INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 4, April 2026 | ISSN: 2456-4184 | IJNRD.ORG

CREDIT CARD FRAUD DETECTION SYSTEM

Peddireddy Venkata Rohith
Department of Artificial intelligence and Data Science,
Dhanalakshmi Srinivasan University ,
Samayapuram-621112.
Trichy, India.
Reg No:1522051231.
Peddireddyvenkata7@gmail.com

Mr.V.V.Shabeer
Assistant Professor
Department of Artificial intelligence and Data Science,
Dhanalakshmi Srinivasan University ,
Samayapuram-621112.
Trichy, India.

Peddireddy Venkata Rohith
Department of Artificial intelligence and Data Science,
Dhanalakshmi Srinivasan University ,
Samayapuram-621112.
Trichy, India.
Reg No: 1522051232.
peddireddyvenkatarevanth@gmail.com

Abstract

This project focuses on the development of a Credit card fraud detection uses machine learning (ML) to identify unauthorized transactions, a critical issue due to rising e-commerce, by training models on historical data to classify new transactions as genuine or fraudulent, often using algorithms like Random Forest, Logistic Regression, or Neural Networks, and addressing the challenge of imbalanced datasets with techniques like SMOTE to achieve high accuracy and reduce financial losses.The goal of this project is to develop a machine learning model that can accurately detect fraudulent credit card transactions using historical data. By analyzing transaction patterns, the model should be able to distinguish between normal and fraudulent activity, helping financial institutions flag suspicious behavior early and reduce potential risks.Credit Card Fraud Detection cares with the illegal use of master card information for purchases.Credit Card transactions are often accomplished either physically or digitally.In the manual transactions,the credit card is included during the transactions.In digital transactions ,this will happen over the phone or over the web.Cardholders might be providing their card number,expiry date, and the verification of the card number through telephone or the website.Billions of dollars are lost thanks to master card fraud.

KEYWORDS

Credit Card Fraud Detection, Fraud Detection, Fraudulent Transactions, K- Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Trees.

I. INTRODUCTION

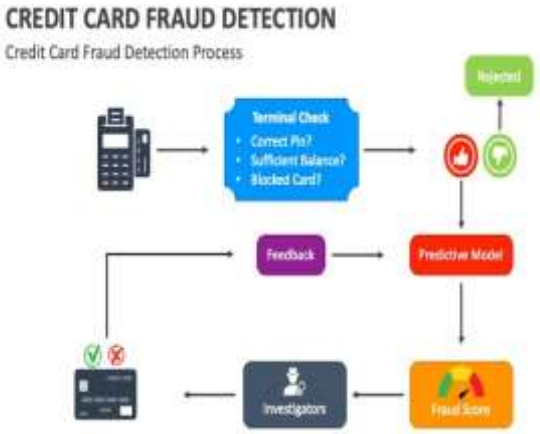
The emergence of electronic transactions at a high growth rate and extensive use of credit cards have completely transformed the financial and banking sectors. Yet,this convenience came at the price of heightened credit card fraud, which proved to be a formidable challenge for banks and consumers as well. With every innovation by fraudsters in their strategies, it became tough for the conventional fraud prevention system to remain current. Accordingly, a more sophisticated and responsive set of solutions to identify and capture fraud effectively is needed in an urgent manner. This paper discusses the use of machine learning (ML) methods to solve the increasing issue of credit card fraud within the banking sector . Supervised learning, unsupervised learning, and deep learning algorithms are quite helpful for fraud transaction identification. Supervised learning algorithms such as logistic regression, decision trees, and random forests can be trained on labeled information to mark the transactions as fraud or real. Unsupervised learning algorithms such as anomaly detection and clustering are most suited to detect unknown fraud patterns. Deep learning techniques, such as neural networks, are most appropriate for highly dimensional and complicated data and therefore they function optimally in real-time fraud detection. Deep learning techniques being used together can enable banks to develop efficient systems that learn to detect changing patterns of fraud and minimize financial loss.

