 ISSN 2809-8587

Comparative Analysis of SMOTE-Based Random Forest and XGBoost Algorithms for Handling Imbalanced Datasets in Credit Card Fraud Detection

Muhamad Lutfi Azizan[1], Yasin Kamil[2], Septiano Alvian Ismau[3], Dede Sandi[4], Ihsan Maulana[5], Ahmad Nursodiq[6]

Informatics Engineering Study Program, Pamulang University, Kota Tangerang Selatan, Indonesia
Email: mlutfiazizan@gmail.com[1] , yasin111kamil@gmail.com[2] , Septianoalvian76@gmail.com[3] , dedesandi2805@gmail.com[4] , ihsanmaulana.app@gmail.com[5] , dosen02526@unpam.ac.id[6]

The rapid growth of digital payment systems has increased the complexity and risk of credit card fraud, particularly due to the highly imbalanced nature of transaction data. This study aims to compare the performance of Random Forest and XGBoost algorithms combined with the Synthetic Minority Over sampling Technique in detecting fraudulent credit card transactions. The proposed approach focuses on improving classification effectiveness by addressing class imbalance and reducing bias toward legitimate transactions. Data preprocessing includes normalization, stratified data splitting, and the application of over sampling techniques on the training dataset. Model performance is evaluated using precision, recall, F score, and the area under the receiver operating characteristic curve, which are more appropriate for imbalanced classification problems. The findings indicate that Random Forest demonstrates more stable and balanced performance, particularly in minimizing false fraud alerts while maintaining adequate fraud detection capability. These results suggest that Random Forest with over sampling provides a practical and reliable solution for real world credit card fraud detection systems.

Keywords: Credit Card Fraud Detection, Imbalanced Dataset, Random Forest, Xgboost, Smote, Machine Learning

This is an open access article under theCC BY-NClicense

Corresponding Author:
Muhamad Lutfi Azizan
Informatics Engineering Study Program, Pamulang University
Jl. Raya Puspitek, Buaran, Kec. Pamulang, Kota Tangerang Selatan, Banten 15310
mlutfiazizan@gmail.com

1. Introduction

The rapid advancement in the field of e-commerce and digital communication systems has significantly increased the use of credit cards for various types of financial transactions [1]. However, this growth has also been accompanied by an increase in financial criminal activities that have adverse impacts on the global economy. Industry reports indicate that global losses due to credit card fraud have reached a staggering figure of $33.45 billion in 2022 and are projected to continue rising, surpassing $43 billion by 2028 [2]. These fraudulent activities not only cause substantial financial losses for banking institutions but also undermine customer trust and the operational stability of financial institutions [3].

The main technical challenge in developing an effective detection system lies in the highly imbalanced nature of transaction data. In real world conditions, fraudulent transactions often represent only a very small percentage, for example around 0.172% of the total transactions [4]. This imbalance poses a serious obstacle for conventional machine learning algorithms, as models tend to be biased toward the majority class, ultimately resulting in a high rate of false negatives or failures to detect actual fraudulent transactions [5].

To address this issue, various data preprocessing techniques have been developed, one of which is the oversampling technique known as the Synthetic Minority Over sampling Technique (SMOTE). The SMOTE

method works by generating synthetic samples for the minority class in order to balance the data distribution, allowing the model to learn fraud patterns more effectively [6]. In addition to data balancing techniques, the use of ensemble learning based algorithms such as Random Forest and XGBoost has also become a popular solution due to their ability to handle high dimensional data and provide more stable classification performance compared to single models [4].

This study aims to improve fraud detection performance by applying a combination of the SMOTE oversampling technique and ensemble learning algorithms to a highly imbalanced credit card transaction dataset. The main focus of this research is to evaluate the extent to which the use of SMOTE can balance data distribution and enhance model sensitivity toward the minority class (fraud) without compromising classification accuracy for normal transactions. Through a methodology that includes preprocessing stages, normalization using StandardScaler, and a comparative analysis between the Random Forest and XGBoost algorithms, this study is expected to identify the most optimal model based on precision, recall, F1-score, and ROC-AUC metrics in order to effectively mitigate the risk of financial losses in the banking sector.

