Perbanas International Conference on Economics, Business, Management, Accounting and IT (PROFICIENT) 2026
*"Integrity-Driven Innovation: Reimagining Business, Finance, and Digital Transformation for Sustainable Impact"*

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

# Online Payment Fraud Detection Using the XGBoost Algorithm in MSMEs

Widya Ayu Purwati[1,a)], Rina Hudaini[2,b)], Mercurius Broto Legowo [3,c)]

*Faculty of Information Technology, Perbanas Institute, Jakarta, Indonesia*

[a)]Corresponding author: widya.ayu09@perbanas.id
[b)]rina.hudaini04@perbanas.id
[c)] mercurius@perbanas.id

**Abstract.** *Rapid growth of online payment transactions among Micro, Small, and Medium Enterprises (MSMEs) in Indonesia has been followed by rising fraud cases that threaten business sustainability. Fraud in digital payments can cause major financial losses and weaken consumer trust in online payment systems. Therefore, MSMEs need an accurate and fast fraud detection system to identify suspicious transactions. This study develops a fraud detection model using the XGBoost (eXtreme Gradient Boosting) algorithm to detect fraudulent transactions in real time within MSMEs payment systems. The model shows high accuracy, precision, recall, and F1-score, outperforming baseline models such as Random Forest and Logistic Regression. The dataset used is the "Fraud Detection" dataset from Kaggle. The research process includes data preprocessing using SMOTE (Synthetic Minority Over-Sampling Technique) to address class imbalance, feature engineering to generate additional attributes such as transaction frequency, and data scaling with RobustScaler to handle outliers. Results show that the XGBoost model performs very strongly in detecting fraudulent transactions. The model also produces predictions quickly, making it suitable for real-time fraud detection in MSMEs online payment platforms. By identifying suspicious transaction patterns early, the system helps MSMEs reduce potential financial losses and maintain trust in digital payment services for users overall.*

***Keywords**: XGBoost, fraud detection, online payment, MSMEs, machine learning, SMOTE, Kaggle dataset.*

## INTRODUCTION

The rapid growth of online payment transactions in the Micro, Small, and Medium Enterprises (MSMEs) sector in Indonesia has become one of the main drivers of national digital economic transformation. Bank Indonesia projects that the value of national digital banking transactions will continue to increase and exceed IDR 555 trillion by 2025, indicating the expanding adoption of digital payment systems among MSMEs (Ibrahim & Alfauzan, 2025a; Permana, Kusuma, et al., 2024). However, this rapid development is also accompanied by increasing cases of digital fraud, including card-not-present fraud, account takeover, and synthetic identity fraud, which may result in substantial financial losses and reduce customer trust in digital platforms (Yunanto & Budiyanto, 2024). This issue is becoming increasingly critical considering the significant contribution of MSMEs to the Indonesian economy in terms of gross domestic product and employment generation.

Conventional rule-based fraud detection approaches are often ineffective in identifying complex anomaly patterns in large-scale transaction data, particularly when the dataset is highly imbalanced, where fraudulent transactions represent only a small portion of the overall data. Therefore, this study proposes the implementation of the XGBoost algorithm as a machine learning solution capable of handling imbalanced datasets through gradient boosting mechanisms, regularization, and class weight optimization (Chen & Guestrin, 2016a). The dataset used in this study is the fraud_payment_UMKM dataset consisting of 10,000

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

transaction records containing transactional, behavioral, and fraud-related features. These features are considered representative for simulating online payment transaction activities within the Indonesian MSME sector.

Based on this background, the study focuses on three main issues. First, the pre-processing of imbalanced transaction data, including handling missing values, categorical encoding, and oversampling using the Synthetic Minority Oversampling Technique (SMOTE). Second, the evaluation of XGBoost model performance using accuracy, F1-score, and Area Under the Curve (AUC) metrics compared with Logistic Regression and Random Forest algorithms. Third, the implementation of real-time fraud detection with an inference time target of less than one second per transaction, along with the identification of important transactional and behavioral features influencing fraud detection.

