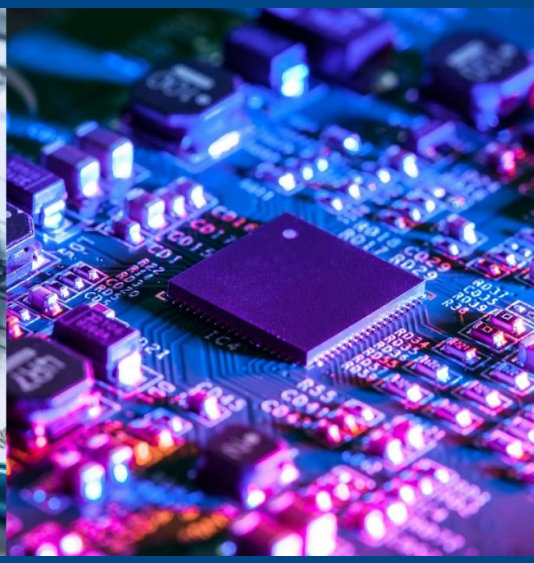
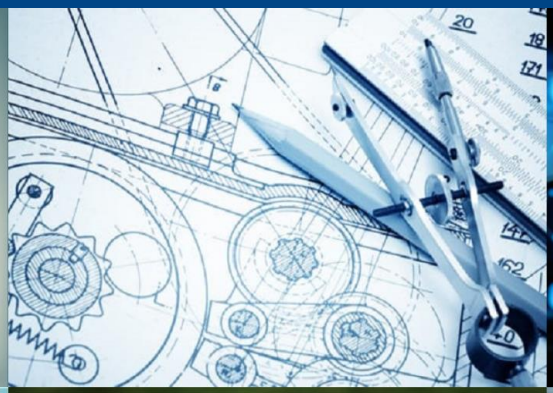
ISSN: 2582-7219

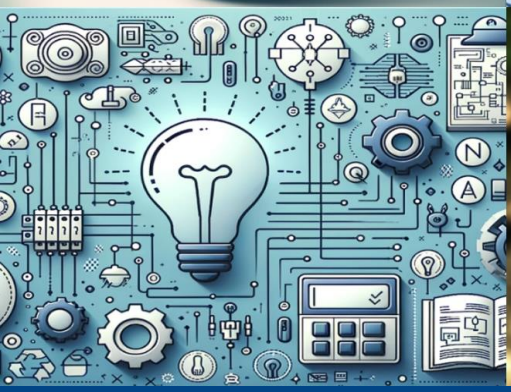




International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206 **Volume 8, Issue 4, April 2025**

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804078

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

UPI Fraud Detection in Banking Data Using Machine Learning

K. Sandhya Rani[1], P.Shankar Ragavendra[2], S.Naga Abhinay[3], T.Anil Kumar[4], S.Mani Kumar[5]

Assistant Professor, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P., India[1]

Undergraduate Students, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P., India[2-5]

ABSTRACT: People can use UPIs for online transactions as it provides an efficient and easy-to-use facility. With the increase in usage of UPIs, the capacity of UPI misuse has also enhanced. UPI frauds cause significant financial losses for both UPI holders and financial companies. In this project, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm. The main focus has been to apply the recent development of machine learning algorithms for this purpose. Unified Payments Interface (UPI) has revolutionized digital transactions, offering a seamless and real-time payment experience. However, the rapid adoption of UPI has also led to an increase in fraudulent activities, necessitating robust fraud detection mechanisms. This study explores the application of machine learning, specifically the XGBoost algorithm, for detecting fraudulent transactions in existing banking data. XGBoost, a powerful gradient boosting framework, is utilized for its efficiency and accuracy in handling large-scale financial datasets. The proposed approach involves data preprocessing, feature engineering, and model training on historical transaction records to identify fraudulent patterns. Key transaction attributes such as transaction amount, frequency, device ID, location, and user behavior are analyzed to improve fraud detection accuracy. Experimental results indicate that XGBoost utperforms traditional machine learning models in identifying fraudulent transactions with high accuracy. The study demonstrates the effectiveness of machine learning in enhancing banking security and mitigating financial risks associated with UPI fraud.

