International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03720
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

SECURITY PROVISION FOR UPI TRANSCATIONS USING MACHINE LEARNING TO DETECT FRAUD TRANSCATIONS

Dr. M.Narender [1],Bopppena Sritej[2], Bolli Chandu[3], Aryaman Verma[4], Andugula Abhishek[5].

[1] Professor, Department of Computer and Science Engineering, TKR

College of Engineering and Technology.

[2,3,4,5]UG Scholars, Department of Computer and Science Engineering, TKR

College of Engineering and Technology, Medbowli, Meerpet.

ABSTRACT:

This project presents a comprehensive fraud detection system using machine learning techniques to identify fraudulent transactions from financial data. The dataset undergoes extensive preprocessing, including column renaming, data type conversion, and feature extraction. Exploratory Data Analysis (EDA) is performed using visualizations to uncover trends and anomalies in the data. Various models, including Decision Trees and XGBoost, are applied to classify transactions as fraudulent or legitimate. Additionally, Synthetic Minority Over- sampling Technique (SMOTE) is used to handle class imbalance, and Principal Component Analysis (PCA) is applied to reduce dimensionality. Performance evaluation using accuracy, confusion matrix, and classification reports demonstrates the effectiveness of the proposed fraud detection system. The model is further saved for deployment using pickle for practical use.

KEYWORDS: Fraud Detection, Machine Learning, XGBoost, Decision Tree, SMOTE, PCA, Financial Data Analysis, Data Preprocessing, Classification, Python.

I. INTRODUCTION:

Fraudulent financial transactions significantly threaten financial institutions, affecting revenues, credibility, and customer trust. With the rise of digital payments and online banking, the risk of fraud has increased. Traditional rule-based detection methods struggle to adapt to evolving fraudulent activities, prompting the adoption of machine learning-based approaches for identifying suspicious patterns and mitigating fraud. This project introduces a comprehensive fraud detection system using advanced machine learning algorithms to classify transactions as fraudulent or legitimate. The dataset is preprocessed through column renaming, data type conversion, and feature extraction to ensure data quality and enhance model accuracy. Exploratory Data Analysis (EDA) employs visualizations to gain insights, identify anomalies, and uncover trends aiding fraud detection. To tackle class imbalance in financial data, the Synthetic Minority Over- sampling Technique (SMOTE) generates synthetic samples, balancing the dataset and improving model training and classification accuracy. Principal Component Analysis (PCA) is used for dimensionality reduction, focusing models on the most informative features. Multiple algorithms, including Decision Trees and XGBoost, are employed to build robust classification models, evaluated using metrics like accuracy, confusion matrix, and classification reports. The final model, showcasing strong classification capabilities, is saved using the pickle library for future deployment, enabling real-time fraud detection in financial systems. This project advances efficient and reliable fraud detection through machine learning and effective data preprocessing.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03720
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

II. LITERATURE SURVEY

M.A Ibrahim [1], This research paper discusses the most common methods of fraud, detection techniques, and recent findings in the field. The researchers used the SMOTE technique to balance the dataset and found that models like Decision Tree, Random Forest, Neural Network, and K-nearest neighbor performed well when fitted and trained with the data. The system allows users to select their preferred model. The Random Forest model achieved an accuracy of 93.58%, but its efficiency decreases when trained with imbalanced transaction datasets.

P. Boulieris [2], The main objective of the work is to present a dataset for online fraud detection that is anonymized and publicly available. We argue that standard evaluation metrics used in existing literature should be complemented with online and offline assessments to evaluate model performance in a real-world business setting. We found that incorporating anomaly detection features improves all metrics except for online detection, highlighting the importance of considering both online and offline evaluation alongside standard metrics. Despite using fewer traditional features, the addition of NLP-based features significantly improved performance compared to a previous study.

