International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

Effective Credit Card Fraud Detection Using Machine Learning Algorithms

Sofiya Afreen Shaik[1], Dr. A. Damodaram[2]

[1]*M.Tech Scholar, Dept of CSE, JNTUH UCESTH, Hyderabad, India*
[2]*Senior Professor, Dept of CSE, JNTUH UCESTH, Hyderabad, India*

Abstract - *Financial transactions have changed as a result of the growing use of digital payment systems, which offer accessibility, speed, and convenience. However, this expansion has also led to an increase in fraudulent activity, posing serious problems for clients and financial institutions. Because fraudulent records only make up a very small percentage of all transactions, creating extremely unbalanced datasets, it is challenging to identify fraudulent credit card transactions. This study offers a machine learning-based system that uses Random Forest, XGBoost, CatBoost, and LightGBM algorithms to detect fraudulent credit card transactions. To enhance model performance, the suggested approach uses class balance, feature scaling, and data preparation. Ensemble learning systems successfully differentiate between fraudulent and legitimate transactions, according to experimental evaluation. The findings show that boosting-based techniques offer enhanced detection capabilities while preserving high. The findings show that boosting-based techniques maintain good classification accuracy while offering enhanced detection capabilities. The suggested methodology can help financial institutions lower financial losses brought on by fraudulent activity and improve transaction security.*

Key Words: Credit Card Fraud Detection, Machine Learning, Random Forest, XGBoost, CatBoost, LightGBM, Fraud Analytics, Classification

1. INTRODUCTION

The way financial transactions are carried out has been profoundly altered by digital transformation. Because of their ease of use and broad acceptance, credit cards have emerged as one of the most popular payment methods. Electronic transactions have significantly increased worldwide as a result of the growing acceptance of mobile payment apps, internet banking, and online shopping.

Even while digital payment systems have many benefits, the increasing volume of transactions has made fraud more likely. Unauthorized transactions made with stolen or compromised card information are referred to as credit card fraud. In addition to causing monetary losses, these occurrences erode consumer trust in electronic payment systems. Therefore, trustworthy systems that can spot suspicious transactions before serious harm is done are necessary for financial institutions.

Conventional fraud detection systems frequently rely on manual verification processes and pre-established rules. These methods are often unsuccessful against recently developed fraudulent schemes, even though they can identify recognized fraud tendencies. Static rule-based systems are hard to update and manage since fraudsters are always changing their tactics.

A potential remedy for these issues is machine learning. Machine learning algorithms can detect hidden trends and automatically discern between genuine and fraudulent activity by examining past transaction data. In contrast to traditional techniques, machine learning models can enhance their forecast ability through training and adjust to changing fraud behaviors.

This study examines how well ensemble learning algorithms such as Random Forest, XGBoost, CatBoost, and LightGBM detect credit card fraud. The main goal is to create an intelligent framework that can minimize false-positive classifications while effectively recognizing fraudulent transactions

2. LITERATURE SURVEY

Credit card fraud detection has been extensively studied due to the increasing reliance on electronic payment systems. Researchers have proposed various computational techniques to improve fraud identification and minimize financial losses.

Early fraud detection systems primarily employed statistical methods and rule-based approaches. Although these methods were relatively easy to implement, they lacked flexibility and adaptability. As fraud patterns became increasingly sophisticated, machine learning techniques emerged as more effective alternatives.

Random Forest has been widely adopted because of its capability to combine multiple decision trees and improve classification accuracy. Several studies reported that Random Forest successfully identifies fraudulent transactions while reducing overfitting issues.

XGBoost introduced an optimized gradient boosting framework capable of handling large-scale datasets efficiently. Research findings indicate that XGBoost consistently outperforms traditional classification algorithms

due to its strong predictive capabilities and robust feature learning mechanisms.

CatBoost further improved classification performance by introducing ordered boosting techniques and reducing prediction bias. The algorithm demonstrated competitive results in fraud detection tasks involving highly imbalanced datasets.

LightGBM has gained popularity because of its computational efficiency and high predictive performance. Its leaf-wise tree growth strategy enables faster training and improved scalability. Recent studies have shown that LightGBM achieves superior results when applied to large transaction datasets.

The findings from existing literature indicate that ensemble and boosting algorithms provide significant advantages in fraud detection applications. Motivated by these observations, this research evaluates and compares the performance of Random Forest, XGBoost, CatBoost, and LightGBM models.

3. PROBLEM STATEMENT

The increasing use of credit cards for online and offline transactions has led to a significant rise in fraudulent activities. Traditional fraud detection systems rely on predefined rules and manual monitoring techniques, which often fail to identify newly emerging fraud patterns. Furthermore, credit card transaction datasets are highly imbalanced because fraudulent transactions represent only a small percentage of the total records. This imbalance affects the performance of classification algorithms and increases the risk of misclassification. Therefore, there is a need for an intelligent fraud detection framework capable of accurately identifying fraudulent transactions while maintaining low false positive rates. The proposed study addresses this challenge by applying advanced machine learning algorithms to improve fraud detection accuracy and enhance financial transaction security.