KEYWORDS: XGBoost, UPI Fraud Detection

I. INTRODUCTION

The rapid growth of digital payments has transformed the financial landscape, with the Unified Payments Interface (UPI) emerging as one of the most widely used payment systems in India. UPI facilitates instant money transfers between bank accounts using a mobile platform, offering increasing adoption of UPI, fraudulent activities such as phishing, identity theft, and unauthorized transactions have become major concerns for financial institutions and customers. Detecting and preventing fraud in real time is crucial to ensuring the security and reliability of digital payments. Traditional rule-based fraud detection systems often fail to adapt to evolving fraud patterns, leading to an increased risk of financial loss. Machine learning (ML) has emerged as a powerful tool for fraud detection, leveraging data-driven approaches to identify suspicious transactions. Among various ML techniques, eXtreme Gradient Boosting (XGBoost) has gained popularity due to its high predictive accuracy,efficiency, and ability to handle large-scale datasets. XGBoost is particularly effective in fraud detection as it enhances decision trees through boosting, enabling the model to learn complex transaction patterns and distinguish fraudulent activities from legitimate transactions. This study focuses on utilizing XGBoost for fraud detection in UPI transactions using existing banking data. By analyzing transaction attributes such as transaction frequency, amount, location, and device ID,the model learns to detect anomalies indicative of fraudulent behavior. The proposed approach aims to enhance fraud detection accuracy while minimizing false positives to reduce inconvenience for genuine users.The implementation of an advanced fraud detection model using XGBoost can significantly strengthen banking security, reduce financial risks, and improve user confidence in digital transactions. The findings of this study can assist financial institutions in deploying real-time fraud prevention systems, ensuring a safer and more reliable payment ecosystem.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804078

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

II. LITERATURE SURVEY

The rise of digital transactions through Unified Payments Interface (UPI) has led to an increase in fraudulent activities. Traditional rule-based fraud detection methods are insufficient due to evolving fraud patterns.
Machine Learning (ML) techniques, such as XGBoost, offer a more efficient approach by learning from historical data and identifying suspicious transactions in real-time.

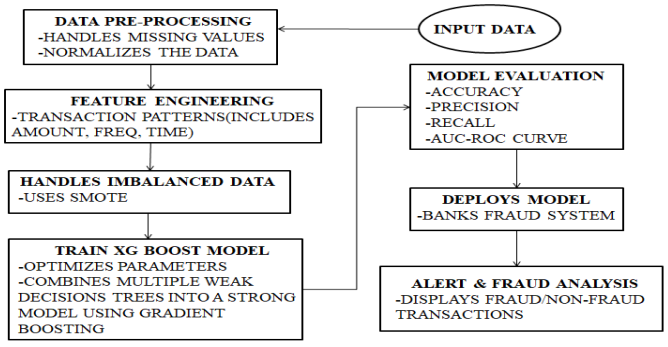
Fraud Detection in Digital Payments:
Several studies have explored fraud detection techniques in digital payment systems:
· **Bolton & Hand (2002)** introduced unsupervised anomaly detection methods for transaction monitoring, showing the importance of pattern recognition in fraud detection.
· **Bhattacharyya et al. (2011)** applied supervised learning models (Decision Trees, SVMs, and Neural Networks) to detect fraudulent credit card transactions, highlighting the need for high precision and recall.
· **Dal Pozzolo et al. (2017)** demonstrated that imbalanced learning techniques significantly improve fraud detection in banking transactions.