Previous studies have demonstrated that the combination of XGBoost and SMOTE is effective in improving fraud detection performance on imbalanced transaction datasets. (Yunanto & Budiyanto, 2024) reported that SMOTE-based XGBoost implementation significantly improved fraud detection capability in the financial services sector. In addition, (Rahmadani et al., 2025) showed that hyperparameter optimization using GridSearchCV enhanced the classification performance of XGBoost on credit card fraud datasets. More recent findings also indicate that optimized XGBoost frameworks for real-time fraud detection can maintain high AUC performance while remaining computationally efficient for operational implementation. Therefore, this study aims to evaluate the effectiveness of XGBoost for fraud detection in MSME digital payment transactions while supporting practical implementation in online payment environments.

The main objective of this research is to develop an XGBoost-based online payment fraud detection model for MSMEs with high classification performance and real-time prediction capability. To achieve this objective, the study conducts data preprocessing, feature engineering relevant to the Indonesian MSME context, hyperparameter optimization using GridSearchCV, and comprehensive evaluation using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics.

Theoretically, this research is expected to contribute to the literature on machine learning-based fraud detection in Indonesian MSMEs, particularly the application of XGBoost on imbalanced datasets. Practically, the proposed model may help MSMEs reduce fraud-related financial losses through efficient real-time detection while supporting the development of more secure and reliable digital payment systems.

# LITERATURE REVIEW

## Online Payment Systems and MSMEs

Online payment systems refer to digital financial infrastructures that facilitate monetary transactions through electronic channels, including mobile banking applications, e-wallets, payment gateways, and point-of-sale terminals integrated with digital networks (Gupta & Arora, 2023a). These systems support real-time transactions and improve operational efficiency in modern commerce. For MSMEs, digital payment adoption offers broader market access and lower transaction costs, but it also increases exposure to cyber-financial threats such as fraudulent transactions and account misuse (Ibrahim & Alfauzan, 2025a).

The adoption of digital payments among Indonesian MSMEs has grown rapidly in the post-pandemic era due to increasing demand for contactless and remote transactions (Permana, Kusuma, et al., 2024). However, many MSMEs still lack adequate cybersecurity infrastructure and fraud prevention mechanisms, creating vulnerabilities within digital payment ecosystems. Limited financial resources and technical expertise further constrain MSMEs from implementing advanced security systems (Permana, Santoso, et al., 2024a).

Recent studies indicate that machine learning-based approaches are more effective than traditional rule-based systems in detecting fraudulent transaction patterns within imbalanced datasets. In particular, XGBoost demonstrates strong classification performance and computational efficiency for real-time fraud detection. The integration of SMOTE also improves minority class detection by balancing fraudulent and legitimate transaction data, making it suitable for MSMEs online payment environments.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

## Fraud Detection in Digital Payment Systems

Fraud detection refers to the process of identifying suspicious or unauthorized financial transactions within digital payment systems. In machine learning, fraud detection is commonly formulated as a supervised binary classification problem that distinguishes legitimate and fraudulent transactions based on transaction features (Awoyemi et al., 2017). Early fraud detection systems relied on rule-based approaches; however, these methods are less effective in adapting to evolving fraud patterns (Carcillo, 2018).

Traditional machine learning models such as Logistic Regression, Decision Trees, and Random Forest have been widely used for fraud detection. Logistic Regression is relatively interpretable but struggles with non-linear fraud patterns, while Decision Trees are prone to overfitting. Random Forest improves classification performance through ensemble learning but may require higher computational resources for large-scale transaction data (Bhatt et al., 2021a; Breiman, 2001a).

Moreover, fraud detection datasets are generally highly imbalanced, where fraudulent transactions represent only a small proportion of total observations. Under these conditions, traditional models often become biased toward the majority class, reducing fraud detection performance. Therefore, more adaptive ensemble-based approaches are needed to effectively handle complex fraud patterns and imbalanced datasets.

## XGBoost in Fraud Detection

XGBoost, introduced by Chen and Guestrin (2016), is an efficient gradient boosting algorithm that achieves strong performance in supervised learning tasks (Chen & Guestrin, 2016a). The algorithm sequentially builds decision trees to correct prediction errors from previous models while applying regularization techniques to reduce overfitting.

