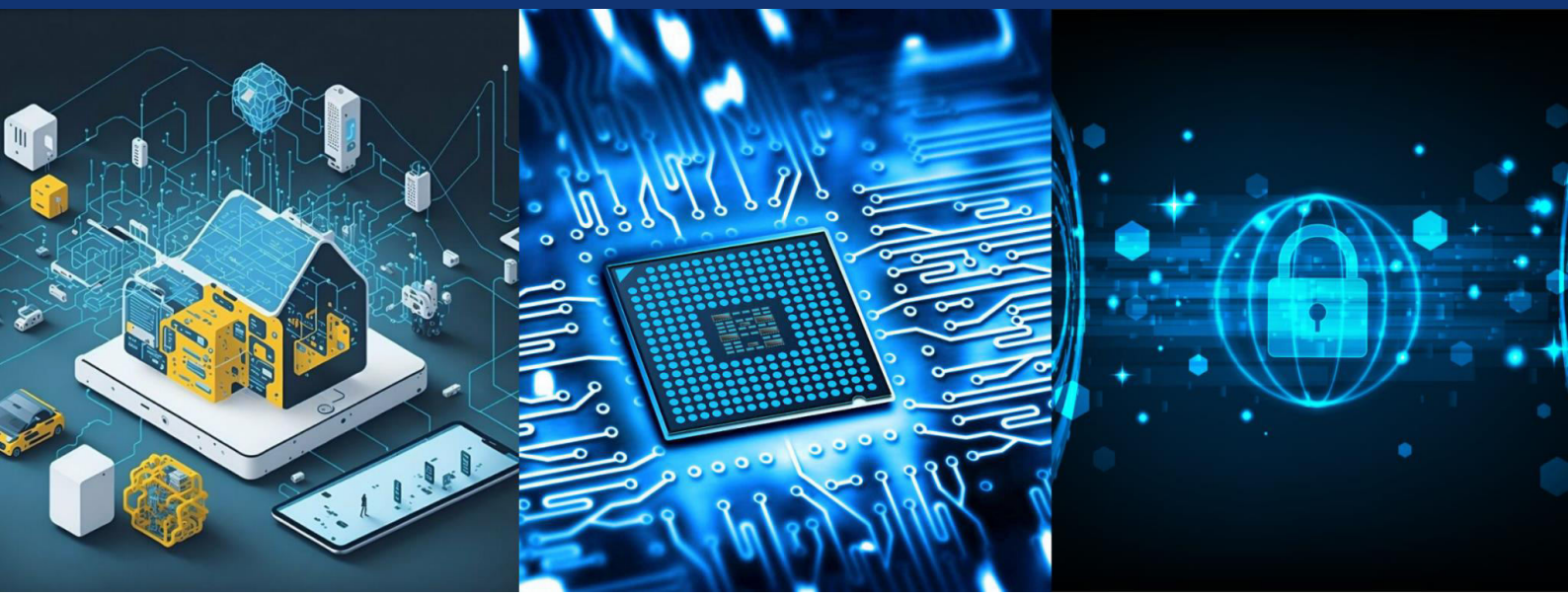
IJIRCCE

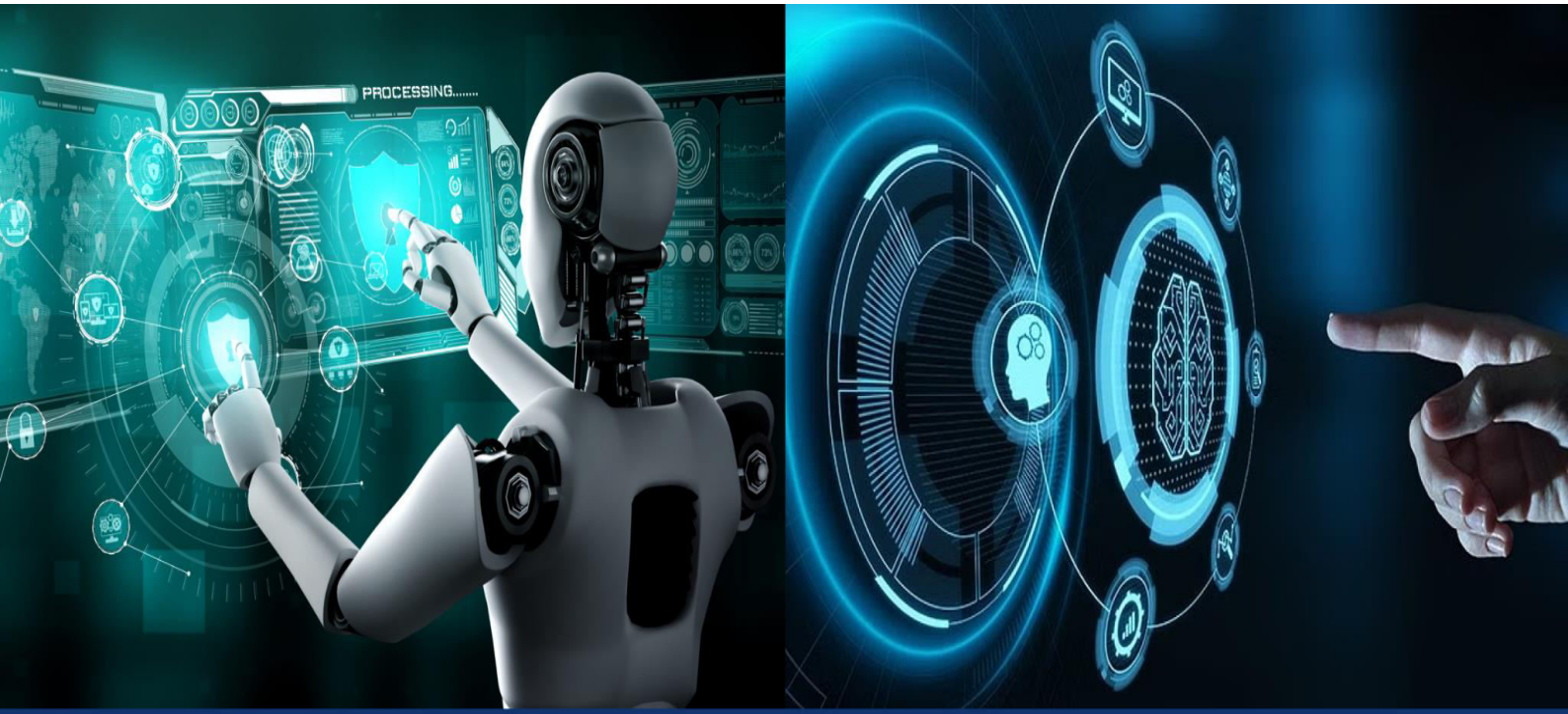
ISSN(O): 2320-9801

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.2025.1304237

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)

Fraud Detection on Bank Transaction using Machine Learning with Python

Dr Latha P H, Shreya K N, Sowjanya N, Chandana K, Dhanushree B

Head of Department, Department of Information Science and Engineering, Sambhram Institute of Technology, Bangalore, India

Bachelor of Engineering, Department of Information Science and Engineering, Sambhram Institute of Technology, Bangalore, India

Bachelor of Engineering, Department of Information Science and Engineering, Sambhram Institute of Technology, Bangalore, India

Bachelor of Engineering, Department of Information Science and Engineering, Sambhram Institute of Technology, Bangalore, India

Bachelor of Engineering, Department of Information Science and Engineering, Sambhram Institute of Technology, Bangalore, India

ABSTRACT: A digital framework designed to enhance the security and accuracy of financial systems is a fraud detection system that uses machine learning techniques. To ensure only legitimate transactions are processed, the system analyzes transaction behavior using intelligent algorithms. Techniques like Logistic Regression, Random Forest, and XGBoost are applied for classification and anomaly detection. Feature selection and data preprocessing are carried out to optimize model performance. Once a transaction is initiated, the system checks it against historical data patterns stored in the database. If the transaction matches typical behavior, it is approved; otherwise, it is flagged as suspicious or fraudulent. This automated approach significantly reduces manual intervention, increases detection speed, and minimizes financial loss. Overall, the use of machine learning in bank fraud detection offers a scalable, adaptive, and secure solution for real-time financial monitoring.

KEYWORDS: Machine Learning, Fraud Detection, Anomaly Detection, Classification, Banking Security.

I. INTRODUCTION

It is a part of artificial intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.For banks has become very difficult for detecting the fraud in bank payments. Machine learning plays a vital role for detecting the financial fraud in the transactions. For predicting these transactions in the proposed system we make use of machine learning methodology, past data has been collected and new features are been used for enhancing the predictive power. At its core, the application employs a pre-trained machine learning model that is designed to classify bank transactionsas either "Fraudulent" or "Benign".

