Machine Learning-Based Fraud Detection in Banking Transactions Using Integrated Behavioral and Transactional Feature Engineering with Weak Labeling Approach

Hongbo Wang[1,*]

[1]Institute of Electrical Engineering, Yanshan University, China

ABSTRACT

Fraud detection in banking transactions is a critical challenge due to the imbalanced nature of data and the lack of labeled fraud instances. This study proposes a machine learning approach for detecting fraudulent transactions by integrating behavioral and transactional features, combined with a rule-based weak labeling strategy to generate fraud labels. The dataset consists of 2,512 banking transactions, with 14.29% labeled as fraud. Three models were evaluated, including Logistic Regression, Random Forest, and XGBoost, using stratified cross-validation and multiple evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The results show that ensemble-based models outperform Logistic Regression, with Random Forest achieving the best balance between precision and recall, and XGBoost obtaining perfect recall and the highest ROC-AUC, indicating its strong ability to detect fraudulent transactions. Feature importance analysis reveals that transaction amount and deviation from typical user behavior are key indicators of fraud. Despite these promising results, the study is limited by the use of rule-based labeling and a relatively small dataset. Future work should focus on validating the proposed approach using real-world labeled data and improving model robustness for practical deployment.

Keywords Fraud Detection, Machine Learning, Behavioral Features, Transactional Data, Weak Labeling

INTRODUCTION

Fraud detection in banking transactions has become an increasingly critical issue with the rapid growth of digital financial services. The widespread use of online banking, mobile payments, and electronic transactions has created new opportunities for fraudulent activities, leading to significant financial losses for both institutions and customers [1]. Detecting fraudulent transactions is inherently challenging due to the dynamic nature of fraud patterns, the high volume of transaction data, and the typically imbalanced distribution between fraudulent and legitimate transactions [2]. As a result, traditional rule-based systems are often insufficient to handle complex and evolving fraud behaviors.

Recent advances in machine learning have enabled the development of more adaptive and data-driven approaches for fraud detection. Various algorithms, including Logistic Regression, Random Forest, and gradient boosting methods such as XGBoost, have been widely applied to identify fraudulent patterns in transaction data [3]. These approaches have demonstrated strong performance, particularly when combined with

Submitted: 25 December 2025
Accepted: 5 February 2026
Published: 23 May 2026

Corresponding author
Hongbo Wang,
Hongbo_w@ysu.edu.cn

Additional Information and Declarations can be found on page 172

DOI: 10.47738/jdmdc.v3i2.66

How to cite this article: H. Wang, "Machine Learning-Based Fraud Detection in Banking Transactions Using Integrated Behavioral and Transactional Feature Engineering with Weak Labeling Approach," *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 2, pp. 159-174, 2026.

effective feature engineering techniques. In particular, ensemble-based models have shown superior capability in capturing non-linear relationships and complex interactions among features, making them well-suited for fraud detection tasks [4].

Despite these advancements, several limitations remain in existing studies. First, many fraud detection models rely on fully labeled datasets, which are often unavailable in real-world scenarios due to privacy constraints and the difficulty of verifying fraudulent transactions. Second, prior research tends to focus primarily on transactional features, such as transaction amount and frequency, while overlooking behavioral aspects such as user activity patterns, login behavior, and transaction timing. Third, limited attention has been given to evaluating model performance under weak supervision settings, where labels are generated using heuristic or rule-based approaches rather than ground truth annotations [5].

To address these gaps, this study proposes a machine learning framework for fraud detection that integrates both behavioral and transactional features while employing a rule-based weak labeling strategy to construct fraud labels. The proposed approach evaluates the performance of three widely used algorithms, namely Logistic Regression, Random Forest, and XGBoost, using stratified cross-validation and multiple evaluation metrics. By incorporating features that capture deviations from normal user behavior, the study aims to improve the detection of anomalous transactions that may not be identified using transactional data alone.

