IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 10, Issue 6, June 2025

Smart Fraud Detection In Online Payments Using Machine Learning

[1] Lokesh N, [2]Srivarthan S, [3]Udhay Rakshith Y, [4]Karthick S, [5]T.Kanchana
[1] [2] [3] [4] Department of Computer Science and Engineering Er.Perumal Manimekalai College of Engineering TamilNadu, India
[5] Assistant Professor, Department of Computer Science and Engineering Er.Perumal Manimekalai College of Engineering TamilNadu, India

Abstract: ***With regards to the straightforwardness of making an installment while settling down anyplace on the planet, online installments have been a wellspring of engaging quality. Throughout the course of recent many years, there has been an expansion in web-based installments. E-installments empower organizations bring in a ton of cash notwithstanding buyers. Nonetheless, on the grounds that electronic installments are so straightforward, there is likewise a gamble of misrepresentation related with them. A customer needs to make sure that the money he or she spends only goes to the right service provider. Online extortion opens clients to the chance of their information being compromised, as well as the burden of announcing the misrepresentation, block their installment strategy, and different things. At the point when organizations are involved, it causes a few issues; every so often, they should give discounts to keep clients. As a result, consumers and businesses alike must be aware of these online scams. This study proposes a model for determining whether an online payment is fraudulent. To decide whether a specific Internet based installment is false or not, a few elements like the sort of installment, the beneficiary's personality, and so on. would be taken into consideration.***

Keywords—Online payment, Fraud detection, Credit card Fraud, Machine Learning, Supervised Learning, Digital Transactions and Data Preprocessing.

I. INTRODUCTION

The rise of e-commerce and digital payment systems has revolutionized the way we conduct transactions, offering convenience and accessibility to consumers worldwide. However, this digital landscape has also given rise to a pervasive threat online payment fraud. Fraudsters exploit vulnerabilities in online payment processes to carry out unauthorized transactions, leading to financial losses for businesses and undermining consumer trust in digital platforms.

Traditional methods of fraud detection rely on rule-based systems that flag transactions based on predefined thresholds and criteria. While effective to some extent, these systems often struggle to keep pace with the evolving tactics of fraudsters and may produce high rates of false positives or negatives. As fraudsters employ increasingly sophisticated techniques such as identity theft, account takeover, and phishing scams, there is a pressing need for more advanced and adaptive fraud detection mechanisms..

Machine learning, a subfield of artificial intelligence, offers a promising approach to addressing the challenges of online payment fraud detection. By leveraging algorithms that can learn from data and detect patterns, machine learning models have the potential to identify fraudulent activities with higher accuracy and efficiency than traditional methods.
These models can analyze vast amounts of transaction data, detect subtle anomalies, and adapt to changing fraud patterns in real-time.In this paper, we present a comprehensive framework for online payment fraud detection using machine learning

techniques. We explore the capabilities of various machine learning algorithms and feature engineering strategies to develop a robust fraud detection model.

Our goal is to demonstrate the effectiveness of machine learning in enhancing online payment security and mitigating the risks associated with fraudulent activities.

Through empirical analysis and experimentation on real-world datasets, we evaluate the performance of our proposed model and compare it to traditional rule-based systems. Additionally, we discuss practical considerations such as model interpretability, scalability, and deployment in real-world environments. By presenting a comprehensive overview of machine learning-based fraud detection, we aim to contribute to the on going efforts to combat online payment fraud and foster trust in digital transactions. Online payments have become more popular during the last few decades. This is because it's so simple to send money from anywhere, but the pandemic has also contributed significantly to the rise in e-payments.

Numerous studies have demonstrated that e-commerce and online payments will continue to grow in popularity in the years to come. The risk of online payment fraud has also increased as a result of this rise in online payments. Online payment fraud has been shown to have increased over the past few years, making it crucial for consumers and service providers to be aware of these frauds. It is crucial for users to be certain that the payments they make are going to the legitimate recipients; otherwise, they run the risk of having to report fraud, freeze their payment method, and run the chance of having their data shared with criminals, which could occasionally result in more crimes. On the other hand, it's crucial for companies to check that their customers aren't giving money to these fraudsters.

Businesses may have to repay money to clients in order to keep their patronage, which puts a strain on them. Even though firms have created and introduced numerous fraud detection programs, only a small number of them are effective in identifying online payment fraud. Although companies make every effort to make the payment method as secure as possible, fraudsters occasionally manage to circumvent security measures and commit these online payment scams.

