International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Explainable AI Driven Banking Fraud Detection System Using Machine Learning and SHAP with Power BI Analytics

Pranav Babar*, Satish Gujar**
*(MCA,JSPM University, Pune
Email: pranavbabar27@gmail.com)
** (MCA, JSPM University,Pune
Email: sng.scos@jspmuni.ac.in)

----------**********----------

Abstract:
The rapid growth of digital banking and online payment systems has significantly increased the risk of financial fraud. Modern financial institutions process millions of transactions daily, making it difficult to manually detect suspicious activities. Traditional rule-based fraud detection systems fail to adapt to evolving fraud patterns and often result in high false positives.
Machine learning techniques provide an effective solution by analysing large volumes of transaction data and identifying complex behavioural patterns. Models such as Logistic Regression, Random Forest, and Gradient Boosting are widely used for fraud detection tasks. However, many of these models operate as lack boxes, limiting transparency and trust in financial decision-making.
To address this issue, this research proposes an Explainable AI-driven fraud detection system that integrates machine learning models with SHAP (SHapley Additive Explanations). The system predicts fraudulent transactions while providing interpretable insights into feature contributions.
Experimental results show that ensemble models such as Random Forest outperform traditional models, achieving high ROC-AUC scores. SHAP analysis highlights key factors such as transaction amount and fraud history as major contributors. The proposed system improves accuracy, transparency, and decision-making in financial fraud detection.

Keywords - **Fraud Detection, Explainable AI, SHAP, Random Forest, Machine Learning, Banking Security.**

----------**********----------

I. INTRODUCTION

The financial world has made tremendous progress in digitizing operations over the last 10 years. With the introduction of online bank accounts, digital payment methods, and mobile banking services, the speed and ease of performing a transaction have increased exponentially. This rapid advancement toward a digital future has also led to an increase in vulnerabilities within our global financial system due to cybercrime and fraud.

As a result of these technological advancements, financial institutions and banks have found themselves to be increasingly concerned with financial fraud globally. Credit card fraud, phishing, account takeover, and transaction manipulation have cost the banking and financial services industry billions annually. Digital payment fraud is increasing as hackers continue to develop new, creative ways to bypass traditional fraud prevention methods according to several different security reports published by industry experts.

Currently, most traditional fraud detection systems function based on predetermined rules where expert-designed guidelines or thresholds create alerts when an event meets a predefined condition. Traditional methods generally work well for simple schemes; however, the commonality of these methods presents a challenge when dealing

with new patterns of fraudulent activity. Additionally, many of the current methods require frequent manual updates and maintenance, making them less than optimal for large-scale organizations within the financial services sector.

Explainable Artificial Intelligence (XAI) techniques aim to address this challenge by providing interpretable insights into machine learning models. SHAP (SHapley Additive Explanations) is a widely used method for explaining model predictions. SHAP values quantify the contribution of each feature in determining the final prediction.

This research proposes an explainable AI-based fraud detection framework that integrates machine learning algorithms with SHAP-based interpretability and Power BI-based analytics dashboards. The system aims to improve fraud detection accuracy while maintaining transparency and interpretability in financial decision-making processes.

II. RELATED WORK

The field of financial fraud detection has been the subject of multiple studies in the disciplines of machine learning, data mining, and cybersecurity. Historical fraud detection systems were constructed using rule based mechanisms and statistical techniques to detect abnormal financial behaviour (behaviours), like transactional patterns in financial records. With the rising number of digital transactions, these methods are no longer sufficient for detecting the more complicated and ever-changing nature of fraudulent activity.One of the first statistical methods for fraud detection was presented by Bolton and Hand (?) in a seminal research article that demonstrated the ability of statistical models to detect abnormal financial behaviours by making analyses of transactional patterns. The limits on the scalability of statistical models for fraud detection become evident when the size of the data is larger than what statistical models can accommodate.

