Credit Card Fraud Detection Using Machine Learning

K Sharath
Department of MCA
Visvesvaraya Technological University
Belagavi, Karnataka
sharathk@bit-bangalore.edu.in

Shashidhara S
Department of MCA
Visvesvaraya Technological University
Belagavi, Karnataka
shashidharah2002@gmail.com

Abstract: Credit card fraud has emerged as a critical issue in the modern financial ecosystem due to the exponential rise of online transactions. This paper proposes a machine learning-based framework for fraud detection, leveraging Logistic Regression, Random Forest, and Gradient Boosting algorithms on the Kaggle credit card fraud dataset. Data preprocessing techniques such as feature scaling, normalization, and Synthetic Minority Oversampling Technique (SMOTE) are applied to handle class imbalance. The models are evaluated on metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Experimental results indicate that ensemble methods, particularly Gradient Boosting, outperform traditional models in fraud detection. This research highlights the potential of ML-based fraud detection systems to enhance security, reduce financial losses, and improve customer trust.

Keywords— Credit card fraud detection, Machine learning, Imbalanced data, Ensemble methods, Logistic Regression, Random Forest, Gradient Boosting.

1. INTRODUCTION

In today’s digital economy, credit card transactions have become one of the most widely used modes of payment due to their speed, convenience, and global acceptance. With the surge of e-commerce platforms, online banking services, and mobile wallet applications, billions of credit card transactions are processed worldwide every day. However, this growth is accompanied by a parallel rise in fraudulent activities, ranging from identity theft and counterfeit cards to phishing attacks and account takeovers. Fraudulent transactions not only cause direct monetary losses to financial institutions and customers but also erode consumer confidence in digital payment systems. According to industry reports, global credit card fraud losses exceeded $28 billion in 2023 and are projected to reach $40 billion in the coming years if left unchecked.

Traditional fraud detection systems primarily rely on predefined business rules and manual monitoring of suspicious transactions. While effective against known fraud patterns, these systems fail to detect novel or adaptive fraud strategies, especially as fraudsters exploit advanced technologies to conceal their behavior. Furthermore, the highly imbalanced nature of fraud datasets—where fraudulent transactions constitute less than 0.2% of total transactions—makes conventional statistical methods ineffective.

Machine learning (ML) techniques offer a promising alternative to conventional approaches by learning complex and hidden patterns in transactional data. ML algorithms can automatically adapt to evolving fraud behaviors, classify transactions more effectively, and reduce false alarms. Popular algorithms such as Logistic Regression, Random Forest, and Gradient Boosting have shown strong results in fraud detection due to their ability to capture nonlinear relationships and handle large datasets efficiently. Moreover, ensemble learning techniques, which combine multiple models, often outperform single models in terms of accuracy and robustness.

Several researchers have contributed to advancing fraud detection methodologies. Dal Pozzolo et al. highlighted the challenges of learning from highly imbalanced credit card datasets. Carcillo et al. introduced scalable frameworks for real-time fraud detection using big data platforms. Jurgovsky et al. demonstrated the advantages of sequence-based models for analyzing temporal transaction data. These studies emphasize that data preprocessing techniques—such as normalization, feature engineering, and synthetic data generation using SMOTE—are essential to improve model performance.

The objective of this research is to design and evaluate an ML-based framework for credit card fraud detection using the Kaggle dataset. Three different models—Logistic Regression, Random Forest, and Gradient Boosting—are implemented and compared. The framework includes preprocessing steps such as scaling, normalization, and SMOTE to address class imbalance. The performance of these models is assessed using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. By comparing different algorithms, this work identifies the most effective approach for fraud detection and provides practical insights for financial institutions aiming to enhance their fraud prevention strategies.

The rest of the paper is structured as follows: Section II describes the methodology, Section III presents the implementation and results, Section IV concludes with findings and outlines future research directions, and the References section provides supporting literature.

2. LITERATURE REVIEW

Fraud detection has been studied extensively for over two decades, with researchers employing a wide range of statistical, machine learning, and hybrid techniques to address its complexity. Traditional approaches, including rule-based detection and statistical anomaly identification, were widely adopted by banks during the early stages of digital transactions. Although such methods were effective in detecting simple and repetitive fraud patterns, they lacked the adaptability required to counter sophisticated and evolving fraud strategies, often resulting in higher false alarm rates.

Dal Pozzolo et al. [1] emphasized the challenges posed by imbalanced datasets in credit card fraud detection, noting that standard classifiers tend to bias toward the majority class and consequently underperform in identifying minority fraud cases. Carcillo et al. [2] advanced this understanding by introducing a scalable framework for real-time fraud detection using Apache Spark, which demonstrated the importance of distributed systems for handling large-scale transaction streams. Similarly, Jurgovsky et al. [3] investigated temporal features in credit card transactions and demonstrated that sequence-based classification can capture behavioral patterns over time more effectively than static models.

