International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-779

FRAUD DETECTION SYSTEM FOR BANKING TRANSACTIONS USING MACHINE LEARNING ALGORITHMS

Pavithra Sri S[1], Sara Maria Priscilla A[2], Sunmitha S J[3] ,Steephan Amalraj J[4]

[1] *Student, Dept. of Computer Science and Engineering, Bannari Amman Institute of Technology, IN*
[2] *Student, Dept. of Computer Science and Engineering, Bannari Amman Institute of Technology, IN*
[3] *Student, Dept. of Computer Science and Engineering, Bannari Amman Institute of Technology, IN*
[4] *Professor, Dept. of Computer Science Engineering, Bannari Amman Institute of Technology, IN*

---***---

Abstract - *The increased digitization of credit card transactions has led to significant growth in fraudulent activities within the banking sector, creating a pressing need for advanced fraud detection mechanisms.Conventional rule-based methods have become insufficient in handling sophisticated and changing fraud schemes as these methods are static and not dynamic. This project will create a sophisticated fraud detection system focused on credit card transactions, utilizing machine learning algorithms for handling high-volume, real-time transaction data. The proposed system will integrate the use of machine learning algorithms such as Random Forest, XGBoost, and Decision Trees in a robust, fault-tolerant framework. The architecture dynamically switches between algorithms to ensure optimal performance, enhance resilience, and improve accuracy in fraud detection. To address the class imbalance issue that is generally a characteristic of fraud detection data, where legitimate transactions significantly outpace fraudulent ones, this system uses SMOTE to synthetically create samples of fraudulent transactions, thereby improving its ability to detect anomalies. The tasks involved in data preprocessing of this system include missing value imputation, duplicate removal, and data normalization, making the input data clean and suitable for training high-performance machine learning models. Furthermore, feature selection is also used to narrow down the database by picking the most salient transaction attributes like amount, location, and frequency that distinguish between legitimate credit card transactions and fraudulent credit card transactions.*

Keywords: Credit card fraud detection, Machine learning, XGBoost, Decision Tree, Random Forest, SMOTE.

1. INTRODUCTION

Online banking has revolutionized the financial market dramatically by the mushrooming of digital transactions, especially payment through credit cards. Credit card transactions have evolved into everyday commerce since they are convenient and fast. Millions of transactions take place each day, and the use of credit cards for instant payment has increased exponentially. While credit card systems have changed the face of the global financial transaction world, significant security concerns have also emanated from them since fraudulent activities exploit vulnerabilities in those systems. Increased transaction volumes and values have made credit cards vulnerable to a host of fraudulent activities, such as access by unauthorized persons, identity thefts, and phishing attacks, that often incur huge financial losses. Apart from the monetary impact, fraudulent transactions damage the reputation of financial institutions, erode customer trust, and cause long-term reputational harm.

As fraud tactics become evermore sophisticated, traditional rule-based methods of fraud detection have not been delivered. These static systems do not keep pace with these developments in fraud schemes, but modern fraud schemes require the involvement of adaptive, real-time approaches to detection. Machine learning can be considered a highly potent approach in the quest for fraud, allowing for enhanced mechanisms of fraud detection that can be both pattern-sensitive and new pattern-adaptive, in tandem with immediate responses to suspicious activity. By having the capability to harness machine learning, financial institutions can proactively prevent credit card fraud and provide a more secure environment for digital transactions.

1.1 Resilient Autonomy

Credit card fraud detection has gained major improvements in the past few years, but there are still a few challenges that are very strong, especially in accuracy and reliability in identifying fraudulent transactions against huge volumes of transactions. Most traditional approaches suffer from low adaptability with the changing fraud tactics and difficulties when dealing with imbalanced data, which can cause failure in detecting fraudulent transactions or true transactions as fraudulent. This would result in huge financial loss and reputational damage when failure occurs in even one of the algorithms in banking environments that host high-frequency and high-value transactions.

Second, with credit card transactions, this often represents a unique challenge in relation to diverse spending patterns, different user behavior and the requirement to account for differences that exist between regions in regard to fraud schemes. These often pose the need to produce fraud detection systems that are adaptive and resilient to new as well as sophisticated tactics. On meeting these challenges, the system ensures that fraud detection is made flawless while financial institutions and customers stay comfortable with the trust and security measures afforded by the same.