According to studies Zanin et al. the cumulative losses from fraudulent bank card transcations increased globally. Other studies Kalbande et al. (2021) concentrate on idea drift, Which refer to the possibilities of change in the dataset's underlying distribution over time. Similar to how cardholders or users may alter their purchasing patterns over time, these fraudsters may modify their tactics.

These fraudsters are always aware of the customers' payment methods and behavior, but occasionally their tactics become outdated with time as some professionals work round-the-clock to uncover these scams and shield people from them. Fraud is an illegal technique to obtain somethingYan et al thus it's important to have in place an effective fraud detection system (FDS) that keeps track of all transactions and looks for any indication of fraud. The investigator looks into these potentially fraudulent transactions and reports back on whether or not the transaction was actually fraudulent. Machine learning techniques were used to determine if a transaction is fraudulent or not. Data mining techniques were typically used to analyze the patterns of fraudulent and non-fraudulent transactionsWang et al Therefore, by analyzing the patterns of the data, a combination of data mining techniques and machine learning can be utilized to determine if transactions are fraudulent or real.

II. PROBLEM STATEMENT

The increasing reliance on online payment systems has brought convenience to users but has also created new opportunities for fraudsters. Online fraud, particularly through credit card transactions, has become a major concern for financial institutions, merchants, and customers alike. Attackers use advanced methods such as identity theft, phishing, card skimming, and synthetic fraud to manipulate transaction systems.

These fraudulent activities not only result in financial loss but also damage customer trust and the reputation of payment platforms. Traditional fraud detection methods rely on static, rule-based approaches, which are often incapable of adapting to the dynamic and evolving nature of modern fraud techniques. These systems struggle to detect previously unseen fraud patterns and are easily bypassed by sophisticated attackers.

Furthermore, fraud detection presents significant technical challenges due to the highly imbalanced nature of transactional datasets—where fraudulent transactions make up less than 1% of the total volume. This imbalance often leads to biased machine learning models that fail to correctly identify minority class instances, i.e., fraudulent cases.
Additionally, online transactions are time-sensitive and require real-time or near-real-time detection to prevent unauthorized access and minimize damage. Therefore, there is a critical need for an intelligent, automated, and adaptive solution that leverages machine learning techniques to detect fraudulent activities effectively. The aim of this project is to develop a machine learning-based model that can analyze transaction patterns, handle imbalanced data, and accurately detect online payment fraud in credit card transactions with high precision and low false alarm rates.

III. LITERATURE REVIEW

Several studies have been conducted to address the growing concern of online payment fraud, especially in credit card transactions. Researchers have explored various machine learning algorithms and statistical models to detect fraudulent behavior based on transaction patterns.
A study by Bahnsen et al. (2016) demonstrated the effectiveness of cost-sensitive classification techniques to handle imbalanced datasets, improving fraud detection accuracy without compromising recall. Another paper by Dal Pozzolo et al. (2015) used ensemble learning methods like Random Forest and highlighted the importance of feature engineering and data preprocessing to boost model performance.

Logistic Regression, Decision Trees, and Support Vector Machines have been widely tested due to their simplicity and effectiveness. However, these models often struggle with imbalanced datasets. Recent works have adopted data balancing techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and undersampling to overcome this issue.

IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 10, Issue 3, March 2025

Meanwhile, deep learning models like Artificial Neural Networks (ANN) and LSTM (Long Short-Term Memory) have also been proposed to capture complex fraud patterns, although they require more computational resources and data.

Furthermore, anomaly detection methods have gained attention they do not require labeled data and can effectively detect new fraud types. Hybrid approaches that combine supervised and unsupervised learning have also shown promising results. These studies highlight the potential of machine learning in building adaptive, real-time fraud detection systems. However, challenges like real-time processing, false positives, and data privacy remain open areas for research and improvement.

IV. METHODOLOGY

The methodology involves several phases, all performed in a secure, closed environment:

a) Data Collection

- A real-world online payment transaction dataset (credit card-based) is used.
- The dataset includes anonymized features and labels indicating fraudulent or genuine transactions.

b) Data Preprocessing

- Removal of duplicates or null values if present.
- Feature normalization or standardization is applied.
- Class imbalance is addressed using SMOTE (Synthetic Minority Over-sampling Technique) to oversample the fraud class.

c) Feature Selection

- Relevant features are selected using correlation analysis or feature importance scores.
- Irrelevant or redundant features are eliminated to reduce overfitting and improve accuracy.

d) Model Selection & Training

Machine learning algorithms such as:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)

e) Model Evaluation

Models are evaluated based on the following metrics:

- Accuracy
- Precision
- Recall
- F1-score

- AUC-ROC curve

f) Model Comparison and Finalization

- All models are compared based on performance.
- The best-performing model (with high recall and low false positives) is chosen for final use.