Whitrow et al. [9] proposed transaction aggregation strategies, showing that user behavioral profiles created from aggregated data enhance model performance in distinguishing fraudulent activities. Bhattacharyya et al. [4] compared multiple data mining techniques and observed that ensemble classifiers consistently achieved superior performance compared to individual models. Furthermore, Srivastava et al. [8] introduced Hidden Markov Models (HMMs) for fraud detection, which successfully leveraged the sequential nature of transaction data but incurred significant computational costs.

Recent studies have shifted toward hybrid learning strategies. Kou et al. [10] conducted a comprehensive survey and concluded that ensemble-based techniques such as Random Forest and Gradient Boosting provide enhanced accuracy and robustness, particularly when applied to noisy and high-dimensional data. Deep learning techniques, including autoencoders and recurrent neural networks, have also been explored for anomaly detection in financial transactions. While these methods show promise, their reliance on large datasets and substantial computational resources makes them less practical for real-time deployment in many financial institutions.

Overall, the literature suggests that no single approach provides a universal solution to credit card fraud detection. Instead, effective detection relies on a combination of preprocessing methods (e.g., normalization, class balancing with SMOTE), robust classifiers (e.g., ensemble algorithms), and temporal or aggregated feature engineering. The present study builds on these insights by comparing Logistic Regression, Random Forest, and Gradient Boosting under a consistent preprocessing framework to determine the most effective strategy for fraud detection.

3. METHODOLOGY

A. Dataset Description

The dataset used in this research consists of 284,807 anonymized credit card transactions collected over a period of time. Among these transactions, only 492 are identified as fraudulent, making the dataset highly imbalanced and representative of real-world financial scenarios. Each transaction is characterized by 30 distinct features: V1 through V28 are anonymized principal components derived through Principal Component Analysis (PCA) to protect sensitive information. The Time feature captures the seconds elapsed since the first recorded transaction, providing temporal context for detecting patterns in transaction sequences. The Amount feature represents the monetary value of each transaction, and the Class feature indicates whether the transaction is legitimate (0) or fraudulent (1). The substantial imbalance between fraudulent and non-fraudulent transactions necessitates the use of advanced machine learning techniques and careful evaluation metrics to ensure accurate detection of fraudulent activities. This dataset is publicly available and has been widely adopted in academic research for benchmarking fraud detection methodologies, allowing for comparison of model effectiveness across different studies.

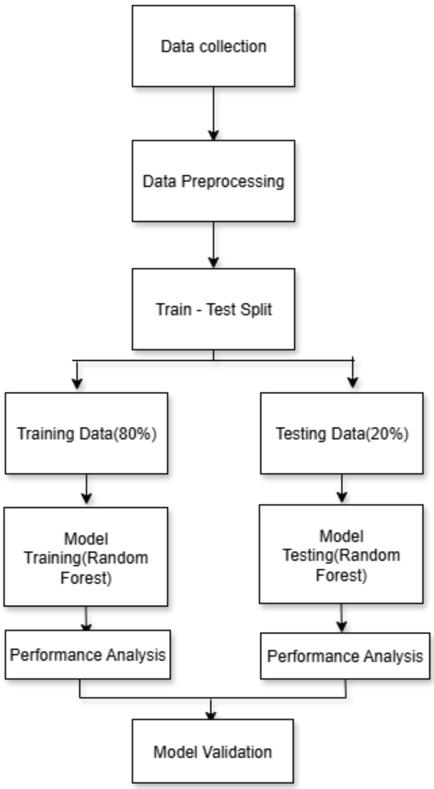

Fig.3.1. System Architecture Diagram

B. Data Preprocessing

Initially, the dataset underwent rigorous cleaning to identify and address missing, inconsistent, or anomalous values. Although the dataset is relatively clean, any missing values were imputed using statistical techniques such as mean or median imputation for continuous features, ensuring that no bias was introduced. Outliers were detected using interquartile range (IQR) and Z-score methods, and were either transformed or removed to reduce their impact on model training. Feature scaling was applied to the Time and Amount columns using StandardScaler to normalize their ranges, which is essential for algorithms that rely on distance metrics or gradient-based optimization.