Compared to traditional machine learning models, XGBoost offers advantages such as efficient handling of missing values, parallel processing, and strong performance in non-linear classification problems. These characteristics make it suitable for fraud detection tasks involving complex and imbalanced transaction data.

Previous studies have demonstrated the effectiveness of XGBoost for fraud detection. Yunanto and Budiyanto (2024) reported that XGBoost achieved higher classification performance than Random Forest and Logistic Regression (Yunanto & Budiyanto, 2024). In addition, Rahmadani et al. (2025) showed that GridSearchCV optimization significantly improved XGBoost performance, particularly in terms of AUC-ROC (Rahmadani et al., 2025). These findings support the use of XGBoost as an effective model for detecting fraudulent transactions in digital payment systems.

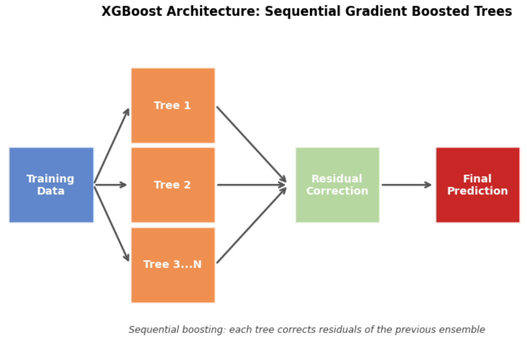



**FIGURE 1. XGBoost Architecture: Sequential Gradient Boosted Trees**

## Handling Class Imbalance using SMOTE

Class imbalance is a major challenge in fraud detection because fraudulent transactions represent only a small proportion of the dataset (Bhatt et al., 2021a). Under these conditions, models often become biased toward the majority class, resulting in poor detection of fraudulent transactions despite high overall accuracy.

To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE), proposed by Chawla et al. (2002), generates synthetic minority-class samples to improve class distribution (Chawla et al., 2002). Compared to simple oversampling, SMOTE reduces overfitting while improving the model's ability to learn fraud patterns.

Previous studies have shown that integrating SMOTE with ensemble learning methods, particularly XGBoost, significantly improves fraud detection performance, especially in terms of recall and F1-score. Therefore, this study applies SMOTE and XGBoost to improve fraud detection performance in MSME digital payment systems.

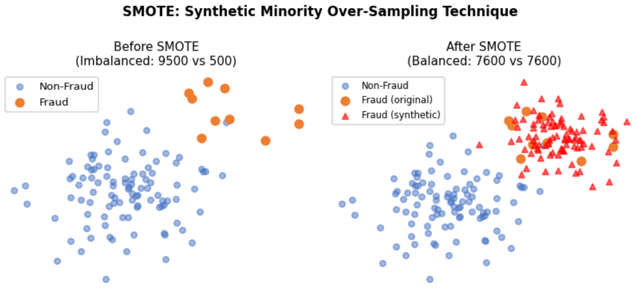



**FIGURE 2. SMOTE: Balancing Class Distribution (Before vs. After)**

The review of prior studies suggests that online payment fraud in MSMEs can be understood from three interconnected perspectivesof. First, information asymmetry theory explains that fraud often occurs because attackers possess transactional advantages over legitimate users, allowing them to exploit weaknesses in digital payment systems (Gupta & Arora, 2023). As fraud patterns continue to evolve, traditional rule-based approaches are often unable to provide adaptive and reliable detection.

Second, the technology adoption and organizational resilience literature suggests that MSMEs in developing economies frequently adopt digital financial technologies faster than they develop adequate cybersecurity capabilities ((Ibrahim & Alfauzan, 2025; Permana, Santoso, et al., 2024). Prior fraud detection studies mainly emphasize predictive accuracy while giving limited attention to the operational constraints of MSMEs, including limited IT infrastructure and financial resources. This creates a need for fraud detection frameworks that are not only accurate but also computationally efficient and practically deployable.

