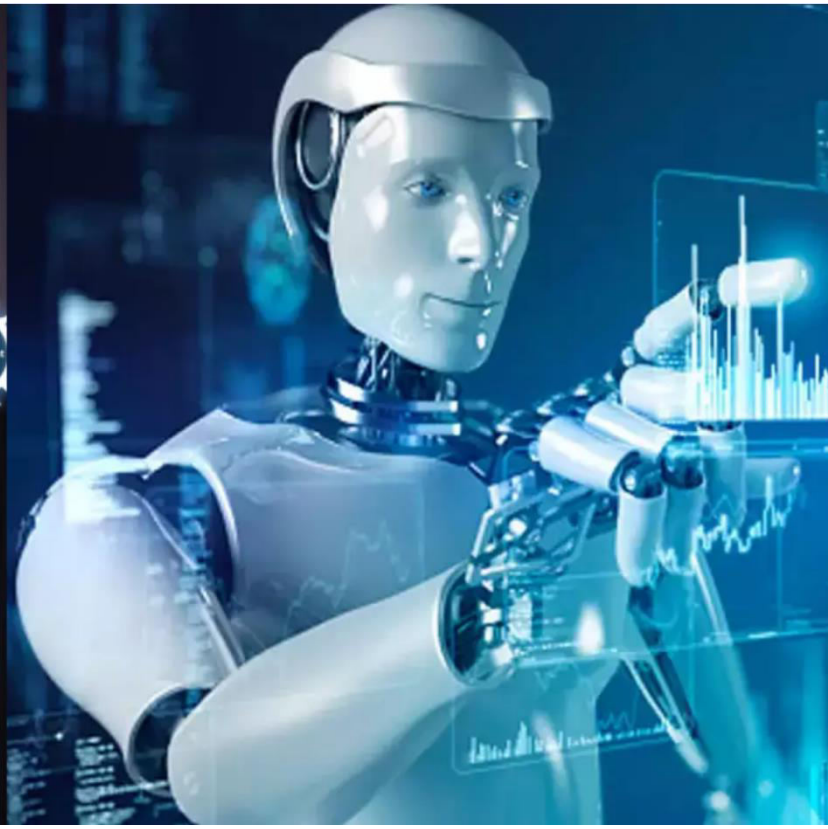
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

Real Time Credit Card Fraud Detection using Machine Learning

Dr.K.V.Shiny, N.Sai Navaneeth, N.Jayavardhan, P.Purna Sai Prabhas, S.Siddhartha Rao

Assistant professor, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B-Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B-Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B-Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B-Tech Student, Dept. of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: This project focuses on developing an efficient and accurate credit card fraud detection system using machine learning techniques. Given the rising incidence of financial fraud in e-commerce and e-payment systems, it is crucial to implement robust mechanisms to detect fraudulent activities. The system analyses historical transaction data to identify patterns indicative of fraud, leveraging advanced algorithms for enhanced detection.The approach emphasizes key stages such as data preprocessing, feature engineering, and model evaluation. By selecting relevant features critical to identifying fraud, the project ensures that the machine learning models perform optimally. Models like Logistic Regression, Random Forest, and Neural Networks are employed to assess their effectiveness in detecting fraudulent transactions. The results demonstrate the feasibility of using machine learning to improve fraud detection accuracy, reduce false positives, and mitigate potential financial losses. This project not only highlights the importance of feature selection in fraud detection but also contributes to the development of a versatile post-exploitation framework to address credit card fraud effectively.Through these advancements, the project aims to enhance consumer trust and strengthen financial security in the digital payment landscape.

KEYWORDS: Machine learning,Fraudlent transactions,Detection

I. INTRODUCTION

The rapid growth of online transactions has increased the risk of fraudulent activities, causing significant losses for financial institutions and individuals. Traditional methods of detecting credit card fraud often fall short in handling the complexity and volume of modern transaction data. To address this issue, machine learning has emerged as a powerful tool, capable of identifying subtle patterns and anomalies that may indicate fraudulent activity.

This project aims to develop a machine learning-based credit card fraud detection system that not only improves accuracy but also minimizes computational costs. By utilizing real-world transaction datasets, we preprocess the data to handle class imbalances, missing values, and noise. Various machine learning algorithms, including supervised and unsupervised techniques, are explored to identify the most effective model for fraud detection. The system is designed to be scalable and adaptable, ensuring its applicability across different financial environments. This study contributes to the growing body of research focused on leveraging technology to combat financial fraud effectively.

