International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Review on Credit Card Fraud Detection Using Machine Learning

Madhuri R. Gangwe[1], Miss. Manisha Mundhe[2]

1Department of Computer Engineering, DIEMS, BATU University, Lonere (M. S) India
2Department of Computer Engineering, DIEMS, BATU University, Lonere (M. S) India

---

Abstract -The banking industry values cyber security due to an increasing risk of cyber attacks and crime. Credit card cyber fraud poses a significant security problem throughout the world. Conventional anomaly detection and rule-based algorithms are popular methods for identifying cyber fraud, but they are time-consuming, resource-intensive, and incorrect. Machine learning is one of the technologies acquiring. Popularity and key influence in this industry. This paper reviews and consists prior research on credit card cyber fraud detection. This review focuses on machine and deep learning methods. Our evaluation found 181 research publications published between 2019 and 2021. This article provides an overview of machine learning and deep learning methods and their use in detecting credit card fraud. When using machine learning to identify credit card fraud, it's crucial to select appropriate features. Credit card fraud detection is a classic example of categorization. We investigated and pre-processed data sets before using anomaly detection techniques like Local Outlier Factor and Isolation Forest on PCA-transformed credit card transaction data.

Key Words: credit card, fraud detection, logistic regression, random forest.

1. INTRODUCTION

Detecting credit card fraud is crucial for ensuring financial security and managing risks in the digital economy. Credit card fraud detection is the process of discovering and blocking unauthorized transactions. Credit card fraud happens when people or criminals use stolen or counterfeit information to make unauthorized transactions or withdrawals. Fraud in credit card transactions refers to the unauthorized and undesired use of an account by someone other than the owner. Preventing misuse and studying fraudulent tactics might help minimize future incidents.

Machine learning algorithms can prevent fraud in credit card transactions. Nowadays, machine learning and artificial intelligence are rapidly developing technologies. Today's technology can enhance our understanding of several societal aspects.

Let's get a basic knowledge of machine learning in the credit card fraud detection project. Algorithms such as forest, genetic, and data-set training may detect fraudulent and non-fraudulent transactions with ease.

Here's an outline of the stages involved in detecting credit card fraud with machine learning.

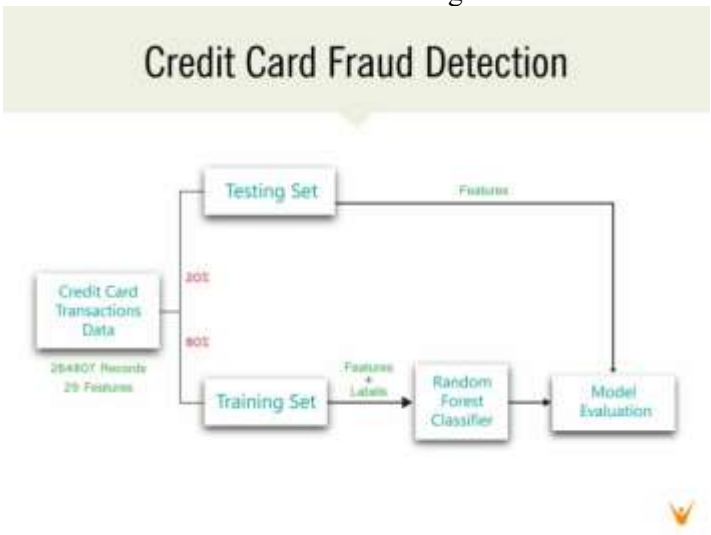

Fig (1). Credit card fraud detection flowchart

The flowchart shows the random forest classifier approach, however linear regression can also be used instead.

Fraud detection systems are always evolving to help criminals adapt to their fraudulent schemes. There are several types of fraud, including credit card fraud, card theft, account bankruptcy, device intrusion, application fraud, counterfeit cards, and telecommunication fraud.

2. Literature Review

Random forest is the most often used supervised technique in the literature for many application domains. Integrating algorithms into a bank's fraud detection system allows for easy identification of fraudulent transactions. Anti-fraud strategies can reduce significant risks and losses for institutions. Unlike previous classification studies, we implemented a variable penalty for misclassification. The recommended system's performance is evaluated using precision, f1score, and accuracy. We analyzed the data, discovering characteristics and imbalances. The proposed system's performance is evaluated using precision, f1-score, and accuracy metrics. [1]

During the detection phase, the credit card holder's buying patterns are examined using the k means clustering approach, and then the sequence is created. In the first stage, a successful score is calculated by analyzing real cardholder transaction history and behavioral changes using sequence alignment. The second point's negative score is generated by using the fraudulent transaction signature from the first. If the difference between the excellent and poor scores exceeds a certain limit, the transaction is illegal; otherwise, it is permitted. [2]