Third, ensemble learning theory, as introduced by Breiman (2001) and later extended through XGBoost by Chen & Guestrin (2016), provides a robust methodological foundation for handling highly imbalanced fraud datasets through the aggregation of weak learners.

Synthesizing these perspectives, this study positions XGBoost-based fraud detection as a socio-technical mechanism that reduces informational asymmetry, strengthens MSME organizational resilience, and supports secure digital financial adoption. This integrated framework provides the theoretical justification for selecting XGBoost, applying SMOTE to address class imbalance, and emphasizing real-time inference capability for MSME payment systems. Based on this theoretical synthesis, the conceptual framework of this study is presented in Figure 3

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

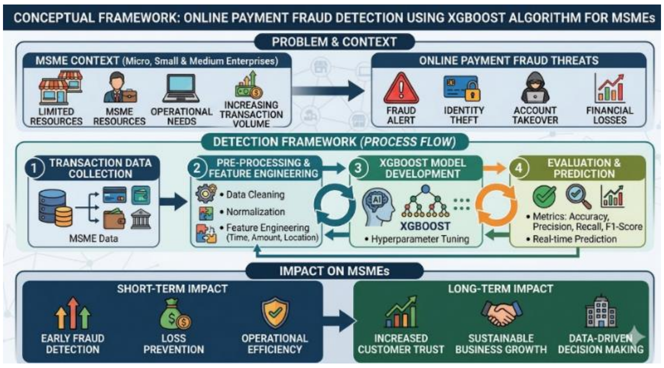



**FIGURE 3. Conceptual Framework of XGBoost-Based Online Payment Fraud Detection for MSMEs**

As illustrated in Figure 3, the conceptual framework consists of three main components: problem context, fraud detection process, and projected outcomes for MSMEs. The framework begins with increasing digital payment activities that expose MSMEs to fraud risks such as account takeover and financial loss. The detection process includes transaction data collection, preprocessing and feature engineering, XGBoost model development, and performance evaluation using metrics such as accuracy, precision, recall, and F1-score. The framework also highlights the expected impact of the proposed system, including improved fraud detection capability, reduced financial losses, and enhanced operational efficiency for MSMEs.

# METHODS

## Research Design

This study employs a quantitative experimental research design, wherein a machine learning-based fraud detection model is developed, optimized, and evaluated under controlled experimental conditions. The experimental workflow follows a standard ML pipeline: data acquisition, preprocessing, feature engineering, model training and tuning, and performance evaluation (Witten et al., 2017). Comparative analysis against baseline models (Logistic Regression, Random Forest) is conducted to contextualize the proposed model's performance.

## Dataset

The dataset used in this study is the fraud_payment_UMKM dataset sourced from Kaggle, consisting of 10,000 transaction records. The dataset was selected because it simulates Indonesian MSME digital payment environments through features such as transaction amounts in Indonesian Rupiah, geographic transaction information, and behavioral indicators including failed_login_attempts and transaction_frequency, which are relevant to fraud patterns reported in Indonesian fintech contexts (Ibrahim & Alfauzan, 2025). In addition, the dataset exhibits a realistic 95:5 imbalance ratio between non-fraud and fraud transactions, providing an appropriate setting for evaluating fraud detection performance under minority-class conditions (Bhatt et al., 2021). Although the dataset is synthetic and may not fully represent real-world MSME payment ecosystems, its MSME-oriented feature design and realistic imbalance characteristics provide a suitable experimental foundation for evaluating the proposed XGBoost-based fraud detection framework.

## Data Preprocessing and Feature Engineering

Preprocessing operations were applied sequentially. Categorical features were encoded using LabelEncoder, while identifier-related attributes were removed to prevent data leakage and overfitting. Feature engineering was conducted to generate additional temporal and transactional attributes relevant to

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

fraud detection. RobustScaler was applied to reduce the influence of outliers, and SMOTE was used on the training data to balance the minority and majority classes.

## Model Development and Hyperparameter Tuning

Three classification models were evaluated: Logistic Regression, Random Forest, and XGBoost. The dataset was split into 80% training and 20% testing sets using stratified sampling. Hyperparameter optimization for XGBoost was performed using GridSearchCV with 5-fold cross-validation to obtain the optimal model configuration.