2. Literature Review and Problem Statement

Literature Review

The rapid expansion of digital financial services, including online payments, e-wallets, and decentralized finance platforms, has significantly increased the volume and complexity of financial transactions [7]. Alongside these developments, financial fraud, particularly credit card fraud and blockchain-based scams, has emerged as a critical challenge for financial institutions and regulatory bodies. Fraudulent transactions typically constitute only a very small proportion of total transactions, creating highly imbalanced datasets that complicate detection and increase the risk of financial losses [8].

Traditional rule-based fraud detection systems have proven inadequate in modern financial environments due to their rigidity, inability to adapt to evolving fraud patterns, and high false positive rates [9]. As a result, machine learning-based approaches have become the dominant paradigm in fraud detection research. Supervised learning models such as logistic regression, decision trees, and support vector machines have been widely applied; however, their performance often degrades when faced with extreme class imbalance [10].

Recent studies emphasize the effectiveness of ensemble learning techniques, particularly Random Forest and gradient boosting-based models, in capturing complex, non-linear relationships within transaction data [11]. Random Forest has been shown to be robust against noise and overfitting while providing stable performance across different datasets. XGBoost, on the other hand, has gained popularity due to its high predictive power, scalability, and optimization efficiency, making it suitable for large-scale and real-time fraud detection systems [12].

Despite their strengths, ensemble models remain vulnerable to class imbalance, where the minority (fraud) class is often misclassified [13]. To address this issue, data-level resampling techniques have been widely adopted. One of the most prominent methods is the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class to balance class distributions. Empirical evidence suggests that SMOTE can significantly improve recall and F1-score for fraud detection tasks, particularly when combined with ensemble classifiers.

However, prior studies also report mixed findings regarding the comparative effectiveness of different ensemble models when combined with SMOTE. Some research indicates that gradient boosting models achieve higher recall but at the cost of substantially lower precision, resulting in increased false positive rates that may be impractical for real-world financial systems. Conversely, Random Forest-based

approaches often demonstrate a more balanced trade-off between precision and recall, which is critical in operational fraud detection environments where excessive false alarms can undermine system credibility and increase operational costs [14].

Moreover, much of the existing literature focuses primarily on accuracy or ROC-AUC as the main evaluation metrics, despite growing consensus that such metrics may be misleading for imbalanced classification problems. Recent studies increasingly advocate the use of precision, recall, and F1-score as more appropriate performance indicators for fraud detection systems [15]. Nevertheless, comparative empirical evidence that systematically evaluates these metrics across ensemble models under identical preprocessing conditions remains limited.

In addition, most prior research emphasizes algorithmic performance without sufficiently addressing practical deployment considerations, such as the trade-off between fraud detection sensitivity and the operational burden caused by false positives. This gap highlights the need for comparative studies that not only evaluate predictive performance but also assess model suitability for real-world financial applications.

Problem Statement

Financial fraud detection presents a persistent and complex challenge due to the extreme imbalance between legitimate and fraudulent transactions, dynamic fraud patterns, and the high cost associated with misclassification [12]. Although ensemble learning models such as Random Forest and XGBoost have demonstrated strong predictive capabilities, their effectiveness in imbalanced fraud detection scenarios remains inconsistent, particularly when evaluated using operationally relevant performance metrics.

Existing studies show that oversampling techniques such as SMOTE can improve minority class detection; however, there is no clear consensus on which ensemble model achieves the most optimal balance between precision and recall when combined with SMOTE. Models with high recall may detect more fraudulent transactions but often generate excessive false positives, while models with higher precision may fail to identify a significant portion of actual fraud cases. This trade-off poses a critical problem for financial institutions that require both accuracy and operational efficiency.

Furthermore, prior research frequently relies on limited evaluation metrics or focuses on single-model performance, providing insufficient comparative insight into how different ensemble algorithms behave under identical data balancing conditions [16]. As a result, practitioners lack clear empirical guidance in selecting fraud detection models that are not only statistically effective but also practically viable.