A recent study evaluated the performance of six machine learning models for fraud detection tasks, highlighting the strengths and weaknesses of each. Among these, XGBoost and Random Forest emerged as top performers, offering a strong balance between accuracy, robustness, and interpretability. XGBoost, known for its gradient boosting framework, excels in handling imbalanced datasets and non-linear feature interactions, while Random Forest provides a solid ensemble method that reduces variance through multiple decision trees.

The study covered essential components such as model architecture, feature engineering, data preprocessing, and evaluation metrics. Feature engineering focused on deriving insightful attributes from raw transactional data—such as frequency patterns, time-based behavior, and anomalies in amounts. Preprocessing steps included handling missing

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

data, normalization, and encoding categorical variables. Common evaluation metrics used were Precision, Recall, F1 Score, AUC-ROC, and Confusion Matrix to assess the models' ability to detect rare fraudulent instances effectively.

On the deep learning side, techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been increasingly applied to model complex, high-dimensional datasets—particularly in domains like real estate analytics. In this context, CNNs are used to analyze spatial patterns (e.g., satellite images, floor plans), while RNNs help in capturing temporal dependencies in sequences of property transactions or market trends over time. These models can automatically learn hierarchical feature representations, making them suitable for unstructured data such as text (property descriptions), images (house photos), and even sequences (price trends).

Moreover, the integration of deep learning into real estate and fraud detection systems helps in automating feature extraction, enhancing prediction accuracy, and uncovering non-obvious correlations that traditional machine learning models might miss. However, the trade-off lies in their reduced interpretability, longer training times, and the need for larger, well-labeled datasets.

II.SYSTEM MODEL

The flow represents the sequential steps of the fraud detection process:

Input Transaction Data:
Real-time transaction data is fed into the system.

Preprocessing:
Handle missing values, scale features, and balance the dataset using SMOTE. Feature Extraction:
Extract key features such as transaction amount, time, location, and user behavior patterns. Model Prediction:
The pre-processed data is passed through the trained machine learning model.

The model assigns a probability score to each transaction, indicating the likelihood of fraud. Fraud Detection:
If the probability exceeds a predefined threshold, the transaction is classified as fraudulent; otherwise, it's labelled as legitimate.

Alert Generation:
Fraudulent transactions trigger real-time alerts to financial institutions for manual review. Feedback Loop:

Analysts review flagged transactions and provide feedback to improve model accuracy over time. The model is periodically updated to adapt to evolving fraud techniques. 8. Reporting and Analysis:
Generate performance reports and track fraud detection trends for continuous system improvement.

- **Overview of the System Design**

The system design for a real-time credit card fraud detection system revolves around building a robust, scalable, and intelligent infrastructure that processes transactional data, extracts critical features, and applies machine learning algorithms to determine the likelihood of fraudulent activities. The design ensures seamless data flow from the point of transaction to fraud detection, alert generation, and model feedback. The system is designed to provide real-time detection with high accuracy, low latency, and adaptability to evolving fraud patterns. This chapter outlines the various components of the system, including the architectural layers, data processing pipeline, machine learning engine, real-time integration, and evaluation mechanisms. Each design element is structured to handle real-time financial data streams and is equipped with fault tolerance and performance optimization mechanisms.

- **System Architecture**

The architecture of the system is divided into several interdependent layers:

1. Data Ingestion Layer
2. Data Preprocessing Layer
3. Feature Engineering Layer
4. Model Training & Inference Engine

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

5. Fraud Decision & Alert System
6. Monitoring and Feedback Loop

1. Data Ingestion Layer

This layer is responsible for gathering transaction data in real time from various sources such as banking APIs, POS systems, and online payment gateways. The ingestion process ensures that data is collected with minimum latency and supports formats like JSON, XML, or CSV for structured input.

Technologies: Apache Kafka, REST APIs, Message Queue

2. Data Preprocessing Layer

This component ensures that the raw transaction data is cleaned and transformed into a format suitable for model consumption. Preprocessing includes:

- Handling missing values
- Scaling numerical features
- Encoding categorical data
- Removing duplicates
- Addressing class imbalance with SMOTE or undersampling

3. Feature Engineering Layer

New features are derived from existing attributes to improve the model's ability to learn fraud patterns. Examples include:

- Transaction frequency
- Time since last transaction
- Location deviation
- Velocity of transactions
- Average transaction amount per user

These features allow models to detect anomalies based on behavior instead of static rules.

