SSR Journal of Economics, Business and Management (SSRJEBM)

Homepage: https://ssrpublisher.com/ssrjebm/

Volume 3, Issue 2, 2026 ISSN: 3049-0405

Transaction Behavior Analysis for Early Fraud Detection Using Interpretable Models

A.O. Akinade[1]*, S.O. Olukumoro[2], I.O Ogundele[3], A.A Adeniran[4] & S.A Aborisade[5]

[1,2,3&4]Department of Computer Technology, Yaba College of technology Yaba Lagos.

[5]Nigeria Revenue Service

Received: 20.01.2026 | **Accepted:** 06.02.2026 | **Published:** 13.02.2026

***Corresponding author:** A.O. Akinade

DOI: *10.5281/zenodo.18627424*

Abstract **Original Research Article**

Detecting irregularities in financial transactions has become easier with the development of machine learning (ML) tools. The usefulness of machine learning algorithms in detecting credit card fraud is investigated in this paper, which also highlights important developments in the field. A significant worry now is credit card fraud as the financial sector embraces more digital transactions. To ensure fair and unbiased decision-making, ethical issues with algorithmic bias, data privacy, and transparency must still be addressed. This study attempts to provide an approach to fraud detection that use interpretable models to analyze transaction behavior patterns in order to identify fraud early through the use of SHAP-explained Gradient Boosting, Decision Trees, and Logistic Regression. As a result, the goal of the study is to determine how well interpretable machine learning models can detect fraud in transactional datasets. to ascertain efficient methods for maximizing the trade-off between Explainability and model performance. Results from the study showed that the logistic regression model produced good results on the five-fold cross validation, with an accuracy of 99.90%, precision of 100%, recall of 90%, and F1-score of 93.3%. In order to meet the operational needs of fraud detection systems in real-time financial contexts, the study's technique enhanced explainability and efficiency.

Keywords: SHAP, Explainable AI, Interpretability, Gradient Boosting, fraud detection.

1. INTRODUCTION

Machine learning (ML) techniques have emerged as an effective approach for detecting anomalies in financial transactions as fraudulent activities evolve. This study examines the effectiveness of ML algorithms in credit card fraud detection, highlighting significant advancements in the field. As the financial industry adopts more digital transactions, credit card fraud has become a major concern. Traditional fraud detection methods, such as rule- based systems, have limitations in handling complex fraud patterns.

By allowing for the real-time examination of massive transactional data, the incorporation of machine learning techniques has transformed the identification of fraud. ML algorithms learn from past data to identify novel fraudulent activities, in contrast to rule-based approaches that depend on preset patterns.

Conventional fraud detection systems mostly use statistical techniques and manually created criteria to find questionable activity. These techniques offer a fundamental strategy for detecting fraud, but they have a number of drawbacks. Rule-based systems frequently mistakenly identify valid transactions while missing complex fraud patterns, leading to a high rate of false positives and false negatives. By integrating several ML models, ensemble learning and hybrid techniques improve fraud detection performance even further. By combining the advantages of several methods, XGBoost and LightGBM increase classification accuracy. According to studies, these models are quite useful in real-world applications since they greatly increase fraud detection rates when combined with anomaly detection approaches. [7]. Numerous machine learning techniques, including logistic regression, decision trees, support vector machines, and ensemble algorithms, have been used recently to identify credit card fraud. Despite their middling performance, these models are frequently constrained by their high false positive rates and inability to adjust to changing fraud patterns. [1]. To find coordinated fraudulent activity across financial networks, AI-powered fraud detection technologies use anomaly detection and network analysis approaches. AI models outperform traditional rule-based systems in terms of detection rates, according to empirical research, which improves regulatory supervision and financial security. Even with the encouraging developments, there are still a number of obstacles to overcome before AI can be used to prevent and comply with financial crime. It is still imperative to address ethical concerns about algorithmic bias, data privacy, and transparency in order to guarantee impartial and equitable decision-making [10].

This study aims to design a method for detecting fraud that uses interpretable models to examine transaction behavior patterns in order to spot fraud early by utilizing Logistic Regression, Decision Trees, and SHAP-explained Gradient Boosting. Hence the study seeks to identify to what extent interpretable machine learning models can effectively identify fraud in transactional datasets. Also to determine effective ways to optimize the trade-off between model performance and Explainability.

2. RELATED WORKS

[11] Proposed an enhanced hybrid model combining Lstm, Resnet, and an attention mechanism for credit card fraud detection. The study utilized a novel hybrid model that integrates resnet for spatial feature extraction, long short-term memory (Lstm) networks for capturing temporal dependencies, and an attention mechanism to prioritize significant features. Results from the study showed that the proposed framework achieves superior results, including a precision of 96%, recall of 92%, and an F1-score of 93.97%, outperforming benchmark models by a significant margin. The study establishes a strong foundation for improving fraud detection systems and contributes to advancing machine learning methodologies in financial security applications. However, the validation process should additionally examine the model's performance on datasets with high-dimensional features and varying degrees of class imbalance in order to guarantee adaptability and resilience in practical settings.

