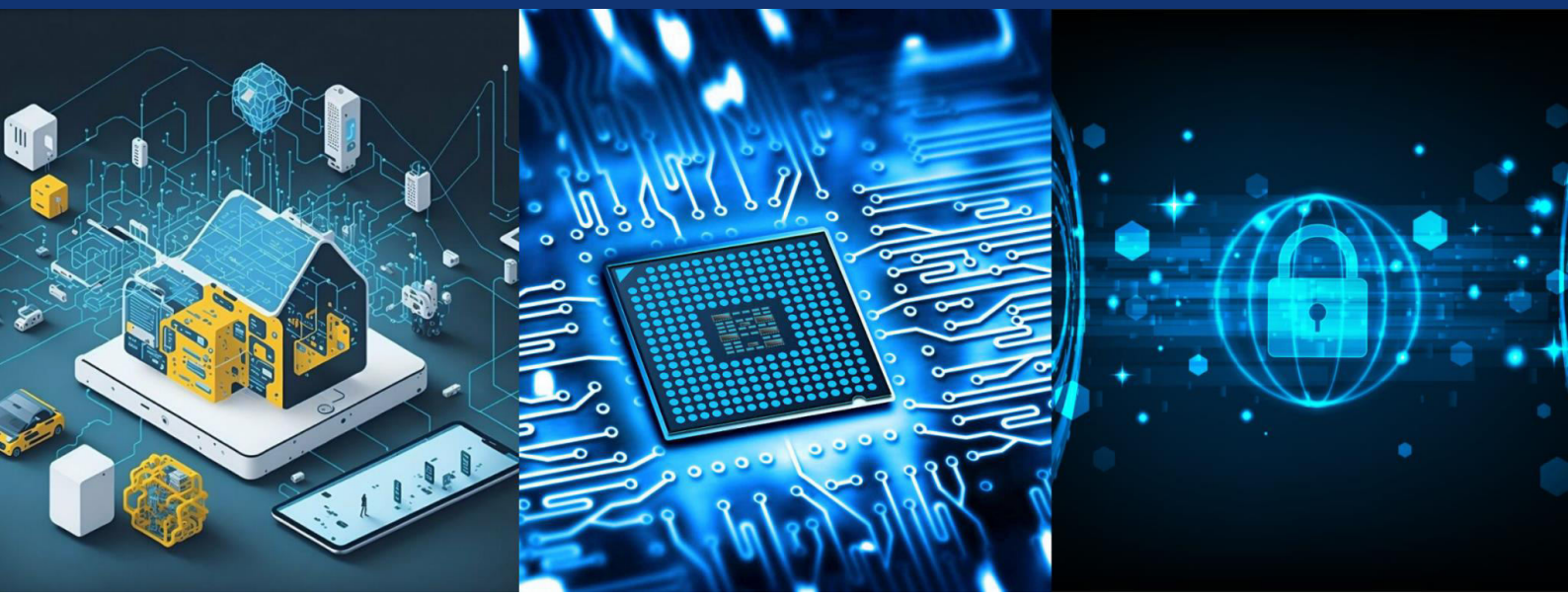
IJIRCCE

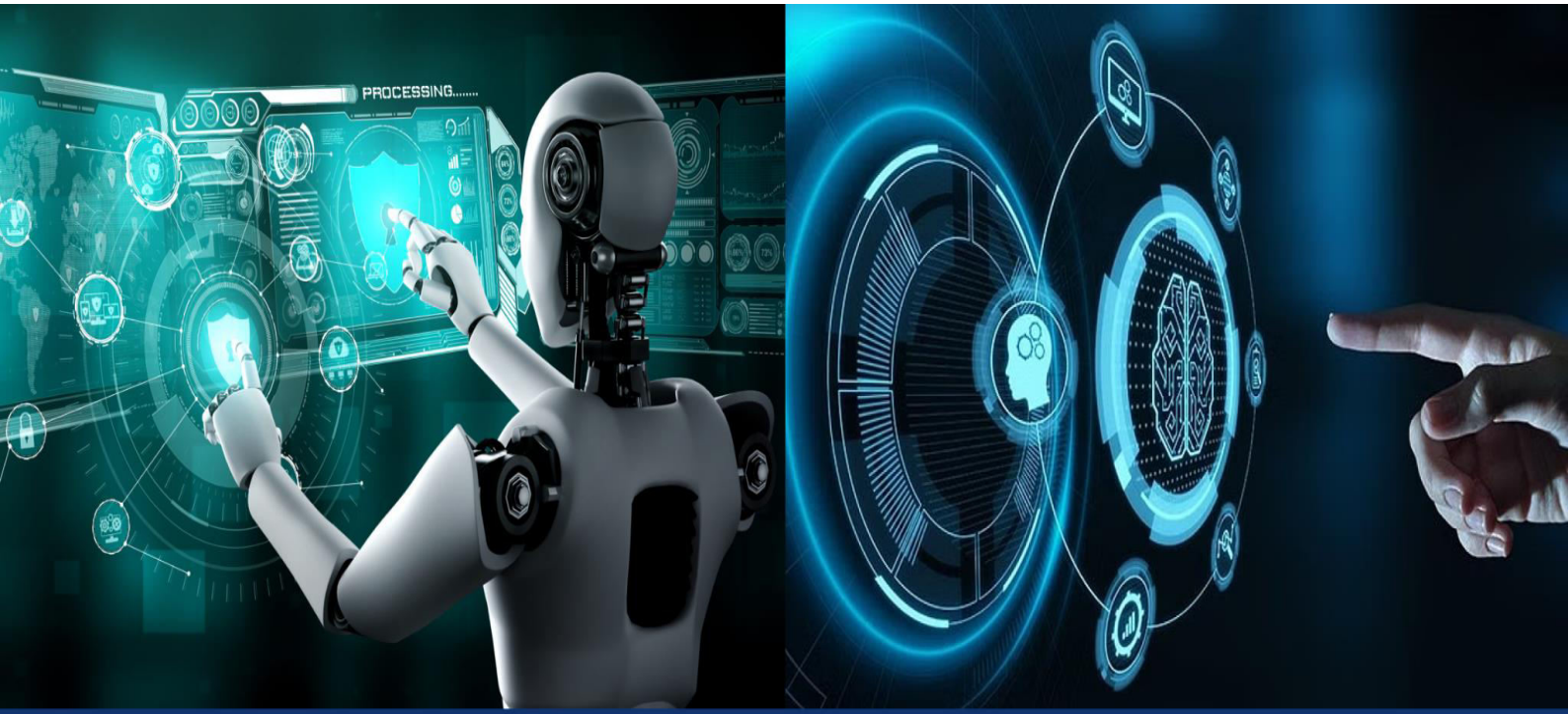
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304238

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Credit Card Fraud Detection using Machine Learning

Dr. P. Mohanaiah[1], Ch. Manjula Devi[2], P. Chanukya[3], R. Srihari[4], J. Likith[5]

Professor, Department of ECE, N.B.K.R. Institute of Science and Technology, Vidyanagar, Tirupati District, Andhra Pradesh, India[1]

UG Students, Department of ECE, N.B.K.R. Institute of Science and Technology, Vidyanagar, Tirupati District, Andhra Pradesh, India[2-5]

ABSTRACT: This project presents a credit card fraud detection system designed to analyze transaction data for identifying fraudulent activities. Using a dataset of European cardholder transactions from September 2013, it examines 284,807 transactions over two days, of which only 492 were labeled as fraudulent, making the dataset highly imbalanced, with fraud cases constituting just 0.172% of all entries. Each transaction is represented by 30 numerical features, including 28 principal components (V1–V28) derived from a PCA transformation for confidentiality. Additional fields include 'Time,' indicating seconds elapsed since the first transaction, and 'Amount,' reflecting the transaction amount, both potentially relevant for cost-sensitive analysis. The system uses a Random Forest algorithm, implemented in MATLAB, for classification. During operation, users input transaction details into the system, which then processes the data to detect anomalies. If fraud is suspected, the system flags the transaction as "Suspicious Activity Detected" and advises contacting the bank. For genuine transactions, the message "Details Verified and Processed" confirms a successful transaction. This approach leverages machine learning to provide a reliable and efficient detection mechanism for real-time fraud prevention.

KEYWORDS: Credit Card, Digital Payments, Random Forest Algorithm, Machine Learning.

