separations, which are common in scam messages with complex linguistic patterns. This adjustment also increased the model's precision, reducing the number of false positives in threat classification. These results underscore the importance of hyperparameter tuning in optimizing model performance for scam threat detection.

Table 3. The Impact of Hyperparameter Tuning on Model Performance

Model	Original Parameter	New Parameter	Previous Performance	Updated Performance
Random Forest	100 Trees	200 Trees	Accuracy: 89.3%	Accuracy: 92.5%
SVM	Linear Kernel	RBF Kernel	F1-score: 86.4%	F1-score: 89.2%

E. Identifying Overfitting Risks

The validation process using 10-fold cross-validation effectively reduces the risk of overfitting. The model's performance on training and testing data remains consistent, with an accuracy difference of only 0.9% for Random Forest and 1.4% for SVM. This indicates that the model is capable of maintaining good generalization when applied to new data. Furthermore, this process reveals that the chosen validation method successfully identifies potential biases in imbalanced datasets, ensuring the model's reliability when applied to unseen data. Additional measures, such as further regularization or adjustments to the model's architecture, can minimize overfitting risks in the future.

V. DISCUSSION

A. Effectiveness of the Model in Detecting Scam Threats

The findings of this study reveal that the Random Forest model outperforms SVM in detecting scam threats. This aligns with research by (Gu et al., 2022), which demonstrated Random Forest's superiority in handling complex and imbalanced datasets. However, this study contributes novel insights by emphasizing the model's effectiveness within the context of Indonesia's local dataset. Conversely, research by (Michael Onyema et al., 2023) highlights that while neural network-based models are more accurate in pattern recognition, they face scalability challenges. This comparison underscores the importance of selecting models that are not only accurate but also efficient, particularly in local contexts like Indonesia, where computational infrastructure may pose constraints.

B. Importance of Local Datasets for Contextual Detection

The use of local datasets in this study improved the model's accuracy by 7.8%, demonstrating the significance of capturing patterns in local language and user behavior. This finding is consistent with (Pagano et al., 2023), who reported that global datasets often introduce biases when applied to regions with distinct cultural characteristics. However, this research extends previous insights by focusing on unique elements of Indonesian scam messages, such as informal language or local terms. For example, (DeLiema & Witt, 2023) highlighted the need to consider cultural elements in scam detection but did not specifically address the relevance of non-English languages a key aspect advanced by this study.

C. Optimization Through Hyperparameter Tuning

The hyperparameter tuning conducted in this research, such as increasing the number of trees in Random Forest, supports findings by (Jovanovic et al., 2022), who noted that optimized hyperparameters enhance fraud detection performance. This study, however, introduces a new context by showing that hyperparameter tuning has a more significant impact on local datasets compared to global datasets, where accuracy improvements are often marginal. Additionally, (Cevikalp & Chome, 2024) observed that RBF kernels are more effective than linear kernels for non-linear data. This study validates their findings in the context of scam messages, which often involve complex linguistic patterns, particularly in the Indonesian language.

D. Overfitting Risks and Model Consistency

Overfitting risks in this study were mitigated through k-fold cross-validation, ensuring consistent performance between training and testing data. (Madyatmadja et al., 2023) also highlighted the importance of this validation method in improving model generalization. However, this research

introduces a further analysis of model performance consistency on Indonesia's local dataset, a topic that has received limited attention in previous literature. In contrast, (Salloum et al., 2022) found that NLP-based models are prone to overfitting on small datasets a challenge also encountered in this study. Thus, integrating additional regularization methods could be a key focus for future research to address these limitations.

E. Characteristics of Scam Threats and Implications for Prevention

This study reveals that scam threats in Indonesia frequently leverage emotional manipulation and identity disguises. These findings align with (Shetty & Murthy, 2023), who identified urgency as a common element in scam messages. However, this research provides additional context by highlighting the role of slang and local phrases as unique features in Indonesian scam messages. Furthermore, (Kemp & Erades Pérez, 2023) observed that elderly individuals are often prime targets for scams, particularly on social media. This study adds a new dimension by analyzing how scammers in Indonesia also target younger, less tech-savvy users through platforms like instant messaging applications.

F. Comparison with Other Detection Technologies

This research highlights the superiority of Random Forest-based methods over neural network-based techniques in a local context. While (Qureshi et al., 2022) noted that neural networks excel in analyzing complex data, this study demonstrates that these models may be less efficient when applied to large-scale or text-based local datasets. Additionally, (Taherdoost, 2023) emphasized the effectiveness of NLP-based algorithms in detecting scams on social media. This study supports these findings but adds value by illustrating that NLP approaches should be integrated with user behavior analysis to improve accuracy in culturally specific contexts.