[7] Proposed Advancing Machine Learning for Financial Fraud Detection: A Comprehensive Review of Algorithms, Challenges, and Future Directions. The study examines how machine learning (ML) algorithms can be used to detect fraud, emphasizing the usefulness of models like Random Forest, LightGBM, and Artificial Neural networks have a high recall and accuracy rate when detecting fraudulent activity. Additionally, the study looks at how feature engineering, ensemble learning, and data augmentation (SMOTE, KCGAN) might improve fraud detection skills. The study's findings indicate that Random Forest, LightGBM, and Artificial Neural Networks are the most successful algorithms in detecting fraudulent transactions, with greater accuracy and recall. To increase the flexibility and effectiveness of fraud detection, however, attention must be paid to deep learning architectures, reinforcement learning, and cross-domain data integration.

[2] Proposed FraudX AI: An Interpretable

Machine Learning Framework for Credit Card Fraud Detection on Imbalanced Datasets. The study utilized an ensemble-based framework addressing the challenges in fraud detection, including imbalanced datasets, interpretability, and scalability. The FraudX AI combines random forest and XGBoost as baseline models, integrating their results by averaging probabilities and optimizing thresholds to improve detection performance. Results from the study showed that FraudX AI achieved a recall value of 95% and an AUC-PR of 97%, effectively detecting rare fraudulent transactions and minimizing false positives. However, Prioritizing the implementation of adaptive learning mechanisms and verifying FraudX AI's applicability across various datasets is necessary in order to improve fraud detection techniques in response to new fraud trends.

[4] proposed Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications. In the study, the Kitchenham framework was employed in the investigation. It draws attention to the expanding application of models like transformers, ensemble approaches, Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and others in fields including financial statements, insurance, and credit card transactions. For researchers, practitioners, and policymakers looking to improve fraud mitigation in the dynamic financial world, the study was able to offer practical insights. However, in order to turn research into practical applications, organizations should create pipelines that combine detection models with fundamental transaction systems, create modular designs for usage in certain industries, and put in place ongoing idea drift monitoring.

[1] proposed Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models. The study used machine learning (ML) algorithms to detect fraudulent transactions in a methodical manner. using deep learning models (Multilayer Perceptron (MLP), Artificial Neural Network (ANN), ensemble learning (EL) models (Random Forest (RF)), Extreme Gradient Boosting (XGBoost), and Adaptive Boosting (Adaboost), as well as machine learning models (Support Vector Machine (SVM), Decision Tree (DT), and Extreme Gradient Boosting). The findings show notable success, with the DT, RF, and MLP models obtaining a high accuracy of 0.99, highlighting their potential for precise financial transaction fraud detection. However, it is necessary to attempt integrating DL architectures in order to identify more challenging patterns in transactional data.

3. METHODOLOGY

The basic idea used in the study is briefly explained in order to accomplish the study's goal. Following that, a full discussion of the model used is explained. For this study, three models were chosen on the basis of their performance profile and interpretability. First, the Linear and interpretable by feature coefficients is logistic regression. Similarly, a rule-based decision tree classifier that may be used to extract decision routes and then the XGBoost with SHAP: Post-hoc explanation using an ensemble model and SHAP values.

3.1 Logistic Regression

This classification technique guarantees that the output is a probability between 0 and 1 by fitting an S-shaped "logistic function" to the data. Traditional statistically based fraud detection techniques are represented by logistic regression models. [9] Based on the predicted probability, the model can classify the outcome into one of two categories. A common threshold is 50 % or 0.5, where a probability above this threshold is classified as one category, and a probability below is the other.

3.2 Decision Tree

A decision tree illustrates a decision-making process by displaying several options, their possible results, and the repercussions of each choice. It begins with a root node and branches out, providing a visual representation of every path that could lead to a solution, which aids in problem analysis. These trees are employed as a machine learning model for categorization and prediction as well as for informal decision-making. A tree-structured model is produced using the Decision Tree (DT) technique. By depicting alternatives as branches and outcomes

as leaf nodes, it separates data based on attributes and makes an effort to forecast a target variable using simple decision rules. [6]

3.3 Random Forest

Due to its ease of implementation and configuration, Random Forest is the most often used method for fraud detection in supervised learning. Furthermore, it produces good results with little parameter adjustment. When dealing with complex, unbalanced data that contains patterns that are hard to identify directly, Random Forest is frequently utilized [12].

3.4 Gradient Boosting

An ensemble machine learning method called gradient boosting turns several weak models usually decision trees into one powerful model. It improves the overall predicted accuracy over a large number of iterations by training new models to fix the mistakes caused by the older ones. This is accomplished by using the "gradient" of a loss function to build models one after the other with the goal of minimizing the combined model's mistakes. Through learning from prior misclassifications, these tree-based boosting techniques improve detection accuracy. [8]

3.5 Shapely Additive Explanations (SHAP)

By allocating a contribution score to each input feature for a particular prediction, SHAP provides an explanation for the results of any machine learning model. It employs Shapley values from game theory, in which features are viewed as "players" in a game, and their value is determined by how much they contribute to the "payout" (the forecast made by the model). Both local explanations (the reasons behind a specific prediction) and global explanations (the overall operation of the model) can be obtained from SHAP. In order to overcome the challenge of the "black box" nature of AI models, which often lack transparency and interpretability, Explainable AI (XAI) offers human-interpretable insights into how AI models make decisions. Using strategies like Shapely Additive Explanations (SHAP) [8].