The main contributions of this study are as follows. First, it introduces a weak labeling approach for fraud detection in the absence of ground truth data, enabling the application of supervised learning techniques in practical scenarios. Second, it demonstrates the effectiveness of combining behavioral and transactional features in improving model performance. Third, it provides a comprehensive comparative analysis of different machine learning models under imbalanced data conditions. The findings of this study are expected to contribute to the development of more robust and practical fraud detection systems in the banking sector.

Literature Review and Related Works

Fraud detection in financial transactions has been widely studied due to its critical role in minimizing financial losses and enhancing system security. Traditional approaches primarily relied on rule-based systems, which use predefined patterns to identify suspicious activities. While these methods are easy to implement, they lack adaptability and struggle to detect new or evolving fraud patterns. As a result, machine learning techniques have been increasingly adopted to address these limitations by learning patterns directly from data [6], [7].

Supervised machine learning models such as Logistic Regression, Decision Trees, and Support Vector Machines have been commonly applied in fraud detection tasks. These models are capable of distinguishing between fraudulent and legitimate transactions based on historical data [8], [9]. However, their performance is often affected by the highly imbalanced nature of fraud datasets, where fraudulent transactions represent only a small fraction of the total data. This imbalance can lead to biased models that favor the majority class, reducing the effectiveness of fraud detection [10], [11].

To address this issue, various techniques have been proposed, including data resampling methods such as Synthetic Minority Over-sampling Technique (SMOTE) and cost-sensitive learning. These approaches aim to improve the representation of the minority class and enhance model performance in detecting fraud [12], [13]. In addition, evaluation metrics beyond accuracy, such as precision, recall, F1-score, and ROC-AUC, have been widely used to provide a more comprehensive assessment of model performance in imbalanced datasets [14].

Ensemble learning methods, including Random Forest and gradient boosting algorithms such as XGBoost, have demonstrated superior performance in fraud detection tasks. These models combine multiple weak learners to improve predictive accuracy and robustness [15], [16]. Their ability to capture complex, non-linear relationships make them particularly effective in identifying fraudulent patterns that may not be detected by simpler models [17]. Recent studies have shown that XGBoost, in particular, achieves high performance due to its regularization mechanisms and efficient handling of structured data [18].

Feature engineering also plays a crucial role in enhancing fraud detection performance. Many studies have focused on transactional features such as transaction amount, frequency, and location [19]. However, recent research highlights the importance of incorporating behavioral features, including user activity patterns, login attempts, and transaction timing, to better capture anomalies and deviations from normal behavior [20]. Combining transactional and behavioral features has been shown to significantly improve the ability of machine learning models to detect fraudulent activities.

Despite these advancements, several challenges remain. One of the main limitations is the lack of publicly available labeled datasets, which restricts the development and evaluation of supervised models. In many real-world scenarios, fraud labels are incomplete or unavailable, leading to the need for alternative approaches such as weak labeling or semi-supervised learning. However, the application of rule-based weak labeling in combination with machine learning models remains relatively underexplored.

Based on these observations, this study aims to address the identified gaps by integrating behavioral and transactional features within a weak labeling framework and evaluating multiple machine learning models for fraud detection. This approach is expected to provide a more practical and adaptable solution for real-world financial systems.

Methodology

Dataset Description and Research Framework

This study utilizes a banking transaction dataset consisting of 2,512 records and 16 attributes, which include transactional, behavioral, and customer-related features. Due to the absence of ground truth fraud labels, a weak labeling approach was applied to generate binary classification targets. The overall research framework adopted in this study is illustrated in figure 1, which outlines the end-to-end process starting from data preprocessing, feature engineering, weak labeling, model training, and performance evaluation.