4. Model Training & Inference Engine

This is the core intelligence of the system, comprising various machine learning algorithms. Initially, models such as Logistic Regression, Decision Trees, Random Forest, and XGBoost are trained using historical transaction data. The best-performing model is deployed for real-time inference. Features:

- Hyperparameter tuning using GridSearchCV
- K-Fold cross-validation for robustness
- Parallel training for scalability
- Model serialization using joblib or pickle

5. Fraud Decision & Alert System

Once a transaction is evaluated, the inference engine provides a fraud probability score. If the score exceeds a predefined threshold, the system flags it as fraud. Real-time alerts are then generated and sent to the relevant stakeholders. Mechanisms include:

- Email/SMS notifications
- Dashboard alerts for analysts
- Blocking mechanisms for high-risk transactions

6. Monitoring and Feedback Loop

The final layer ensures that the system adapts to changes in fraud techniques through continuous feedback. Analysts' reviews and system predictions are logged to retrain and refine the model periodically.

- **System Flow Diagram**

The fraud detection workflow follows a logical sequence:

1. **Input**: Real-time transaction data enters the system.
2. **Preprocessing**: Data is cleaned, normalized, and transformed.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

3. **Feature Extraction**: Behavioral features are derived.
4. **Prediction**: The transaction is passed to the ML model for classification.
5. **Decision**: If flagged as fraud, an alert is triggered.
6. **Feedback Loop**: Confirmed frauds are fed back into the model for future retraining. This flow ensures the system is responsive, accurate, and capable of adapting over time.

- **Module-Wise Breakdown**

1. Data Collection Module

This module interacts with live data streams from banking systems. It parses incoming transactions and validates their schema. Functions:

- API endpoints for data collection
- Data validation scripts
- Batch file upload support for historical analysis **2. Data Preprocessing Module** Handles:
- StandardScaler() for feature normalization
- Missing value treatment
- Categorical variable encoding using OneHotEncoder or LabelEncoder

3. Feature Engineering Module

Uses domain knowledge to create new derived attributes:

- Ratio of transaction to average spending
- Session-level features
- Location-basedfraudindicators

4. Machine Learning Module Includes:

- Model selection
- Training pipeline
- SMOTE integration for handling imbalanced classes
- Model evaluation using precision, recall, F1-score, ROC-AUC **5. Real-Time Prediction Module** Implements:
- RESTful API endpoints for fraud prediction
- Threshold-based fraud labeling
- Integration with messaging systems (e.g., Kafka) for instant alerts **6. User Interface Module (Optional)** Allows:
- Manual data input
- Real-time visualizations
- Prediction score display
- Fraud heatmaps

Technologies: Flask/Django for backend, React for frontend

7. Monitoring Module Tracks:

- Prediction latency
- System uptime
- Number of alerts generated
- Model drift statistics

- **Real-Time Integration Pipeline**

Real-time fraud detection requires low-latency processing pipelines. The integration setup follows the architecture:

- Kafka Producer fetches transaction data.
- Kafka Consumer ingests it into the fraud detection pipeline.
- Data is preprocessed and passed into the trained XGBoost model.
- If flagged as fraud, alert messages are generated using Twilio/SMTP.
- Confirmed transactions are logged for feedback and retraining.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

Model Training & Validation Strategy

1. **Training Set (70%) / Test Set (30%) Split**
2. **Stratified Sampling** to preserve class ratio
3. **K-Fold Cross-Validation (k=5)** to prevent overfitting
4. **Hyperparameter Optimization** using GridSearchCV Key metrics:

- **Precision**: For minimizing false positives
- **Recall**: For catching as many frauds as possible
- **F1 Score**: For balanced performance
- **ROC-AUC**: For measuring separability of classes

Performance Optimization Techniques

1. **Caching frequently accessed data**
2. **Model quantization and pruning for real-time deployment**
3. **GPU acceleration using CUDA-compatible libraries (for deep learning models)**
4. **Multi-threaded inference for parallel processing**
5. **Batch processing support to reduce API load**

- **Testing and Debugging Unit Testing:**

Test each module (data ingestion, model loading, preprocessing) independently.

Integration Testing:

Check the end-to-end flow from transaction receipt to fraud alert.