3.6 Datasets.

For the study the following dataset were utilized to evaluate the SHAP-explained Gradient Boost in fraud detection. The This study utilizes Credit Card Fraud Detection dataset from Kaggle, which contains 1000 datasets with transaction amount and the class detected. This can be further explained mathematically as $let\ A = \{a_1, a_2, - - - a_n\}$ represent the users transactions and $z \in \{0,1\}$ be the class detected, then $z = 0$ represents the legitimate transactions while $z = 1$ represents the fraudulent ones. The transaction amount is explained given $a \in A = \{\mathbb{R}\}$. In order to improve accuracy and speed, the data was also preprocessed to guarantee proper cleaning, transformation, and preparation into a high-quality format appropriate for analysis and machine learning.

3.7 Conceptual Framework

This section describes the framework as shown in Figure 1 which is designed for detecting fraud using interpretable models and it divided in three phases. The first phase comprises of the AI Models chosen for the study which are the Logistic regression, Decision Tree and the XGboost. This models help to classify and accurately predict each user's transaction from the dataset. The second phase involves the use of the Explainable AI specifically the Shapley Additive Explanations (SHAP) which would be used to interpret correctly the gradient boosting model giving better interpretability and enhanced transparency. The third phase shows the accurate predictions from the system in which it is able to distinguish a legitimate transaction from a fraudulent one.

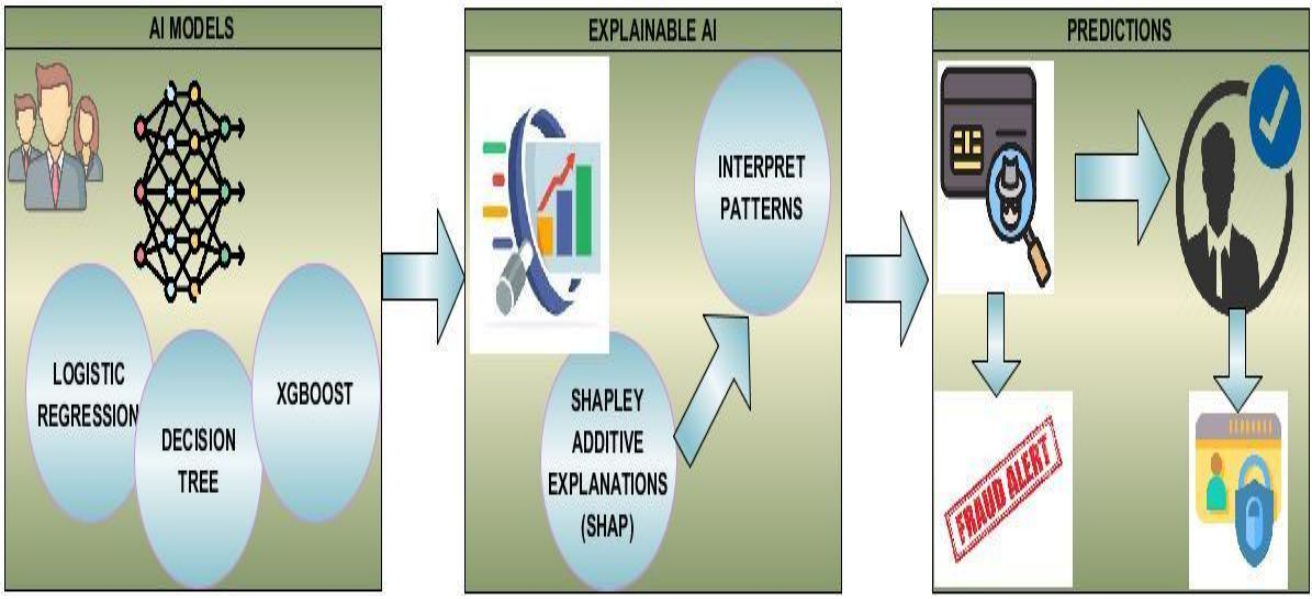

Figure 1: Conceptual Framework for the detecting of fraud using interpretable models

4 RESULTS AND DISCUSSION

The interpretable models used to examine the transaction behavior patterns are the Logistic Regression, Decision Trees, and SHAP-explained Gradient Boosting. Table 1 shows that the logistic regression model performed well on the cross validation of 5 fold with accuracy of 99.90%, precision of 100%, Recall of 90% and F1-score of 93.3%.

Table 1: Cross Validation Scores of each Metric for Logistic Regression

METHOD	ACCURACY	PRECISION	RECALL	F1-SCORE
LOGISTIC REGRESSION	99.90%	100%	90%	93.3%

4.1 Cross Validation Metric for Logistic Regression

The figure 2 shows how the model's performance varied across the different folds for each metric. The analysis shows that, most scores are very high, close to 1.0, indicating excellent performance across the folds for most metrics. The Recall score in Fold 2 is slightly lower than the others, suggesting that in that particular fold, the model had a bit more difficulty identifying all the positive cases.