The model is stored in a serialized format using Pickle and the application interacts with this model to make real-time predictions based on user input.The web application enables users to upload a dataset for review, preview the data, and enter specific transaction details through a form for fraud detection. Users can also access performance metrics and a visual representation of the results. This system provides an accessible and interactive platform for detecting fraud in bank transactions aiming to prevent financial losses caused by fraudulent transactions. The entire workflow from file upload, data preview, prediction, and result display is powered by Flask and scikit-learn**,** leveraging machine learning for efficient fraud detection.

The proposed system processes the input transaction, extracts key features, and compares it against patterns stored in the system database. If the transaction aligns with legitimate behavior, it is approved. Otherwise, it is flagged for further investigation or automatically rejected. By using intelligent algorithms that can learn and adapt, the system

enhances the speed and accuracy of fraud detection, minimizes financial losses, and reduces dependency on manual intervention.

ImplementationStrategies

A structured pipeline including data collection, preprocessing, feature engineering, model training, and real-time prediction. Various machine learning algorithms such as Logistic Regression, Decision Trees, Random Forest, XGBoost, and Neural Networks were tested. Python with libraries like pandas, scikit-learn, and TensorFlow was used for development. Model performance was optimized using evaluation metrics like Precision, Recall, F1-score, and AUC-ROC.

Challengesandconcerns

A major challenge is the class imbalance, as fraudulent transactions are significantly fewer than legitimate ones. High false positive rates can lead to customer dissatisfaction and unnecessary operational costs. The dynamic nature of fraud patterns requires continuous model updates to maintain accuracy. Privacy and data protection concerns arise due to the sensitive nature of financial data. Ensuring ethical and transparent AI usage is also a critical concern.

GlobalCaseStudies

European and Latin American banks using machine learning to enhance fraud detection accuracy and reduce false positives. Studies emphasize the effectiveness of supervised, unsupervised, and cost-sensitive learning approaches on real-world financial data.

PotentialImpactonDemocracy

Effective fraud detection strengthens public trust in financial institutions, which is vital for a stable democracy. It promotes economic security and protects citizens from financial exploitation. Ethical AI use ensures individual privacy rights are respected. Overall, it supports transparency, fairness, and accountability in digital financial systems.

FutureDirections

- Incorporating CNNs and RNNs for complex pattern recognition in sequential transaction data.
- Real-time model updating to combat adaptive fraudulent behavior.
- Enabling banks to train models collectively without sharing raw data, enhancing privacy.
- Making fraud predictions transparent to improve regulatory compliance and user trust.

II. PROBLEMSTATEMENT

Enhancing User Experience with the increasing digitization of financial transactions, banks and financial institutions face a growing threat of fraudulent activities. Fraudulent transactions not only lead to financial losses but also undermine trust in the banking system. Traditional methods of fraud detection, such as manual reviews and rule-based systems, have limitations in detecting sophisticated fraud patterns in real-time. As a result, there is a need for more advanced and efficient fraud detection systems that can quickly and accurately identify fraudulent transactions.

III. LITERATUREREVIEW

- **Halvaiee, H., & Akbari, M. (2014): An Effective Fraud Detection Model Based on Artificial Immune System**

This paper presents a novel fraud detection model based on artificial immune systems (AIRS). It improves detection accuracy by 25%, reduces cost by 85%, and speeds up response time by 40% compared to traditional algorithms. The model mimics biological immunity to adaptively detect fraud patterns. The study also emphasizes the importance of robust and flexible algorithms that can handle evolving fraud scenarios, making it highly applicable to real-time banking systems.

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)

- **Z Bahnsen, A. C., Aouada, D., & Ottersten, B. (2016): Feature Engineering Strategies for Credit Card Fraud Detection**

This paper introduces a feature engineering strategy using transaction time behavior based on the von Mises distribution. It also proposes a cost-sensitive evaluation approach for model performance. Using real-world credit card data, the study shows how engineered features can enhance detection precision. It highlights the role of time-based behavior and cost-aware evaluation metrics in building more accurate and financially viable fraud detection systems.