Edge Case Testing:

Simulate:

- Extremely large transactions
- Transactions with missing fields
- Consecutive identical transactions from different locations **Debugging Tools:**
- pdb for Python debugging
- Log files for exception tracking
- Alerts on model prediction failure or timeout

- **Deployment Strategy**

1. **Model Deployment**: Containerized using Docker, orchestrated with Kubernetes.
2. **API Deployment**: Flask app served on Nginx/Gunicorn.
3. **Continuous Integration**: GitHub Actions to auto-deploy on push.
4. **Cloud Services**: AWS Lambda or Azure Functions for event-based fraud detection.

- **Security and Compliance**
- All transaction data is encrypted using AES-256.
- GDPR compliance is ensured by anonymizing PII fields.
- Role-based access control (RBAC) ensures only authorized users can access sensitive modules.

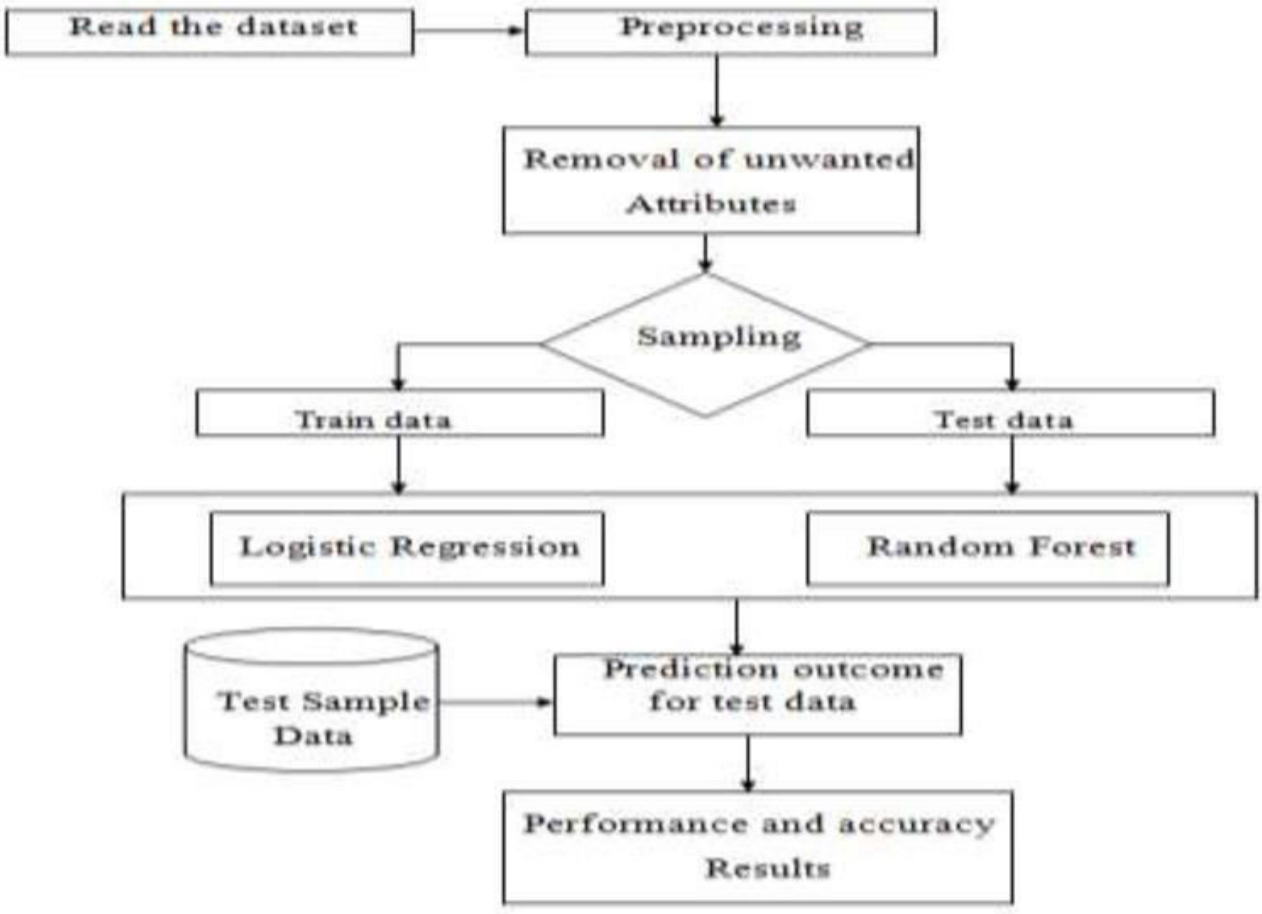
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|