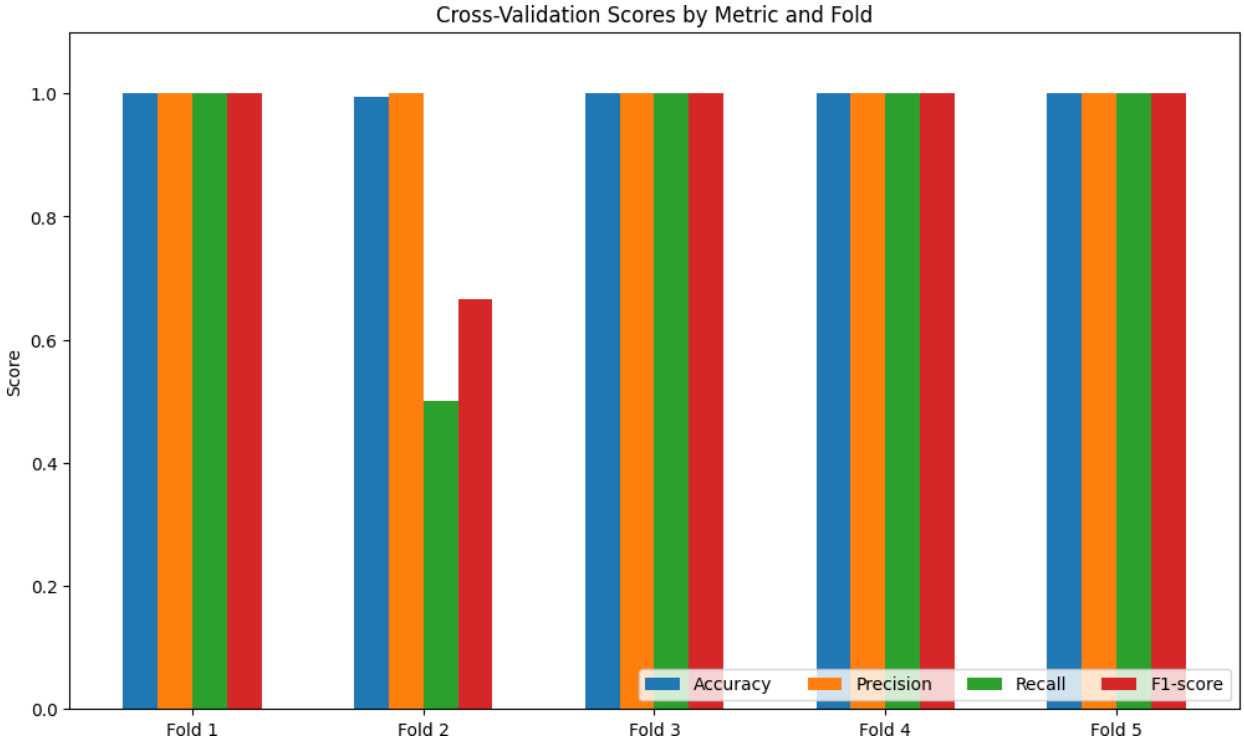

Figure 2: Cross-Validation Scores by Metric and Fold

4.1.1 ROC Curve (Receiver Operating Characteristic Curve) for Logistic Regression

The figure 3 shows the ROC curve which is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It plots two parameters; True Positive Rate (TPR) which is also known as Sensitivity or Recall, this is the proportion of actual positive cases that are correctly identified as positive. Also the False Positive Rate (FPR) which is the proportion of actual negative cases that are incorrectly identified as positive. The ROC curve shows the trade-off between TPR and FPR at various threshold settings. A perfect classifier would have a curve that goes straight up from (0,0) to (0,1) and then straight across to (1,1), indicating a TPR of 1 and an FPR of 0 at some threshold. A completely random classifier would have a diagonal line from (0,0) to (1,1), represented by the 'Random Guess' line in your plot.

4.1.2 AUC (Area Under the Curve) for Logistic Regression:

The AUC is a single scalar value that summarizes the overall performance of a binary classifier across all possible classification thresholds. It represents the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. An AUC of 1.0 represents a perfect classifier that can perfectly distinguish between the positive and negative classes. Similarly, an AUC of 0.5 indicates a classifier that performs no better than random guessing. Furthermore, an AUC less than 0.5 suggests that the classifier is performing worse than random, likely due to errors in the model or data. The figure 3 shows that the ROC curve goes directly to the top-left corner, and the AUC is 1.0. This confirms the earlier evaluation metrics, indicating that the Logistic Regression model was able to perfectly separate the two classes in the dataset based on the 'Amount TRANSACTED' feature.

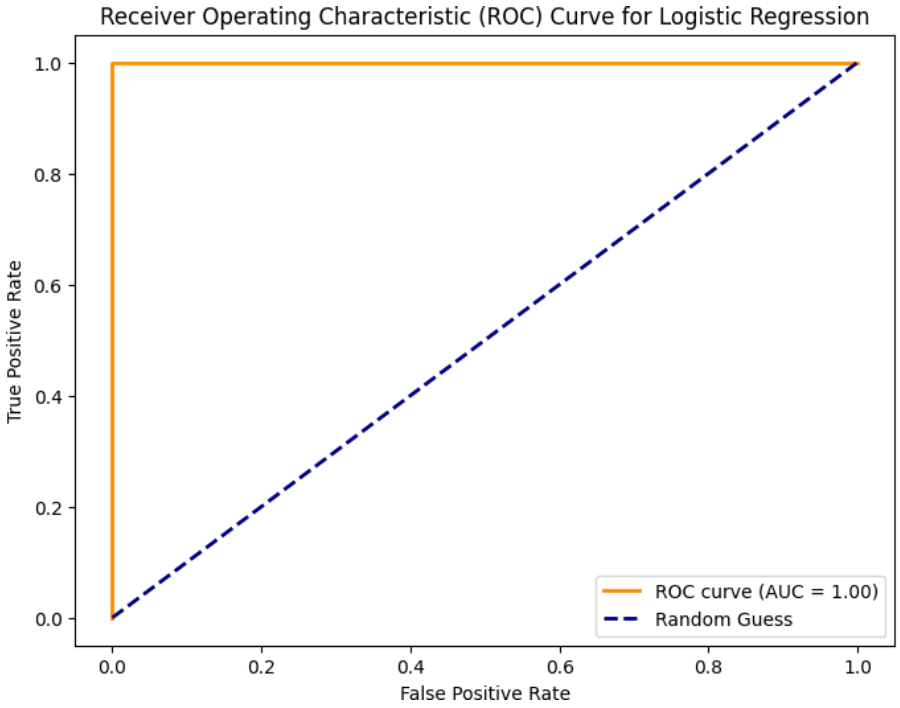

Figure 3: Receiver Operating Characteristic (ROC) Curve for Logistic Regression

Table 2: Performance Metrics for the decision Tree Model

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
DECISION TREE	1.0	1.0	1.0	1.0