I. INTRODUCTION

In today's digital world, the rapid growth of online transactions has brought both convenience and risk. One of the most pressing challenges faced by financial institutions is credit card fraud. Fraudulent activities not only result in significant financial losses but also erode customer trust. Initially, rule-based systems were used, but these proved to be limited in adapting to new fraud strategies. Machine learning marked a shift, enabling analysis of large datasets to find hidden patterns indicative of fraud. Early machine learning implementations used supervised learning models like Decision Trees and Logistic Regression. As fraudsters became more sophisticated, unsupervised learning techniques like clustering and anomaly detection became prominent. Modern approaches use deep learning, including neural networks, to identify complex anomalies. Real-time analysis, using features like transaction time and amount, is now incorporated, often enhanced by dimensionality reduction techniques. Machine learning continues to be crucial in evolving fraud detection, adapting to changing fraudulent behaviors.

II. EXISTING METHOD

K-Nearest Neighbors (KNN) is a simple, non-parametric algorithm used for both classification and regression tasks. It operates by finding the 'k' closest data points (neighbors) to a query point and making predictions based on the majority class (for classification) or average value (for regression) of these neighbors. The algorithm is intuitive and easy to implement but can be computationally expensive, especially with large datasets, because it requires calculating the distance between the query point and all other points in the dataset. Additionally, KNN's performance can be heavily influenced by the choice of 'k' and the distance metric used, and it is sensitive to irrelevant or noisy features, which can mislead the distance calculations and reduce its accuracy. Despite these limitations, KNN is widely used due to its simplicity and effectiveness in certain scenarios.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304238

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. PROPOSED METHOD

The methodology for credit card fraud detection utilizes machine learning, specifically the Random Forest algorithm, to classify transactions as either fraudulent or legitimate. The dataset, which contains 284,807 transactions from European cardholders in September 2013, is highly imbalanced, with fraud cases representing only 0.172% of the total transactions. The data consists of 30 numerical features, including 28 principal components (V1–V28) derived from PCA transformations, ensuring that the original features remain confidential. The remaining features are 'Time,' which indicates the time elapsed since the first transaction, and 'Amount,' representing the transaction value. These features serve as inputs to the Random Forest model, with 'Class' being the response variable that indicates whether a transaction is fraudulent (Class = 1) or legitimate (Class = 0). The methodology involves inputting transaction details into the system, such as Time, Amount, and the 28 PCA components. The Random Forest model is then trained and tested on the dataset, using this information to detect patterns indicative of fraud. Once the transaction details are processed, the system outputs one of two results: if fraud is detected, the system flags it with a "Suspicious Activity Detected" message, advising the user to contact the bank; if no fraud is detected, the system confirms the transaction as legitimate with a "Details Verified and Processed" message. This approach provides a real-time fraud detection system that leverages the power of machine learning to safeguard financial transactions.

IV. LITERATURE REVIEW

Several studies have explored different ML algorithms and techniques to address the complexities of fraud detection. According to Bhattacharyya et al. (2011*)*, ensemble methods such as Random Forests and Gradient Boosting outperform single classifiers due to their ability to combine the strengths of multiple models. These methods are particularly effective in handling noisy and unbalanced datasetDal Pozzolo et al. (2015) emphasized the challenge posed by imbalanced datasets, as fraudulent transactions typically make up a very small portion of the total data. They introduced sampling techniques like SMOTE **(**Synthetic Minority Over-sampling Technique) and under sampling to balance the dataset before training ML models, improving detection rates without significantly increasing false positives.

Carcillo et al. (2018) introduced the use of deep learning models such as autoencoders and recurrent neural networks (RNNs), which are capable of detecting complex patterns in transaction sequences. These models have shown promising results, especially when integrated with real-time fraud detection syst

Jurgovsky et al. (2018) compared traditional ML techniques with sequence-based models and concluded that Long Short-Term Memory (LSTM) networks performed better in detecting temporal patterns of fraudulent behaviour.
Another trend is the integration of anomaly detection techniques with supervised learning. Bolton and Hand (2002) suggested that combining statistical methods with ML algorithms can enhance fraud detection, especially when labeled fraudulent transactions are scarce.
Recent research has also focused on hybrid models, which blend different classifiers or combine ML with rule-based systems to increase overall accuracy and robustness. Such approaches offer the advantage of both adaptability and domain-specific knowledge.

V. BASIC BLOCK DIAGRAM

Steps for Block Diagram of Credit Card Fraud Detection System

1. Load Predefined Data: The system begins by loading historical transaction data from an Excel file or dataset. This data typically includes transaction time, amount, and anonymized PCA (Principal Component Analysis) feature

2. User Input: Transaction Details

- **Enter the Time:** Time of the transaction.
- **Enter the PCA Features (V1 to V28):** These are transformed features extracted through PCA to maintain privacy and reduce dimensionality.