The use of machine learning to identify credit card fraud is accompanied by a number of challenges. One of the significant problems is the class imbalance, where the fraudulent transactions constitute a very

small percentage of all transactions. This can cause biased models towards the majority class, leading to ineffective fraud detection. Oversampling, under sampling, and creating synthetic data (e.g., SMOTE) are utilized to avoid this problem. For avoiding this issue, oversampling, under sampling, and synthetic data generation (e.g., SMOTE) are applied. Feature engineering is utilized to select and manipulate feature attributes from transactional data for optimal model performance. Since any delay in detection can lead to massive financial loss, real- time detection is required. Through data analysis of transactions and testing several ML algorithms, we aim to develop a low- false-positive, effective, and accurate real-time model. The outcome will help banks use ML-based solutions to combat fraud. Lastly, machine learning in fraud detection will enhance financial security and boost the trust of customers in the banking sector. Credit card fraud detection has been a problem explored at very long lengths in the last two years with numerous techniques being employed to increase the efficiency and effectiveness of fraud detection mechanisms. Statistics and rule - based mechanisms were among early solutions that were most dependent on predefined thresholds and patterns in identifying suspicious transactions. Whereas these techniques had helped in the detection of well - known fraud patterns, they were unable to match the latest and refined means of operating by fraudsters. Therefore, emphasis was laid on newer techniques like machine learning, which can process enormous amounts of transaction data and detect sophisticated, non - linear patterns. The outcome will help banks use ML-based Solutions to combat fraud.

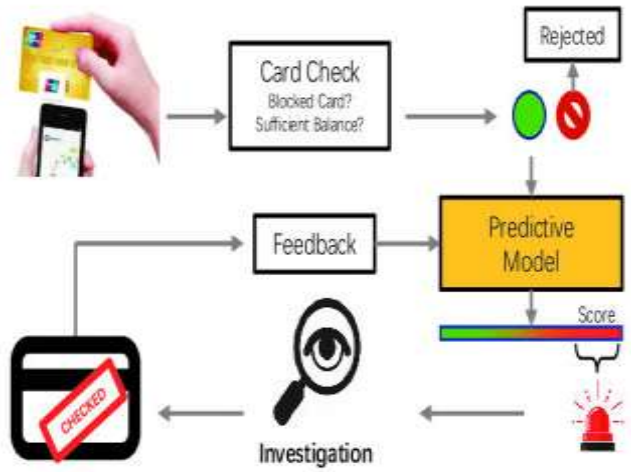
(i) II. Related Works

Credit card fraud detection has been a problem explored at very long lengths in the last two years with numerous techniques being employed to increase the efficiency and effectiveness of fraud detection mechanisms. Statistics and rule - based mechanisms were among early solutions that were most dependent on predefined thresholds and patterns in identifying suspicious transactions. Whereas these techniques had helped in the detection of well - known fraud patterns, they were unable to match the latest and refined means of operating by fraudsters. Therefore, emphasis was laid on newer techniques like machine learning, which can process enormous amounts of transaction data and detect sophisticated, non - linear patterns . Supervised learning algorithms have been widely used for credit card fraud detection due to the ability to label the transactions as fraud based on labelled data.



It has been demonstrated through research that logistic regression, SVM, and decision trees are effective algorithms for detecting fraud. For instance, ensemble algorithms like random forests and gradient boosting have been shown to be very accurate in aggregating strengths of multiple models. However, one of the largest supervised learning challenges is class imbalance in fraud datasets, where fraudulent transactions significantly outnumber legitimate transactions. To address this, techniques such as oversampling, under sampling , and synthetic data creation (e.g., SMOTE) have been employed to balance the dataset as well as to improve model performance Unsupervised learning techniques have also been identified as having the ability to detect unknown patterns of fraud without relying on labeled data. Clustering methods such as k - means and DBSCAN have been used to cluster similar transactions and separate out outliers that could be indicative of fraud.

Algorithms such as isolation forests and auto encoders have proven to detect well rare and unusual transactions which do not represent usual behavior. These unsupervised algorithms are particularly beneficial for discovering new types of fraud that are not based on well - known patterns and therefore complement supervised learning algorithms. Deep learning, which is a machine learning method, has been found to be a highly effective approach in detecting credit card fraud due to its ability to handle high - dimensional and complex data. Some of the neural networks utilized for handling sequences of transactions and extracting important features to detect fraud are recurrent neural networks (RNNs) and convolutional neural networks (CNNs). Research has established that deep learning models are capable of state - of- the - art performance on fraud detection problems, particularly when augmented with methods such as transfer learning and attention mechanisms. Yet, the computational complexity and resource needs of deep learning models pose a problem for There have also been recent experiments on integrating real - time fraud detection systems into banking infrastructure. Stream processing platforms such as Apache Kafka.



