International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Identifying Fraudulent Credit Card Transactions Using Ensemble Learning

[1] Mrs. RENUKHA B.N [2] PUSHPA SAHANA M
[1]Assistant Professor, Department of MCA, BIET, Davanagere
[2] Student,4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

Identifying deceitful praise postcard relations remains a significant challenge for financial institutions. Even after passing authentication checks, fraudulent activities can still occur if attackers successfully impersonate legitimate cardholders. This research explores the effectiveness of ensemble learning techniques in categorizing praise card scam using two different datasets: a synthetic dataset generated by Sparkov and a real-world dataset from European Union consumers. The models employed include XGBoost, Random Forest, and Naive Bayes classifiers. These models are evaluated based on accuracy, precision, recall, and F1 score. Outcomes signpost that collective means yield high performance on the real-world dataset but show limited effectiveness on the synthetic one. The findings suggest that deterministic patterns in real-world transaction data can be easily learned by machine learning models, while the lack of randomness may inadvertently increase the risk of card information exposure.

Keywords: Credit card fraud detection, ensemble learning, XGBoost, Random Forest, Naive Bayes, synthetic dataset, real-world dataset, fraudulent transactions, machine learning, financial security

1. INTRODUCTION

With the rapid growth of digital financial services, electronic payments have become a fundamental part of everyday transactions. Technologies such as credit cards, digital wallets, and online banking offer convenience, speed, and transparency. However, these developments have also opened new avenues for cybercriminals to exploit, leading to a significant rise in credit card fraud. As fraudulent activities become more sophisticated, detecting and preventing unauthorized transactions is now a top priority for financial institutions, including banks, insurers, and global payment networks.

Praise card fraud (CCF) typically involves unauthorized access to cardholder data, which allows attackers to initiate transactions without the owner's consent. Despite continuous efforts to secure digital payments, the frequency and financial impact of such frauds continue to grow. Traditional methods of fraud detection, including statistical techniques and rule-based systems, often fall short in adapting to evolving fraud patterns. As a result, more intelligent solutions, especially those based on machine learning and data-driven models, are being explored to enhance fraud detection capabilities.

The objectives of this study are to investigate how well ensemble models learn from different data types, identify possible weaknesses in current fraud detection systems, and explore how deterministic transaction processing scripts might expose vulnerabilities. The research also proposes practical strategies to enhance the security of credit card approval mechanisms. Through comparative analysis and performance evaluation using The study aims to offer significant insight into developing systems for detecting fraud in financial institutions through investigation of significant metrics like accuracy, precision, recall, and F1-score.

2. LITERATURE REVIEW

“SEISENSE J. supervise, vol. 5, no. 1, pp. 49–59, 2022 February review of artificial intelligence algorithms for detecting credit card fraud with a case study. The utilization of different supervision artificial intelligence approaches for identifying credit card scams is investigated in this paper. Applying data from the real world, it focuses on evaluating the performance of algorithms like XGBoost, Random Forest, Decision Tree, K-nearest-neighbors algorithm (KNN), and Logistic Regression. Faraji, Z. [1]

New deep learning and machine learning methods for detecting credit card fraud," Access, IEEE, vol. 10, pp. 39700–39715, 2022. The present research offers a thorough comparison of 22 methods for credit card fraud detection using contemporary machine learning techniques and deep learning. Using a sizable dataset, the study implements CNNs, LSTMs, Random Forest, Gradient Boosting, Logistic Regression, and 22 decision tree methods. Precision, recalled, AUC-ROC, and F1-score are the basis for evaluation. N. Almusallam, F. K. Alarfaj, I. Malik, H. U. Khan, M. Ramzan, and M. Ahmed. [2]

“ This industry report offers updated statistics on global praise card scam ”, highlighting trends in information breaches, fraud rates, and regional variations. It emphasizes the rising costs of fraud worldwide and the increasing necessity for hearty scam prevention tools, including AI-based solutions. The report helps as a foundational reference for identifying the scale of the problem addressed in research studies on fraud detection. Nilson Report, Card Fraud Worldwide. Accessed: May 2023. [Online]. Available: https://nilsonreport.com/. [3]