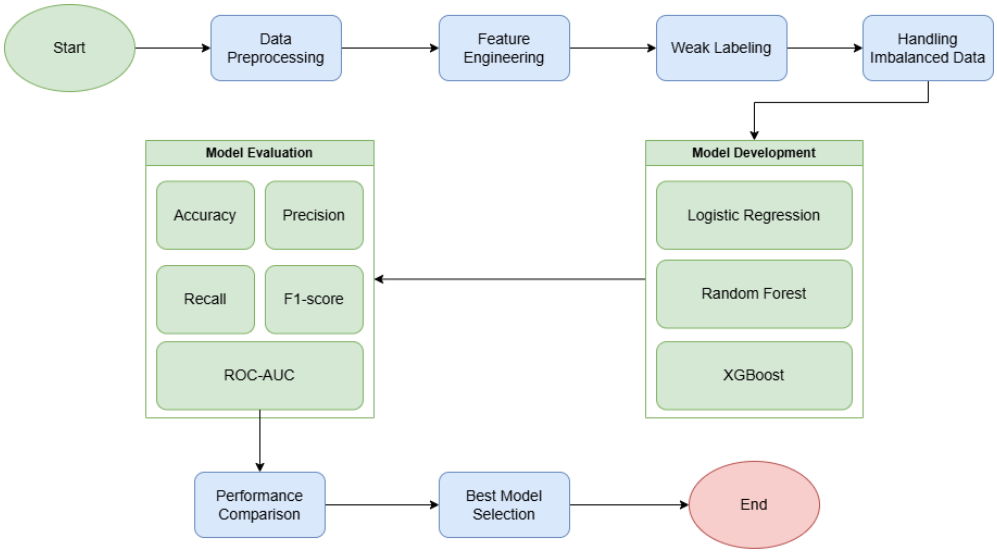


Figure 1 Research Framework

As shown in figure 1, the proposed framework consists of several key stages. The raw transaction data is first preprocessed and transformed into structured features. Subsequently, feature engineering is applied to capture both transactional characteristics and behavioral patterns. A rule-based mechanism is then used to generate fraud labels, followed by model training using multiple machine learning algorithms. Finally, model performance is evaluated using various metrics to determine the most effective approach.

Data Preprocessing

Data preprocessing ensures that the dataset is clean, consistent, and suitable for machine learning algorithms. Let a dataset be defined as:

$$X = \{x_1, x_2, \ldots, x_n\} \tag{1}$$

each instance $x_i$represents a transaction with multiple features.

Missing values in numerical features are handled using median imputation:

$$x_{ij} = \text{median}(X_j) \tag{2}$$

$X_j$represents all values of feature j.

For feature scaling, numerical features are standardized using z-score normalization:

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

$\mu$is the mean and $\sigma$is the standard deviation.

Categorical variables are transformed using one-hot encoding:

$$x_{cat} \rightarrow [0,1,\ldots,0] \tag{4}$$

This transformation allows categorical features to be processed by machine learning models.

Feature Engineering

Feature engineering is performed to enhance the representation of user behavior and transaction patterns. A key feature introduced is the deviation from the average transaction:

$$D_i = \mid T_i - \bar{T}_{acc} \mid \tag{5}$$

$T_i$is the transaction amount

$\bar{T}_{acc}$is the average transaction amount for a given account

This feature captures abnormal behavior, which is critical for fraud detection.

Additionally, transaction frequency per account is defined as:

$$F_{acc} = \sum_{i=1}^{n} I(AccountI\ D_i = acc) \tag{6}$$

$I$is an indicator function.

These engineered features allow the model to capture both transactional magnitude and behavioral deviation.

Weak Labeling Strategy

Due to the absence of labeled data, a rule-based scoring function is used to generate fraud labels. Each transaction is assigned a fraud score:

$$S_i = \sum_{k=1}^{m} w_k \cdot f_k(x_i) \tag{7}$$

$f_k(x_i)$is a binary function representing rule k

$w_k$is the weight (set to 1 in this study)

The fraud label is defined as:

$$y_i = \begin{cases} 1, & \text{if } S_i \geq \theta \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$\theta$is the threshold.

This formulation allows heuristic knowledge to be incorporated into supervised learning.