The experiments are presented and discussed in two stages. The first stage involves a comparison of eight categorization systems. The comparison included three factors: sensitivity, accuracy, and area under the precision recall curve (AUPRC). After comparing algorithms, SVM and ANN were identified as the best choices. The second phase analyses imbalance classification approaches, such as Random Oversampling, One Class Classification, and Cost Sensitive, utilizing selected algorithms. The SVM's performance is compared to that of the One-Class Classification SVM and the Cost Sensitive SVM as a binary classification tool. The Auto Associative Neural Network is compared and contrasted with the ANN. [3]

This work proposes a feature selection strategy for ML-based credit card fraud detection engines. It utilizes the genetic algorithm (GA). The recommended fraud detection engine's performance was evaluated using a dataset of European cards. The proposed detection engine employs ML classifiers such as Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Artificial Neural Network (ANN), and Naive Bayes to choose the best features (NB).The researcher used the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalances in the dataset. The researcher assessed the efficacy of each ML approach based on classification accuracy. The study authors used a hybrid sampling strategy to resolve the imbalance in the dataset. [4]

This study used ANN, Decision Trees, SVM, Logistic Regression, and Random Forest to detect fraudulent transactions. The accuracy, precision, and false alarm rate are used to determine how well each approach performs. The experiment used data from the Kaggle repository. The assembled dataset contains 150000 Tran's activities. The data set has several fields. Principal component analysis was used to reduce dimensions and extract required attributes such as transaction time, amount, and class. However, using these approaches may not yield consistent results in all instances. The amount and type of the dataset influence how effectively the tactics operate. While the ANN model is dependable, training may be time-consuming and costly. SVM produces high-quality results even with little datasets. Decision Trees perform best with preprocessed and sampled data, whereas Logistic Regression works best with raw data. Random forest is effective with both categorical and continuous data. [5]

The remaining columns are time, amount, and class. The time interval between the first and second transactions in a sequence is known as time. The amount refers to the total money sent. A fraudulent transaction is one that has a legitimate class 0.After analyzing the datasets; a histogram is generated for each column considered. The datasets are visualized as a graph. This will ensure that no data values are missing. A heat map graph is used to show the association between output assumptions and class variables. [6]

The dataset was divided into three sets: training, validation, and testing. The 70/30 rule was followed, with test data comprising 15%, validation data at 15%, and training data at 70%. This ratio was chosen because to the large dataset and lack of additional data points for training, reducing the risk of classification bias. The dataset was used to train various machine learning models, including logistic regression, support vector machine, and random forest. After training using several machine learning approaches, a comparison analysis is conducted to evaluate the performance of each model. The macro averages of the F1 score, recall, and accuracy are utilized. [7]

The recommended method classifies the credit card dataset using the random forest technique. The Random Forest method is a classification and regression strategy. This is essentially A collection of decision tree classifiers. Random forests outperform decision trees due to their ability to avoid over fitting the training set. Training entails sampling from a subset. A decision tree is created, with each node based on a randomly picked feature from the full collection. Training each tree individually allows for fast training, even for large data sets with multiple features and instances. The method is resistant to over fitting and provides solid estimates. [8]

The recommended gadget To address the lack of minority class data in the dataset, SMOTE (Synthetic Minority Oversampling Technique) combines existing minority class elements. It works by selecting any arbitrary point from the minority class and determining its k-nearest neighbors. The synthetic points were inserted between the
Chosen spot and its neighbors. Another approach is to use the random forest algorithm. The method selects "k" features from the total "m" features in the international dataset. The produced nodes are separated based on the chosen characteristics. A decision tree normally uses the optimal split function. Recursively splitting the nodes generates the number of daughter nodes. There should be

a restriction on the number of nodes generated for each tree. [9]

This study examines the use of machine learning and neural networks to identify potential fraudsters by analyzing their prior actions and data on previous criminals. Support Vector Machines, Logistic Random Forest Regression, Multinomial Naive Bayes, and a Simple Neural Network are all utilized. ULB's Machine Learning Group compiled and made the dataset available to the public. There were no missing values, only numerical inputs, and the data was imbalanced, with a significant difference between positive and negative data series. The Class column specifies whether a transaction was fraudulent, with 1 indicating fraud and 0 indicating otherwise. The Sklearn programme created confusion matrixes and classification reports. [10]

3. BODY OF THE PAPER

Imbalanced Data: Credit card fraud datasets are often skewed, with a tiny fraction of fraudulent transactions vs. valid ones. Addressing the issue of unbalanced data and providing appropriate solutions is a research gap. The goal is to enhance fraud detection performance using sampling, data augmentation, and ensemble methodologies.

Feature Engineering and Selection: A research gap exists in identifying and optimizing key aspects for credit card fraud detection. This involves studying sophisticated feature selection methods, dimensionality reduction techniques, and domain expertise to improve feature discrimination.