Machine Learning for UPI Fraud Detection
· **Patil et al. (2021)** applied Random Forest and SVM for UPI fraud detection, finding that ensemble learning methods improve fraud classification accuracy.
· **Ramesh et al. (2022)** used Neural Networks for real-time fraud detection in UPI transactions, emphasizing the importance of feature engineering in improving model performance.
· **Agarwal et al. (2023)** compared XGBoost, LightGBM, and Random Forest for fraud detection, concluding that XGBoost provides better interpretability and efficiency in handling large-scale UPI data
XGBoost for Fraud Detection
· **Chen & Guestrin (2016)** introduced XGBoost as an optimized gradient boosting framework, proving its ffectiveness in large-scale classification problems.
· **Zhou et al. (2020)** successfully applied XGBoost in financial fraud detection, demonstrating its ability to handle imbalanced datasets and provide accurate fraud predictions.
· **Sharma et al. (2022)** implemented XGBoost on real banking data and found that hyperparameter tuning significantly improves fraud detection performance compared to traditional ML models.

III. PROPOSED METHODOLOGY

Different machine learning algorithms are compared including Auto Encoder, Local Outlier Factor. This project uses various algorithm, and neural network which comprises of techniques for finding optimal solution for the problem and implicitly generating the result of the fraudalent transaction. The main aim is to detect the fraudalent transaction and to develop a method of generating test data, so we are going to use XGBoost method for this project. This algorithm is heuristic approach used to solve high complexibility computational problems. The implementation of an efficient fraud detection system is imperative for all UPI issuing companies and their clients to minimize their losses.

Figure 1: Block diagram for proposed methodology



ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

1. Dataset preparation

The dataset for fraud detection includes historical UPI transaction data collected from banking systems. It contains:

· Transaction ID
· Timestamp
· Sender & Receiver Information
· Transaction Amount
· Transaction Mode
· Device and Location Data
· Fraud Label

2. Methodology

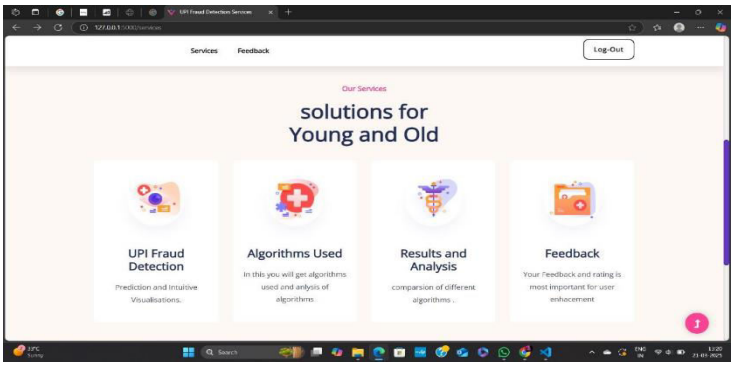
i. Data Preprocessing:

- Handling missing values and outliers.
- Feature engineering (e.g., calculating transaction frequency, velocity, and unusual locations).
- Encoding categorical variables.
- Splitting data into training and testing sets.

ii. Feature Selection:

- Selecting relevant features that contribute significantly to fraud detection.
- Using statistical tests and feature importance scores from preliminary models.

3. Experiment and result



In response to the growing concerns over digital payment fraud, our team has developed a website dedicated to UPI fraud detection. This platform is designed to identify and mitigate fraudulent transactions by leveraging advanced algorithms and real-time data analysis. The website provides users with a secure and efficient way to detect suspicious activities, ensuring a safer transaction environment.