## Evaluation Metrics

Model performance was assessed using: Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Additionally, average inference time per transaction was measured across the test set to assess real-time deployment feasibility. Confusion matrix analysis was also conducted to identify False Negative (undetected fraud) and False Positive (false alarm) rates.

# RESULTS AND DISCUSSION

## Exploratory Data Analysis

The exploratory data analysis (EDA) identified several important characteristics of the dataset. The fraud distribution indicates a severe class imbalance, where fraudulent transactions account for only 5% of total observations. Transaction amounts exhibit a right-skewed distribution with the presence of outliers, supporting the application of RobustScaler during preprocessing. Fraud cases were observed across multiple payment types, indicating that fraudulent activity is not limited to a specific payment channel. In addition, fraudulent transactions were more frequent during late-night and early-morning hours, suggesting that transaction time may serve as an important fraud indicator.

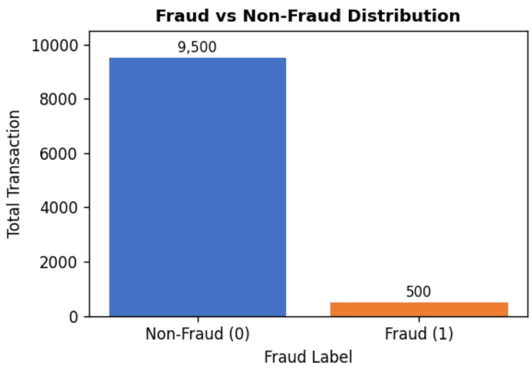



**FIGURE 4. Fraud vs Non-Fraud Distribution**

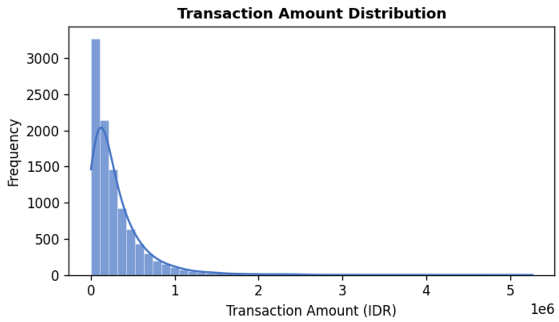



**FIGURE 5. Transaction Amount Distribution (Right-Skewed)**

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

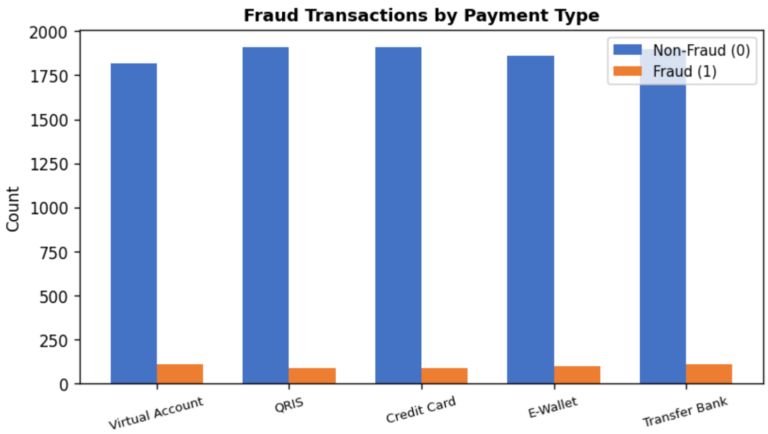



**FIGURE 6. Fraud Transactions by Payment Type**

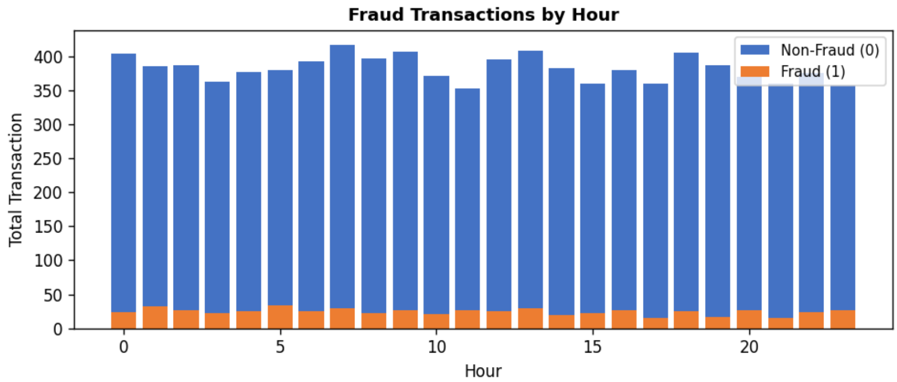



**FIGURE 7. Fraud Transactions by Transaction Hour**

## Feature Correlation Analysis

The correlation analysis indicates that features such as distance_km, amount_idr, failed_login_attempts, and transaction_frequency show positive relationships with fraud activity. This suggests that transactions involving unusual distances, larger transaction amounts, high transaction frequency, and repeated failed login attempts tend to be associated with fraudulent behavior. In contrast, account_age_days demonstrates a negative relationship with fraud activity, indicating that newer accounts may be more vulnerable to fraud. In addition, transaction_hour also shows a relationship with fraud activity, suggesting that fraudulent transactions are more likely to occur during specific transaction periods.

