CREDIble-A smart model to train, detect and prevent credit card fraud and provide visual insights

Arpudha Soundararajan[1], Sivasangari S[2]

Department of Information Science and Technology – College of Engineering
Guindy, Anna University, India, Tamil Nadu, Chennai – 600025.
*Corresponding author's email: arpudha.s@gmail.com

Abstract

The Credit Card Fraud Detection system represents a significant advancement in financial security technology, offering an innovative solution to the growing challenges of fraud in digital transactions. This intelligent system integrates machine learning algorithms with real-time data analytics to revolutionize fraud detection practices. At its core, the system utilizes advanced models such as XGBoost, Random Forest, and Isolation Forest to analyze transaction patterns and identify anomalies indicative of fraudulent activity. Its strength lies in its ability to process vast amounts of transaction data in real-time, enabling proactive detection and prevention of fraud. By leveraging predictive analytics, the system evaluates parameters such as transaction amount, location, frequency, and user behavior to make informed decisions about potential threats. This data-driven approach ensures minimal false positives while maintaining high detection accuracy, thereby enhancing trust and reliability in financial systems. The system's user-friendly interface, including dashboards and mobile applications, empowers financial institutions and users with real-time monitoring and actionable insights. Initial trials have demonstrated promising results, showcasing significant reductions in financial losses due to fraud while improving operational efficiency. By addressing the dual challenges of fraud prevention and transaction security, the Credit Card Fraud Detection system emerges as a timely and practical solution for safeguarding digital payments. As global concerns about cybersecurity and financial integrity continue to grow, this system paves the way for more secure and technologically advanced methods of protecting sensitive financial ecosystems.

Keywords: *fraud detection, machine learning, credit card security , real-time analytics, financial fraud*

1 Introduction

In the rapidly evolving landscape of digital finance, credit card transactions have become a cornerstone of modern commerce, offering convenience, speed, and accessibility to consumers worldwide. However, this widespread adoption of electronic payments has also given rise to an alarming increase in fraudulent activities. Credit card fraud, which encompasses unauthorized transactions, identity theft, and account takeovers, poses a significant threat to both financial institutions and consumers. According to industry reports, billions of dollars are lost annually due to fraudulent activities, underscoring the urgent need for robust detection mechanisms. As fraudsters continually adapt their tactics, leveraging sophisticated techniques such as synthetic identities and phishing schemes, traditional rule-based systems have proven insufficient in addressing the dynamic nature of these threats. This has paved the way for advanced technologies like machine learning and artificial intelligence to play a pivotal role in combating credit card fraud. Machine learning, with its ability to analyze vast amounts of transaction data and identify intricate patterns, has emerged as a powerful tool in fraud detection. Unlike conventional methods that rely on predefined rules, machine learning models can learn from historical data to detect anomalies and predict potential fraudulent activities. Techniques such as supervised learning, unsupervised learning, and hybrid approaches have demonstrated remarkable success in distinguishing legitimate transactions from fraudulent ones. For instance, ensemble methods like Random Forest and XGBoost excel in handling imbalanced datasets, where fraudulent transactions constitute a small fraction of the overall data. Similarly, unsupervised algorithms like Isolation Forest are adept at identifying outliers, making them valuable for detecting previously unseen fraud patterns. By integrating these advanced models into fraud detection systems, financial institutions can achieve higher accuracy, reduce false positives, and respond to threats in real-time. The importance of timely and accurate fraud detection cannot be overstated. Fraudulent transactions not only result in direct financial losses but also erode consumer trust, damage institutional reputations, and impose regulatory penalties. Moreover, the growing adoption of e-commerce, mobile banking, and contactless payments has expanded the attack surface for fraudsters, further complicating the detection process. In response, researchers and practitioners have turned to innovative solutions that leverage real-time analytics, behavioral biometrics, and anomaly detection to stay ahead of emerging threats. These systems continuously monitor transactional behavior, flagging suspicious activities based on deviations from established patterns. For example, unusual spending habits, irregular transaction locations, or abnormal purchase frequencies can trigger alerts for further investigation. By combining these capabilities with user-friendly interfaces, fraud detection systems empower both financial institutions and consumers to take proactive measures against fraud. This research focuses on developing a comprehensive credit card fraud detection system that integrates cutting-edge machine learning techniques with real-world applicability. The primary objective is to design a model capable of accurately identifying fraudulent transactions while minimizing false positives, ensuring a seamless user experience. To achieve this, we explore various machine learning algorithms, evaluate their performance using metrics such as precision, recall, F1-score, and ROC-AUC, and identify the most effective approach for real-world deployment. Additionally, we emphasize the importance of addressing challenges such as class imbalance, evolving fraud patterns, and scalability to ensure the system's long-term viability. By leveraging real-time data analysis and predictive modeling, our proposed solution aims to enhance financial security, reduce operational costs, and foster trust in digital payment ecosystems. Credit card fraud detection represents a critical frontier in the fight against financial crime. As technology continues to reshape the financial landscape, the integration of machine learning and advanced analytics offers a transformative approach to safeguarding sensitive transactions. This study contributes to this ongoing effort by exploring innovative methodologies, addressing existing limitations, and paving the way for more secure and resilient financial systems. By bridging the gap our work underscores potential of intelligent systems to combat fraud.