III. Comparison With Previous Methodology

For the comparison of supervised and unsupervised machine learning algorithms for predicting credit card fraudulent detection, we followed this types of methodology:

1. Data set collection
2. Dataset preparation
3. Supervised learning
4. Unsupervised learning

Comparison and analysis: Compare the performance of supervised and unsupervised machine learning algorithms based on their ability to detect fraudulent transactions accurately. Consider factors such as precision, recall, F1 score, and computational efficiently.

TABLE 1. Accuracy result for un-sampled data distribution

Metrics	Classifiers		
	Naïve Bayes	*k-Nearest Neighbour*	*Logistic Regression*
Accuracy	0.9737	0.9691	0.9824
Sensitivity	0.8072	0.8835	0.9767
Specificity	0.9741	0.9711	0.9824
Precision	0.0505	0.4104	0.0873
Matthews Correlation Coefficient	+0.1979	+0.5903	+0.2893
Balanced Classification Rate	0.8907	0.9273	0.9796

IV. Proposed Framework

1. Data Collection: Gather historical transaction data, including legitimate and fraudulent transactions.

2. Data Preprocessing: Clean and preprocess data, handling missing values, outliers, and normalization.

3. Feature Engineering: Extract relevant features, such as transaction amount, location, time, and cardholder behavior.

4. Data Split: Split data into training and testing sets.

5. Model Selection: Choose suitable algorithms, such as:
   - Traditional: Logistic Regression, Decision Trees
   - Machine Learning: Random Forest, SVM, k-NN
   - Hybrid: Ensemble methods, Stacking

6. Model Training: Train models on the training set.
7. Model Evaluation: Evaluate models using metrics like Accuracy, Precision, Recall, F1-score, and AUC-ROC.
8. Hyperparameter Tuning: Optimize model parameters for better performance.
9. Model Deployment: Deploy the best-performing model in a production environment.
10. Real-time Detection: Use the deployed model to detect fraudulent transactions in real-time.

11. Feedback Loop: Continuously update the model with new data and feedback to improve performance.

A. Algorithm

Bank credit card fraud detection is based on a set of machine learning algorithms to identify fraud transactions in the right way. Supervised machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, Support Vector Machines (SVM), and Gradient Boosting(XGBoost, LightGBM, CatBoost) are the very most popular algorithms for fraud classification. They are trained on labeled samples of known fraud historical transactions, and therefore they can detect true as well as the fraudulent transactions.

B. Module List and Descriptions

Data Collection and Preprocessing: This module is for gathering credit card transaction data samples from banking statements and open sources. Preprocessing data operations such as handling missing values, removing duplicates, transforming categorical variables into numerical variables, and normalizing numeric features are employed. Also, methods such as Synthetic Minority Over - sampling Technique (SMOTE) are employed to solve the imbalance in class and provide efficient training set.

Data Set: Downloading datasets from Kaggle can be beneficial for data analysis, machine learning, and research. This dataset has 4,850 records and 11 fields, with a size of around 319KB. It seems to deal with credit card transactions, with one record per transaction. The fields have some information about the transaction and cardholder. The "Unnamed: 0" column likely is an index or auto - indexed ID of each row. The "cc_num" column has the credit card number or its masked form, and "category" has the category of the transaction, e.g., grocery shopping, gas, online shopping.The "amt" column captures the amount spent on each transaction, and the "gender" column captures the gender of the cardholder. The "is_fraud" column is a binary flag, with 1 representing a fraudulent transaction and 0 representing a valid one. The "age" column contains the age of the cardholder, and the "trans_month" and "trans_year" columns detail the date of the transaction. Lastly, the "lat_dis" and distance (in latitude and longitude) between the transaction location and the cardholder's known location, which may aid in spotting suspicious activity or fraud due to location irregularities.