## Data Balancing with SMOTE

SMOTE was applied to balance the minority and majority classes in the training dataset, resulting in 7,600 samples for both fraud and non-fraud classes. This balancing process enabled the model to better learn fraud patterns while reducing bias toward the majority class.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

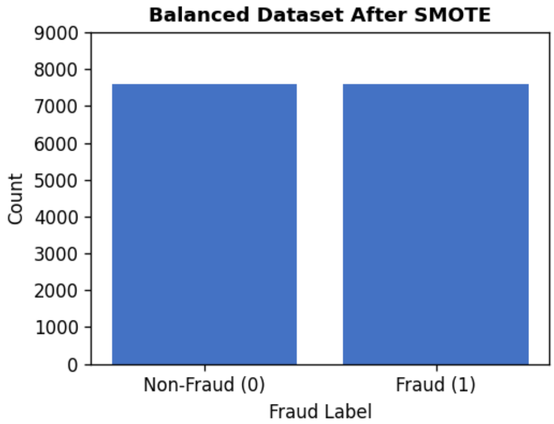



**FIGURE 8. Balanced Dataset After SMOTE (7,600 vs 7,600 samples)**

## Model Performance Comparison

Table 1 presents the comparative performance of the three evaluated models on the test set (2,000 transactions: 1,900 non-fraud and 100 fraud). Among the evaluated models, XGBoost achieved the best overall performance in terms of accuracy, precision, F1-score, and AUC-ROC.

**TABLE 1. Performance Comparison of Fraud Detection Models**

| Model | Accuracy | Precision | Recall | F1-Score | AUC-ROC |
|---|---|---|---|---|---|
| Logistic Regression | 90.50% | 32.95% | **87.00%** | 47.80% | N/A |
| Random Forest | 97.30% | 77.38% | 65.00% | 70.65% | 97.65% |
| XGBoost (Proposed) | **97.35%** | **79.01%** | 64.00% | **70.72%** | **98.03%** |

Note: Bold values indicate the best performance for each metric.

Logistic Regression achieved the highest recall (87.00%) but exhibited very low precision (32.95%), indicating a high number of false positives. Random Forest and XGBoost produced comparable F1-scores; however, XGBoost achieved higher precision (79.01%) and AUC-ROC (98.03%), demonstrating superior discriminative capability and more balanced fraud detection performance. Consequently, XGBoost was selected as the final model for real-time fraud detection implementation.

Beyond technical performance, these findings carry important managerial and operational implications for MSMEs. The low precision of Logistic Regression would result in many legitimate transactions being incorrectly flagged as fraudulent, potentially disrupting customer experience and increasing operational burden for MSMEs with limited human and financial resources. In contrast, XGBoost substantially reduces false alarms, enabling MSME operators to respond to fraud alerts more efficiently and with greater confidence. The model's high AUC-ROC further indicates reliable fraud discrimination across varying transaction conditions, supporting more effective real-time risk mitigation in digital payment environments. These findings suggest that XGBoost offers not only strong predictive capability but also practical operational value for MSMEs seeking lightweight and scalable fraud protection solutions.