Handling Imbalanced Data

To address class imbalance, SMOTE is applied to generate synthetic minority samples. Given two minority samples $x_i$and x_{nn}, a new synthetic sample is generated as:

$$x_{new} = x_i + \lambda(x_{nn} - x_i) \tag{9}$$

$\lambda \in [0,1]$is a random number

This approach increases the diversity of minority class samples and improves model learning.

Model Development

Three models are used in this study:

The probability of fraud is modeled as:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}} \tag{10}$$

Random Forest builds multiple decision trees:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b\,(x) \tag{11}$$

$T_b$represents each decision tree.

XGBoost minimizes an objective function:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{12}$$

lis the loss function

Ωis the regularization term

Model Evaluation

Model performance is evaluated using the following metrics:

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Precision:

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

Recall:

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

F1-score:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{16}$$

ROC-AUC measures the area under the ROC curve:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \tag{17}$$

Algorithm 1: Fraud Detection Framework

Input: banking transaction dataset D

Output: predicted fraud label $\hat{y}$and evaluation metrics

Process:

Start

Load dataset $D = \{x_1, x_2, \ldots, x_n\}$

Data Preprocessing

Convert TransactionDate to datetime

Extract transaction hour h_i

If $h_i \geq 22$or $h_i \leq 5$then

$IsNight_i = 1$

Else

$IsNight_i = 0$

End if

Impute missing values
Standardize numerical features
Encode categorical features
Feature Engineering
For each account $a$in D:
    Compute transaction count
    $C_a = \sum I(AccountID_i = a)$
    Compute average transaction
    $\bar{T}_a = \frac{1}{C_a}\sum T_i$
End for
For each transaction x_i:
    Compute deviation
    $D_i = \mid T_i - \bar{T}_{a(i)} \mid$
End for
Weak Labeling
Compute thresholds $q_{0.90}$for key features
For each transaction x_i:
    Initialize score $S_i = 0$
    If $T_i \geq q_{0.90}^{amt}$then $S_i = S_i + 1$
    If $D_i \geq q_{0.90}^{dev}$then $S_i = S_i + 1$
    If $Login_i \geq threshold$then $S_i = S_i + 1$
    If $Duration_i \geq q_{0.90}^{dur}$then $S_i = S_i + 1$
    If Channel = Online then $S_i = S_i + 1$
    If $IsNight_i = 1$then $S_i = S_i + 1$
    If $S_i \geq 2$then
        $y_i = 1$
    Else
        $y_i = 0$
    End if
End for
Train-Test Split
Split dataset into training and testing sets
Handling Imbalance
If minority class sufficient then
    Apply SMOTE
End if
Model Training
Train Logistic Regression
Train Random Forest
Train XGBoost
Evaluation
For each model:
    Predict $\hat{y}$
    Compute Accuracy, Precision, Recall, F1-score, ROC-AUC
End for
Model Selection
Select best model based on F1-score or ROC-AUC
End

Results

Data Distribution

The dataset used in this study consists of 2,512 banking transactions, where 2,153 transactions (85.71%) are labeled as non-fraud and 359 transactions (14.29%) are labeled as fraud. This distribution reflects a clear imbalance between the majority and minority classes, with non-fraud transactions significantly outnumbering fraudulent ones. Such imbalance is commonly observed in real-world financial datasets, where fraudulent activities represent only a small portion of overall transactions.

This class imbalance introduces challenges for machine learning models, as algorithms may become biased toward the majority class and fail to adequately detect fraudulent cases. In this context, relying solely on accuracy can be misleading, since a model could achieve high accuracy by predominantly predicting the majority class. Therefore, additional evaluation metrics such as precision, recall, F1-score, and ROC-AUC are essential to provide a more comprehensive assessment of model performance, particularly in identifying minority class instances.