Based on these gaps, the central problem addressed in this study is the lack of comprehensive comparative analysis between Random Forest and XGBoost models integrated with SMOTE in handling imbalanced credit card fraud datasets, particularly when assessed using precision, recall, F1-score, and ROC-AUC [17]. Addressing this problem is essential to identify a fraud detection approach that achieves an optimal trade-off between detection accuracy and false positive control, thereby supporting more reliable and efficient fraud prevention systems in modern financial environments.

3. Method

This study adopts a quantitative and experimental research design to evaluate the effectiveness of machine learning models in detecting credit card fraud under highly imbalanced data conditions. The methodological framework is structured to ensure that model comparisons are fair, reproducible, and aligned with real world fraud detection scenarios. A comparative approach is employed to examine the performance differences between ensemble based algorithms when combined with an oversampling strategy, with particular attention given to the minority class representing fraudulent transactions.

The dataset used in this research consists of credit card transaction records that reflect real transaction behavior and inherent class imbalance. Prior to model development, the data undergo an initial inspection process to identify potential issues such as missing values, duplicated records, or inconsistencies. This step is essential to ensure data integrity and to prevent bias or noise from influencing the learning process. Only clean and validated data are retained for subsequent analysis.

Data preprocessing plays a critical role in this study. All numerical features are normalized using the StandardScaler method to ensure that variables are placed on a comparable scale, thereby improving model convergence and stability. The dataset is then divided into training and testing subsets using stratified sampling, which preserves the original class distribution in both subsets. This approach is particularly important in fraud detection tasks, where minority class representation must be maintained during evaluation.

To address the severe class imbalance problem, the Synthetic Minority Over sampling Technique is applied exclusively to the training data. This technique generates synthetic samples for the minority class by interpolating between existing observations, allowing the models to better learn fraud related patterns without altering the original distribution of the testing data. Applying oversampling only to the training set is intended to prevent data leakage and to ensure that model performance reflects genuine generalization capability.

Two ensemble learning algorithms, Random Forest and XGBoost, are implemented as the primary classification models. Random Forest is selected due to its robustness against overfitting and its ability to capture complex, non linear relationships through multiple decision trees. XGBoost is chosen for its gradient boosting framework, which emphasizes learning from previous errors and is known for strong predictive performance in structured data problems. Both models are trained under comparable conditions to enable an objective performance comparison.

Model evaluation is conducted using metrics that are appropriate for imbalanced classification tasks. Instead of relying solely on accuracy, this study emphasizes precision, recall, F score, and the area under the receiver operating characteristic curve. These metrics provide a more comprehensive understanding of each model's ability to correctly identify fraudulent transactions while controlling the rate of false alarms, which is a critical consideration in financial applications.

4. Results And Discussion

Data Distribution and Characteristics

The Credit Card Fraud Detection dataset exhibits a highly imbalanced data distribution. Out of a total of 284,807 transactions, 284,315 transactions, representing 99.83 percent, are classified as non fraud transactions, while only 492 transactions, or 0.17 percent, fall into the fraud category. This condition clearly indicates a significant imbalanced dataset problem, which necessitates the application of data balancing techniques prior to the modeling process.

After the dataset was divided into training and testing sets, the class distribution in the training data consisted of 227,451 non fraud transactions and 394 fraud transactions. Subsequently, the application of the Synthetic Minority Over sampling Technique successfully balanced the class distribution in the training data, resulting in an equal number of observations for each class, with 227,451 instances per class. Through this process, the model is able to learn fraud transaction patterns more effectively without being biased toward the majority class.

 ISSN 2809-8587

Classification Results Using the Random Forest Algorithm

The fraud detection model based on the Random Forest algorithm was evaluated using the testing dataset after the application of the SMOTE oversampling technique on the training data. Model performance was assessed using precision, recall, F score, and the area under the receiver operating characteristic curve to ensure reliable evaluation under imbalanced data conditions.

The evaluation results indicate that the Random Forest algorithm performs exceptionally well in classifying non fraud transactions, as reflected by precision, recall, and F score values of one. For the fraud class, the model achieves a precision value of 0.87, a recall value of 0.83, and an F score of 0.85.