"Improved neural network model for spotting credit card fraud," electronics vol. 11, no. 4, p. 662, Feb. 2022. This research offers an enhanced approach for detecting credit card fraud through using sophisticated selection of features and tuning of hyper parameter methods. Assessments of performance are made comparing technologies like Random Forest (RF), Support Vector Machines (SVM), and gradient-boosting. S. M. Fati and N. S. Alfaiz. [4]

"A look at methods for detecting credit card fraud," With an emphasis on statistical in nature, artificial intelligence, and hybrid methods, this review reviews the traditional and modern identification of fraud techniques. Important issues like feature engineering, data imbalance, and evolving fraud

trends are addressed in the paper. B. Sourabh and Arora. [5]

Internet Analysis and Programs, "Automation of bank identification of fraud utilizing group model." 211–224 (Springer Nature, 2022). The present section introduces an approach to fraud detection that enhances detection efficiency by utilizing a combination of classifier which includes random forests, AdaBoost as well and XGBoost. K. Sowmya is, S. Karthika, and S. Srinidhi. [6]

Expert Syst. Appl., vol. 192, Apr. 2022, Art. no. "A unique framework for recognizing mobility and appearance-based anomaly utilizing group learning with LSTMs." The work illustrates how deep learning may be used to efficiently identify variations in spatial and visual anomalies. This strategy has implications for 1 theft, namely detection technologies, in modeling recurring repeated behavior suggestive of illicit behavior. D. K. Vishwakarma and M. Sabih. [7] This accessible dataset, which has been extensively utilized in scholarly research, comprises European cardholder' anonymised transactions with credit cards. It has a major class imbalance, which makes it an among the best methods of detection. To evaluate fraud on Kaggle (2022). Data of European Cardholders. recovered in May 2023.
[Online].
The next information is available for download: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. [8]

“ This GitHub project provides synthetic data generation tools for simulating credit card transactions ”. It is specifically valuable for examiners demanding large-scale, balanced datasets for training and evaluating fraud detection algorithms. Sparkov Data Generation on Github, Sparkv Simulator, 2020. [9]

"Detection of fraud with credit cards using artificial intelligence techniques," in Abstracts of the 18th Global Symposium on Computer Technology-Jahorina (INFOTEH), March 2019, pp. 1–5. Popular computer learning methods for spotting trends, like supported vector machines, Random Forests, and Decision Trees, are evaluated in this workshop paper in relation to illicit transactions. M. Karanovic, S. Sladojevic, M. Arsenovic, A. Anderla, and D. Varmedja. [10]

3. METHODOLOGY

In this research, two different datasets were rummage-sale to appraise the efficiency of ensemble learning methods in detecting credit card fraud. One is a real-world dataset containing anonymized transaction records of European cardholders, while the other is a synthetic dataset generated using the Sparkov simulator. Both datasets are inherently imbalanced, with a significantly higher number of genuine transactions compared to fraudulent ones. To speech this subject and improve model performance, the Manmade Underground Oversampling Technique (SMOTE) was applied to balance the dataset by generating synthetic examples of the minority class. Furthermore, preprocessing techniques such as handling missing data, normalizing feature values using standard scaling, and converting categorical

data where necessary were employed to warrant that the information was ready for efficient training and evaluation. The study implements three ensemble learning algorithms—Random Forest, Naive Bayes, and XGBoost—due to their proven success in handling classification tasks, especially with skewed data distributions. These models were trained on both datasets using a stratified train-test split to maintain the class distribution across the subsets. Each model was developed and fine-tuned using Python's machine learning libraries, particularly Scikit-learn and XGBoost. During training, attention was given to how well each algorithm could distinguish between fraudulent and legitimate transactions, and whether the type of dataset (real or simulated) impacted the model's ability to learn and generalize patterns associated with fraudulent behaviour.