4.2 Performance Metrics for Decision Tree

The Table 2 shows the performance metrics for the decision Tree which indicate that the model achieved perfect performance on the testing data. The decision tree model is shown in Figure 4 and it uses the single feature 'Amount TRANSACTED' to make classifications. Looking at the tree visualization, the decision boundary is determined by a single threshold value for this feature. All data points with 'Amount TRANSACTED' less than or equal to this threshold are classified into one class (likely class 0), and all data points with 'Amount TRANSACTED' greater than this threshold are classified into the other class (likely class 1). The exact threshold value is shown at the root of the decision tree plot. Since the model achieved perfect accuracy, this single split based on 'Amount TRANSACTED' is sufficient to perfectly separate the two classes in the dataset.

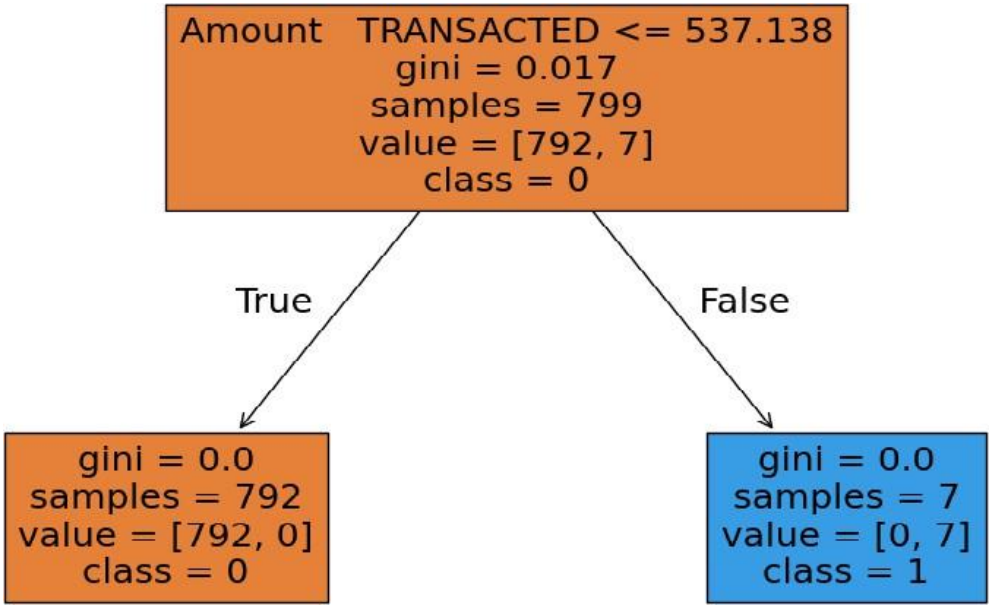

Figure 4: Decision Tree Model for detecting fraud in transaction behavior patterns Table 3: Performance

Metrics for the Gradient Boosting Model

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
GRADIENT BOOSTING	1.0	1.0	1.0	1.0

4.3 Performance Metrics for Gradient Boosting

The Table 3 shows the performance metrics for the gradient boosting which indicate that the model achieved perfect performance on the testing data making it effective for detecting fraud in transaction behavior patterns. The Figure 5 illustrates the performance of the gradient boosting across each metric achieving accuracy of 1.0, precision of 1,0, recall of 1.0 and F1- score of 1.0.