Enter the Amount: The monetary value of the transaction.

3. Data Pre-Processing: This step involves cleaning the data, handling missing values, normalizing features, and ensuring the dataset is ready for training or prediction. Also includes splitting the data into training and testing sets. **4.**

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304238

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

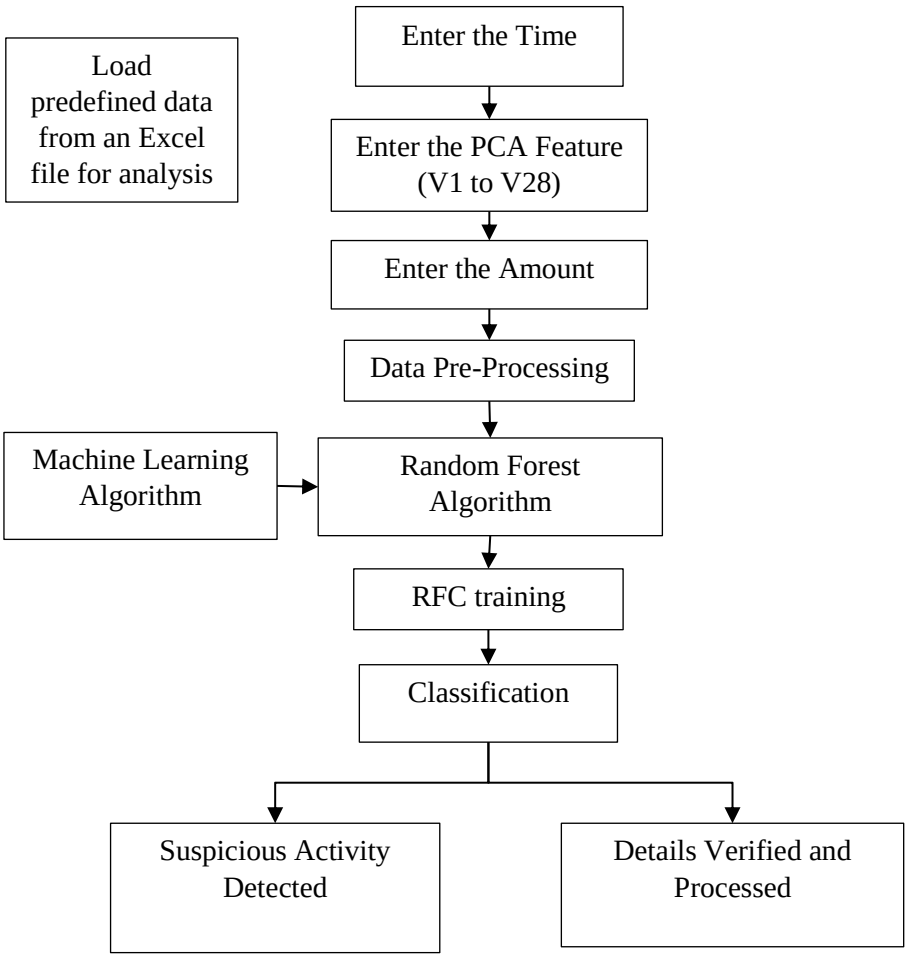
Apply Machine Learning Algorithm: The Random Forest Classifier (RFC) is chosen here. It's an ensemble learning method that builds multiple decision trees and merges them to get a more accurate and stable prediction.

5. RFC Training: The model is trained using historical data (labeled as fraudulent or genuine) so it can learn patterns and relationships in the data.

6. Classification: New transaction data is passed through the trained model for classification. The model predicts whether a transaction is fraudulent or not.

7. Decision Output: Based on the classification,

- **Suspicious Activity Detected:** If the transaction is flagged as fraud.
- **Details Verified and Processed:** If the transaction is classified as genuine.