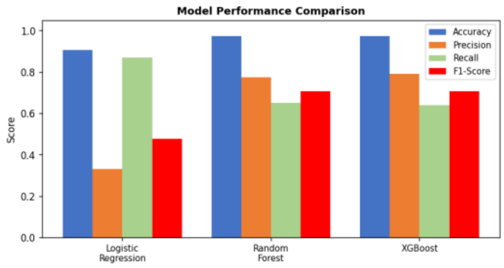
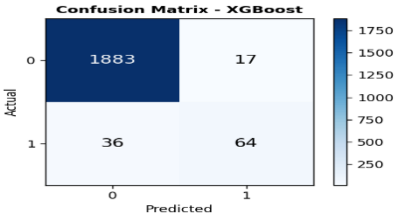



**FIGURE 9. Model Performance Comparison (Accuracy, Precision, Recall, F1-Score)**

## XGBoost Evaluation: Confusion Matrix and ROC Curve

The confusion matrix results for the XGBoost model show 1,883 true negatives (TN), 17 false positives (FP), 36 false negatives (FN), and 64 true positives (TP). Although 36 fraudulent transactions were not detected, the model maintained high precision by producing only 17 false alarms. Furthermore, the ROC curve demonstrates excellent discriminative performance with an AUC value of 0.980, indicating that the model performs effectively across different classification thresholds.



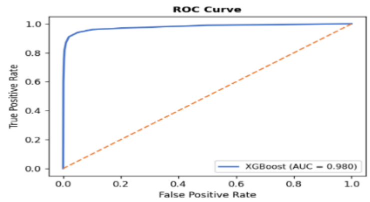




**FIGURE 10. Confusion Matrix – XGBoost** **FIGURE 11. ROC Curve - XGBoost (AUC   0.980)**

## Feature Importance Analysis

Feature importance scores from the optimized XGBoost model indicate that account_age_days and failed_login_attempts are among the most influential predictors in fraud detection. This suggests that newer accounts and repeated failed login attempts are strongly associated with fraudulent activity. In addition, transaction_frequency, device_type, and distance_km also contribute significantly to the model, highlighting the importance of behavioral and transactional patterns in identifying fraudulent transactions.

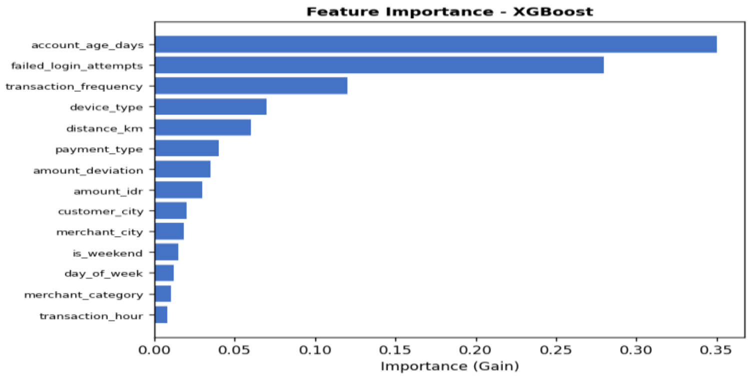



**FIGURE 11. Feature Importance - XGBoost (Top 14 Features by Gain)**

## Real-Time Inference Performance

The XGBoost model achieved an average inference time of approximately $5.38 \times 10^{-5}$ seconds (0.054 ms) per transaction on the test set. This result demonstrates that the proposed model is suitable for real-time fraud detection deployment in MSME online payment system.