4.2 Cross-Validation Performance

The performance of the evaluated models using stratified cross-validation is summarized in table 1. Among the three models, XGBoost consistently achieved the highest recall (0.986), F1-score (0.873), and ROC-AUC (0.998), indicating its strong ability to correctly identify fraudulent transactions and effectively distinguish between fraud and non-fraud classes. This high recall suggests that XGBoost is highly sensitive to fraud cases, minimizing the risk of missed detections. In contrast, Random Forest obtained the highest precision (0.965), which indicates that the model produces fewer false positive predictions and is more reliable in classifying legitimate transactions. While Logistic Regression demonstrated stable performance across all metrics, its overall results were lower compared to the ensemble-based models, suggesting limited capability in capturing more complex patterns within the data.

Table 1 Cross-Validation Performance

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.930	0.724	0.829	0.773	0.974
Random Forest	0.956	0.965	0.721	0.824	0.982
XGBoost	0.959	0.783	0.986	0.873	0.998

Test Set Performance

The evaluation results on the test set are presented in table 2. Random Forest achieved the highest F1-score (0.874) and accuracy (0.966), indicating that it provides the most balanced performance in terms of correctly identifying both fraudulent and non-fraudulent transactions. Its

high precision value also suggests that the model produces very few false positives, making it reliable for practical deployment. In comparison, XGBoost achieved perfect recall (1.000), meaning that all fraudulent transactions in the test set were successfully detected without any false negatives. This reflects its strong sensitivity to fraud cases, which is critical in minimizing financial risk. Additionally, XGBoost obtained the highest ROC-AUC (0.999), demonstrating an almost perfect ability to distinguish between classes across different thresholds. However, this performance comes with a lower precision compared to Random Forest, indicating a higher number of false positive predictions.

Table 2 Test Set Performance

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.932	0.738	0.819	0.776	0.973
Random Forest	0.966	0.937	0.819	0.874	0.985
XGBoost	0.954	0.758	1.000	0.862	0.999

Logistic Regression

The classification performance of Logistic Regression is illustrated in figure 2 and figure 3. Based on the confusion matrix, the model correctly classified 410 non-fraud transactions and 59 fraud transactions. However, it also produced 21 false positives and 13 false negatives. The presence of false negatives indicates that some fraudulent transactions were not detected, which can be critical in real-world applications where undetected fraud may lead to financial losses. At the same time, the number of false positives suggests that some legitimate transactions were incorrectly flagged as fraud, which could impact user experience.

The ROC-AUC value of 0.973 indicates that Logistic Regression has strong discriminative ability in separating fraud and non-fraud classes across different thresholds. This suggests that the model is capable of learning general patterns within the data. However, compared to ensemble-based methods such as Random Forest and XGBoost, Logistic Regression demonstrated lower overall effectiveness in terms of precision, recall, and F1-score. This limitation may be attributed to its linear nature, which restricts its ability to capture more complex relationships present in transactional and behavioral features.

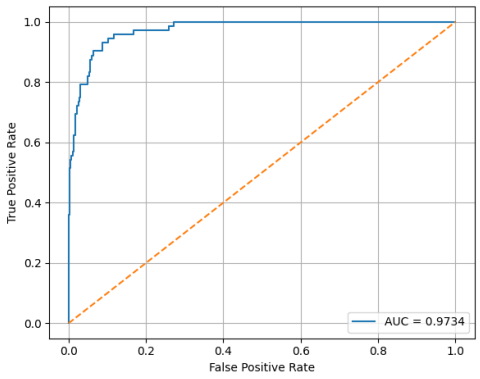


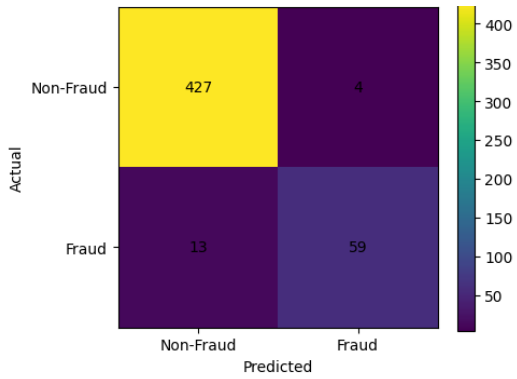
Figure 2 Confusion Matrix of Logistic Regression