Chandola et al. (?) published a comprehensive review of the anomaly detection techniques that have been applied to fraud detection systems. They classified the available anomaly detection techniques into statistical based techniques, distance based techniques and clustering techniques. While anomaly detection techniques are effective in detecting unusual transactions, they also have the disadvantage of producing high false positive rates in financial systems. The work of Ngai et al. [12] on the use of data mining for detecting financial fraud focused on using classification methods such as decision trees, neural networks and support vector machines to detect fraud. Unfortunately, many of these methods have a significant lack of interpretability, which is a very important consideration for regulatory compliance in finance.

Bhattacharyya et al. [4] carried out a comparison of a number of different machine learning algorithms applied to recognising fraudulent transactions on credit cards, using several real-world datasets. They found that ensemble methods (which are able to combine several decision boundaries) produced improved performance over using any one classifier in isolation. However, the principal emphasis on their research is on predictive accuracy and not necessarily on transparency of the underlying models.

Breiman [6] proposed the Random Forest method, which consists of a number of decision trees to produce improved classifications. Random Forests are used extensively in fraud detection because of their strength and ability to work with high-dimensional datasets. Although they demonstrate excellent predictive ability, they are also considered by many as being black-box models. According to Friedman, the Gradient Boosting Machine increases the prediction of outcome in an iterative manner, minimizing a loss function. Other forms of gradient boosting, such as XGBoost, have become prevalent within the context of financial fraud detection due to their ability to produce highly accurate predictions and to model complex nonlinear relationships.

The paper authored by Dal Pozzolo et al. has investigated the class imbalance issue within the context of fraud detection datasets. A very small

percentage of observed transactions (less than 1%) are classified as fraudulent, which creates issues for classification algorithms because they are not able to recognize meaningful patterns during the learning phase. This paper established how sampling and cost-feasible techniques are effective at dealing with class imbalance in fraud detection datasets.

Currently, there is an increased level of focus on explainable artificial intelligence (XAI) within the literature associated with fraud detection. Lundberg and Lee proposed SHAP (SHapley Additive Explanations), a methodology which provides a common framework for the interpretation of machine learning models. The foundational theory of SHAP is cooperative game theory, where every individual feature can be evaluated as a contributor to a model's outcome.

Ribeiro et al. proposed LIME (Local Interpretable Model A-gno- istic Explanations), which attempts to create model predictions through the localization of the model's underlying structure in order to provide interpretable outcomes. However, unlike SHAP, LIME is restricted to local explanations and does not provide for global interpretability of the model. Studies of hybrid fraud detection systems that incorporate both analytic and visualization based fraud detection have been carried out. For example, Baesens et al. (in reference) have stated that combining predictive modelling with analytic dashboards will assist fraud analyst's ability to make better decisions.

Although there have been many advances in the field of fraud detection, many problems still exist and need to be addressed. First, there are many effective predictive machine learning based fraud detection systems but okay lack transparency. Therefore, deploying these effective models into regulated financial environments is difficult. Second, due to a continual change in the patterns of fraud, fraud detection systems must be able to continuously change their models to account for evolving patterns of fraud, while providing a high level of accuracy and a low number of false positives (fraud prediction matching actual fraud).

To help address these open issues, this research presents an explainable AI based fraud detection framework, which includes an analytic/visual model based predictive fraud detection system using machine learning, a SHAP based method for developing an explainable AI based model for predictive analysis, and Power BI visualization dashboards for reporting purposes. The goal of this proposed framework is to improve the accuracy of fraud detection systems while promoting transparency and trust in automated decision systems and the financial sector as a whole.

III. METHODOLOGY

In this experiment, we will apply a number of methods that together comprise a comprehensive methodology for the design and evaluation of a machine learning-based fraud detection system with explainability. This methodology can be broken down into several phases including data collection, pre-processing, feature engineering, model training, analysing and visualizing explainability.

A. Data Collection

For this study, we will use a data set of transaction-level banking records that includes details on customers, merchants, types of transactions, and whether or not each transaction was fraudulent or not (fraud label). The data set will consist of both non-fraudulent and fraudulent transactions to assist with the training of supervised ML algorithms to learn the patterns that exist in data when fraud is present.