2 Related Work

Credit card fraud detection is a critical challenge in the financial system due to the increasedvolume of transactions and the complexity of fraud. Machine learning has been developed as a powerful tool to treat this problem by identifying patterns and anomalies in transactional data [1]. The problem is naturally imbalanced because fraudulent transactions represent small parts of the data set that require special techniques to effectively deal with class imbalances [2]. For example, Awoyemi et al. [1] highlighted the importance of preprocessing procedures such as normalization and feature selection to improve model output .Similarly, Dastidar et al. [5] provided a comprehensive overview of machine learning methods as it is necessary to highlight the need for robust evaluation metrics to assess classification performance of unbalanced datasets.

Several algorithms for machine learning have been used to recognize credit card fraud, each with its strengths and limitations:

Logistic regression:
Logistic regression is often used in binary classification tasks due to its simplicity and interpretability. However, it may be difficult to record complex relationships in a high-dimensional dataset [1]. Adepoju et al. [3] showed that logistic regression combined with feature engineering techniques works reasonably well, but often surpasses more sophisticated models.

Random Forest:
Thennakoon et al. [2] used random forests to recognize fraud and reported high levels of accuracy and robustness for overly adaptation. S. Xuan et al .[10] examined its effectiveness in separating fraudulent transactions and erroneously maintaining low-false positive rates.

Xgboost:
Prajapati et al. [4] found that it is very effective when recognizing fraudulent transactions with minimal false positives compared to Xgboost with other classifiers. Ahmed and Saini [12] highlighted the importance of hyperparameter tuning in maximizing Xgboost performance.

Isolation Forest:
Hernandez Aros et al. [8] checked various machine learning techniques and determined the efficiency of insulated forests when identifying outliers. It is effective, but does not always outperform the methods of learning highly disproportionate datasets [5]. Our study was built on these previous work b y using logistic regression, random forests, xg boost, and isolated forests in the fraud test.csv dataset, and their performance was evaluated using ROC-AUC values. The result was that Xgboost achieved the highest ROC-AUC score, consistent with findings of Prajapati et al. [4] and Ahmed and Saini [12]. Random Forest also shows competitive perfor mance, and Thennakoon et al. [2] and S. Xuan et al. [10]. Logistic regression and Isolation Forest are effective , but are not very suitable for the complexity of data records.