- **Randhawa, K., Loo, C. K., Seera, M., Ibrahim, Z., & Lim, C. P. (2018): Credit Card Fraud Detection Using AdaBoost and Majority Voting**

This study evaluates several machine learning models and proposes a hybrid ensemble using AdaBoost and majority voting. It demonstrates that ensemble learning outperforms individual models, especially under noisy data conditions. The paper also explores the robustness of voting systems in fraud detection and shows improved results on public datasets. It concludes that combining classifiers enhances accuracy and reduces false positives in fraud prediction tasks

- **Porwal, U., & Mukund, S. (2018): Credit Card Fraud Detection in E-commerce: An Outlier Detection Approach**

This paper proposes a clustering-based method for fraud detection, assuming stable behavior among legitimate users. Fraud is detected as deviations from this norm. The authors emphasize the importance of outlier detection in the absence of labeled data and argue that precision-recall is more appropriate than ROC for evaluation. Their approach is resistant to evolving fraud patterns, making it suitable for dynamic e-commerce environments.

- **Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2023): Fraud Detection in Banking Data by Machine Learning Techniques**

This paper presents a fraud detection framework using LightGBM, XGBoost, and CatBoost with Bayesian hyperparameter tuning. It also proposes deep learning with class weight balancing to address data imbalance. The system achieves an F1-score of 0.81 and MCC of 0.81 on real data. It emphasizes the benefits of combining deep learning with optimized boosting models for scalable and accurate fraud detection.

IV. OBJECTIVES

The objective of this project is to develop a reliable, intelligent, and real-time fraud detection system that leverages machine learning to identify and prevent fraudulent bank transactions. The system aims to:

- Detect Suspicious Activities Accurately – Apply advanced ML algorithms to identify unusual transaction patterns and flag potential fraud.
- Enhance Banking Security – Integrate predictive analytics to proactively secure banking operations and reduce financial risks.
- Improve Response Time and Efficiency – Automate fraud detection to enable quick responses and reduce the need for manual review.
- Prevent Financial Loss and Fraud – Minimize false positives and ensure timely intervention to block fraudulent transactions.
- Ensure Data Privacy and Security – Implement encryption and privacy-preserving techniques to protect voter data and prevent unauthorized access.

V. ARCHITECTURE

The architecture of the Fraud Detection System using Machine Learning is designed to ensure security, adaptability, and real-time response, where transactions are continuously monitored and validated. When a user initiates a transaction, essential details such as transaction amount, time, location, and device ID are captured. These parameters are processed by a machine learning model that has been trained on historical transaction data to detect suspicious behavior. The system uses behavioral profiling and anomaly detection to evaluate the legitimacy of each transaction.

A fraud detection module processes the real-time transaction data through a series of pre-trained classifiers like Random Forest, XGBoost, or Neural Networks. These models compare the current transaction with the user's past behavior stored in the encrypted database. If a significant deviation is detected, the system flags the transaction as

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304237

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

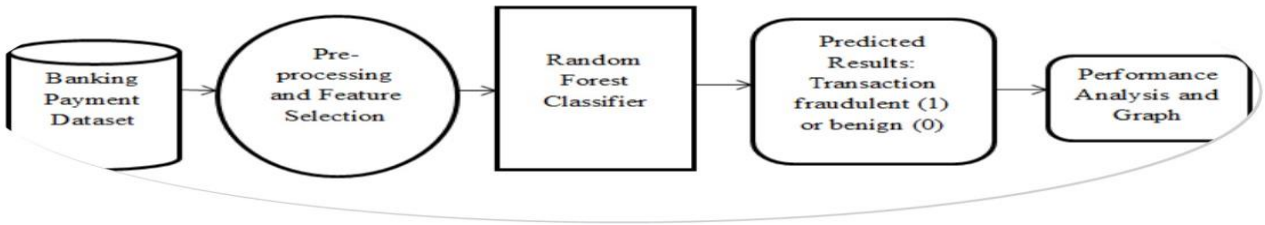
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