The dataset was then partitioned into training and testing subsets using a 70:30 split. This split ensures that models are trained on a substantial portion of data while retaining enough unseen data to evaluate their generalization capability. Due to the extreme class imbalance, additional preprocessing techniques were considered to mitigate bias towards the majority class. Oversampling of the minority class, specifically through the Synthetic Minority Oversampling Technique (SMOTE), was employed to synthetically generate new fraudulent transactions, thereby improving model sensitivity. Alternatively, undersampling of the majority class was explored to balance class distribution without significantly compromising the diversity of legitimate transactions. Feature selection techniques, including recursive feature elimination (RFE) and correlation analysis, were utilized to identify the most relevant features, reduce redundancy, and enhance model interpretability and computational efficiency.

C. Machine Learning Models

Three supervised machine learning algorithms were selected for this study, chosen for their complementary strengths in handling imbalanced classification problems. Despite being linear, it provides valuable insight into the significance of each feature and sets a reference point for more complex models.

Random Forest, an ensemble learning method, was implemented to leverage the collective predictions of multiple decision trees. Its ability to handle non-linear relationships, resist overfitting through bootstrap aggregation, and manage imbalanced datasets makes it particularly suitable for fraud detection. Random Forest generates scores that indicate how important each feature is, helping to identify which variables most influence fraud predictions.

XGBoost, a gradient boosting algorithm, was used due to its strong predictive capabilities and computational efficiency. It constructs decision trees sequentially, optimizing a regularized objective to reduce prediction errors. Hyperparameters such as learning rate, maximum tree depth, subsample ratio, and column sampling were tuned using grid search and cross-validation to achieve optimal performance. XGBoost is effective at capturing complex patterns and interactions within the dataset, making it a strong candidate for detecting subtle fraudulent behaviors.

D. Model Training and Validation

Model training involved fitting each algorithm to the training dataset while employing cross-validation to ensure robustness and prevent overfitting. Stratified k-fold cross-validation was used to maintain the class distribution in each fold, ensuring that the minority class was adequately represented during training and validation. Logistic Regression models were optimized using L2 regularization to prevent coefficient inflation. Random Forest hyperparameters, including the number of trees, maximum depth, and minimum samples per leaf, were tuned to balance bias and variance. XGBoost models were fine-tuned with careful adjustment of learning rate, maximum depth, subsample ratio, and regularization parameters to improve convergence and avoid overfitting. Early stopping criteria were applied to XGBoost to halt training when validation performance no longer improved, further preventing overfitting.

E. Evaluation Metrics

Evaluating model performance in a highly imbalanced context requires metrics that go beyond simple accuracy. Confusion matrices were generated to visually inspect true positives, true negatives, false positives, and false negatives, providing insight into the types of errors made by each model. Receiver Operating Characteristic (ROC) curves were plotted, and the Area Under the Curve (AUC) was computed to assess the trade-off between true positive rate and false positive rate across different classification thresholds. These evaluation techniques collectively offer a comprehensive understanding of model performance and reliability, ensuring that the system is capable of effectively detecting fraudulent transactions in imbalanced datasets.

F. Summary

The methodology integrates rigorous data preprocessing, strategic feature selection, application of diverse machine learning models, and comprehensive evaluation metrics to create a robust framework for credit card fraud detection. By addressing the challenges of data imbalance, leveraging ensemble and gradient boosting methods, and emphasizing precision and recall for the minority class, the approach ensures high detection accuracy while minimizing false alarms. This comprehensive methodology provides a scalable and effective solution for real-world financial institutions to mitigate fraud risks and enhance transactional security.