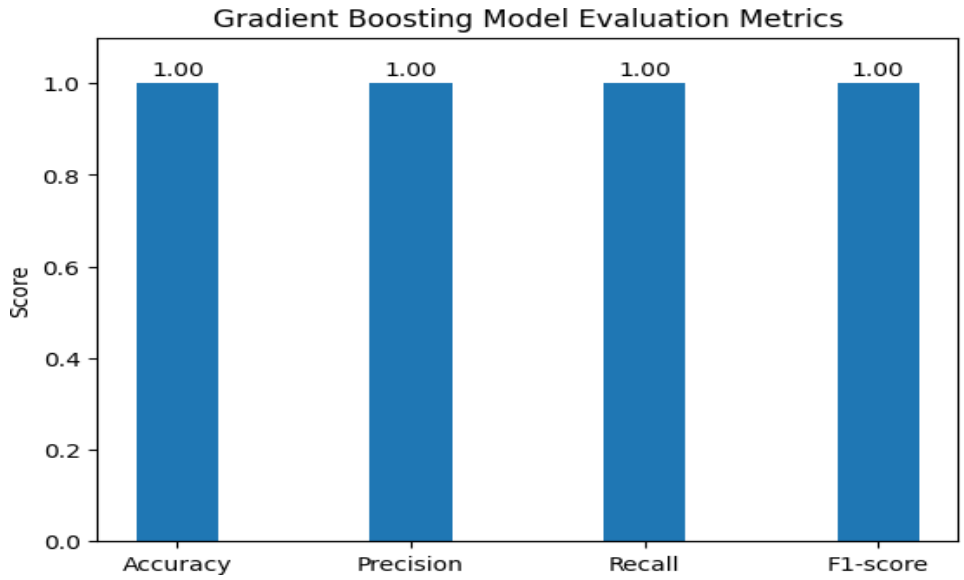

Figure 5: Gradient Boosting Model Evaluation Metrics

4.3.1 Cross Validation on Gradient Boosting

The cross validation was also performed on the gradient boosting model and the results are shown in Table 4. The model yielded an accuracy of 99.9%, precision of 100%, recall of 90% and F1-score of 93.33% indicating an effective performance in the prediction of fraudulent transactions

Table 4: Cross Validation on Gradient Boosting Model

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE
CROSS VALIDATION ON GRADIENT BOOSTING	99.9%	100%	90%	93.33%

The figure 6 shows that average scores across the 5 folds are very high for Accuracy (0.9990), Precision (1.0000), and F1-score (0.9333). The average Recall is slightly lower at 0.9000. Similarly, the scores per fold shows that the Accuracy and Precision scores are consistently high (1.0) across all 5 folds. However, the Recall and F1-score show some variation across the folds. Specifically, in Fold 2, the Recall and F1-score are lower compared to the other folds. This suggests that while the model generally performs very well, there might be some instances in certain subsets of the data (like in Fold 2) where it has a bit more difficulty identifying all the

positive cases (as indicated by the lower Recall).

Overall, the cross-validation results indicate that the Gradient Boosting model is generally robust and performs very well on the dataset, although there is a slight variation in its ability to capture all positive instances across different data splits.

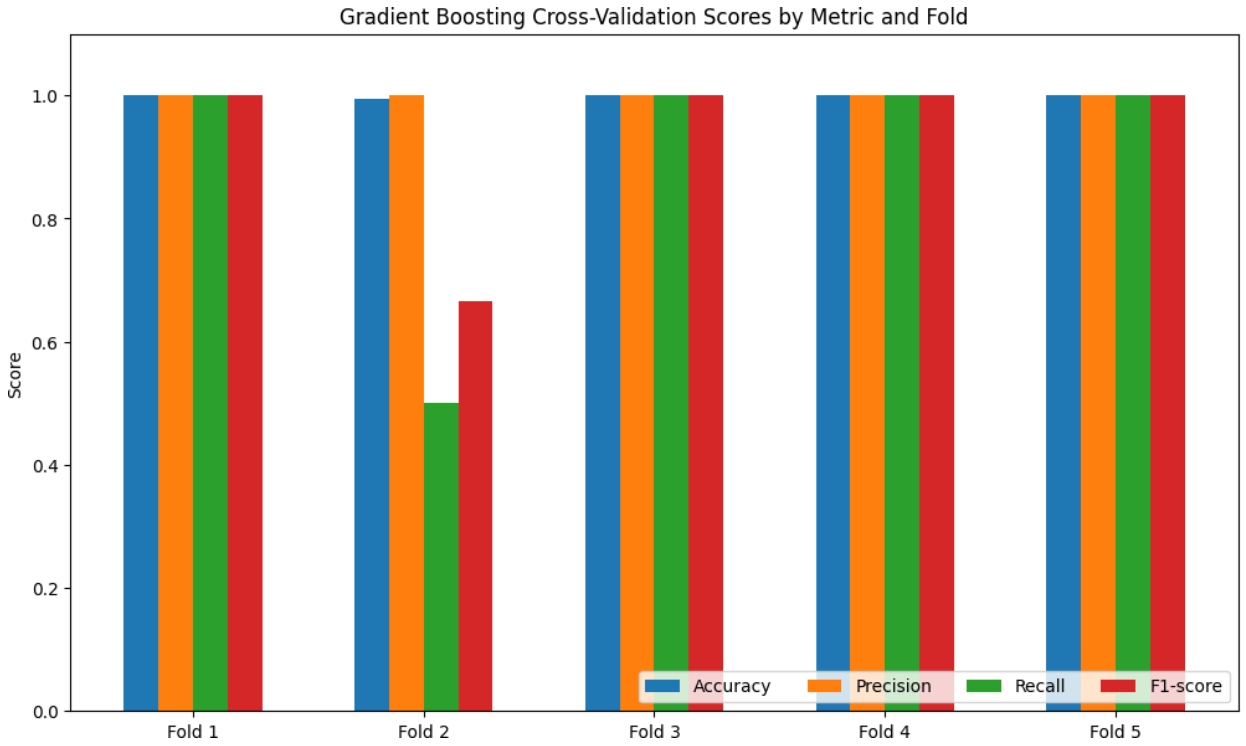

Figure 6: Gradient Boosting Cross Validation Scores

4.4 SHAP-Explained Gradient Boosting

The Shapley Additive Explanation (SHAP) was also applied on the gradient boosting model in an attempt to optimize the trade-off between model performance and Explainability. The Figure 7 shows a single row for 'Amount TRANSACTED', as it's the only feature. The dots are spread along the x-axis, indicating that 'Amount TRANSACTED' has a significant impact on the model's predictions. The color of the dots changes from blue to red moving from left to right along the x-axis. This means that lower values of 'Amount TRANSACTED' represented by the blue dots are associated with negative SHAP values (pushing the prediction towards class 0), and higher values of 'Amount TRANSACTED' represented by the red dots are associated with positive SHAP values (pushing the prediction towards class 1). This plot confirms that 'Amount TRANSACTED' is the most important feature for the Gradient Boosting model's predictions, and it shows a clear relationship as higher transaction amounts are associated with one class (likely the positive class, which represents fraud), and lower amounts with the other.