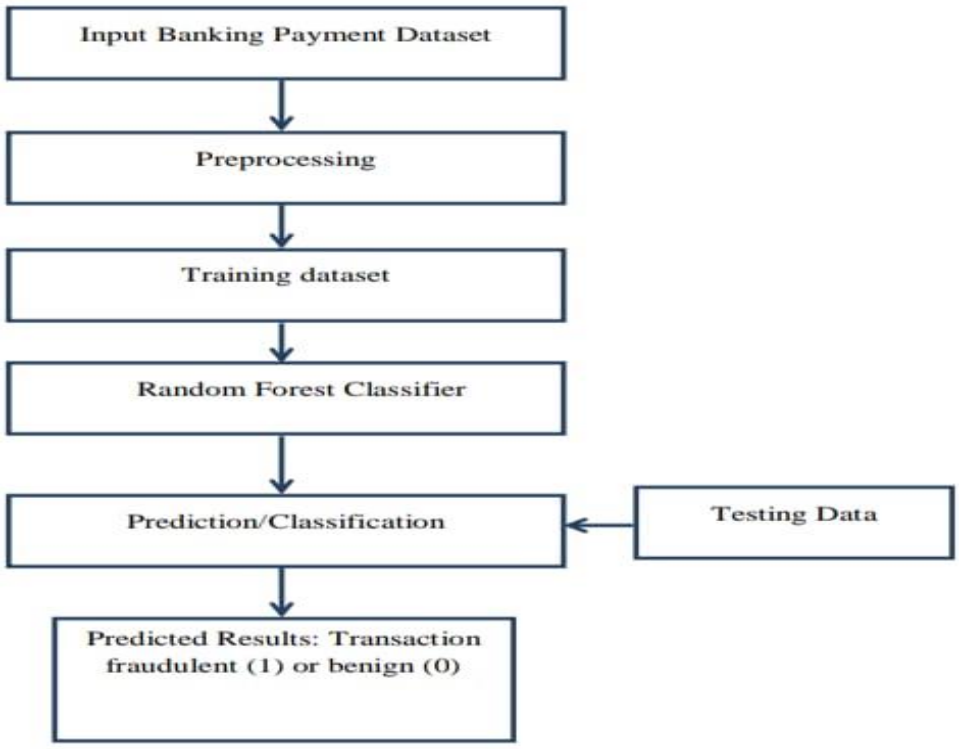
suspicious or fraudulent. The authentication server validates the transaction against the customer's known behavior profile, device fingerprint, and geo-location. If the transaction passes all verification checks, it is approved and stored in the secure transaction ledger; if not, it is either blocked or sent for manual review.

The backend system performs key operations including data preprocessing, feature extraction, prediction, and model evaluation. This architectural framework ensures a streamlined and tamper-resistant fraud detection process by integrating adaptive machine learning algorithms and secure communication protocols. Additionally, the system uses feedback loops to learn from false positives or missed fraud cases, thereby continuously improving detection accuracy over time.

The system includes an admin interface for authorized banking personnel to monitor flagged transactions, retrain models, and generate reports for audits. This allows efficient fraud case management and system transparency. The database plays a critical role in storing encrypted user profiles, transaction logs, and model parameters. Additional technologies include secure APIs, SSL encryption, and authentication tokens to ensure privacy and security in every stage of processing.



VI. IMPLEMENTATION



© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304237

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)

A Fraud Detection System for Bank Transactions using Machine Learning involves integrating several modules to process data, identify patterns, and detect anomalies. These modules include data input, preprocessing, model training, classification, prediction, and result interpretation.

- Pandas/NumPy: For loading the banking dataset and data manipulation.
- Scikit-learn: For implementing the Random Forest Classifier and evaluation metrics.
- Matplotlib/Seaborn: For visualizing transaction patterns and model performance.
- Jupyter Notebook/Colab: For interactive development and testing.
- Python Libraries: For handling preprocessing, encoding, and splitting datasets.
- CSV/Database: For storing transaction records and prediction results.

The implementation starts by importing the bank transactions dataset, usually a CSV file containing transaction details. Preprocessing is then performed to clean the data, handle missing values, normalize numerical features, and encode categorical values. The dataset is split into training and testing subsets to evaluate model performance. A Random Forest Classifier is trained on the training data to learn patterns associated with fraudulent and benign transactions. Once trained, the model is tested on the unseen testing dataset. The classifier makes predictions, assigning each transaction a label: fraudulent (1) or benign (0). Finally, the predicted results are analyzed and stored, providing insights into potentially suspicious activities.

VII. RESULTS

Initially, the system loads the transaction dataset which includes features such as transaction ID, amount, time, origin, destination, and transaction type. This information is stored in a structured format for analysis. The dataset is then divided into training and testing subsets to prepare it for machine learning. Data preprocessing is applied, including feature scaling, outlier removal, and encoding of categorical variables to make the data suitable for model training.