This approach provided a comprehensive comparison of both supervised and unsupervised technologies dealing with gaps identifiedby Mutemi and Bacao [6], highlighting the need for a hybrid model that combines several approaches. Despite significant advances, challenges remain in the realization of credit card fraud. Class imbalance, evolving fraud patterns, and the need for real-time detection are persistent issues [8]. Sadineni [7] highlighted the importance of adaptive models that can be continuously learned from new data, but E. Btoushet al. [11] addressed these challenges, examined a variety of fraud identification techniques and highlighted the possibilities of deep learning and hybrid models. Our work deals with some of these gaps using advanced algorithms such as Xgboost and Isolation Forest. However, future research could investigate deep learning approaches or hybrid models to further improve performance.

The novelty of our work is a systematic evaluation of several machine learning technologies and provides a clear benchmark for future research. By including data visualization and strict assessment metrics, it contributes to an increase in the literature for recognizing credit card fraud. Furthermore, the use of the ROC-AUC curve is consistent with Hernandez Aros et al. [8] ensuring reliable comparison of model output.

3 System Design

System Flowchart

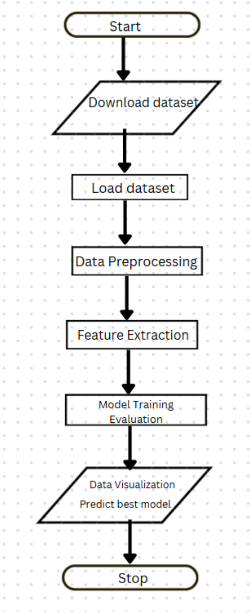


Figure 1: System Flowchart

Figure 1 illustrates the overall flow of the system

4 Implementation Details

Dataset Preparation:
Fraudulent data records are pre-processed to address the lack of value, normalize features, and exceed class levels using techniques such as synthetic minority over sampling technique (SMOTE) [1].

Feature Engineering :
Relevant features were extracted and scaled using standardization

Model Training and Evaluation:
Logistic Regression, Random Forest, XGBoost, Isolation Forest were trained with pre-processed data records . A layered K specialist cross-validation approach was used to ensure robust evaluation [2].

These procedures ensured that the dataset was ready for training and that the model was highly evaluated under consistent conditions.

Technical Aspects:
Programming Language: Python has been selected as the main programming language because of extensive ecosystem and simple application of machine learning libraries and pre-processing steps.

XGBoost: Used for gradient boosting with hyperparameter tuning via grid search.
Matplotlib and Seaborn : Used for data visualization and plotting ROC-AUC curves.
PyOD : Used for implementation of Isolation Forest, a special anomaly detection algorithm.

This section describes the tools and components used to create the system:

Software Environment: The solution was developed in the Jupyter notebook environment in Python 3.12. Dependencies were managed using pip and requirements.txt. Additionally Flask was considered as an optional framework for service prediction via APIs.
These details provide a clear roadmap for replication of the solution or deploying it in real-world scenarios.

5 Results and Discussion

5.1 Spending Distribution Category Analysis

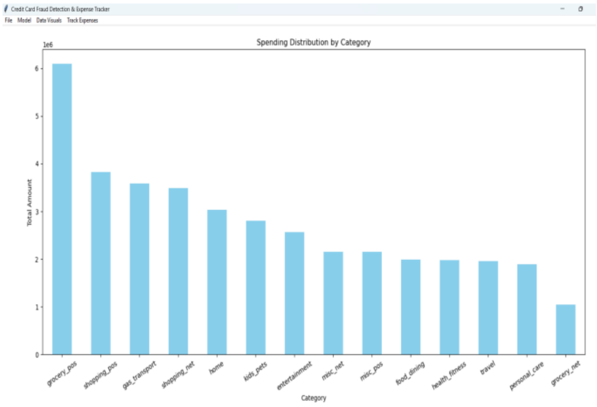


Figure 2: Spending Distribution Analysis

Figure 2 shows the distribution of costs across different categories in a dataset of credit card transactions. The chart shows the dominant role of categories such as grocery_pos and shopping_pos, while simulaneously recording the contributions of other categories like home, entertainment, and travel.