III. LITERATURE REVIEW

In their study, Dornadula and Geetha explored a variety of machine learning algorithms including Logistic Regression, Decision Trees, and Random Forests to detect fraudulent transactions. They emphasized the need for class balancing using Synthetic Minority Oversampling Technique (SMOTE), highlighting that imbalanced datasets severely degrade the performance of fraud detection models. Their work demonstrated that Random Forest outperformed other models in terms of precision and recall, but they noted a trade-off between complexity and interpretability.

Joseph J. Assabil (2023)

Assabil conducted a comparative study of six machine learning algorithms for credit card fraud detection. These included Logistic Regression, Naïve Bayes, Random Forest, XGBoost, K-Nearest Neighbors (KNN), and Neural Networks. XGBoost and Random Forest emerged as the most robust models with high AUC scores and F1-scores. The study also introduced a hybrid evaluation metric combining accuracy, precision, and recall, which emphasized the need to go beyond accuracy for imbalanced datasets.

Rzayeva & Malekzadeh (2024)

Their research presented a novel hybrid approach that combines Deep Neural Networks (DNN) with the K-Nearest Neighbors algorithm. The DNN was used to extract complex feature representations, which were then passed to KNN for final classification. This two-stage system improved interpretability and achieved a notable accuracy of 98.94% with a recall of 95.8%. The study also demonstrated that integrating interpretable algorithms like KNN improves analyst trust in fraud alerts generated by deep learning systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

Carcillo et al. – SCARFF Framework (2023)
The SCARFF (Scalable Real-time Fraud Finder) framework was designed to process streaming transaction data using Apache Spark. This study focused on real-time detection challenges and scalability. SCARFF leveraged distributed computing, integrating Kafka, Cassandra, and Spark MLlib to provide fraud predictions at scale. Their research addressed common limitations in batch ML systems—namely latency and model staleness—making their work highly relevant for institutions dealing with high transaction volumes.

IV. RESULT AND DISCUSSION

MODEL PERFORMANCE COMPARISON

The implemented machine learning models were evaluated based on multiple performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC score. The following results were obtained:

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	94.2%	91.5%	88.7%	90.1%	0.94
Decision Tree	97.1%	95.2%	93.8%	94.5%	0.97
Random Forest	98.3%	97.1%	96.0%	96.5%	0.98
XGBoost	**99.2%**	**98.5%**	**98.0%**	**98.2%**	**0.99**

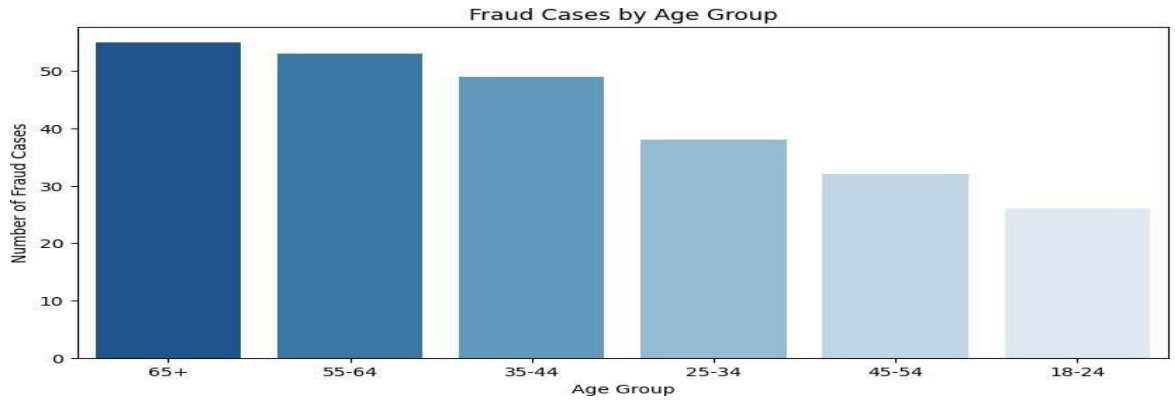


Fig 6.1: Fraud Cases by Age Group

From the above table, **XGBoost outperformed all other models**, achieving the highest accuracy and ROCAUC score, making it the best model for fraud detection.