# CONCLUSIONS

This study proposed an XGBoost-based fraud detection framework for MSME online payment systems by integrating SMOTE for class imbalance handling, feature engineering, and hyperparameter optimization using GridSearchCV. Experimental results demonstrated that the proposed model achieved 97.35% accuracy, 79.01% precision, 70.72% F1-score, and 98.03% AUC-ROC, outperforming Logistic Regression and Random Forest in terms of precision and discriminative capability. In addition, the model achieved an average inference time of 0.054 ms per transaction, indicating its suitability for real-time fraud detection deployment in MSME payment environments.

The findings further revealed that account_age_days and failed_login_attempts were among the most influential fraud indicators, while the integration of SMOTE and RobustScaler effectively improved minority-class detection and reduced the impact of outliers. These results demonstrate that XGBoost provides an effective and computationally efficient solution for fraud detection in digital payment systems, particularly for resource-constrained MSMEs.

From a theoretical perspective, this study contributes to the fraud detection literature in three ways. First, it extends the application of ensemble-based fraud detection to the relatively underexplored context of Indonesian MSME digital payments. Second, it demonstrates that SMOTE combined with XGBoost regularization can effectively mitigate majority-class prediction bias in highly imbalanced datasets. Third, the identification of behavioral indicators such as account_age_days and failed_login_attempts supports behavioral fraud signal theory, suggesting that account-level behavioral patterns provide strong predictive value in MSME fraud detection contexts.

Beyond the technical domain, this study also highlights the broader economic importance of secure digital financial adoption among MSMEs in Indonesia. By providing a lightweight and high-precision fraud detection framework with near-real-time inference capability, the proposed approach offers a practical pathway for MSMEs to strengthen payment security without requiring large-scale IT infrastructure. Improved fraud protection may enhance consumer trust, support digital payment adoption, and contribute to the broader digital economic transformation agenda in Indonesia.

Despite these contributions, this study is limited by the use of a synthetic Kaggle dataset, which may not fully represent real-world MSME transaction environments. Future research is therefore encouraged to validate the proposed framework using real transactional data and explore advanced approaches such as deep learning, federated learning, and streaming-based real-time fraud detection architectures.

# Acknowledgments

The authors would like to express their sincere gratitude to the supervising lecturer for the valuable feedback and guidance provided during the preparation of this manuscript. In addition, AI-based language assistance tools were utilized to support grammar refinement and academic writing improvement during manuscript preparation.

# References

Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques: A comparative analysis. *Proc. ICCNI*, 1–9.

Bhatt, S., Santara, F., & Ghosh, D. (2021a). SMOTE and borderline-SMOTE for fraud detection under class imbalance. *Expert Systems with Applications*, *180*, 115045.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Carcillo, R. (2018). Scarff: A scalable framework for streaming credit card fraud detection with spark. *Information Fusion*, *41*, 182–194.

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Proceedings of the Perbanas International Seminar on Economics, Business, Management, Accounting and IT
[June 2026] [ISSN 3047-0951]

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD*, 785–794.

Gupta, A., & Arora, R. (2023). Architecture and security challenges in digital payment ecosystems. *International Journal of Information Security*, *22*(4), 901–922.

Ibrahim, M., & Alfauzan, R. (2025). Digital payment growth and financial technology adoption in Indonesian MSMEs: Trends and implications. *Journal of Financial Technology and Innovation*, *6*(2), 45–61.

Permana, A., Kusuma, D., & Rahayu, E. (2024). Post-pandemic acceleration of fintech adoption in Indonesian MSMEs. *Bulletin of Monetary Economics and Banking*, *27*(3), 210–228.

Permana, A., Santoso, B., & Wijaya, C. (2024a). Financial technology adoption barriers in micro and small enterprises. *Asian Journal of Business and Management*, *12*(1), 78–93.

Rahmadani, D., Hakim, F., & Santosa, G. (2025). Optimization of XGBoost hyperparameters using GridSearchCV for credit card fraud detection. *Indonesian Journal of Computing and Cybernetics Systems*, *19*(2), 115–128.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques* (4th ed.). Elsevier.

Yunanto, A., & Budiyanto, B. (2024). Fraud detection in financial services using XGBoost and SMOTE: An empirical evaluation. *Journal of Intelligent Systems and Applications*, *9*(1), 33–47.