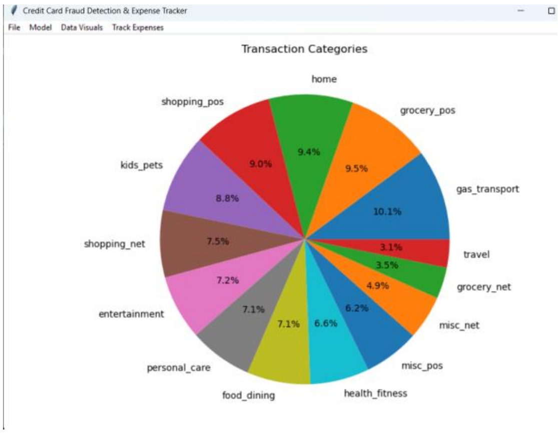


Figure 3: Transaction Categories Distribution

Figure 3 shows a cycle diagram showing the distribution of transaction categories as a percentage of total transactions in credit card data. Each slice represents a specific output category, and the label specifies the corresponding percentage of the total transaction.

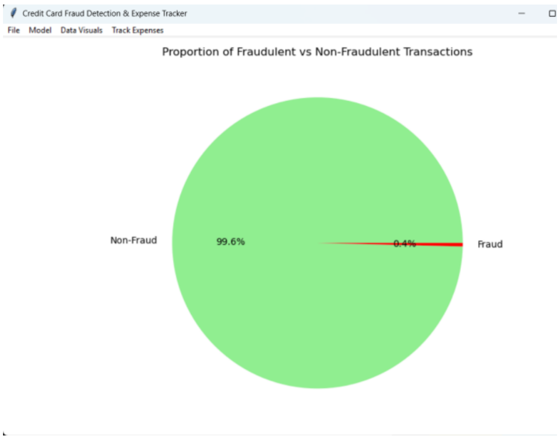

Credit Card Fraud Detection & Expense Tracker
File
Model
Data Visuals
Track Expenses
Proportion of Fraudulent vs Non-Fraudulent Transactions
Non-Fraud
99.6%
0.4%
Fraud

Figure 4: Fraud and Non-Fraud Transactions

Figure 4 shows a pie chart depicting the distribution of fraudulent and non-fraudulent transactions in a credit card dataset. The chart shows a significant class imbalance, with non-fraudulent transactions comprising 99.6% of the total, while fraudulent transactions account for only 0.4% . This visualization highlights the challenge of detecting rare fraudulent activities in highly skewed datasets.

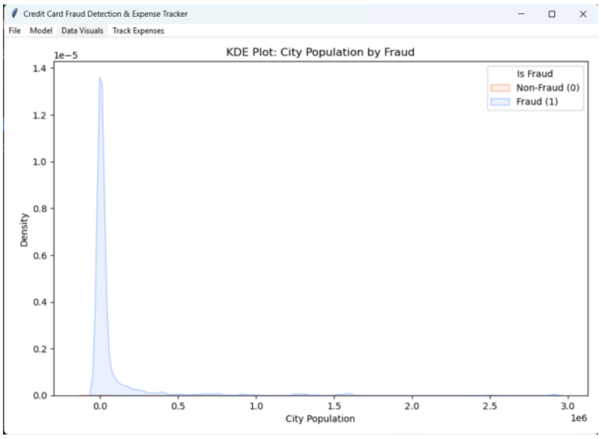


Figure 5: KDE Plot: Fraud Population

Figure 5 depicts a Kernel Density Estimation (KDE) plot titled "KDE Plot: City Population by Fraud." This visualization compares the distribution of city populations for two categories: non-fraud transactions (labeled as "0") and fraud transactions (labeled as "1").