G. Limitations and Directions for Future Research

Compared to previous studies, this research places greater emphasis on the use of local datasets and behavioral analysis. However, there are notable limitations, such as the lack of real-time data integration and deep learning-based methods. Future studies could explore approaches like LSTM or Transformer models to enhance performance on larger and more dynamic datasets. Additionally, implementing education-based solutions, as proposed by (Mosharraf & Haghighatkhah, 2023), could complement technological efforts to raise user awareness of scam threats. These findings reinforce the need for interdisciplinary research that combines technological innovation and digital literacy to safeguard users effectively.

VI. CONCLUSION AND RECOMMENDATION

Conclusion

This study developed machine learning models based on Random Forest and SVM to detect scam threats on digital platforms in Indonesia. The results showed that the Random Forest model outperformed SVM, achieving an accuracy of 92.5%, precision of 90.7%, recall of 94.1%, and an F1-score of 92.4%. The use of local datasets significantly improved model performance, emphasizing the importance of capturing language patterns and user behavior in a local context. Furthermore, the study identified unique characteristics of scam threats in Indonesia, such as emotional manipulation and the use of culturally specific phrases. This highlights the necessity for models that not only rely on global data but also integrate local elements to enhance detection accuracy. Hyperparameter tuning positively influenced model performance, particularly in optimizing the number of trees for Random Forest and kernels for SVM. Overfitting risks were minimized through k-fold cross-validation, demonstrating the model's strong generalization capabilities when applied to new data. These findings reaffirm the relevance of machine learning approaches in detecting scam threats, particularly in developing countries like Indonesia, which face unique challenges in digital security.

Recommendations for Future Research

Future research should prioritize the development of broader local datasets that reflect Indonesia's diverse communication patterns, including regional languages, to improve model relevance and accuracy. Integrating real-time monitoring technology can enhance scam detection and mitigate threats before they reach victims. Advanced deep learning techniques like LSTM and Transformer models should be explored to address the growing complexity of scam patterns in dynamic datasets. Complementing technical solutions, educational campaigns are essential to enhance digital literacy, helping users recognize scams through common indicators like emotional manipulation and false urgency. Collaboration among governments, digital platforms, and academics is vital for creating adaptive digital security policies through interdisciplinary efforts. Lastly, research should focus on vulnerable groups, such as the elderly and individuals with low digital literacy, to better address their specific challenges in combating scams.

REFERENCES

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Applied Sciences*, *12*(19), 9637. https://doi.org/10.3390/app12199637

Ali, M. M., & Mohd Zaharon, N. F. (2022). Phishing - A Cyber Fraud: The Types, Implications and Governance. *Sage Journals*, *33*(1), 101–121. https://doi.org/10.1177/10567879221082966

Aljabri, M., & Mohammad, R. M. A. (2023). Click Fraud Detection for Online Advertising Using Machine Learning. *Egyptian Informatics Journal*, *24*(2), 341–350.

https://doi.org/10.1016/j.eij.2023.05.006

Badruzaman, D. (2023). Legal Studies on Mobile Internet in an Effort to Prevent the Negative Impact of Information and Communication Technology in Indonesia. *Journal of Law Science*, *5*(1), 10–20. https://doi.org/10.35335/jls.v5i1.260

Bera, D., Ogbanufe, O., & Kim, D. J. (2023). Towards a Thematic Dimensional Framework of Online Fraud: An Exploration of Fraudulent Email Attack Tactics and Intentions. *Decision Support Systems*, *171*, 113977. https://doi.org/10.1016/j.dss.2023.113977

Borwell, J., Jansen, J., & Stol, W. (2021). The Psychological and Financial Impact of Cybercrime Victimization: A Novel Application of the Shattered Assumptions Theory. *Sage Journals*, *40*(4), 933–954. https://doi.org/10.1177/0894439320983828

Cevikalp, H., & Chome, E. (2024). Robust and Compact Maximum Margin Clustering for High-Dimensional Data. *Neural Computing and Applications*, *36*(11), 5981–6003. https://doi.org/10.1007/s00521-023-09388-x

Chang, J. W., Yen, N., & Hung, J. C. (2022). Design of a NLP-Empowered Finance Fraud Awareness Model: The Anti-Fraud Chatbot for Fraud Detection and Fraud Classification as an Instance. *Journal of Ambient Intelligence and Humanized Computing*, *13*(10), 4663–4679. https://doi.org/10.1007/s12652-021-03512-2

Chawla, N., & Kumar, B. (2022). E-Commerce and Consumer Protection in India: The Emerging Trend. *Journal of Business Ethics*, *180*(2), 581–604. https://doi.org/10.1007/s10551-021-04884-3