4. EXISTING SYSTEM

In today's digital era, electronic payments (e-payments) have become one of the leading innovations in the Fintech space. These payments refer to the electronic transfer of money through mediums like mobile wallets, credit/debit cards, or online banking systems. They offer frequent assistances, with cost-effectiveness, time efficiency, transaction transparency, and promotion of a cashless economy. Despite these advantages, e-payment systems are vulnerable to security risks. One major concern is the compromise of credit card details, which can result in financial loss. With the growing use of credit cards for transactions, there has been a corresponding rise in recognition card fraud (CCF). Detecting such fraudulent activities is essential for financial institutions like banks and card network providers (e.g., Visa, MasterCard). The main challenge is identifying suspicious transactions that may be initiated by unauthorized users. Fraudulent activities must be flagged before the transactions are finalized to prevent monetary loss.

Though numerous haven procedures consume remained implemented to safeguard credit card data, fraud cases continue to increase. This ongoing threat has driven the development of advanced fraud detection systems using statistical methods and machine learning models. These systems analyze historical transaction data to uncover patterns and anomalies that may indicate fraudulent behaviour.

Machine learning offers a powerful alternative for tackling CCF. A wide range of machine learning algorithms—both for classification and clustering—have been applied depending happening in what way the tricky is modeled. Organisation replicas remain typically skilled by labeled transaction information, such by way of the well-known European cardholder's dataset. This dataset is generally second-hand in research for building and challenging scam detection systems. In our study, we also make use of the European dataset, along with the Sparkov simulated fraud dataset. Other datasets, such as the Brazilian financial dataset and transaction data from commercial banks in China, have also been explored in existing work. One key challenge with CCF datasets is that they are highly imbalanced, containing many more legitimate transactions than fraudulent ones. To address this issue, preprocessing methods like SMOTE, oversampling,

and under sampling are commonly used. The effectiveness of classification models is typically measured using performance metrics such as precision, recall, F1-score, accuracy, and ROC-AUC. For clustering-based models, evaluation metrics include homogeneity, completeness, and V-measure. Ensemble learning methods, like random forests and boosting algorithms, have generally shown superior performance compared to basic models such as decision trees and Naive Bayes classifiers.

5. PROPOSED SYSTEM

The increasing usage of online payment systems such as credit cards, digital wallets, and cryptocurrencies has led to a significant rise in fraudulent transactions and unauthorized fund transfers. These incidents often highlight vulnerabilities in the security frameworks of financial service providers. This research aims to address these issues by evaluating the performance of ensemble learning algorithms on both real and synthetic datasets, assessing their capability to accurately detect fraudulent behavior. It further investigates the root causes of failure in existing fraud detection mechanisms and analyzes the weaknesses in transaction processing scripts. Based on these insights, the study proposes practical improvements to enhance the decision-making process for approving or rejecting credit card transactions. The key features incorporated in the proposed system include the use of advanced ensemble models such as Random Forest, XGBoost, and Voting Classifier, real-time monitoring of transaction patterns, anomaly detection through statistical thresholds, and continuous model training for adapting to new fraud strategies. Together, these elements aim to build a more robust and intelligent fraud detection framework.