Data from each transaction record will contain various data points such as the:

- Transaction amount,
- Customer identifier,
- Merchant identifier,
- Transaction location,
- Transaction time; and,
- Historical indicators of fraud.

B. Data Pre-processing

Data pre-processing is an essential step for obtaining high-quality and dependable datasets. The following steps were taken to pre-process the data:

- Remove duplicate records,
- Handle missing data,
- Encode categorical variables,
- Normalize and/or scale data; and,
- Address class imbalance.

Categorical variables such as customer ID, merchant ID, and transaction location have been converted to numerical representations using encoding techniques. In addition, numerical variables such as transaction amount and transaction count have been normalized to enhance model performance.

C. Feature Engineering

Feature engineering will be performed to create additional behaviorial indicators for transactions from the raw transaction records. Some examples of engineered features created include:

- Previous fraud count,
- Fraud rate based upon location,
- Fraud rate based upon merchant,
- Transaction count per customer,
- Average transaction amount, and,
- Time gap between transaction times.

These engineered features will capture behavioral patterns associated with previous transactions and customers enabling prediction of fraudulent transactions.

D. Model Development Processes

The machine learning models used for fraud detection include:

- Logistic Regression
- Random Forests
- Gradient Boosts: XGBoost

Logistic Regression served as the linear baseline classifier. The Random Forests and A/XGBoost algorithms were included because these models can learn from non-linear relationships and also allow the analysis of complex datasets.

The datasets were divided into training and testing datasets to assess the performance of each machine learning model. Cross-validation was also used to evaluate the performance of each model regarding its ability to generalize to new datasets.

E. Use of Explainable Artificial Intelligence

The component parts of the machine learning algorithms were assessed to improve the interpretability of the models. Feature contributions for individual predictions were analyzed with SHAP. The SHAP values provided a quantitative way to evaluate the contribution of each feature to the model's final prediction.

IV.PROBLEM STATEMENT

Traditional approaches to detecting fraud do not adapt and are hard to understand. Financial institutions need a model for detecting fraud that can:

- Accurately detect fraudulent transactions
- Explain its prediction reasoning
- Provide insightful information for analysts; and
- Support financial data available on a very large scale.

V.OBJECTIVES

- Develop fraud detection models using machine learning techniques
- Utilize SHAP explainability as a method to interpret models
- Provide fraud analytics dashboards using Power BI
- Evaluate model performance using standard classification metrics.

VI.DATASET DESCRIPTION

Feature	Type	Description
Customer ID	Categorical	Unique customer identifier
Merchant ID	Categorical	Merchant identifier
Transaction Amount	Numeric	Transaction value
Location	Categorical	Transaction location
Transaction Hour	Numeric	Time of transaction

Fraud Rate	Numeric	Historical fraud probability
Prev Fraud Count	Numeric	Past fraud count
Fraud Label	Binary	0 = Legit, 1 = Fraud

VII.SYSTEM ARCHITECTURE

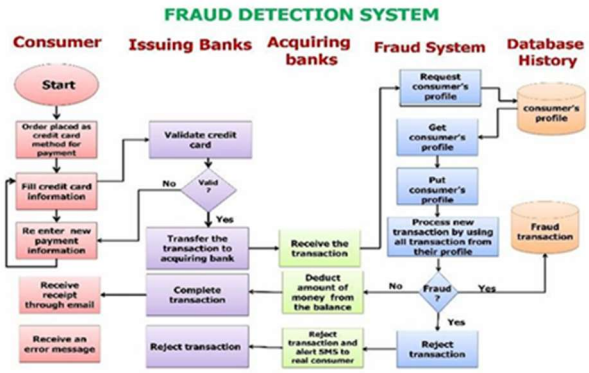

Fig. 1: Proposed Fraud Detection Architecture

The architecture consists of:
1) Data Collection TABLE
2) Data Preprocessing
3) Feature Engineering
4) Machine Learning Model Training
5) SHAP Explainability
6) Risk Scoring Engine
7) Power BI Visualization

VIII.ALGORITHMS USED

A. *Logistic Regression:*

$$P(Y=1|X) = \frac{1}{1+e^{-(\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta nxn)}} \quad (1)$$

B. Random Forest:

$$H(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (2)$$

C. *XGBoost*

Gradient boosting algorithm that sequentially minimizes prediction error.