5.2 Analysing Best Model Performance

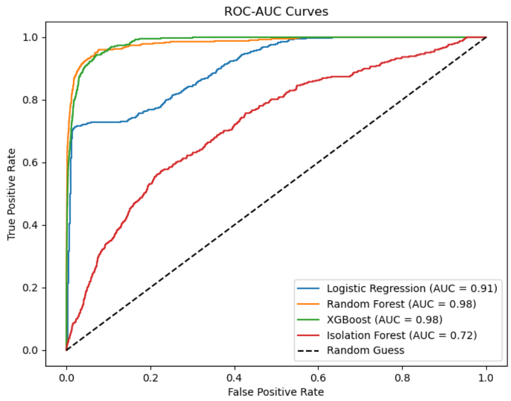
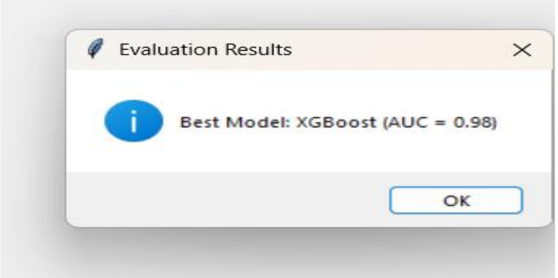


Figure 6: ROC-AUC Curve

Figure 6 shows ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) curves for four machine learning models: Logistic Regression, Random Forest, XGBoost, and Isolation Forest. The ROC-AUC curve is a standard metric used to evaluate the performance of binary classification models, particularly in imbalanced datasets like fraud detection. X-axis indicates the False Positive Rate (FPR) which measures the proportion of negative instances correctly classified as positive. Y-axis indicates the True Positive Rate (TPR) , also called recall or sensitivity which measures the proportion of positive instances correctly classified as positive.

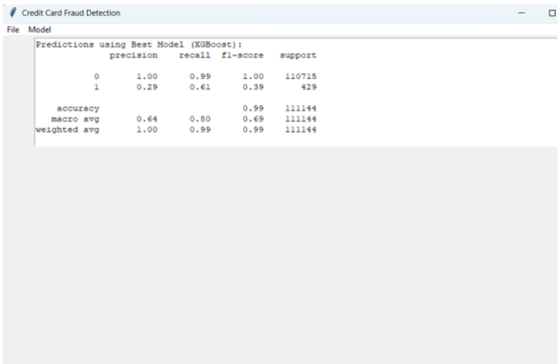


Best Model: XGBoost (AUC = 0.98)

Figure 7: Best Model - XGBoost

Figure 7 depicts that XGBoost has achieved high performance compared to other three models namely Logistic Regression, Random Forest and Isolation Forest. The performance was evaluated using AUC (Area Under the Curve) with a score of 0.98, indicating strong predictive capabilities.

5.3 Prediction Report of Best Model



```
Predictions using Best Model (XGBoost):
              precision    recall  f1-score   support

           0       1.00      0.99      1.00    110715
           1       0.29      0.61      0.39       429

    accuracy                           0.99    111144
   macro avg       0.64      0.80      0.69    111144
weighted avg       1.00      0.99      0.99    111144
```

Figure 8: Best Model Prediction Report

Figure 8 shows a classification report for a machine learning model, generated using Python's sklearn.metrics.classification_report function. This prediction report provides metrics for evaluating the performance of a binary classification model. The model performs exceptionally well on the majority class for non-fraud transactions, with near-perfect precision and recall. The high weighted average metrics indicate that the model effectively handles the imbalanced dataset while maintaining good performance across both classes.

6 Conclusion

The Credit Card Fraud Detection project showcases the transformative potential of integrating machine learning and advanced analytics to enhance financial security. By leveraging real-time data analysis, predictive modeling, and automated decision-making, this system addresses the persistent challenges of fraud detection in the rapidly evolving digital payment landscape. The project demonstrates how cutting-edge technology can protect consumers and financial institutions from fraudulent activities while ensuring efficient and secure transactions.