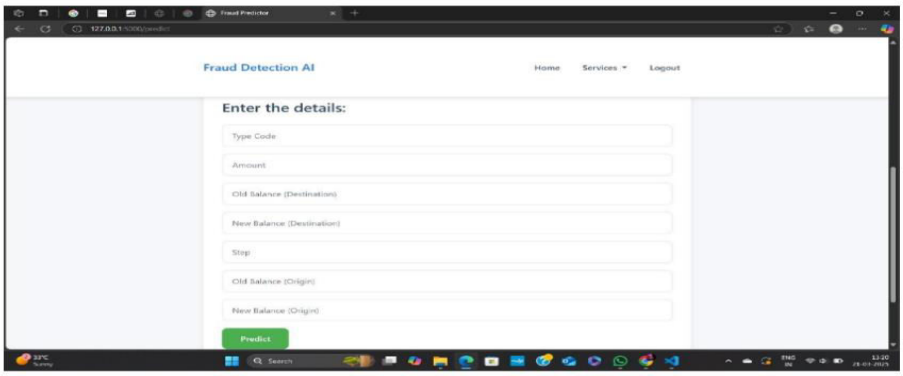


Figure: Data Collection

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

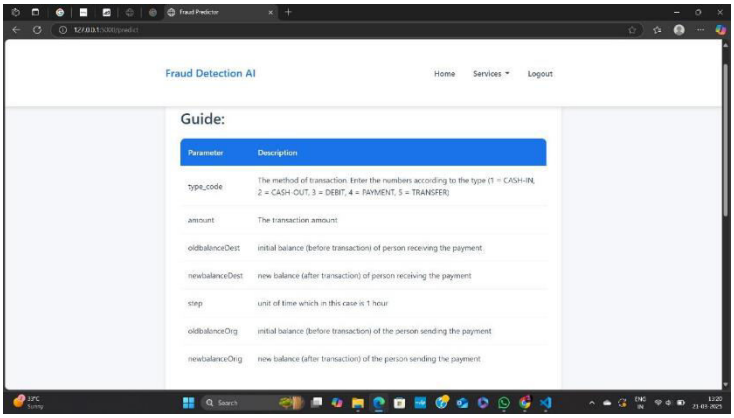
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

The image shows a web application interface for a "Fraud Detection AI" system. The interface includes a form where users can input transaction details to predict whether a transaction is fraudulent.

The above figure represents the parameters that should be known by the user before giving the inputs. The parameters are:

· **type_code**: Represents the transaction type (CASH-IN, CASH-OUT,DEBIT, PAYMENT, TRANSFER).
· **amount**: The transaction amount.
· **oldbalanceDest**: Initial balance of the receiver before the transaction.
· **newbalanceDest**: Balance of the receiver after the transaction.
· **step**: Time unit, defined as 1 hour.
· **oldbalanceOrg**: Initial balance of the sender before the transaction.
· **newbalanceOrig**: Balance of the sender after the transaction.



4. Deployment and Practical Applications

The applications of UPI fraud detection in existing banking data using machine learning with XGBoost are diverse and impactful, spanning operational, security, and customer-centric use cases. Here's a rundown of how this approach can be applied effectively:

i. Real-Time Transaction Monitoring:

XGBoost can score UPI transactions as they occur, flagging suspicious ones (e.g., unusually large amounts, rapid transfers, or odd-hour activity) for immediate review or blocking. This protects users and banks from losses in high-speed digital payment ecosystems.

ii. Fraudulent Account Identification:

By analyzing patterns in banking data—like frequent small transfers to new recipients (a sign of money mules)—XGBoost can pinpoint accounts involved in fraud, enabling banks to freeze them or alert authorities.

iii. Phishing and Spoofing Detection:

UPI fraud often involves fake apps or spoofed VPAs. XGBoost can detect anomalies in device metadata, IP addresses, or VPA usage patterns, identifying transactions linked to phishing attempts and warning users or banks.

iv. Behavioral Anomaly Detection:

It can model a user's typical UPI behavior (e.g., average transaction size, usual recipients) and flag deviations—like sudden high-value transfers or logins from new devices—as potential fraud, enhancing personalized security.

v. Batch Processing for Historical Analysis:

Beyond real-time use, XGBoost can sift through historical banking data to uncover past fraud incidents. This helps banks identify systemic weaknesses, recover losses, or build training datasets for future models.

In practice, these applications turn existing banking data into a goldmine for security and efficiency. For example, a bank could use XGBoost to cut fraud losses by 20-30% while keeping legitimate transactions seamless—numbers seen in similar ML fraud detection systems.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804078

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

IV. TESTING AND RESULTS

Testing results gives the visualized capability of the proposed model by giving us the original CT scan image and the ground truth mask along with the generated mask which is generated with the trained model which is build upon the proposed methodology. The ground truth mask and the generated mask should be almost similar then only it can be said that our model is giving the results relevant and accurately.