These results demonstrate that the majority of fraudulent transactions are successfully detected with a relatively low misclassification rate. Furthermore, the ROC AUC value of 0.9782 indicates that the model has a very high discriminative ability in distinguishing between fraud and non fraud transactions. Therefore, the Random Forest model can be considered reliable and suitable for implementation in credit card transaction fraud detection systems.

Table 1. Classification Evaluation Results of the Random Forest Algorithm

Class	Precision	Recall	F Score	Support
Non Fraud (0)	1	1	1	56,864
Fraud (1)	0.87	0.83	0.85	98
ROC AUC				0.9782

Table 1 presents the classification performance of the Random Forest algorithm in detecting credit card fraud after the application of the SMOTE oversampling technique. The evaluation results summarized in this table provide a quantitative basis for the visual analysis of model performance, which is further illustrated in the following figure.

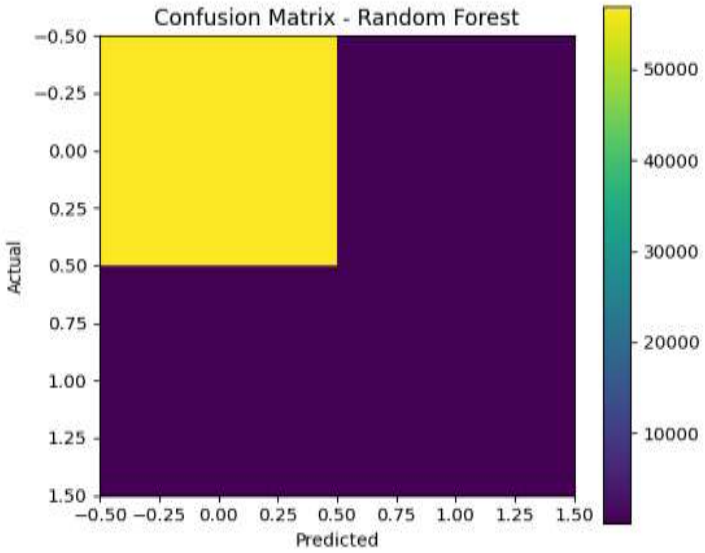

Figure 1. Confusion Matrix of Random Forest Classification Results

Figure 1 illustrates the confusion matrix obtained from the classification results using the Random Forest algorithm. The figure shows that the majority of fraudulent transactions are successfully detected, with a relatively small number of false negatives and false positives. This indicates that the Random Forest model achieves a good balance between sensitivity and precision in detecting fraudulent transactions.

 ISSN 2809-8587

Classification Results Using the XGBoost Algorithm

Subsequent testing was conducted using the XGBoost algorithm on the same testing dataset in order to compare the performance of ensemble models. The evaluation was carried out using the same metrics, namely precision, recall, F score, and the area under the receiver operating characteristic curve.

The testing results indicate that the XGBoost algorithm performs very well in identifying non fraud transactions, as reflected by precision and recall values of one. For the fraud class, XGBoost achieves a high recall value of 0.86, indicating that most fraudulent transactions are successfully detected. However, the precision value is relatively low at 0.25, which suggests an increased number of non fraud transactions being incorrectly classified as fraud.

The ROC AUC value of 0.9806 demonstrates that, overall, XGBoost has an excellent class separation capability. Nevertheless, the results also highlight a trade off between the level of fraud detection and the misclassification of legitimate transactions.

Table 2. Classification Evaluation Results of the XGBoost Algorithm

Class	Precision	Recall	F Score	Support
Non Fraud (0)	1	1	1	56,864
Fraud (1)	0.25	0.86	0.38	98
ROC AUC				0.9806

Table 2 summarizes the classification performance of the XGBoost algorithm on the credit card fraud detection task after the application of the SMOTE oversampling technique. These results serve as a quantitative reference for interpreting the classification behavior of the model, which is further visualized in the following figure.