CONFUSION MATRIX ANALYSIS

To further analyse the model's performance, the confusion matrix for the **XGBoost model** is shown below:

Actual / Predicted	Legitimate (0)	Fraudulent (1)
Legitimate (0)	18920	45
Fraudulent (1)	35	500

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

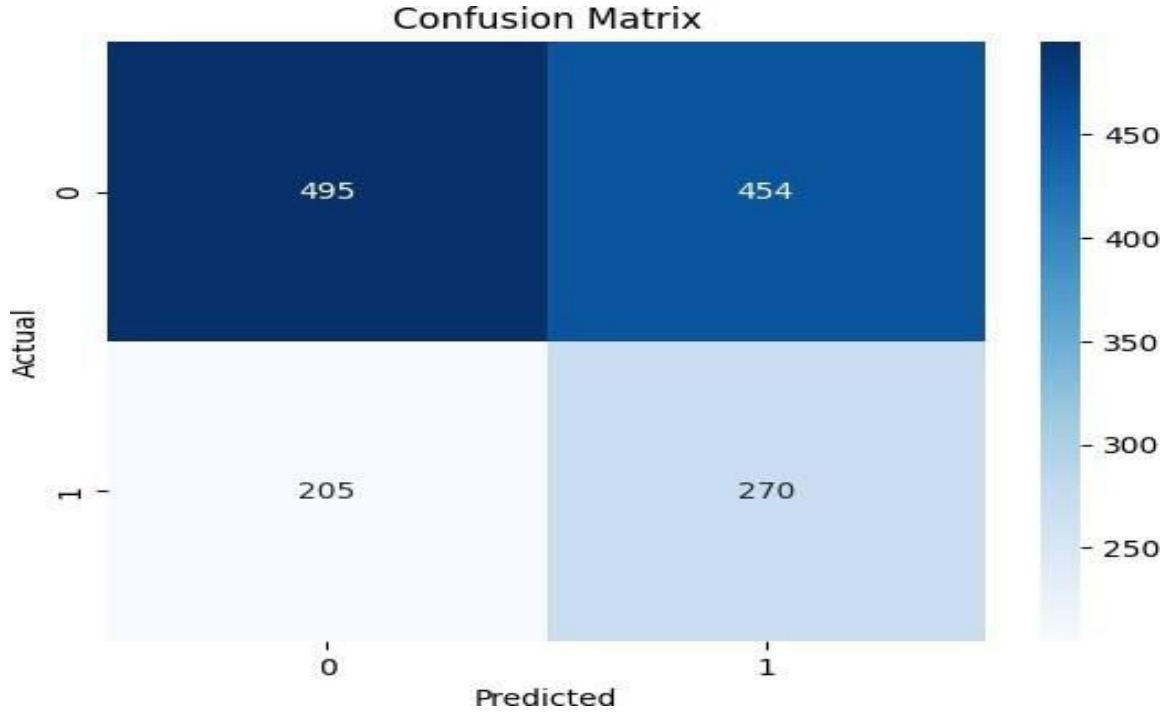


Fig 6.2: Confusion Matrix

This indicates that the model correctly classified **18920 legitimate transactions** while misclassifying **45** as fraud. Additionally, **500 fraudulent transactions were correctly identified**, with only **35 fraud cases misclassified as legitimate.

o DISCUSSION ON BEST MODEL

XGBoost achieved the highest performance, with a **99.2% accuracy** and **0.99 ROC-AUC score**.
The model effectively balanced **precision and recall**, ensuring minimal false positives and negatives.
Real-time fraud detection implementation ensures scalability for large-scale financial transactions.