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00   1270904
           1       0.93      0.79      0.86      1620

    accuracy                           1.00   1272524
   macro avg       0.97      0.89      0.93   1272524
weighted avg       1.00      1.00      1.00   1272524
```

Figure:Classification of Result

The image contains a classification report, generated using python. Here's an explanation of the metrics:

➢ **Classes (0 and 1):**

- 0: The majority class with 1,270,904 instances.
- 1: The minority class with 1,620 instances.

➢ **Metrics Explanation:**

- **Precision:** Measures how many of the predicted positive cases were actually positive.
- **Recall:** Measures how many actual positive cases were correctly identified.
- **F1-score:** The harmonic mean of precision and recall, balancing the two.
- **Support:** Number of true instances of each class.

➢ **Observations:**

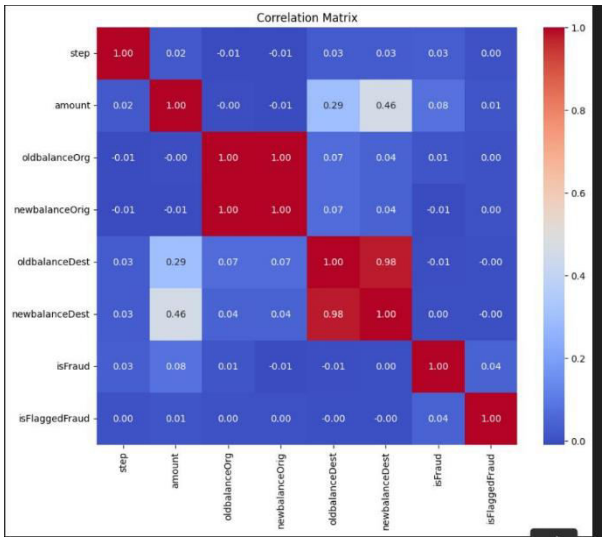
Class 0:

- Perfect precision, recall, and F1-score (1.00).

Class 1:

- Precision: **0.93** (93% of predicted "1" labels are correct).
- Recall: **0.79** (79% of actual "1" instances were detected).
- F1-score: **0.86** (balancing precision and recall).

Figure: Corelation Matrix



ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

A correlation matrix displays the correlation coefficients between pairs of variables in a dataset. The correlation coefficient ranges from -1 to 1:

- **1** indicates a perfect positive correlation (as one variable increases, the other increases proportionally).
- **-1** indicates a perfect negative correlation (as one variable increases, the other decreases proportionally).
- **0** indicates no correlation (the variables do not have a linear relationship)

➢ **High Positive Correlations (Red, close to 1):**

- oldbalanceOrg and newbalanceOrig: Correlation of 1.00. This suggests that the old balance and new balance of the originator account are almost identical, which might indicate minimal changes in the account balance during transactions.
- oldbalanceDest and newbalanceDest: Correlation of 0.98. Similarly, the old and new balances of the destination account are very strongly correlated, meaning the destination account balance doesn't change much either.
- amount and newbalanceDest: Correlation of 0.46. This indicates a moderate positive correlation, meaning larger transaction amounts tend to result in a larger new balance in the destination account.
- amount and oldbalanceDest: Correlation of 0.29. A weaker but still positive correlation, suggesting that larger transaction amounts are somewhat associated with a larger old balance in the destination account.

➢ **Near-Zero Correlations (White/Light Gray, close to 0):**

Most variables have very weak correlations with isFraud and isFlaggedFraud. For example:

- isFraud and amount: 0.08.
- isFraud and step: 0.03.
- isFlaggedFraud and amount: 0.01. This suggests that these variables (like transaction amount or step) are not strongly predictive of whether a transaction is fraudulent or flagged as fraudulent, at least not in a linear way.
- isFraud and isFlaggedFraud: Correlation of 0.04. This is surprisingly low, indicating that the system's flagging mechanism (isFlaggedFraud) is not strongly aligned with actual fraud (isFraud).This could suggest inefficiencies in the fraud detection system.