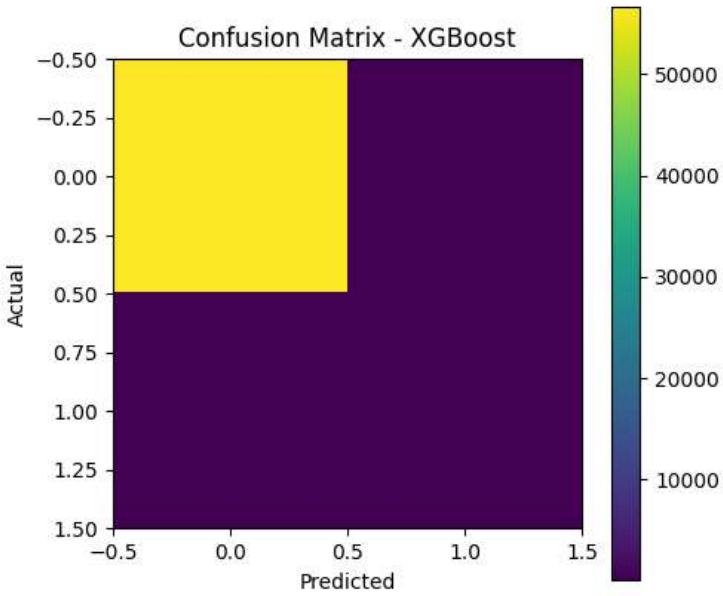

Figure 2. Confusion Matrix of XGBoost Classification Results

Figure 2 presents the confusion matrix obtained from the classification results using the XGBoost algorithm. The model demonstrates a very strong ability to detect fraudulent transactions; however, it also produces a higher false positive rate compared to the Random Forest model. This indicates that XGBoost tends to be more aggressive in fraud detection, suggesting that further adjustments are required to reduce the misclassification of non fraud transactions.

Model Analysis and Comparison

Based on the testing results, both Random Forest and XGBoost algorithms exhibit very strong performance in detecting non fraud transactions, as indicated by near perfect precision and recall values. This finding suggests that both ensemble algorithms are capable of effectively learning normal transaction patterns following the application of the SMOTE oversampling technique.

Nevertheless, a clear performance difference emerges in the detection of fraudulent transactions. The Random Forest algorithm achieves a precision value of 0.87 and a recall value of 0.83 for the fraud class, indicating a well balanced trade off between the ability to detect fraudulent transactions and the minimization of misclassification of legitimate transactions. The F score value of 0.85 further confirms the high stability of the Random Forest model in handling imbalanced data.

In contrast, the XGBoost algorithm demonstrates a high recall value of 0.86 for the fraud class, but this is accompanied by a relatively low precision value of 0.25. This condition indicates that XGBoost tends to be more aggressive in detecting fraudulent transactions, resulting in a larger number of false positive cases. Although the ROC AUC value of XGBoost, at 0.9806, is slightly higher than that of Random Forest, which is 0.9782, this difference does not directly reflect superior classification quality for the fraud class.

From a fraud detection system implementation perspective, misclassification in the form of false positives has significant operational consequences, such as disruption of legitimate customer transactions and an increased burden of manual verification. Therefore, models that provide a better balance between precision and recall are generally preferred.

Based on this analysis, it can be concluded that Random Forest is the most optimal model in this study, as it delivers more balanced, stable, and practical performance for implementation in credit card transaction fraud detection systems. While XGBoost demonstrates strong potential, it requires further adjustments, such as threshold optimization or the application of cost sensitive learning, to reduce the false positive rate.

5. Conclusion

This study demonstrates that addressing class imbalance is a critical factor in improving the effectiveness of credit card fraud detection systems. The application of an oversampling strategy enables machine learning models to better recognize fraudulent transaction patterns that are typically underrepresented in real world datasets. By combining the oversampling technique with ensemble learning approaches, this research provides empirical evidence that model performance can be substantially enhanced without sacrificing generalization capability. The comparative analysis reveals that both ensemble models are capable of accurately identifying legitimate transactions, indicating their strong ability to learn normal behavioral patterns. However, notable differences emerge in their handling of fraudulent transactions. The Random Forest model exhibits a more balanced performance, successfully detecting fraudulent activity while maintaining a relatively low rate of misclassification for legitimate transactions. This balance reflects its robustness and stability when applied to highly imbalanced data environments.