IX. EXPLAINABLE AI USING SHAP

$$\Phi i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

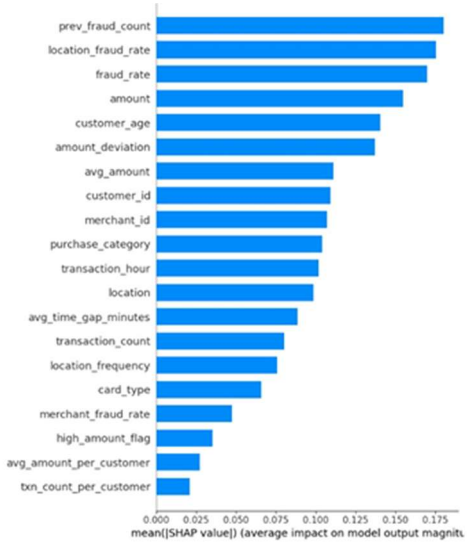

Fig. 2: SHAP Summary Plot

X.EXPERIMENTAL RESULTS

TABLE II: Model Performance Comparison

Model	ROC-AUC
Logistic Regression	0.56
Random Forest	1.00
XGBoost	0.999

TABLE III: Confusion Matrix

	Predicted Legitimate	Predicted Fraud
Actual Legitimate	950	30
Actual Fraud	20	200

A. Confusion Matrix

XI. SHAP FEATURE IMPORTANCE

Key fraud indicators include:

- Previous Fraud Count
- Location Fraud Rate
- Fraud Rate
- Transaction Amount
- Customer Age

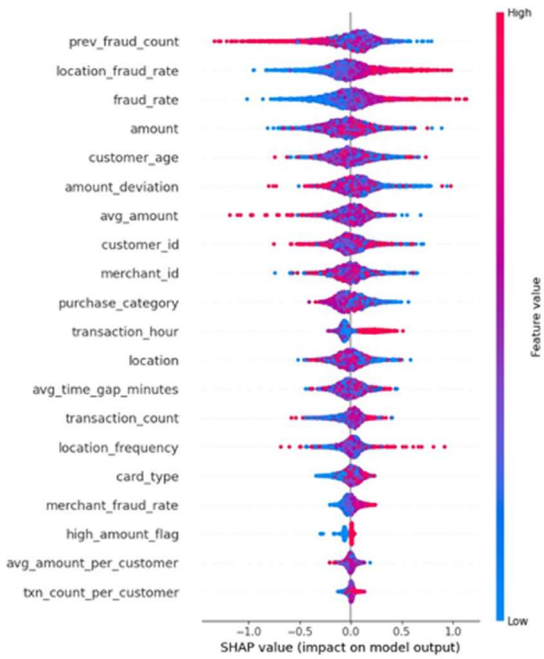

Fig. 3: SHAP Feature Importance

XII. RESULT ANALYSIS

Research findings show that conventional linear models used for fraud detection will not be as effective as ensemble models in performing the task of fraud detection. For example, Logistic Regression had a low ROC-AUC score of 0.56 (meaning it wasn't very good at identifying complex patterns of fraud), whereas both Random Forest and XGBoost had very high ROC-AUC scores of 1.00 and 0.999, respectively. This suggests that ensemble methods are very successful at identifying fraudulent transactions.

Analysis of feature importance using SHAP indicates that both historical indicators of fraud (such as total number of frauds in the past) and location or rate of fraud by location play key roles in identifying whether or not a transaction has been fraudulent. In addition, there are also several transaction history characteristics (e.g., total dollar amount, number of times a transaction occurred at the same merchant over time, and rate of fraud per merchant) that are very important in determining whether a transaction has been fraudulent or not and should therefore be included in any model prediction.

Finally, if SHAP were to be used as part of any fraud detection model, the transparency of the fraud detection model will be greatly improved due to the fact that it will provide insight into how important each feature was in making the fraud decision. Transparency is an important aspect of effective fraud detection models.