A Random Forest Classifier is trained on the cleaned training data. It builds multiple decision trees to learn patterns in fraudulent and legitimate transactions. During testing, the model evaluates the unseen data by comparing actual outcomes with predicted values. The classifier also provides probabilities, enhancing the reliability of the fraud prediction.

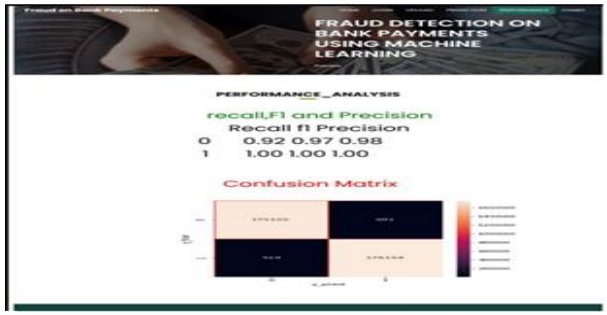
The model captures transaction behavior by analyzing multiple features in parallel, improving detection accuracy. These parallel evaluations, much like frames in a video, allow the model to gather stronger evidence before classifying a transaction. Fraud detection is a machine learning task that focuses on classifying whether a financial activity is legitimate or suspicious. It can work with real-time or batch transaction data.The system classifies each transaction as either fraudulent (1) or benign (0). The final results can be visualized through confusion matrices, accuracy scores, and precision-recall graphs. These metrics help in assessing the model's performance and ensuring secure, data-driven banking.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI:10.15680/IJIRCCE.2025.1304237

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)



VIII. CONCLUSION

The implementation utilized a structured dataset of banking transactions, including both fraudulent and legitimate examples. Essential preprocessing steps—such as data cleaning, normalization, and encoding—prepared the dataset for training. The Random Forest Classifier was chosen for its robustness and ability to handle large datasets with high dimensionality. By training the model on labeled data, it was able to learn complex patterns that distinguish fraudulent behavior from normal transactions.

The model's performance was evaluated using key metrics such as accuracy, precision, recall, and the F1-score. These metrics ensure that the classifier not only correctly identifies most fraudulent transactions but also minimizes false positives, which could otherwise lead to unnecessary account freezes or customer inconvenience. Visual tools such as confusion matrices and ROC curves further validated the model's effectiveness and helped in fine-tuning its performance.

One of the critical advantages of using a machine learning approach is its adaptability. As new types of fraud emerge, the model can be retrained on updated datasets to improve detection accuracy. Additionally, using ensemble techniques like Random Forests enhances the reliability and stability of the prediction results by combining multiple decision trees and reducing the risk of overfitting.

While the system has shown promising results, it also highlights areas for future development. Incorporating real-time detection capabilities, integrating additional data sources (such as user behavior analytics), and deploying the model into a secure banking environment are essential next steps. Moreover, employing deep learning models or anomaly detection techniques could further improve fraud detection in highly unbalanced datasets.

In summary, this project successfully demonstrates that Machine Learning, particularly with models like Random Forest, can significantly enhance the detection of fraudulent bank transactions. The results support the feasibility of deploying such models in real-world banking environments to improve security, reduce financial losses, and enhance customer trust. With continued advancements and integration into existing banking systems, machine learning-powered fraud detection has the potential to revolutionize the way financial crimes are identified and prevented.

REFERENCES

ResearchPapersandArticles:

1. Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602-613.
2. Pozzolo, A. D., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784-3797.
3. Thulasiram, P. P. (2025). EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI): ENHANCING TRANSPARENCY AND TRUST IN MACHINE LEARNING MODELS.
4. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569.

Books:

5. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Chapman and Hall/CRC.
6. West, J., & Bhattacharya, M. (2016). Intelligent Financial Fraud Detection: A Comprehensive Guide. Springer.

Frameworks and Libraries:

7. Scikit-learn – Scikit-learn Documentation (https://scikit-learn.org)
8. Pandas – Pandas Documentation (https://pandas.pydata.org)
9. Matplotlib/Seaborn – Visualization Libraries (https://matplotlib.org, https://seaborn.pydata.org)
10. XGBoost – XGBoost Documentation (https://xgboost.readthedocs.io)
11. TensorFlow/Keras – TensorFlow for ML Modeling (https://www.tensorflow.org)

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