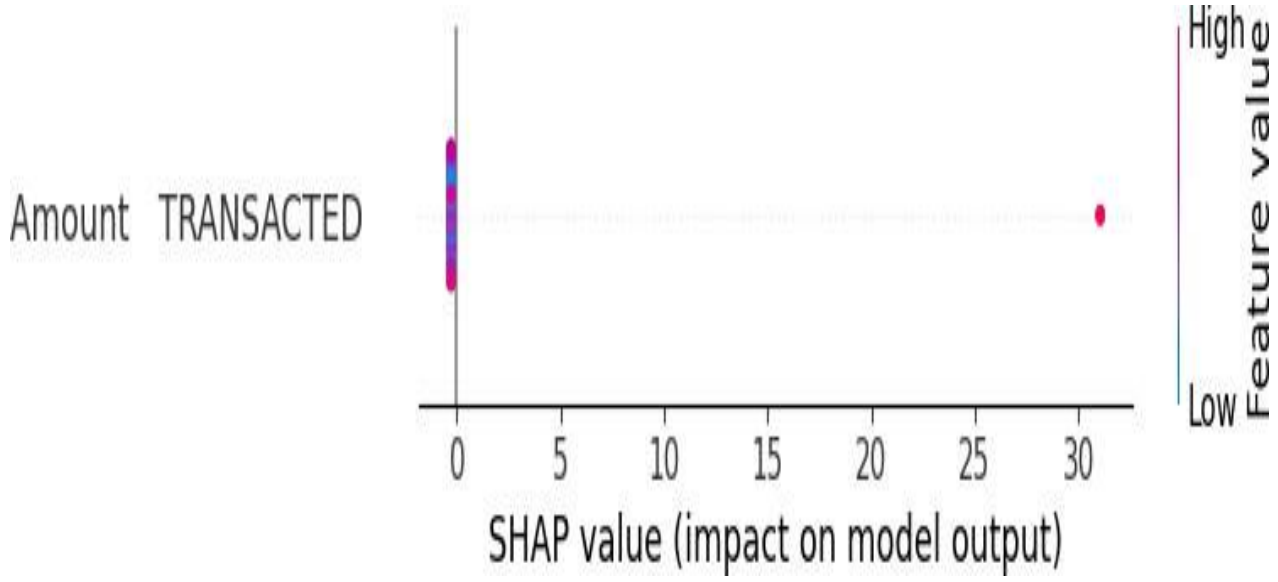

Figure 7: SHAP Plot for the Gradient Boosting Model

4.4.1 SHAP Dependency Plot for the Gradient Boosting Model

In figure 8 the x-axis represents the actual value of the 'Amount TRANSACTED' feature while the y-axis represents the SHAP value for 'Amount TRANSACTED'. Each dot represents a single data instance from the test set. For lower values of 'Amount TRANSACTED' (on the left side of the x-axis), the SHAP values are generally negative. This means that lower transaction amounts push the model's prediction towards the base value (the average prediction over the training data), and in the context of a binary classification as in this case, likely towards predicting the negative class (Class 0). As the value of 'Amount TRANSACTED' increases, the SHAP values become more positive. This indicates that higher transaction amounts push the model's prediction away from the base value and towards predicting the positive class (Class 1). The plot visually demonstrates the strong positive relationship between the 'Amount TRANSACTED' and its impact on the Gradient Boosting model's prediction. Higher transaction amounts are associated with a higher likelihood of being classified as the positive class. Since there is only one feature, there are no interaction effects with other features shown in this plot. This framework has been able to enhance both efficiency and explainability, aligning with the operational needs of fraud detection systems in real-time financial environments.