Figure 3 ROC Curve of Logistic Regression

Random Forest

The performance of Random Forest is presented in figure 4 and figure 5. Based on the confusion matrix, the model correctly classified 427 non-fraud transactions and 59 fraud transactions. It produced only 4 false positives, indicating that very few legitimate transactions were incorrectly identified as fraud. In addition, the model resulted in 13 false negatives, which means that some fraudulent transactions were not detected. Overall, these results demonstrate that Random Forest is highly effective in correctly identifying non-fraud transactions while maintaining a strong capability to detect fraud cases.

The ROC-AUC value of 0.985 indicates excellent discriminative performance, showing that the model can effectively separate fraud and non-fraud classes across different classification thresholds. The high precision value of 0.937 further confirms that the model is robust against false alarms, making it suitable for real-world deployment where minimizing incorrect fraud alerts is important. Compared to Logistic Regression, Random Forest provides improved overall performance, likely due to its ability to capture non-linear relationships and complex interactions among transactional and behavioral features.



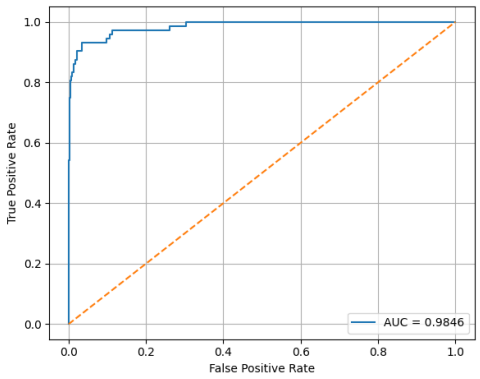


Figure 4 Confusion Matrix of Random Forest

Figure 5 ROC Curve of Random Forest

XGBoost

The performance of XGBoost is shown in figure 6 and figure 7. Based on the confusion matrix, the model correctly classified all 72 fraud transactions, resulting in a recall value of 1.000. This means that no fraudulent transactions were missed, which is highly desirable in fraud detection systems where undetected fraud can lead to significant financial losses. In addition, the model correctly identified 408 non-fraud transactions. However, it produced 23 false positives, indicating that a number of legitimate transactions were incorrectly classified as fraud. This reflects a trade-off in the model's behavior, where improving fraud detection sensitivity leads to an increase in false alarms.

The ROC-AUC value of 0.999 indicates near-perfect discriminative performance, showing that XGBoost is highly effective in distinguishing between fraud and non-fraud classes across various thresholds. This strong performance can be attributed to its ability to model complex, non-linear relationships within the data and to handle feature interactions effectively. Despite its slightly lower precision compared to Random Forest, the model's ability to detect all fraud cases makes it particularly suitable for applications where maximizing detection rate is more critical than minimizing false positives.

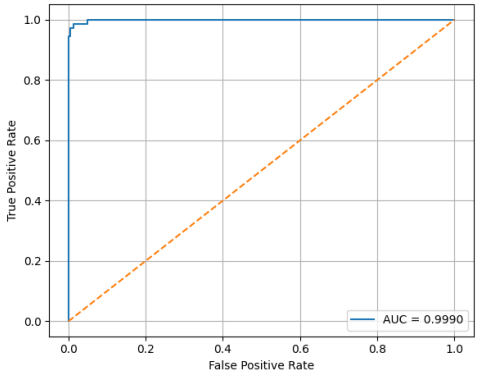


Figure 6 Confusion Matrix of XGBoost

Figure 7 ROC Curve of XGBoost

Feature Importance

The feature importance derived from the Random Forest model is shown in figure 8. The results indicate that TransactionAmount and DeviationFromAccountAvg are the most influential features, suggesting that both the magnitude of a transaction and its deviation from a user's typical behavior play critical roles in identifying fraudulent activity. In addition, features such as Channel and TransactionDuration also contribute significantly, highlighting the importance of contextual and behavioral information in the detection process. Transactions conducted through certain channels, particularly online platforms, may carry higher risk, while unusually long transaction durations can signal abnormal

activity. Overall, these findings demonstrate that combining transactional characteristics with behavioral patterns provides a more comprehensive representation of fraud-related signals, thereby improving the effectiveness of the detection model.