In contrast, the XGBoost model demonstrates higher sensitivity toward fraudulent transactions but tends to produce a greater number of false alerts. Although this aggressive detection behavior may be beneficial in scenarios where minimizing missed fraud is the primary objective, it can lead to operational challenges, such as increased verification efforts and potential disruption to legitimate customer transactions. As a result, this model requires further refinement before practical deployment. Overall, the findings indicate that Random Forest provides a more practical and reliable solution for credit card fraud detection under imbalanced data conditions. The study also highlights the importance of aligning model selection with

operational priorities and suggests that future research should explore threshold optimization and cost sensitive learning to further enhance detection performance.

6. References

[1] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido, "A neural network ensemble with feature engineering for improved credit card fraud detection," *IEEE access*, vol. 10, pp. 16400–16407, 2022.

[2] I. Almubark, "Advanced Credit Card Fraud Detection: An Ensemble Learning Using Random Under Sampling and Two-Stage Thresholding," *IEEE Access*, 2024.

[3] I. E. Eteng, U. L. Chinedu, and A. E. Ibor, "A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets," *J. Niger. Soc. Phys. Sci.*, p. 2066, 2025.

[4] V. Sinap, "Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets," *Turkish J. Eng.*, vol. 8, no. 2, pp. 196–208, 2024.

[5] R. Bounab, K. Zarour, B. Guelib, and N. Khlifa, "Enhancing medicare fraud detection through machine learning: Addressing class imbalance with SMOTE-ENN," *IEEE Access*, vol. 12, pp. 54382–54396, 2024.

[6] B. Ahmed, S. Hussain, D. Shakir, N. ur Rehman, and G. Nadeem, "Identifying Credit Card Fraud with Machine Learning: Evaluation of Algorithms and Oversampling Techniques," *Asian Bull. Big Data Manag.*, vol. 4, no. 3, pp. 33–50, 2024.

[7] N. Yathiraju and B. Dash, "Gamification Of E-Wallets With The Use Of Defi Technology-A Revisit To Digitization In Fintech," *Int. J. Eng. Sci.*, vol. 3, no. 1, pp. 2582–9734, 2023.

[8] E. M. Al-dahasi, R. K. Alsheikh, F. A. Khan, and G. Jeon, "Optimizing fraud detection in financial transactions with machine learning and imbalance mitigation," *Expert Syst.*, vol. 42, no. 2, p. e13682, 2025.

[9] C. D. Ikemefuna, O. Okusi, A. C. Iwuh, and S. Yusuf, "Adaptive fraud detection systems: Using ML to identify and respond to evolving financial threats," *Int. Res. J. Mod. Eng.*, vol. 6, pp. 2077–2092, 2024.

[10] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," *Int. J. Mach. Learn. Cybern.*, vol. 4, no. 1, pp. 13–24, 2013.

[11] T. A. Shaikh, T. Rasool, P. Verma, and W. A. Mir, "A fundamental overview of ensemble deep learning models and applications: systematic literature and state of the art," *Ann. Oper. Res.*, pp. 1–77, 2024.

[12] L. Theodorakopoulos, A. Theodoropoulou, A. Tsimakis, and C. Halkiopoulos, "Big data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost," *Electronics*, vol. 14, no. 9, p. 1754, 2025.

[13] A. Ayodele, "A comparative study of ensemble learning techniques for imbalanced classification problems," *World J. Adv. Res. Rev.*, vol. 19, no. 1, pp. 1633–1643, 2023.

[14] H. Y. J. Lam, "Reducing Fraud with Anomaly Detection Algorithms," 2025.

[15] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, p. 268, 2025.

[16] Z. Wang, Y. Hong, L. Huang, M. Zheng, H. Yuan, and R. Zeng, "A comprehensive review and future research directions of ensemble learning models for predicting building energy consumption," *Energy Build.*, p. 115589, 2025.

[17] M. Imani, A. Beikmohammadi, and H. R. Arabnia, "Comprehensive analysis of random forest and XGBoost performance with SMOTE, ADASYN, and GNUS under varying imbalance levels," *Technologies*, vol. 13, no. 3, p. 88, 2025.