Key achievements of the project include:

- Development of robust Machine Learning model
- Real time monitoring and alert system.
- Implemented graphical user-friendly interfaces for enhanced accessibility
- Ensures improved Accuracy and reduced False Positives
- Supports safer transactions , reduces financial losses and fosters trust among customers and businesses.

The success of the Credit Card Fraud Detection system highlights the most important role of advanced analytics in safeguarding financial systems. By using machine learning to analyze vast amounts of transaction data, the system not only addresses immediate concerns of fraud prevention but also sets a foundation for more secure financial practices. This approach ensures that financial institutions can adapt to emerging threats and maintain robust protection mechanisms. In conclusion, this system represents a significant leap towards addressing the growing threat of financial fraud. As we continue to refine and expand this technology, we envision a future where fraud detection systems play a vital role in protecting global ecosystems ensuring secure and reliable transactions. This system will remain at the forefront of securing financial integrity and fostering sustainable growth in the digital economy.

7 Future Works

Looking forward, the Credit Card Fraud Detection system presents numerous opportunities for further development and broader application:

- Integration of Behavioral Biometrics: Enhancing system by incorporating biometrics like device usage patterns, to provide authentication.
- Adaptive Learning for Evolving Threats.
- Collaboration with financial experts: Establishing partnerships with industry to refine system requirements and use cases.
- Designing a scalable infrastructure capable of handling large volumes of transactions.
- Enhanced User Privacy and compliance.
- Multi- Channel Fraud Detection: Enabling system to monitor fraud across multiple channels like mobile transactions and online payments.

8 References

J.O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI) , Lagos, Nigeria, 2017.

A. Thennakoon, C. Bhagyani, S. Premadasa, S. Mihiranga and N. Kuruwitaarachchi, "Real-time Credit Card Fraud Detection Using Machine Learning," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence) , Noida, India, 2019.

O.Adepoju, J. Wosowei, S. lawte and H. Jaiman, "Comparative Evaluation of Credit Card Fraud Detection Using Machine Learning Techniques," 2019 Global Conference for Advancement in Technology (GCAT) , Bangalore, India, 2019.

Prajapati, A. Tripathi, J. Mehta, K. Jhaveri and V. Kelkar, "Credit Card Fraud Detection Using Machine Learning," 2021 International Conference on Advances in Computing, Communication, and

Control (ICAC3) , Mumbai, India, 2021.

K. G. Dastidar, O. Caelen and M. Granitzer, "Machine Learning Methods for Credit Card Fraud Detection A Survey," in IEEE Access.

A. Mutemi and F. Bacao, "E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review," in Big Data Mining and Analytics .

P. K. Sadineni, "Detection of Fraudulent Transactions in Credit Card using Machine Learning Algorithms," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) , Palladam, India.

Hernandez Aros, L., Bustamante Molano, L.X., Gutierrez-Portela, F. et al. Financial fraud detection through the application of machine learning techniques: a literature review. Humanit Soc Sci Commun 11, 1130 (2024)

Ogundunmade, T. P., & Adepoju, A. A. (2024). Modelling Credit Card Fraud Data using Machine Learning Algorithms. International Journal on Computational Engineering .

S. Xuan, G. Liu, Z. Li, L. Zheng, S. Wang and C. Jiang, "Random forest for credit card fraud detection," 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) , Zhuhai, China, 2018.

E. Btoush, X. Zhou, R. Gururaian, K. Chan and X. Tao, "A Survey on Credit Card Fraud Detection Techniques in Banking Industry for Cyber Security," 2021 8th International Conference on Behavioral and Social Computing (BESC) , Doha, Qatar, 2021.

A. N. Ahmed and R. Saini, "Detection of Credit Card Fraudulent Transactions Utilizing Machine Learning Algorithms," 2023 2nd International Conference for Innovation in Technology (INOCON) , Bangalore, India, 2023.