Feature Selection and Engineering: This module emphasizes identifying and selecting essential transaction attributes for fraud detection. Significant attributes including transaction amount, location, timing, device information, and behavioral patterns of a user are used to improve the accuracy of a model. Reduction of dimension and improvement in computing efficiency are attained through techniques like Principal Component Analysis (PCA) and Recu rsive Feature

Model Training and Classification: Various machine learning algorithms, including Decision Trees, Random Forest, Support Vector Machines (SVM) are implemented in this module. Supervised learning is used for labeled transaction data, while unsupervised techniques identify anomalies without predefined fraud .

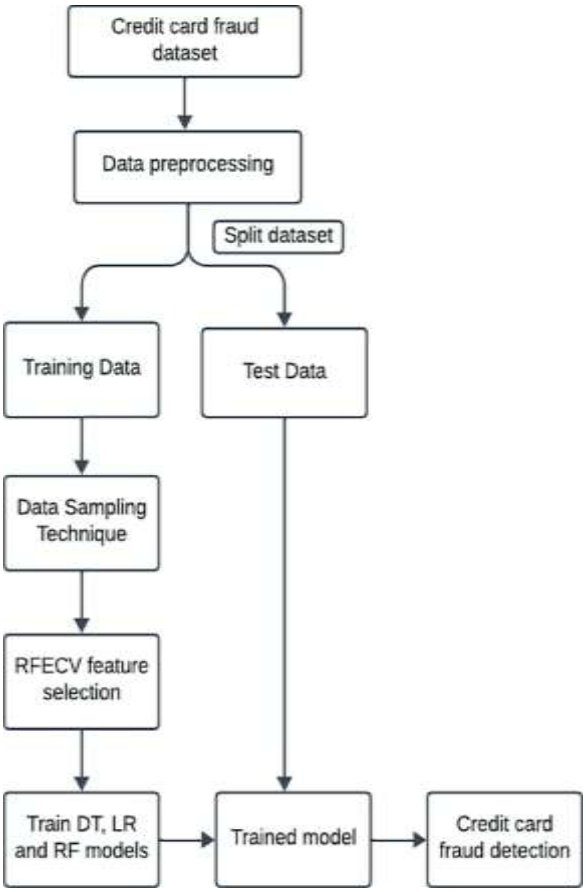
Anomaly Detection and Fraud Identification: This module utilizes unsupervised learning algorithms like Isolation Forest, One - Class SVM, and Clustering algorithms (K - Means, DBSCAN) to identify abnormal transaction patterns that could suggest fraudulent activity. These models assist in detecting new fraud patterns which could escape supervised models .

Real- time Fraud Detection and Alert System: The framework combines stream data processing platforms to identify fraud in real - time. In case of a transaction suspected of being fraudulent, an alert is triggered and subsequent verification processes, like multi- factor authentication, are invoked to block unauthorized transactions.

Performance Evaluation and Optimization: The last module assesses model performance on measures like accuracy, precision, recall, F1 - score, and ROC- AUC curves. Hyper parameter tuning methods like Grid Search and Random Search are employed to optimize the model and enhance credit.

1. Data Acquistion and the Preprocessing
2. Proposed methodology/Architecture
3. Implementation and Experimental setup
4. Performance Evaluation
5. Results and its Discussion
6. Conclusion and future Work

This study proposes a machine learning - driven fraud detection system that integrates supervised and unsupervised learning approaches to identify fraudulent transactions with high accuracy. The proposed model will employ a mix of feature engineering, anomaly detection, and ensemble learning to increase fraud detection effectiveness. Raw transaction data will be preprocessed for the first time by handling missing values, encoding categorical variables, and solving data imbalance through techniques like Synthetic Minority Over - sampling Technique (SMOTE)



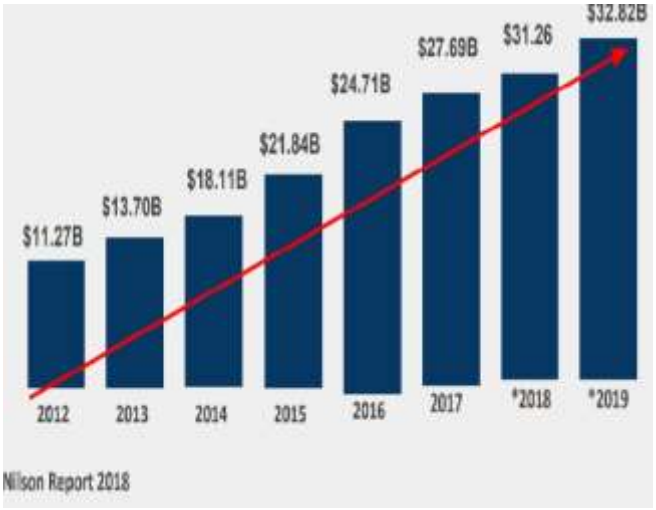
V. Results And Discussion

The machine learning approach proposed for detecting credit card fraud in banking is made up of several stages for the purpose of achieving accurate and an efficiently.