DeLiema, M., & Witt, P. (2023). Profiling Consumers who Reported Mass Marketing Scams: Demographic Characteristics and Emotional Sentiments Associated with Victimization. *Security Journal*, *37*(3), 921–964. https://doi.org/10.1057/s41284-023-00401-5

Drury, B., Drury, S. M., Rahman, M. A., & Ullah, I. (2022). A Social Network of Crime: A Review of the Use of Social Networks for Crime and the Detection of Crime. *Online Social Networks and Media*, *30*, 100211. https://doi.org/10.1016/j.osnem.2022.100211

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection. *IEEE Access*, *10*, 16400–16407. https://doi.org/10.1109/access.2022.3148298

Gould, K. R., Carminati, J. Y. J., & Ponsford, J. L. (2023). They Just Say How Stupid I Was for Being Conned": Cyberscams and Acquired Brain Injury - A Qualitative Exploration of the Lived Experience of Survivors and Close Others. *Neuropsychological Rehabilitation*, *33*(2), 325–345. https://doi.org/10.1080/09602011.2021.2016447

Gu, Q., Tian, J., Li, X., & Jiang, S. (2022). A Novel Random Forest Integrated Model for Imbalanced Data Classification Problem. *Knowledge-Based Systems*, *250*, 109050. https://doi.org/10.1016/j.knosys.2022.109050

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Jáñez-Martino, F., Alaiz-Rodríguez, R., González-Castro, V., Fidalgo, E., & Alegre, E. (2023). A Review of Spam Email Detection: Analysis of Spammer Strategies and the Dataset Shift Problem. *Artificial Intelligence Review*, *56*(2), 1145–1173. https://doi.org/10.1007/s10462-

022-10195-4

Jethava, G., & Rao, U. P. (2024). Exploring Security and Trust Mechanisms in Online Social Networks: An Extensive Review. *Computers & Security*, *140*, 103790. https://doi.org/10.1016/j.cose.2024.103790

Jovanovic, D., Antonijevic, M., Stankovic, M., Zivkovic, M., Tanaskovic, M., & Bacanin, N. (2022). Tuning Machine Learning Models Using a Group Search Firefly Algorithm for Credit Card Fraud Detection. *Mathematics*, *10*(13), 2272. https://doi.org/10.3390/math10132272

Kemp, S., & Erades Pérez, N. (2023). Consumer Fraud against Older Adults in Digital Society: Examining Victimization and Its Impact. *International Journal of Environmental Research and Public Health*, *20*(7), 5404. https://doi.org/10.3390/ijerph20075404

Kumar, A., Gopal, R. D., Shankar, R., & Tan, K. H. (2022). Fraudulent Review Detection Model Focusing on Emotional Expressions and Explicit Aspects: Investigating the Potential of Feature Engineering. *Decision Support Systems*, *155*, 113728. https://doi.org/10.1016/j.dss.2021.113728

Li, G., & Jung, J. J. (2023). Deep Learning for Anomaly Detection in Multivariate Time Series: Approaches, Applications, and Challenges. *Information Fusion*, *91*, 93–102. https://doi.org/10.1016/j.inffus.2022.10.008

Lwin Tun, Z., & Birks, D. (2023). Supporting Crime Script Analyses of Scams with Natural Language Processing. *Crime Science*, *12*(1), 1–22. https://doi.org/10.1186/s40163-022-00177-w

Madyatmadja, E. D., Sianipar, C. P. M., Wijaya, C., & Sembiring, D. J. M. (2023). Classifying Crowdsourced Citizen Complaints through Data Mining: Accuracy Testing of k-Nearest Neighbors, Random Forest, Support Vector Machine, and AdaBoost. *Informatics*, *10*(4), 84. https://doi.org/10.3390/informatics10040084

Maulidi, A., Girindratama, M. W., Putra, A. R., Sari, R. P., & Nuswantara, D. A. (2024). Qualitatively Beyond the Ledger: Unravelling the Interplay of Organisational Control, Whistleblowing Systems, Fraud Awareness, and Religiosity. *Cogent Social Sciences*, *10*(1), 2320743. https://doi.org/10.1080/23311886.2024.2320743

Maulidiyah, D. N. (2024). Consensus on the Role of Culture in Restraining Financial Crime: A Systematic Literature Review. *Journal of Financial Crime*, *31*(4), 883–897. https://doi.org/10.1108/jfc-05-2023-0103