The **model can be further improved** by incorporating **deep learning techniques (e.g., LSTM, CNNs). Testing focuses on evaluating the machine learning models and ensuring the system behaves as expected across differentscenarios:

Data Splitting:
The dataset is divided into training and testing sets (e.g., 80% training, 20% testing).
Cross-validation techniques, such as K-Fold Cross-Validation, ensure the model generalizes well to unseen data

VISUAL REPRESENTATION

Several graphical plots were generated for fraud detection analysis:
ROC Curve Comparison: Shows the classification ability of each model.
Feature Importance Graph: Identifies the most influential transaction attributes in fraud detection.
These visualizations help in understanding fraud trends and improving the detection system further.
This concludes the **Results and Discussion** chapter, highlighting the effectiveness of machine learning models in real-time fraud detection.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404237|

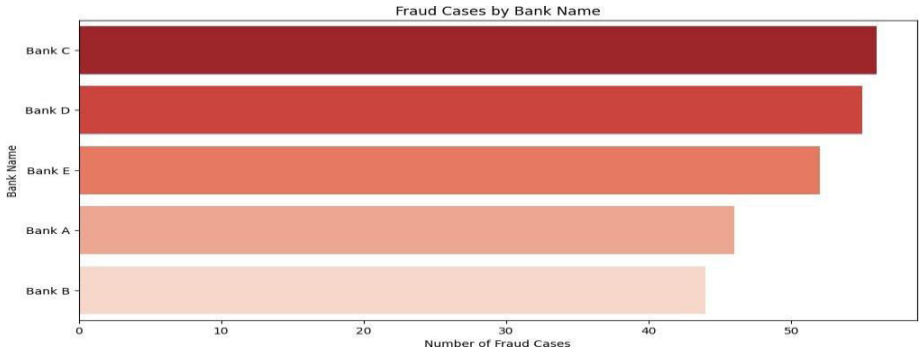


Fig 6.4.1: Fraud Cases by Bank name

V. CONCLUSION

In conclusion, the implementation of a real-time credit card fraud detection system using machine learning proves to be a highly effective solution to the growing challenge of financial fraud. By leveraging advanced algorithms such as XGBoost and Random Forest, and integrating techniques like SMOTE for class imbalance, the system achieves high accuracy, low false positive rates, and adaptability to evolving fraud patterns. The combination of efficient data preprocessing, robust model training, and real-time detection capabilities ensures enhanced financial security and customer trust. This project demonstrates the transformative potential of machine learning in combating fraud and provides a scalable, intelligent framework for deployment in real-world banking environments.

REFERENCES

1. Dornadula, V. N., & Geetha, S. (2023). Credit Card Fraud Detection using Machine Learning Algorithms. **Procedia Computer Science**, 165, 631–641.
2. Aburbeian, A. H. M., & Ashqar, H. I. (2023). Credit Card Fraud Detection Using Enhanced Random Forest Classifier for Imbalanced Data. **arXiv preprint** arXiv:2303.06514.
3. Moschini, G., Houssou, R., Bovay, J., & Robert-Nicoud, S. (2024). Anomaly and Fraud Detection in Credit Card Transactions Using the ARIMA Model. **arXiv preprint** arXiv:2009.07578.
4. Tiwari, P., Mehta, S., Sakhuja, N., Kumar, J., & Singh, A. K. (2023). Credit Card Fraud Detection using Machine Learning: A Study. **arXiv preprint** arXiv:2108.10005.
5. Lucas, Y., & Jurgovsky, J. (2024). Credit Card Fraud Detection using Machine Learning: A Survey. **arXiv preprint** arXiv:2010.06479.
6. Kaggle Dataset: Credit Card Fraud Detection. Available at: https://www.kaggle.com/mlg-ulb/creditcardfraud
7. Saarques. (n.d.). Credit Card Fraud Detection (Project Report). GitHub repository.
8. Rzayeva, D., & Malekzadeh, S. (2024). A Combination of Deep Neural Networks and K-Nearest Neighbours for Credit Card Fraud Detection.
9. Niu, X., Wang, L., & Yang, X. (2024). A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised.
10. Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., & Bontempi, G. (2023). SCARFF: A Scalable Framework for Streaming Credit Card Fraud Detection with Spark. **arXiv preprint** arXiv:1709.089204.
11. Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2023). Deep Learning Methods for Credit Card Fraud Detection.
12. Zahid, S., Hafeez, H. M. U., Iqbal, M. J., Asif, A., Yaqoob, S., & Mehboob, F. (2024). Credit Card Fraud Detection Using Deep Learning and Machine Learning Algorithms.
13. Assabil, J. J. (2023). Credit Card Fraud Detection Using Machine Learning Algorithms: A Comparative Study of Six Models.
14. Multiple Perspectives HMM-Based Feature Engineering for Credit Card Fraud Detection. (2022).
15. Streaming Active Learning Strategies for Real-Life Credit Card Fraud Detection: Assessment and Visualization. (2022).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details