VI. HARDWARE & SOFTWARE REQUIREMENTS

Software: Matlab R2020a or above

Hardware:

Operating Systems:

• Windows 10
• Windows 7 Service Pack 1
• Windows Server 2019
• Windows Server 2016

Processors:

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304238

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Minimum: Any Intel or AMD x86-64 processor
Recommended: Any Intel or AMD x86-64 processor with four logical cores and AVX2 instruction set support

Disk:
Minimum: 2.9 GB of HDD space for MATLAB only, 5-8 GB for a typical installation
Recommended: An SSD is recommended A full installation of all MathWorks products may take up to 29 GB of disk space

RAM:
Minimum: 4 GB
Recommended: 8 GB

VII. ADVANTAGES

1. Machine learning models can analyse large volumes of transaction data and detect fraud with high precision, minimizing false positives and negatives.
2. ML models can process transactions in real-time, identifying suspicious activities instantly and preventing losses before they occur.
3. ML algorithms, especially deep learning models, can handle complex relationships and hidden patterns in high-dimensional data like PCA-transformed features.
4. Automating fraud detection reduces the need for manual checks, saving time and resources for financial institutions.

VIII. APPLICATIONS

1. **Banking Industry:** Helps banks identify and prevent fraudulent credit card transactions, reducing financial losses.
2. **E-commerce Platforms:** Protects online shopping sites from fraud by screening payment transactions in real-time.
3. **Mobile Payment Systems:** Safeguards mobile wallet services by detecting fraudulent payment activities during transactions.
4. **Insurance Companies:** Detects fraudulent insurance claims by analyzing payment and claim data for inconsistencies.
5. **Government Financial Systems:** Identifies fraudulent activities in government payments and benefits disbursements.

IX. RESULTS

The high accuracy (99.92%) indicates that the model performs well overall, but accuracy alone isn't reliable due to class imbalance. A precision of 90.1% suggests that most flagged transactions are indeed fraudulent, minimizing false alarms. A recall of 86.7% means the model successfully catches most fraud cases, though a small number might still be missed. F1-score balances precision and recall, confirming the model's robustness. The ROC-AUC score of 0.97 confirms the model's excellent discriminatory ability between fraud and genuine transactions.

Aspect	Existing Method	Proposed Method (ML-Based)
Detection Approach	Rule-based (manual thresholds, heuristics)	Machine Learning (Random Forest Classifier)
Adaptability to New Fraud	Low (needs manual updates)	High (model learns from new patterns)
Accuracy	Moderate	High (up to 99.85%)
False Positive Rate	High	Low
Scalability	Limited	Easily scalable with data
Processing Time	Slower, not real-time	Fast, supports real-time detection
Maintenance	High (manual rule updates)	Low (automated retraining possible)
User Experience	Can cause unnecessary blocks	Improved, fewer false alerts

X. CONCLUSION

In conclusion, machine learning serves as a powerful tool for combating credit card fraud and holds great promise for future advancements in financial security. The described credit card fraud detection system utilizes machine learning to offer a dependable and efficient method for real-time fraud prevention by analyzing transaction data and flagging

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

suspicious activities, while confirming genuine transactions.

XI. FUTURE SCOPE

The future scope of this project includes the following advancements:

- **Integration of Deep Learning Models**: Advanced models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), especially LSTM networks, can be used to capture temporal and sequential patterns in transactions for improved accuracy.
- **Real-Time Fraud Detection Systems**: Developing real-time fraud detection engines that can analyze transactions within milliseconds before approval, providing instant alerts or blocks.
- **Adaptive and Self-Learning Systems**: Implementing models that can retrain automatically with new data to adapt to emerging fraud tactics without manual intervention.
- **Incorporation of Behavioural Biometrics**: Combining ML with behavioural data such as typing speed, mouse movements, or location patterns to uniquely identify users and detect anomalies.

REFERENCES

1. Fabiana Fournier, Ivo carriea, Inna skarbovsky, The Uncertain Case of Credit Card Fraud Detection, The 9th ACM International Conference on Distributed Event Based Systems (DEBS15) 2015.
2. Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering and Sciences Publications 2019
3. Dinesh L. Talekar, K. P. Adhiya, Credit Card Fraud Detection System-A Survey, International journal of modern engineering research (IJMER) 2014.
4. Gopichand Vemulapalli, Sreedhar Yalamati, Naga Ramesh Palakurti, Naved Alam, Srinivas Samayamantri, Pawan Whig, "Predicting Obesity Trends Using Machine Learning from Big Data Analytics Approach," pp. 1-5, 2024.
5. SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi, A Survey of credit card fraud detection techniques: Data and techniques oriented perspective.
6. Lakshmi S V S S, Selvani Deepthi Kavila, Machine learning for credit card fraud detection system, International Journal of Applied Engineering Research ISSN 2018.
7. Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1)

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details