Real-time Fraud Detection: This remains a huge research gap. This includes creating efficient algorithms and models to manage enormous amounts of streaming data and detect fraud quickly and accurately.

Transferability and Generalization: Improving the adaptability and extension of machine learning models across domains and datasets remains a major challenge. This entails studying strategies to improve fraud detection algorithms for different banks, geographies, and time periods.

Explain ability and Interpretability: Building confidence and understanding in credit card fraud detection systems requires explaining and interpreting the decision-making processes of machine learning models. Improving the explain ability and interpretability of machine learning models in fraud detection is a significant research gap.

Adversarial Attacks: Exploring the susceptibility of machine learning-based fraud detection systems to adversarial assaults remains a research need. Our research focuses on improving model robustness and resilience to adversarial manipulations, as well as providing real-time detection and mitigation strategies.

Privacy and Ethical Considerations: Privacy and ethical considerations: Addressing privacy and ethical considerations in credit card fraud detection with machine learning is a research need. This entails creating tools to identify fraud while protecting client privacy and assuring fair and unbiased decision-making.

Deployment and implementation: It's vital to bridge the gap between research and practical use of machine learning models for detecting credit card fraud. This involves examining obstacles and solutions for integrating machine learning systems into current fraud detection infrastructures, including scalability, resource limits, and real-world implementation challenges.

3. CONCLUSIONS

The application of machine learning and deep learning techniques to credit card fraud detection has yielded encouraging results. This technique effectively distinguishes between authentic and fraudulent transactions, improving financial security. The study indicated that deep learning models, namely deep neural networks, outperformed typical machine learning algorithms in identifying credit card fraud. Deep neural networks excel in extracting complicated patterns and relationships from data, resulting in greater performance. This emphasizes the need for improved strategies to capture complex linkages in transactional data. The study emphasized the importance of feature engineering and preprocessing strategies in enhancing fraud detection accuracy. Techniques including outlier identification, feature scaling, and dimensionality reduction were critical in improving the models' performance. The models improved their ability to distinguish between legal and fraudulent transactions by selectively modifying important characteristics. This study found that combining machine learning and deep learning methodologies may considerably improve credit card fraud detection. Further research and development in this subject can enhance the accuracy and effectiveness of fraud detection systems, helping financial institutions and clients protect against fraudulent actions.

REFERENCES

[1] Saiju, Sanisa, S. Akshaya Jyothy, Christeena Sebastian, Liss Mathew, and Tintu Sabu. "Credit Card Fraud Detection Using Machine Learning." International Journal of Recent Advances in Multidisciplinary Topics 2, no. 4 (2021): 31- 34

[2] Arafath, Yeasin, Animesh Chandra Roy, M. Shamim Kaiser, and Mohammad Shamsul Arefin. "Developing a Framework for Credit Card Fraud Detection." In Proceedings of the International Conference on Big Data, IoT, and Machine Learning, pp. 637-651. Springer, Singapore, 2022.

[3] Awoyemi, John O., Adebayo O. Adetunmbi, and Samuel A. Oluwadare. "Credit card fraud detection using machine learning techniques: A comparative analysis."In 2018 international conference on computing networking and informatics (ICCNI), pp. 1-9. IEEE, 2018.

[4] Ileberi, Emmanuel, Yanxia Sun, and Zenghui Wang. "A machine learning based credit card fraud detection using the GA algorithm for feature selection." Journal of Big Data 9, no. 1 (2022): 1-17.

[5] Sadineni, Praveen Kumar. "Detection of fraudulent transactions in credit card using machine learningalgorithms." In 2020 Fourth International Conference on I- SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 659- 660. IEEE, 2020.

[6] Sanobar khan, Sanovar, Suneel Kumar , Mr Hitesh Kumar(2021);Credit Card Fraud Detection Using ML; International Journal of Scientific and Research Publications(IJSRP).

[7] Sharma, Pratyush, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni. "Machine learning model for credit card fraud detection-a comparative analysis." Int. Arab J. Inf. Technol. 18, no. 6 (2021): 789-796.

[8] Meenakshi, B. Devi, B. Janani, S. Gayathri, and N. Indira. "Credit card fraud detection using randomforest." International Research Journal of Engineering and Technology (IRJET) 6, no. 03 (2019).

[9] Priya, G. Jaculine, and S. Saradha. "Fraud detection and prevention using machine learning algorithms: a review." In 2021 7th International Conference on Electrical Energy Systems (ICEES), pp. 564-568. IEEE, 2021.

[10]Azhan, Mohammed, and Shazli Meraj. "Credit card fraud detection using machine learning and deep learning techniques." In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 514-518. IEEE, 2020.