1.1 Background of the Work

Machine learning algorithms have the unique capability to identify subtle patterns within vast amounts of historical transaction data, learning to distinguish between legitimate and fraudulent transactions. This adaptability enables ML-based fraud detection systems to remain effective even as fraud tactics evolve, surpassing the capabilities of traditional rule-based systems, which often struggle to keep up with new fraud patterns. With the rapid growth in digital transactions, especially through Credit Card (CC) systems and other online banking formats, the demand for intelligent, adaptive fraud detection mechanisms has become crucial.The objective of this project is to develop a fraud detection system that analyzes transaction data using machine learning to identify anomalies, enabling the classification of transactions as fraudulent or non-fraudulent. This system is specifically tailored for credit card transactions, given their widespread use and the high frequency of associated fraud cases. The increased adoption of credit cards has made them a convenient yet vulnerable target for fraud, underscoring the need for a robust detection mechanism. By focusing on credit card fraud, this system addresses a significant area of concern for financial institutions and offers insights that may be applicable to other transaction types where similar fraud tactics are employed. Upon analysis, the system generates a comprehensive report detailing the nature of detected fraudulent activities. This report includes the count and percentage of flagged transactions and provides a breakdown of fraud patterns, such as unusual transaction amounts, out-of-pattern spending, or high-frequency transactions within short timeframes. This information enables financial institutions to identify emerging fraud trends, empowering them to act proactively in securing their systems and protecting customer assets. Overall, the proposed fraud detection system is both scalable and high-performing, meeting the modern demands of financial institutions. It enhances fraud detection capabilities with machine learning-driven real-time analysis, addresses data imbalances using SMOTE, and supports adaptive learning to stay resilient against evolving fraud tactics. This system offers banks a powerful tool to reduce financial losses and bolster customer trust in their digital services. Unlike traditional rule-based systems, machine learning provides unparalleled flexibility and adaptability. Rule-based systems rely on predefined criteria to flag suspicious transactions, such as high-value amounts or activity from high-risk locations, which limits their ability to adapt to new fraud patterns. In contrast, machine learning models continually learn from data, adapting to the evolving nature of fraud.

2. METHODOLOGY

This chapter outlines the pipeline structure that integrates machine learning algorithms in a hierarchical framework to ensure redundancy and reliability for credit card fraud detection. It processes transaction data from diverse sources, beginning with risk evaluation and mitigation. Data preprocessing includes cleaning, feature selection, and handling class imbalance through techniques like SMOTE to ensure accurate fraud detection. Fraud identification dynamically employs machine learning models such as Random Forest and XGBoost, depending on dataset complexity and performance requirements. These models are optimized to detect anomalies and classify transactions as fraudulent or non-fraudulent effectively. This multi-layered design enhances the system's robustness, maintaining operational accuracy under various transaction scenarios.

SI. No	Feature	Description	Benefits
1.	Analysis of Historical Data	Analyzes past transaction patterns to identify trends and anomalies.	Provides insights into emerging fraud tactics, improving the system's adaptability over time.
2.	User-Friend Reporting	Generates detailed reports in PDF or Excel formats.	Assists analysts with clear, actionable insights for fraud prevention and response.
3.	Risk Scoring and Alerts	Assigns risk scores to flagged transactions and generates alerts.	Prioritizes high-risk cases for review, enabling efficient resource allocation for investigation.

Table -1: Features of the fraud detection

2.1 Feature selection from Multiple attributes

Feature selection is one of the major steps in designing a fraud detection machine learning algorithm. This is because to maximize model performance, effectiveness and interpretability, it consists of finding out the most relevant feature out of the available attributes from a set of attributes. The following attributes can be used to spot a fraudulent transaction: Amount, time of transaction, number of transactions, location, device and/or browser information. Other irrelevant information is filtered out to improve accuracy and reduce complexity. The method improves computational efficiency and accuracy in the detection of fraudulent transactions by selecting the most useful attributes. This stage is critical in the development of strong, trustworthy models that can handle the complexities of real fraud scenarios.

2.2 Handling Imbalance data

Especially in banking transactions, fraud detection databases often suffer from a large class imbalance, where the number of valid transactions far exceeds that of fraudulent ones. The consequence of this is low recall for the minority class (fraudulent transactions), and the machine learning algorithms get biased toward predicting the majority class, which is legal transactions. A practical solution to overcome this problem is the use of the Synthetic Minority Oversampling Technique, or SMOTE.