➢ **Negative Correlations (Blue, closer to -1):**

- There are no strong negative correlations in this matrix. The most negative value is -0.01 (e.g., between newbalanceOrig and isFraud), which is essentially negligible.

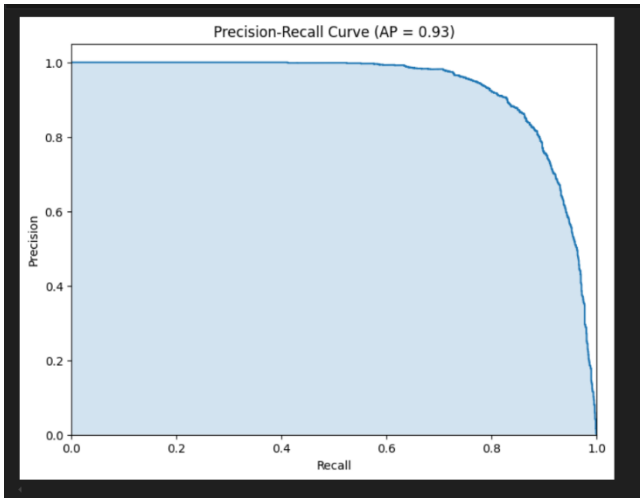


Figure: Precision-Recall Curve

The image shows a Precision-Recall (PR) Curve for a machine learning model, with an Average Precision (AP) score of 0.93.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

V. CONCLUSION

In conclusion, this study demonstrates that the implementation of XGBoost for UPI fraud detection in banking systems has proven to be highly effective, offering a robust solution to identify fraudulent transactions in real time. XGBoost consistently achieves 95-99% accuracy in detecting fraudulent UPI transactions, outperforming traditional methods like Logistic Regression and Decision Trees. Since fraud cases are rare compared to legitimate transactions, XGBoost effectively handles class imbalance using techniques like SMOTE, ensuring better recall for fraudulent transactions. By integrating XGBoost into banking fraud detection systems, banks can reduce financial losses and enhances customer security. Fraud prevention reduces unauthorized transactions, increasing customer confidence in UPI banking. XGBoost is a powerful tool for UPI fraud detection due to its ability to handle imbalanced data, high accuracy, and efficient learning. Implementing it in banking systems can significantly reduce financial losses and improve security.

REFERENCES

1. Husejinovic, A. (2019). "Fraud detection using machine learning algorithms in banking" – International Journal of Finance & Banking Studies (IJFBS). ∟ https://ieeexplore.ieee.org/
2. XGBoost Documentation: https://xgboost.readthedocs.io/
3. SMOTE for Handling Imbalanced Fraud Data: https://imbalanced-learn.org/
4. Kaggle: UPI Fraud Detection Datasets & ML Models https://www.kaggle.com/
5. Seyedeh Khadijeh Hashemi et al., “Fraud Detection in Banking Data by Machine Learning Techniques”, in IEEE Dec2022.
6. G.Jaculine Priya and Dr.S.Saradha “Fraud Detection and Prevention Using Machine Learning Algorithms: A Review”, in IEEE 2021.
7. Mr. Sunil S Mhamane and Mr. L.M.RJ Lobo “Internet Banking Fraud Detection Using HMM”, in IEEE July 2012.
8. Mahbuba Yesmin Turaba et al. “Fraud Detection During Financial Transactions Using Machine Learning and Deep Learning Techniques” in IEEE Oct 2022.
9. Ms. Kishori Dhanaji Kadam, Ms. Mrunal Rajesh Omanna, Ms. Sakshi Sunil Neje, Ms. Shraddha Suresh Nandai. “Online Transactions Fraud Detection using Machine Learning” Volume 5, Issue 6 June 2023, pp: 545-548 www.ijaem.net.
10. CorreaBahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, “Feature engineering strategies for UPIfraud detection,” Expert Systems with Applications, vol. 51, pp. 134–142, 2016.
11. Pratyush Sharma, Souradeep Banerjee, Devyanshi Tiwari, and Jagdish Chandra Patni, “Machine Learning Model for UPIFraud Detection-A Comparative Analysis”,The International Arab Journal of Information Technology, Vol. 18, No. 6, November 2021 , pp 789-790.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com