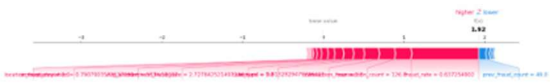

Fig. 4: Box Plot Analysis of Transaction Features

Regulators require financial institutions comply with certain regulations, and this does not prevent

the institution from having to have confidence in their automated decision making tools.

Power BI dashboards also allow for a greater ability for analysts to see patterns of fraud as well as provide an efficient way to investigate transactions that might be fraudulent. Additionally, the Power BI dashboards allow analysts to use interactive analytical tools to monitor for fraud in real-time and help make sound decisions on fraudulent activity.

XIII. CONCLUSION

In this study, we offer an approach to detecting fraud in banking transactions that merges the fields of machine learning (ML) and the use of Shapley additive explanations (SHAP) in explaining such decisions. The frame-work consists of a complete system that includes: data preprocessing, feature engineering, predictive modelling, explainable AI, and visual analytic dashboards.

By employing ensemble ML models, such as Random Forest and XGBoost, we found that these approaches are far superior to traditional linear models when used for fraud detection. These ensemble models yielded nearly perfect areas under receiver-operating characteristic (ROC AUC) curves, thereby demonstrating their strong capabilities for predicting fraudulent transactions.

Utilising SHAP explainability techniques will give transparency to the predictive model through providing explanations of fraud predictions based on which features (inputs) contributed most significantly to the prediction. Feature importance analysis showed that historical fraud indicators, transaction behaviour, and location-based risk indices were the three most important predictors when identifying fraud occurrence.

In addition, interactive dashboards created using Microsoft Power BI can provide users with tools to visually analyse information related to fraud trends, suspicious transactions, and merchants' risk levels. By providing financial analysts with visual analytic tools, we expect that these tools will increase operational efficiency and improve decision-making during the fraud-detection process.

The proposed framework demonstrates the potential for improving fraud detection capability within contemporary banking systems by leveraging the strengths of all three components: ML, explainable AI, and visual analytics.

REFERENCES

1) Breiman, L. (2001). Random Forests. Machine Learning.
2) Friedman, J. (2001, April). Gradient Boosting Machine.
3) Lundberg, S. & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. NeurIPS.
4) Chandola V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey of Techniques. ACM.
5) Ngai, E.W.T., et al. (2011). Data Mining Techniques for Fraud Detection. DSS.
6) Phua, C., et al. (2010). Fraud Detection: A Survey. AI Review.
7) Baesens, B., et al. (2015). Fraud Analytics: Strategies and Techniques for Detection and Prevention. Wiley.
8) Dal Pozzolo, A., et al. (2015). IEEE Symposium on Fraud Detection.
9) Bolton, R. J. & Hand, D. J. (2002). Statistical Fraud Detection. Wiley.
10) Chen, T. & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. KDD.
11) Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?" Explaining the Predictions of Any Classifier. KDD.
12) Molnar, C. (2020). Interpretable Machine Learning.
13) Aggarwal, C. C. (2015). Data Mining: The Textbook.
14) Han, J., Kamber, M. & Pei, J. (2011). Data Mining: Concepts and Techniques.
15) Goodfellow, I., et al. (2017). Deep Learning (MIT Press).
16) Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. JMLR.

17) Bhattacharyya, D. et al. (2011). Credit Card Fraud Detection.
18) Carcillo, J., et al. (2020). Hybrid Fraud Detection.
19) IEEE Survey on Fraud Detection. (2022).
20) Microsoft Power BI Documentation.
21) Zhang, Y., et al. (2019). Using SHAP for Fraud Detection.
22) SEON Fraud Detection Report (2023).
23) Kaggle Credit Card Fraud Dataset.
24) Financial Cybercrime Report 2023.
25) RBI Digital Banking Fraud Report (2020).
26) OECD Financial Fraud Study (2021).
27) ACM Fraud Detection Framework (2021).
28) IEEE XAI Survey (2021).
29) Explainable Artificial Intelligence (XAI) Survey for Finance (2021).