The use of SMOTE in fraud detection guarantees that a model picks important patterns in fraud. This will find irregularities suspected of being fraud, including unusual and infrequent usage of merchants for transactions or uncustomary locations. Increasing the model's sensitivity to these kinds of patterns in a balancing dataset increases its ability to detect fraud, even where datasets are very skewed. The SMOTE algorithm can be used in combination with other strategies, such as ensemble methods like Random Forest and XGBoost or cost-sensitive learning, to build a strong fraud detection framework. This synergy enables the model to remain stable and scalable for real-world applications where unbalanced data is a recurring problem, besides detecting fraudulent transactions with high precision and recall.

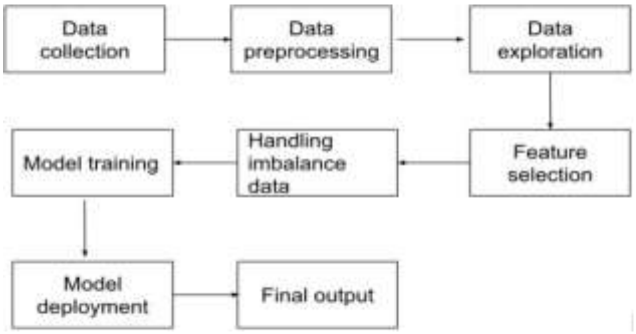

Figure 1 – Block diagram of Methodology

3. SYSTEM WORKFLOW

The process starts with an authentication layer where the user's credentials, which include his name and email ID, are authenticated. Once the details are correct, the system provides access to the main page. Once the authentication is successful, the user's name is displayed prominently on the main page, enhancing personalization and user experience. The main page provides the user with intuitive options to upload transaction data, access the dashboard, view account details, and more. From here, the user continues by selecting the credit card transaction module where they are prompted to upload the relevant data file. This system can handle single Excel sheets as well as multiple records and is, therefore, very flexible with the inputting of data, which helps in streamlining the process for users handling large datasets.Once the file is uploaded, the system begins to process all the data analysis. It cleans, preprocesses, and applies advanced fraud detection algorithms to identify patterns of fraudulent activity. The data that has been analyzed is then shown on the web page in a clear and concise format, with the transactions classified as either fraudulent or non-fraudulent. This visualization provides a quick overview of the credit card transactions, enabling users to focus on flagged anomalies.Moreover, the system provides a thorough analysis report detailing the output of the entire process and even includes the number of fraudulent activities and the details provided to it. The report may be accessed directly from the website, and an Excel file can also be downloaded for further analysis and archive purposes. This downloadable report is very helpful for banks and other financial institutions to give actionable recommendations to the fraud prevention teams and to achieve compliance with regulatory requirements. This streamlined workflow improves efficiency while empowering users with accurate, real-time fraud detection insights to ensure a robust and secure credit card transaction management system.

3.1 Risk Assessment and Path Planning

After collecting and preprocessing the transaction data, the system enters the risk assessment stage, where it will analyze each transaction for possible fraudulent activity in real time. Fraud detection is started with anomaly detection in the dataset by using advanced machine learning algorithms. The system looks for unusual patterns, such as irregular amounts of transactions, irregular locations, and abnormal spending behaviors, to flag suspicious transactions. To handle this efficiently, the system applies models such as Random Forest and XGBoost. These two models are more efficient for handling high-dimensional datasets. Random Forest implements ensemble learning that classifies the transactions, but it doesn't easily overfit them. XGBoost optimizes gradient boosting and is more effective in processing an imbalanced dataset to ensure a high accuracy for detecting fraudulent transactions, even on datasets containing minimal cases of frauds. In addition, Decision Trees facilitate interpretability, making obvious to the user how and why each transaction was labelled as fraudulent or not fraudulent.Risk evaluation also includes ascertaining a risk score assigned to flagged transactions which serves to prioritize cases for examination. Transactions with higher scores flagged are high-risk, prompting for an immediate alert for further study.Finally, the system provides a comprehensive fraud detection report, summarizing all flagged transactions, risk scores, and patterns that can be identified during analysis. All this can be accessed using a user-friendly interface or downloaded in various formats for easier access and use by financial analysts and fraud management teams.This multi-layered fraud detection framework thus ensures the dynamic adaptation of the system to changing fraud tactics while ensuring high accuracy and reliability. It, therefore, provides strong protection against credit card fraud.