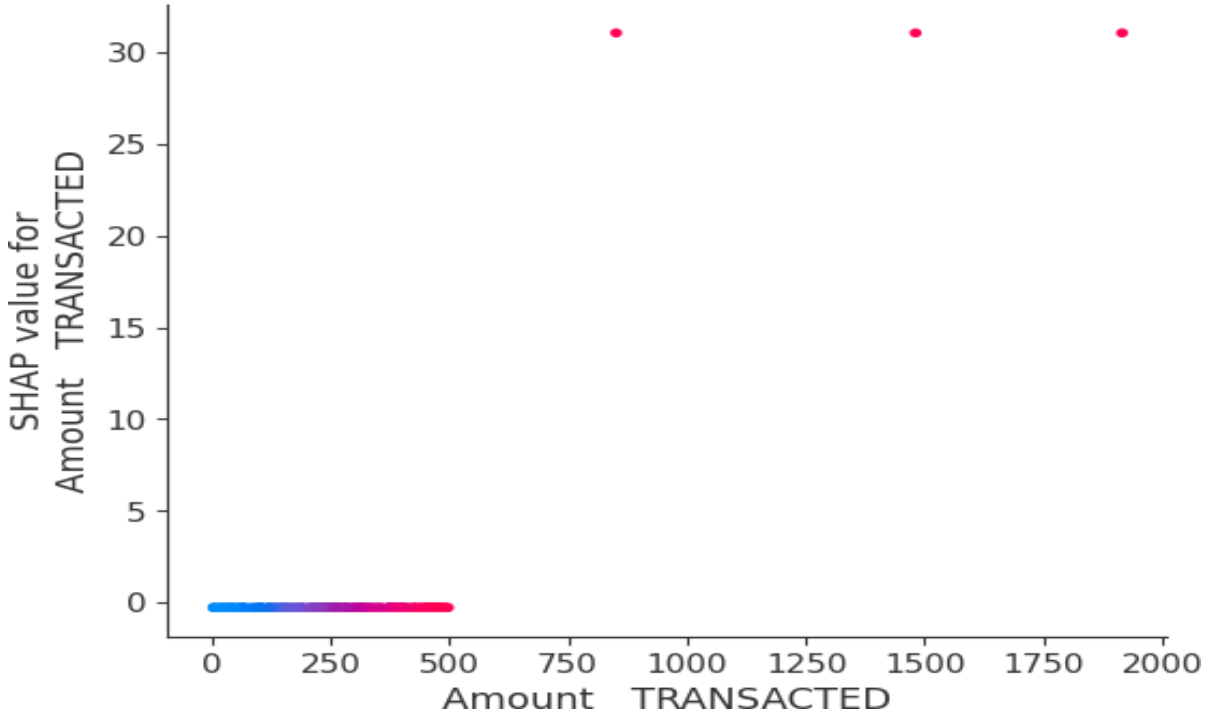

Figure 8: SHAP Dependency Plot for the Gradient Boosting Model

5 CONCLUSION

This study looked at how to identify fraud early by analyzing transaction behavior patterns using interpretable models. A rule-based decision tree classifier was utilized which was used to extract decision routes and then the gradient boosting with SHAP: Post-hoc explanation using an ensemble model and SHAP values. Results from the study showed that the Gradient Boosting model is generally robust and performed very well on the dataset. Similarly, the result of the AUC-ROC showed that the Logistic Regression model was able to perfectly separate the two classes in the dataset. The study's methodology improved explainability and efficiency, meeting the operational requirements of fraud detection systems in real-time financial settings. This study shows that early fraud detection may be accomplished with interpretable models without noticeably sacrificing performance. More visible, responsible, and efficient fraud prevention systems are made possible by combining interpretable algorithms with visualization strategies like SHAP. In order to improve contextual relevance, future work might use behavioral biometrics, integrate real-time streaming, and test on Nigerian fintech transaction data.

Acknowledgement

The authors acknowledge the efforts of the reviewers of this paper. We also appreciate their meaningful contribution, valuable suggestions, and comments to this paper which helped us in improving the quality of the manuscript.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the author.

Conflicts of Interest

The authors declare that they have no conflicts ofinterest to this work.

Data Availability Statement

The data that support the findings of this study are openly available online

REFERENCE

[1] Alrasheedi, M. A. (2025). Enhancing Fraud Detection in Credit Card Transactions: A Comparative Study of Machine Learning Models. *Computational Economics*, 1-27.

[2] Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova,

K. (2025). FraudX AI: An Interpretable Machine Learning Framework for Credit Card Fraud Detection on Imbalanced Datasets. *Computers*, *14*(4), 120.

[3] Baratzadeh, F., & Hasheminejad, S. M. (2022). Customer Behavior Analysis to Improve Detection of Fraudulent Transactions using Deep Learning. *Journal of AI and Data Mining*, *10*(1), 87-101.

[4] Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications. *Data Science and Management*.

[5] Darwish, S. M., Salama, A. I., & Elzoghabi, A. A. (2025). Intelligent approach to detecting online fraudulent trading with solution for imbalanced data in fintech forensics. *Scientific Reports*, *15*(1), 17983.

[6] Hafez, I. Y., Hafez, A. Y., Saleh, A., Abd El-Mageed, A. A., & Abohany, A. A. (2025). A systematic review of AI-enhanced techniques in credit card fraud detection. *Journal of Big Data*, *12*(1), 6.

[7] Herath, H. M. M. N. (2025). Advancing machine learning for financial fraud detection: A comprehensive review of algorithms, challenges, and future directions. *ASEAN Journal of Economic and Economic Education*, *4*(1), 49-68.

[8] John, A., & Ahsun, A. (2025). Explainable AI (XAI) for Fraud Detection: Building Trust and Transparency in AI-Driven Financial Security Systems Authors. *Available at SSRN 5285281*.

[9] Li, W., Liu, X., Su, J., & Cui, T. (2025). Advancing financial risk management: A transparent framework for effective fraud detection. *Finance Research Letters*, *75*, 106865.

[10] Paul, A. A., & Ogburie, C. (2025). The Role of AI in preventing financial fraud and enhancing compliance.

[11] Umaru, I. A., Aliyu, A. A., Ibrahim, M., Abdulkadir, S., Ahmed, M. A., Abubakar,

M. A., ... & Tanko, S. A. (2025). AN ENHANCED HYBRID MODEL COMBINING

LSTM, RESNET, AND AN ATTENTION MECHANISM FOR CREDIT CARD FRAUD DETECTION. *FUDMA JOURNAL OF SCIENCES*, *9*(2), 42-48.

Wahyono, T., & David, F. (2025). A systematic review of machine learning-based approaches for financial fraud detection. Journal of System and Management Sciences, 15(1), 69–84. https://doi.org/10.33168/JSMS.2025.0105