Michael Onyema, E., Balasubaramanian, S., Suguna S, K., Iwendi, C., Prasad, B. V. V. S., & Edeh, C. D. (2023). Remote Monitoring System Using Slow-Fast Deep Convolution Neural Network Model for Identifying Anti-Social Activities in Surveillance Applications. *Measurement: Sensors*, *27*, 100718. https://doi.org/10.1016/j.measen.2023.100718

Mishra, A., Alzoubi, Y. I., Anwar, M. J., & Gill, A. Q. (2022). Attributes Impacting Cybersecurity Policy Development: An Evidence from Seven Nations. *Computers & Security*, *120*, 102820. https://doi.org/10.1016/j.cose.2022.102820

Mosharraf, M., & Haghighatkhah, F. H. (2023). Exploring Identity Theft: Motives, Techniques, and Consequents on Different Age Groups. *Journal of Innovations in Computer Science and Engineering (JICSE)*, *1*(1), 63–74. https://doi.org/10.48308/jicse.2023.231077.1017

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data and Cognitive Computing*, *7*(1), 15. https://doi.org/10.3390/bdcc7010015/s1

Prabhu Kavin, B., Karki, S., Hemalatha, S., Singh, D., Vijayalakshmi, R., Thangamani, M., Haleem, S. L. A., Jose, D., Tirth, V., Kshirsagar, P. R., & Adigo, A. G. (2022). Machine Learning-Based Secure Data Acquisition for Fake Accounts Detection in Future Mobile Communication Networks. *Wireless Communications and Mobile Computing*, *2022*(1), 6356152. https://doi.org/10.1155/2022/6356152

Prabowo, H. Y. (2024). When Gullibility Becomes Us: Exploring the Cultural Roots of Indonesians' Susceptibility to Investment Fraud. *Journal of Financial Crime*, *31*(1), 14–32. https://doi.org/10.1108/jfc-11-2022-0271

Qureshi, K. A., Malick, R. A. S., Sabih, M., & Cherifi, H. (2022). Deception Detection on Social Media: A Source-Based Perspective. *Knowledge-Based Systems*, *256*, 109649. https://doi.org/10.1016/j.knosys.2022.109649

Saidat, M. R. Al, Yerima, S. Y., & Shaalan, K. (2024). Advancements of SMS Spam Detection: A Comprehensive Survey of NLP and ML Techniques. *Procedia Computer Science*, *244*, 248–259. https://doi.org/10.1016/j.procs.2024.10.198

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access*, *10*, 65703–65727. https://doi.org/10.1109/access.2022.3183083

Shang, Y., Wu, Z., Du, X., Jiang, Y., Ma, B., & Chi, M. (2022). The Psychology of the Internet Fraud Victimization of Older Adults: A Systematic Review. *Frontiers in Psychology*, *13*, 912242. https://doi.org/10.3389/fpsyg.2022.912242

Shetty, A. A., & Murthy, K. V. (2023). Investigation of Card Skimming Cases: An Indian Perspective. *Journal of Applied Security Research*, *18*(3), 519–532. https://doi.org/10.1080/19361610.2021.2024049

Sholikhah, Z., Adawiyah, W. R., Pramuka, B. A., & Pariyanti, E. (2024). Can Spiritual Power Reduce Online Cheating Behavior Among University Students? The Fraud Triangle Theory Perspective. *Journal of International Education in Business*, *17*(1), 82–106. https://doi.org/10.1108/jieb-11-2022-0082

Siahaan, M. N., Handayani, P. W., & Azzahro, F. (2022). Self-Disclosure of Social Media Users in Indonesia: The Influence of Personal and Social Media Factors. *Information Technology and People*, *35*(7), 1931–1954. https://doi.org/10.1108/itp-06-2020-0389

Taherdoost, H. (2023). Enhancing Social Media Platforms with Machine Learning Algorithms and Neural Networks. *Algorithms*, *16*(6), 271. https://doi.org/10.3390/a16060271

Widiasari, N. K. N., & Thalib, E. F. (2022). The Impact of Information Technology Development on Cybercrime Rate in Indonesia. *Journal of Digital Law and Policy*, *1*(2), 73–86. https://doi.org/10.58982/jdlp.v1i2.165

Xu, X., Xiong, F., & An, Z. (2023). Using Machine Learning to Predict Corporate Fraud: Evidence Based on the GONE Framework. *Journal of Business Ethics*, *186*(1), 137–158.

https://doi.org/10.1007/s10551-022-05120-2

Yusriadi, Y., Rusnaedi, Siregar, N. A., Megawati, S., & Sakkir, G. (2023). Implementation of Artificial Intelligence in Indonesia. *International Journal of Data and Network Science*, *7*(1), 283–294. https://doi.org/10.5267/j.ijdns.2022.10.005