System Architecture

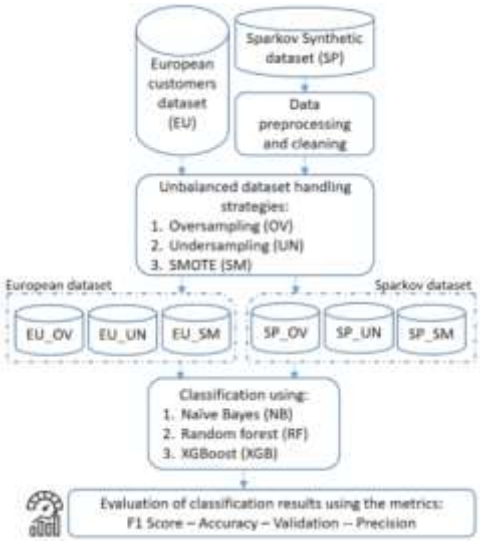

Fig:5.1 System Architecture

6. RESULTS

The project is implemented using Python and the Django web framework, transforming conceptual modules into functional software components. Data preprocessing is carried out using Pandas and NumPy, with Standard Scaler applied for feature scaling, and SMOTE used to balance the imbalanced dataset. Machine learning models such as Random Forest and Naive Bayes and Gradient Boosting Classifier are trained using train-test split and cross-validation, while XGBoost is implemented using its dedicated package. The frontend is developed using HTML.

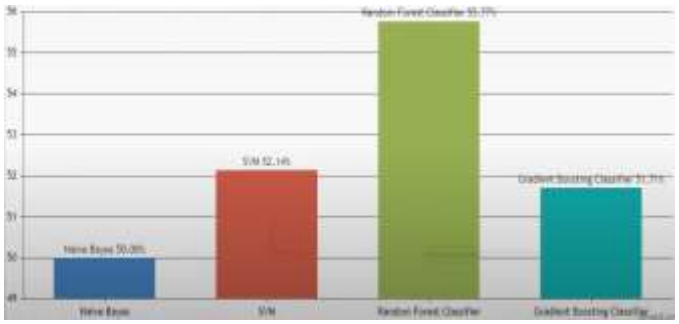

Fig:6.1 Resultant graph

7. CONCLUSION

With the rise of automated imbursement systems, praise card fraud has become a critical concern for financial institutions. This study evaluates the effectiveness of ensemble learning techniques—namely XGBoost, Random Forest, and Naive Bayes—in identifying fraudulent transactions using both actual and synthetic datasets. By analyzing performance indicators such as accurateness, exactness, memory, and F1-score, it was experiential that collective replicas perform better with real transaction data. However, the deterministic nature of real-world transaction processes can create vulnerabilities, as attackers may exploit predictable patterns. The research highlights the necessity of integrating adaptive and dynamic elements into fraud detection systems to counter such threats. Additionally, it points out current system limitations and suggests enhancements, including the use of behavioural analytics and improved ensemble techniques to better manage data imbalance and complexity. Overall, the study underscores the potential of ensemble learning to strengthen fraud detection systems and adapt to the rapidly evolving challenges in the financial sector.

8. REFERENCES

[1] "A evaluation of kit erudition solicitations for tribute letter fraud detection with a case study," SEISENSE J. Manage., vol. 5, no. 1, pp. 49–59, Feb. 2022. Z. Faraji.

[2] "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," IEEE Access, vol. 10, pp. 39700–39715, 2022. F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed.

[3] "Card Fraud Worldwide". Accessed: May 2023. [Online]. Available: https://nilsonreport.com/ Nilson Report.

[4] "Enhanced credit card fraud detection model using machine learning," Electronics, vol. 11, no. 4, p. 662, Feb. 2022. N. S. Alfaiz and S. M. Fati.

[5] "A review of praise postcard fraud detection techniques," Recent Innov. Comput., pp. 485–496, 2022. B. Arora and Sourabh.

[6] "Involuntary acknowledgment fraud exposure using communal model," in ICT Analysis and Applications. Springer, 2022, pp. 211–224. S. Srinidhi, K. Sowmya, and S. Karthika.

[7] "A novel framework for detection of motion and appearance-based anomaly using ensemble learning and LSTMs," Expert Syst. Appl., vol. 192, Apr. 2022, Art. no. 116394. M. Sabih and D. K. Vishwakarma.

[8] "European Cardholders Dataset". Accessed: May 2023. [Online].
Available:https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud. Kaggle. (2022).

[9] Sparkov Data Generation on Github, Sparkv simulator, 2020.

[10] "Credit card fraud detection–machine learning methods," in Proc. 18th Int. Symp. INFOTEH-JAHORINA (INFOTEH), Mar. 2019, pp. 1–5. D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, and A. Anderla.

*****