V. IMPLEMENTATION

The implementation phase begins with setting up the development environment using Python along with essential libraries such as Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, and Imbalanced-learn. The dataset used for this project is a real-world online payment dataset that contains anonymized features along with a binary label indicating whether a transaction is fraudulent or legitimate. The initial step involves loading the dataset and performing exploratory data analysis to understand its structure, feature distributions, and class imbalance.

Data preprocessing is then carried out, which includes handling missing values, scaling the feature values using normalization techniques like Standard Scaler, and balancing the dataset using SMOTE (Synthetic Minority Over-sampling Technique) to generate synthetic samples of the minority fraud class. The dataset is then split into training and testing sets to evaluate the model's performance on unseen data. Various supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), are trained and fine-tuned using grid search and cross-validation techniques. Each model is evaluated using performance metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. Visualizations such as confusion matrices, ROC curves, and feature importance plots are used to better understand the performance and behavior of each model.

VI. RESULTS AND DISCUSSION

1.Model Performance Comparison

- Random Forest outperformed other models (Logistic Regression, Decision Tree, and SVM) in terms of key metrics like accuracy, precision, and recall.
- Random Forest showed the highest true positive rate and lowest false negative rate, indicating its strong ability to detect fraudulent transactions.

2. Effect of Data Balancing (SMOTE)

- Applying SMOTE to address class imbalance significantly improved the recall for the fraud (minority) class.
- This technique helped the model better identify fraud cases without compromising overall model accuracy.

3. Evaluation Metrics

- The models were evaluated using accuracy, precision, recall, F1-score, and AUC-ROC curve.

- The Random Forest classifier showed the most balanced performance, providing an optimal trade-off between high recall and low false positives.

4. Confusion Matrix and AUC-ROC Analysis

- The confusion matrix revealed that Random Forest minimized false positives and false negatives, crucial for fraud detection systems.
- The AUC-ROC curve for Random Forest indicated excellent discrimination between fraudulent and non-fraudulent transactions.

5. Real-Time Deployment Consideration

- Although the model showed promising results in batch testing, real-time deployment would require further optimization for faster
- prediction and integration with payment gateways.

6. Potential for Future Work

- Future improvements could involve model optimization, feature selection refinement, and deployment in real-world environments to ensure scalability and robustness.

VI. CONCLUSION

In this project, an effective online payment fraud detection system was developed using machine learning techniques, specifically focusing on credit card fraud detection. The Random Forest algorithm demonstrated superior performance compared to other models like Logistic Regression, Decision Trees, and SVM, with the highest accuracy, precision, and recall scores. By employing SMOTE to balance the imbalanced dataset, the system significantly improved its ability to correctly identify fraudulent transactions, while maintaining a low false positive rate.

The evaluation of the models using various performance metrics such as accuracy, precision, recall, and the AUC-ROC curve confirmed that the Random Forest classifier is well-suited for detecting online payment fraud. However, for real-time fraud detection in live payment systems, further optimizations, such as reducing prediction latency and ensuring system scalability, are necessary. This project demonstrates the potential of machine learning in the online payment industry and provides a foundation for future research aimed at enhancing fraud detection systems, integrating them into secure payment gateways, and improving the user experience by minimizing false alerts

VII. ACKNOWLEDGEMENT

I would like to thank my supervisor, [Kanchana T], for their guidance and support throughout this project on Online Payment Fraud Detection Using Machine Learning. I also appreciate [Er. Perumal Manimekalai College of Engineering] and the [Computer Science and Engineering] for providing the necessary resources. My heartfelt thanks to my family and friends for their continuous encouragement and support.

REFERENCES

[1] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating Probability with Undersampling for Unbalanced Classification," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, 2015, pp. 159–166.

[2] C. Sahin and I. Duman, "Detecting credit card fraud by decision trees and support vector machines," Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, 2011, pp. 442–447.

[3] A. Jurgovsky et al., "Sequence classification for credit-card fraud detection," Expert Systems with Applications, vol. 100, pp. 234–245, June 2018.

[4] P. Ghosh and D. Reilly, "Credit card fraud detection with a neural-network," Proceedings of the Twenty-Seventh Hawaii International Conference on System Sciences, 1994, pp.