4. PROPOSED METHODOLOGY

The proposed credit card fraud detection system utilizes machine learning techniques to accurately identify fraudulent transactions and minimize financial losses. The methodology begins with collecting and preprocessing transaction data, including data cleaning, handling missing values, feature scaling, and balancing the dataset to address class imbalance. After preprocessing, the dataset is divided into training and testing sets, and advanced machine learning algorithms such as Random Forest, XGBoost, CatBoost, and LightGBM are trained to learn transaction patterns and distinguish between legitimate and fraudulent activities. The trained models are then evaluated using performance metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC Score to determine the most effective model. Finally, the best-performing model is used to predict fraudulent transactions in real time, enabling financial institutions to detect suspicious activities quickly, improve transaction security, and reduce the risk of fraud.

4.1 System Architecture

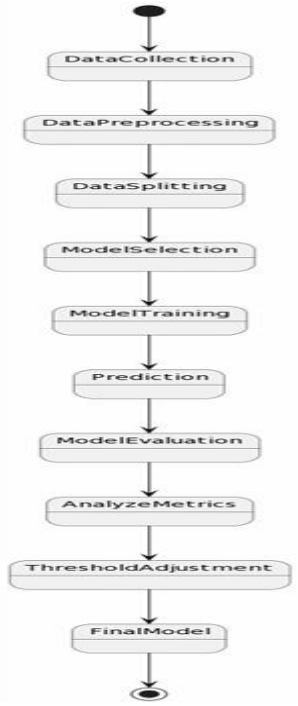

Fig-1 :System Architecture

4.2 Dataset Description

The dataset used in this study contains credit card transaction records collected from real-world financial activities. Each transaction is labeled as either legitimate or fraudulent. The dataset includes transaction time, transaction amount, transformed feature variables, and class labels.

A major challenge associated with this dataset is class imbalance. Fraudulent transactions constitute only a small percentage of the total records. Therefore, appropriate preprocessing techniques are required to ensure effective model learning.

4.3 Data Pre-processing

Data pre-processing plays a critical role in improving model performance. The dataset is examined for missing values, inconsistencies, and anomalies. Feature scaling is performed to normalize numerical attributes and ensure consistent value distributions. Data cleaning procedures improve the overall quality of the dataset and facilitate effective model training.

4.4 Class Balancing Using SMOTE

Since fraudulent transactions are significantly underrepresented, Synthetic Minority Oversampling Technique (SMOTE) is applied to balance the dataset. SMOTE generates synthetic samples for the minority class, enabling machine learning algorithms to learn fraud patterns more effectively and reducing classification bias.

4.5 Random Forest

Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and combines their predictions through majority voting. The algorithm improves classification accuracy while minimizing overfitting and variance.

4.6 XGBoost

XGBoost is a gradient boosting algorithm that sequentially builds decision trees to correct prediction errors from previous models. Its optimization capabilities and regularization mechanisms contribute to improved predictive performance and robustness.

4.7 CatBoost

CatBoost utilizes ordered boosting techniques to reduce prediction shift and enhance model stability. The algorithm effectively handles complex feature interactions and achieves high classification accuracy in fraud detection tasks.

4.8 LightGBM

LightGBM is a gradient boosting framework designed for speed and efficiency. By employing a leaf-wise growth strategy, it reduces computational complexity while maintaining excellent predictive performance on large datasets.

5. RESULT

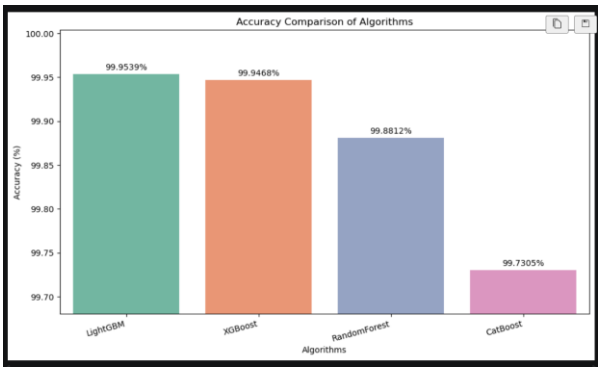

Fig -2: Accuracy Comparison of Machine Learning Models

Table 1: Accuracy

Algorithm	Accuracy (%)
Random Forest	99.8812
XG Boost	99.9468
Cat Boost	99.7305
Light GBM	99.9539

The accuracy comparison results indicate that all machine learning models achieved high fraud detection performance. Among the evaluated models, LightGBM achieved the highest accuracy, followed by XGBoost, Random Forest, and CatBoost.