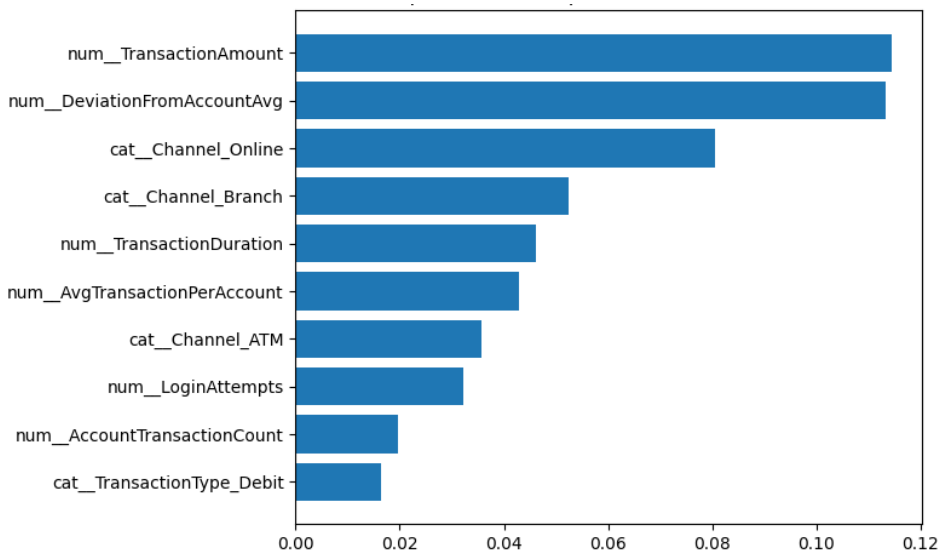


Figure 8 Feature Importance Ranking

Discussion

The experimental results indicate that ensemble-based models, particularly Random Forest and XGBoost, provide superior performance compared to Logistic Regression in detecting fraudulent transactions. This can be attributed to their ability to capture complex, non-linear relationships within the data, which are common in fraud patterns that involve both transactional and behavioral characteristics [21]. Logistic Regression, while demonstrating stable and relatively strong baseline performance, is inherently limited by its linear nature, making it less effective in modeling interactions between multiple features [21]. In contrast, Random Forest and XGBoost leverage multiple decision trees to learn intricate patterns, resulting in improved classification outcomes across key evaluation metrics [22]. The difference in performance highlights the importance of selecting models that align with the complexity of the underlying data in fraud detection tasks.

A notable observation from the results is the trade-off between precision and recall among the models. Random Forest achieved the highest precision and F1-score, indicating a well-balanced performance with relatively few false positive predictions, which is advantageous in reducing unnecessary fraud alerts. On the other hand, XGBoost achieved perfect recall, successfully identifying all fraudulent transactions in the test set. This demonstrates its strong sensitivity to fraud cases, which is critical in minimizing undetected fraud. However, this improvement comes with an increase in false positives, reflecting a

common trade-off in classification problems involving imbalanced data [23]. Additionally, the high ROC-AUC values observed across all models confirm their strong ability to distinguish between fraud and non-fraud classes. The results also emphasize the importance of feature engineering, as features capturing transaction magnitude and deviations from normal behavior play a significant role in enhancing model performance [22].