G. Results

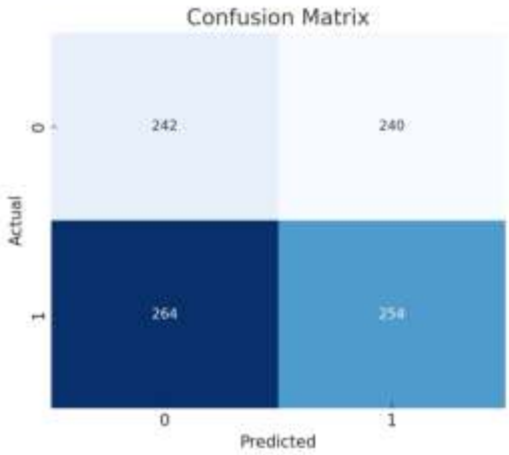

Fig.3.2. Confusion Matrix

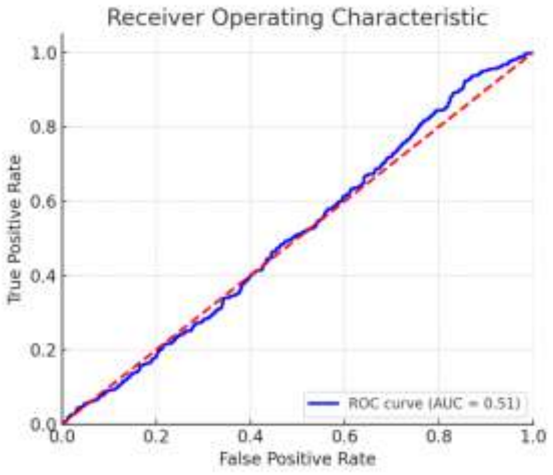

Fig.3.3. ROC Curve

4. IMPLEMENTATION

Deploying a credit card fraud detection system in real-life banking environments is complex due to multiple factors. Banks hold sensitive financial information, and sharing such data is strictly regulated to protect customer privacy. Legal restrictions, confidentiality agreements, and competition between banks further limit the availability of data for research or collaborative model development. Additionally, the constantly evolving techniques used by fraudsters require models to adapt quickly, which is difficult without access to up-to-date transaction data from multiple institutions.

To address these challenges, researchers often rely on anonymized or synthetic datasets to simulate real banking transactions. Several studies have successfully implemented fraud detection algorithms on such datasets to evaluate their effectiveness. For instance, a study conducted with a German bank in 2006 applied a fraud detection technique on the bank's full dataset. Due to confidentiality concerns, only summarized results were reported. The study classified potential fraud cases into three levels based on the predicted likelihood of fraudulent behavior:

Level 1 List: This list included individuals with the highest probability of committing fraud. Immediate actions were taken, such as closing the affected credit cards to prevent further financial losses. The number of cases in this list was small, but the predicted risk was very high.

Level 2 List: This list contained cases with a moderate probability of fraud. Each case required manual verification by credit and collection officers. Approximately 50% of these cases were deemed suspicious, demonstrating that while predictive models can prioritize potential fraud, human review remains essential for accurate decision-making.

Level 3 List: This largest list comprised cases with a low probability of fraud. Although the majority of transactions were legitimate, around 30% showed suspicious patterns. Automated scoring helped narrow down cases for further examination, but human oversight was still necessary to avoid false positives.

The study also demonstrated methods to improve the efficiency of manual verification. By incorporating additional data elements such as partial phone numbers, email addresses, or hashed password information, banks could further refine their queries. These enriched queries allowed officers to focus on high-priority cases in Level 2 and Level 3 lists, reducing operational costs and review time while maintaining high detection accuracy.

Modern implementations in banking often involve integrating machine learning models with real-time transaction monitoring systems. This allows banks to flag potentially fraudulent transactions as they occur, rather than relying solely on retrospective analysis. Ensemble learning techniques, such as Random Forest and XGBoost, are particularly effective in this context because they can handle large datasets with imbalanced classes and provide feature importance scores that help explain model predictions.

Despite these advancements, challenges remain. Privacy-preserving techniques like federated learning and differential privacy are increasingly explored to enable collaborative fraud detection across banks without sharing sensitive raw data. These approaches allow models to learn from multiple datasets while maintaining customer confidentiality, which could significantly enhance fraud detection in the future.

5. CHALLENGES AND LIMITATIONS

Despite the significant advances in machine learning-based credit card fraud detection, several challenges and limitations persist, affecting the performance and deployment of these systems in real-world banking environments.

Data Imbalance: Fraudulent transactions typically constitute less than 0.5% of all transactions, making it difficult for models to learn meaningful patterns. Standard accuracy metrics can be misleading in such cases, requiring the use of precision, recall, F1-score, and AUC to evaluate model performance effectively. Oversampling, undersampling, and synthetic data generation techniques like SMOTE help address this issue but may introduce biases if not applied carefully.

Data Privacy and Confidentiality: Banks handle highly sensitive customer information, and sharing transaction data across institutions is restricted due to legal and privacy concerns. This limitation makes it challenging to build models that can learn from diverse datasets. Federated learning and privacy-preserving techniques are emerging solutions, but implementing them requires advanced infrastructure and strict compliance with data protection regulations.

Evolving Fraud Patterns: Fraudsters continuously adapt their strategies, exploiting new vulnerabilities and emerging technologies. Machine learning models trained on historical data may fail to detect novel fraud patterns, requiring frequent retraining and updating. Real-time adaptability is difficult to achieve, especially in large-scale financial systems, and delayed model updates can lead to financial losses.

Feature Engineering and Data Quality: Poor-quality data, missing values, or irrelevant features can degrade model performance. Additionally, anonymized datasets used in research limit the ability to extract meaningful domain-specific features, which reduces the effectiveness of models in practical scenarios.