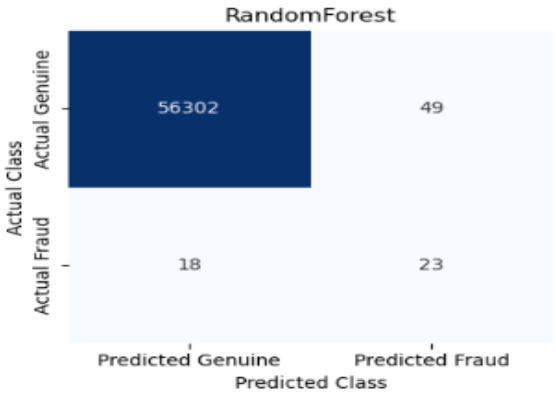

Fig-3: Confusion Matrix of Random Forest

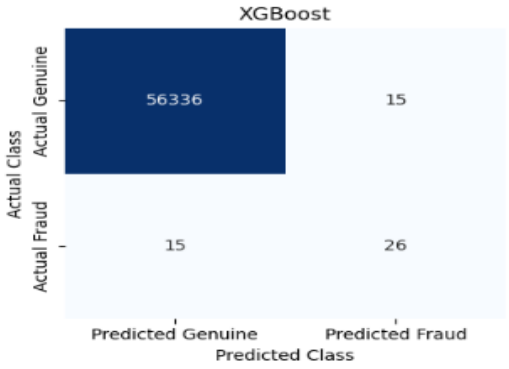

Fig-4: Confusion Matrix of XGBoost

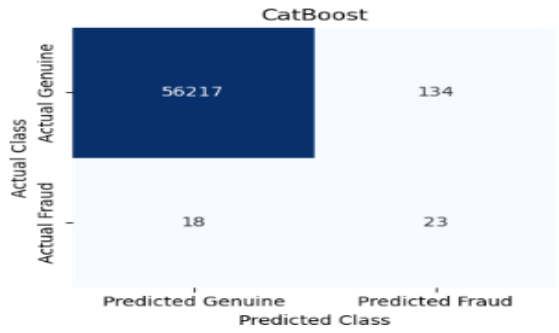

Fig-5: Confusion Matrix of CatBoost

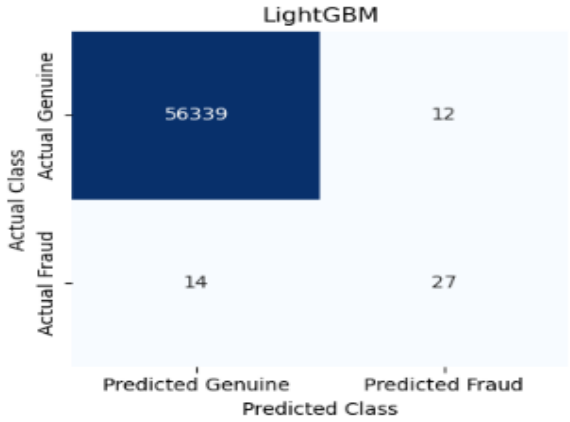

Fig -6: Confusion Matrix of LightGBM

The confusion matrices illustrate the classification performance of each model. The majority of transactions were correctly classified, indicating strong fraud detection capability. LightGBM demonstrated the highest number of correctly classified fraudulent transactions, resulting in superior overall performance.

6. CONCLUSION

This research presented a machine learning-based framework for credit card fraud detection using Random Forest, XGBoost, CatBoost, and LightGBM algorithms. Data preprocessing and SMOTE balancing techniques were employed to address dataset imbalance and improve classification performance. Experimental evaluation demonstrated that ensemble learning algorithms effectively distinguish fraudulent transactions from legitimate activities. Among the evaluated models, LightGBM achieved the highest performance and demonstrated superior capability in identifying fraudulent transactions. The proposed framework can assist financial institutions in strengthening transaction security and reducing financial losses. Future research may focus on integrating deep learning techniques and real-time fraud monitoring systems to further enhance detection performance.

REFERENCES

[1] Neha Ahirwar, Divakar Singh, and Kamini Maheshwar, "Efficient Credit Card Fraud Detection Based on Multiple ML Algorithms."

[2] Dingling Ge, Shunyu Chang, Jianyang Gu, and JingHui Cai, "Credit Card Fraud Detection Using LightGBM Model."

[3] Nghia Nguyen, Truc Duong, Tram Chau, Van-Ho Nguyen, Trang Trinh, Duy Tran, and Thanh Ho, "A Proposed Model for Card Fraud Detection Based on CatBoost and Deep Neural Network."

[4] Xinyi Hu, Haiwen Chen, and Ranxin Zhang, "Credit Card Fraud Detection Using LightGBM with Asymmetric Error Control," 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), 2019.

[5] Suni Jose, Deepa Devassy, and Anly Antony M, "Detection of Credit Card Fraud Using Resampling and Boosting Technique."