Conclusion

This study proposed a machine learning approach for detecting fraudulent banking transactions by integrating behavioral and transactional features, supported by a rule-based weak labeling strategy to address the absence of ground truth labels. The experimental results demonstrated that ensemble-based models, particularly Random Forest and XGBoost, consistently outperformed Logistic Regression across multiple evaluation metrics, highlighting their ability to capture complex patterns in fraud-related data. Random Forest achieved the best overall balance between precision and recall, making it suitable for practical applications that require reliable detection with minimal false alarms, while XGBoost achieved perfect recall and the highest ROC-AUC, indicating its effectiveness in identifying all fraudulent transactions. The findings also emphasize the critical role of feature engineering, where features such as transaction amount and deviation from user behavior significantly contributed to model performance. Despite these promising results, the study is limited by the use of rule-based labeling and a relatively constrained dataset, which may affect generalizability. Future work should focus on validating the approach using real-world labeled datasets, improving feature robustness, and assessing model deployment in operational environments.

Declarations

Author Contributions

Conceptualization: H.W.; Methodology: H.W.; Software: H.W.; Validation: H.W.; Formal Analysis: H.W.; Investigation: H.W.; Resources: H.W.; Data Curation: H.W.; Writing Original Draft Preparation: H.W.; Writing Review and Editing: H.W.; Visualization: H.W.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. D. Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *2015 IEEE Symposium Series on Computational Intelligence (SSCI)*, Cape Town, South Africa, vol. 2015, no. January, pp. 159–166, 2015, doi: 10.1109/SSCI.2015.33.

[2] J. Jurgovsky, M. Granitzer, K. Ziegler, S. Calabretto, P.-E. Portier, L. He-Guelton, and O. Caelen, "Sequence classification for credit-card fraud detection," *Expert Systems with Applications*, vol. 100, no. June, pp. 234–245, Jun. 2018, doi: 10.1016/j.eswa.2018.01.037.

[3] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, Feb. 2011, doi: 10.1016/j.dss.2010.08.008.

[4] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, vol. 2016, no. August, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[5] L. P. G. Evans, N. M. Adams, and C. Anagnostopoulos, "When does active learning work?," in *Advances in Intelligent Data Analysis XII*, A. Tucker, F. Höppner, A. Siebes, and S. Swift, Eds., *Lecture Notes in Computer Science*, vol. 8207. Berlin, Heidelberg: Springer, 2013, pp. 174–185, doi: 10.1007/978-3-642-41398-8_16.

[6] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, Feb. 2011, doi: 10.1016/j.dss.2010.08.006.

[7] C. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," arXiv preprint arXiv:1009.6119, vol. 2010, no. September, pp. 1-14, 2010, doi: 10.48550/arXiv.1009.6119.

[8] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[10] C. X. Ling and V. S. Sheng, "Class imbalance problem," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2011, p. 171, doi: 10.1007/978-0-387-30164-8_110.

[11] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. June, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[13] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[14] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[16] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/aos/1013203451.

[17] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences*, vol. 557, pp. 317–331, May 2021, doi: 10.1016/j.ins.2019.05.042.

[18] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188, Dec. 2012, doi: 10.2307/41703503.

[19] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, no. June, pp. 134–142, Jun. 2016, doi: 10.1016/j.eswa.2015.12.030.

[20] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Mining and Knowledge Discovery*, vol. 18, no. 1, pp. 30–55, Feb. 2009, doi: 10.1007/s10618-008-0116-z.

[21] W. Xu, "Interpretable Machine Learning-Based Fraud Detection Model and Knowledge Discovery in Financial Transactions," in *Recent Developments in Computational Finance and Business Analytics*, R. Gupta, F. Bartolucci, V. N. Katsikis, and S. Patnaik, Eds. Cham, Switzerland: Springer, 2025, pp. xxx–xxx, doi: 10.1007/978-3-031-99477-7_28.

[22] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection," *Information Sciences*, vol. 479, pp. 448–455, Mar. 2019, doi: 10.1016/j.ins.2017.12.030.

[23] F. Carcillo, A. Dal Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection," *Information Sciences*, vol. 557, pp. 317–331, May 2021, doi: 10.1016/j.ins.2019.05.042.