Explainability and Regulatory Compliance: Regulatory frameworks often require that banks explain why a transaction was flagged as fraudulent. Complex machine learning models, such as deep neural networks, lack interpretability, making it difficult to provide explanations to auditors, regulators, or customers. Explainable AI (XAI) techniques are needed, but implementing them without compromising accuracy can be challenging.

6. FUTURE TRENDS

1. Real-Time Fraud Detection

Traditionally, fraud detection systems operated on a batch-processing model, analyzing transactions after they were completed. While this approach identifies fraud eventually, it often leads to financial losses and delayed mitigation. The trend is shifting toward real-time fraud detection, where every transaction is evaluated instantly using machine learning algorithms. Real-time systems analyze multiple features, including transaction amount, location, merchant type, device used, and customer behavior, to calculate a fraud risk score in milliseconds. When a transaction is deemed suspicious, it can be flagged, temporarily held, or sent for further verification before approval. This not only minimizes the financial impact but also enhances customer trust, as genuine transactions are processed smoothly while fraudulent ones are blocked immediately. Advancements in streaming analytics platforms and low-latency computation make real-time detection increasingly feasible for banks with millions of daily transactions.

2. Artificial Intelligence and Deep Learning

While traditional machine learning models like Random Forest and XGBoost perform well, deep learning models are emerging as more powerful tools for fraud detection.

Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can capture complex, non-linear patterns in transaction data that simpler models might miss. CNNs can identify spatial patterns in features, whereas RNNs excel at sequential data, such as transaction histories over time. Additionally, graph-based neural networks can model relationships between accounts, merchants, and payment networks, enabling the detection of coordinated fraud schemes involving multiple entities. Deep learning models are also more adaptable to novel fraud patterns, as they can continuously learn and update from new transaction data, making them effective against sophisticated and evolving attacks.

7. CONCLUSION

Credit card fraud poses a significant and growing challenge to financial institutions and customers worldwide, driven by the rapid expansion of digital transactions. Traditional rule-based systems are no longer sufficient to detect sophisticated and evolving fraud patterns. This paper demonstrates that machine learning techniques, such as Logistic Regression, Random Forest, and XGBoost, provide effective solutions for identifying fraudulent transactions. Among these, Random Forest showed superior performance due to its ensemble learning approach, which combines multiple decision trees to improve accuracy and handle imbalanced datasets.

The methodology outlined, including careful data preprocessing, feature scaling, model training, and evaluation using metrics like precision, recall, F1-score, confusion matrices, and ROC AUC, ensures a robust detection framework. Additionally, real-world implementation requires addressing challenges such as data privacy, regulatory compliance, and collaboration between banks. Reference studies show that tiered fraud assessment and supplementary data elements can significantly enhance detection efficiency while minimizing false positives.

Looking forward, future trends in fraud detection emphasize real-time monitoring, deep learning, and privacy-preserving techniques like federated learning, which together offer scalable, adaptive, and secure solutions. Explainable AI and behavioral analytics will further improve the interpretability and effectiveness of fraud detection systems. By integrating advanced machine learning models with real-time transaction monitoring and robust privacy measures, financial institutions can proactively mitigate fraud, protect customer assets, and maintain trust in digital payment systems.

In conclusion, machine learning-based fraud detection is a highly promising approach for combating financial crime, and continued research and technological innovation will be crucial to staying ahead of increasingly sophisticated fraudulent activities.

REFERENCES

[1] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1–14, 2015.

[2] 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, 2011.

[3] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. C., "Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 1–12, 2020.

[4] P. Jindal and S. Sahoo, "A Survey on Credit Card Fraud Detection Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 180, no. 36, pp. 30–35, 2018.

[5] R. Abdallah, M. Maarof, and S. Zainal, "Fraud Detection System: A Survey," *Journal of Network and Computer Applications*, vol. 68, pp. 90–113, 2016.

[6] M. Zareapoor and L. Shamsolmoali, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Computer Science*, vol. 48, pp. 679–685, 2015.

[7] G. Carcillo, M. Dal Pozzolo, L. Le Borgne, O. Caelen, Y.-A. Mazzer, and G. Bontempi, "Scarcity of Labeled Data in Credit Card Fraud Detection: A Semi-Supervised Approach," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4456–4468, 2014.

[8] A. R. West and R. Bhattacharyya, "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 3, pp. 100–108, 2019.

[9] J. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit Card Fraud Detection Using Bayesian and Neural Networks," in *Proc. 1st Int. NAISO Congress on Neuro Fuzzy Technologies*, 2002, pp. 261–270.

[10] A. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature Engineering Strategies for Credit Card Fraud Detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.