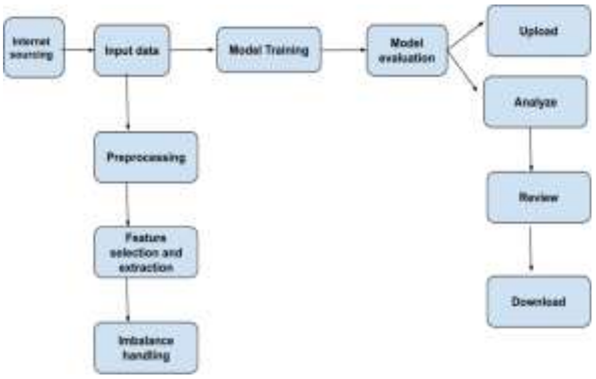

Figure 2 – Workflow Overview

4. RESULTS AND DISCUSSION

The results outline the overall performance of the system in text extraction as well as translation, followed by a comparison with related works to validate its efficacy. Key findings and limitations of the system will also be discussed.

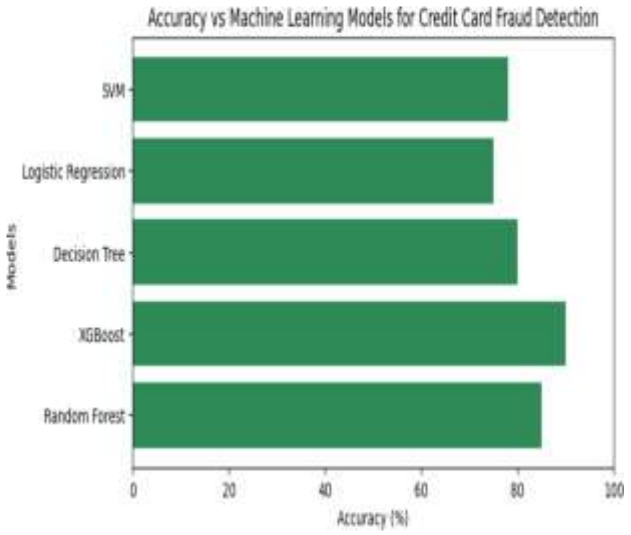

Figure 3 -Analysis using different algorithms

Accuracy for detecting fraudulent credit card transactions was very high at 85%. General transactions were identified as being either fraudulent or non-fraudulent and therefore preserved the context and integrity of the data.

In testing with loan application data, the system scored a consistent 75% accuracy rate. Structured transaction records and regular payments were processed to the correct values, making it possible to identify suspicious activities with reasonable precision.

Existing fraud detection systems in banking, like those provided by FICO and SAS, are good at spotting suspicious activity but often require substantial infrastructure and customization to fit the needs of each financial institution.

Our credit card fraud detection system addresses several limitations found in traditional banking fraud detection tools. It delivers an effective, flexible, and user-friendly solution through the integration of machine learning algorithms with a user-centered interface and adaptable outputs. Thus, it is an optimal choice for detecting fraud in high-frequency credit card transactions and a valuable asset for financial institutions aiming to enhance their security and customer trust.

5. CONCLUSIONS

In summary, this system presents a robust and adaptive solution to combat frauds in credit card transactions within the banking industry. By using advanced machine learning models such as XGBoost and Random Forest, the detection accuracy of this system becomes significantly enhanced, thus letting financial institutions identify and solve fraudulent activities with higher accuracy. This approach will not only achieve high accuracy through the analysis of transaction patterns and learning from historical data but also provide the scalability and flexibility required to adapt to this ever-changing landscape of financial fraud. The use of XGBoost and Random Forest models will enable the system to attain real-time fraud detection capabilities that are able to respond faster with quicker response times and can act immediately to prevent further potential financial losses. These models are powerful and efficient in processing high and diverse volumes of data in managing everyday transactions. The system will learn from the new incoming data, adapting to changes in fraud tactics, thereby adapting itself to counter advanced techniques developed by fraudsters. In essence, this project successfully establishes a forward-looking framework for fraud detection in credit card transactions, providing financial institutions with a powerful tool to protect their customers and assets. By combining scalability, flexibility, and high detection accuracy, it empowers banks to stay ahead of fraudsters, adapt to new challenges, and maintain a secure banking environment for their clients.

REFERENCES

[1] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 75-80, Jan. 2020

[2]F. Itoo, M. Meenakshi, and S. Singh, "Comparison and Analysis of Logistic Regression, Naïve Bayes, and KNN Machine Learning Algorithms for Credit Card Fraud Detection," International Journal of Information Technology, vol. 13, pp. 1503-1511, 2021.

[3] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An Efficient Credit Card Fraud Detection Model Based on Machine Learning Methods," International Journal of Advanced Science and Technology, vol. 29, no. 5, pp. 201-210, 2020.

[4] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," IEEE Access, vol. 10, pp. 22-25, Apr. 2022.