Data Acquisition & Preprocessing: The procedure begins with the collection of transactional data from bank statements or publicly available datasets. The data is preprocessed by handling missing values, removing duplicates, and encoding categorical features. Since fraud detection datasets are usually highly imbalanced, techniques like SMOTE and under sampling are used to balance the dataset so that the model does not lean towards non - fraudulent transactions.

Feature Selection & Engineering: Relevant transaction attributes, including transaction value, frequency, device usage, geographical location, and time - based patterns, are derived. Feature scaling and transformation methods, like Min - Max scaling or Standardization, are used to maintain consistency in data distribution

Model Selection & Training: The data set is separated into training and testing sets, and machine learning models are trained. It is classified through supervised learning models such as Logistic Regression, Decision Trees, Random Forest, SVM. Anomaly detection is performed through the unsupervised learning algorithms such as Auto encoders and Isolation Forest. **Anomaly Detection & Fraud Identification** : The trained models categorize transactions into fraudulent or legitimate using learned patterns. The blended techniques that use both supervised and unsupervised techniques help accuracy and fraud discovery improvement. Other ensemble learning approaches like Boosting and Stacking help increase the performance of Credit card fraudulent Detection.



Real- Time Detection & Alert Generation: A real - time fraud detection system is implemented in a banking system, where streaming data is scanned in real time for dubious fraudulent transactions. In case a doubtful transaction is detected, the system initiates automatic alerts and security measures like OTP verification or locking of an account to avert fraud .

Model Evaluation & Performance Optimization: The value of fraudulent detection models is measured by Precision, Recall, F1 - score, Confusion Matrix, and AUC- ROC curve. Methods of hyper parameter tuning such as Grid Search, Bayesian Optimization, and Neural Architecture Search (NAS) are applied for improving model performance .

VI. Conclusion

Machine learning- based credit card fraud detection in banking has been an effective tool for detecting and preventing fraudulent transactions in real time. Through the use of supervised, unsupervised, banks are able to identify sophisticated fraud patterns that rule - based systems are unable to detect. The application of techniques such as anomaly detection, ensemble learning, and feature engineering enhances the accuracy and reliability of fraud detection systems. In addition, real - time processing and automatic alerts allow financial institutions to take prompt action against suspicious activities, minimizing financial losses and boosting customer trust. As the pace of innovative fraudulent techniques accelerates, robust and elastic machine learning algorithms will become even more crucial in future years. Explainable AI, federated learning, blockchain protection, and quantum computing will steadily make fraud detection capabilities more user - friendly with assurances of privacy and

transparency. As banks continue the evolution and consolidation of new technology, they are able to develop an even more secure and efficient payment process, making fraud risk obsolete and further increasing overall transaction security.

REFERENCES

[1] M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.

[2] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson, 2021.

[3] T. Brown et al., "Language Models are Few-Shot Learners," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[4] OpenAI, "GPT-4 Technical Report," 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[5] J. Wei et al., "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models," in *Proc. NeurIPS*, 2022.

[6] H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models," 2023. [Online]. Available: https://arxiv.org/abs/2302.13971

[7] Y. Yao et al., "ReAct: Synergizing Reasoning and Acting in Language Models," 2023. [Online]. Available: https://arxiv.org/abs/2210.03629

[8] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Proc. NeurIPS*, 2020.

[9] LangChain Documentation, "Building applications with LLMs through composability," 2023. [Online]. Available: https://www.langchain.com

[10] LM Studio Documentation, "Local LLM Deployment Interface," 2024. [Online]. Available: https://lmstudio.ai

[11] M. Wooldridge and N. R. Jennings, "Intelligent Agents: Theory and Practice," *The Knowledge Engineering Review*, vol. 10, no. 2, pp. 115–152, 1995.

[12] D. Silver et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, 2016.