III. METHODLOGY

The methodology for this project centers on developing an efficient and scalable fraud detection system for UPI transactions using machine learning. It commenced with data collection from publicly available financial transaction datasets. Following data acquisition, a comprehensive preprocessing phase was undertaken. This involved renaming columns for clarity, converting data types, especially timestamps, handling missing values with suitable imputation methods, and extracting new features like transaction hours. Categorical features such as merchant names, transaction categories, and user jobs were encoded using Ordinal and One-Hot Encoding to make them machine-readable. After cleaning and structuring the data, Exploratory Data Analysis (EDA) was performed to identify trends and anomalies. Visualization tools like Matplotlib, Seaborn, and Plotly were used to gain key insights, including the distribution of fraudulent transactions by time of day, merchant, and user occupation. These insights guided feature selection and enhanced the understanding of fraud patterns. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic minority class samples, preventing model bias toward the majority class.

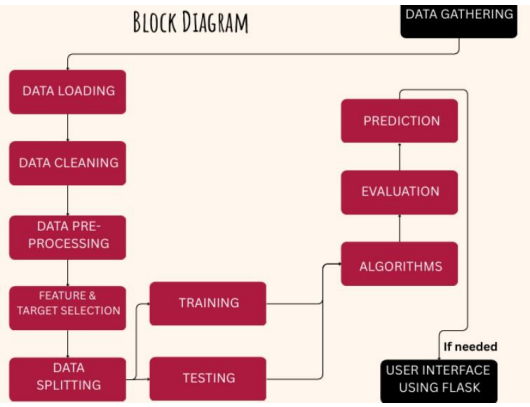

Fig 1: BLOCK DIAGRAM

For dimensionality reduction and to improve model performance and training efficiency, Principal Component Analysis (PCA) was applied. PCA helped retain 95% of the variance in the data while reducing the number of input features, thus enhancing computational efficiency. Feature selection was also performed using statistical methods like the Chi-Squared test to retain the most relevant attributes for model training. The core machine learning algorithms used for classification were Decision Trees and XGBoost. Decision Trees were chosen for their simplicity and interpretability, while XGBoost was selected for its high performance, regularization capabilities, and scalability.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03720
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The models were trained and evaluated using standard metrics such as accuracy, precision, recall, F1-score, and ROC AUC score. Evaluation results showed that the XGBoost classifier significantly outperformed the Decision Tree model in detecting fraudulent transactions, especially after applying SMOTE and PCA. Once the best-performing model was finalized, it was saved using the Pickle library to enable deployment in real-world applications. Along with the model, the preprocessing pipeline—including scalers and PCA transformers—was also saved to maintain consistency during inference. To support real-time predictions, a simple command- line interface was developed where users could input transaction details and receive immediate feedback on whether the transaction was fraudulent, along with a confidence score. This end-to-end methodology—from data preprocessing to model deployment—ensures the system is robust, scalable, and practical for financial institutions looking to enhance security in UPI transactions using AI-powered solutions.

IV. FLOW CHART

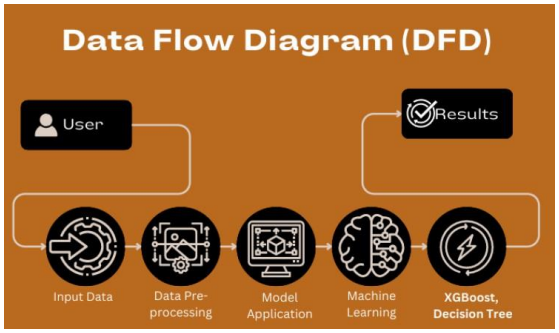

Fig 2 : DATA FLOW DIAGRAM

The Data Flow Diagram (DFD) provided represents the systematic flow of operations involved in building a fraud detection system for UPI (Unified Payments Interface) transactions using machine learning techniques. It begins with the user, who inputs transaction data into the system. This input data typically includes key attributes such as transaction ID, amount, merchant name, category, time of transaction, and customer profession. These inputs are essential for identifying behavioral patterns associated with both legitimate and fraudulent transactions. Next, the data f lows into the data preprocessing stage. This step is critical to ensure data quality and consistency. It involves renaming columns, converting data types (e.g., dates into timestamps), handling missing values, and encoding categorical features into numerical formats. It also includes feature extraction, such as deriving ‘hour of transaction’ from timestamps, which helps in better pattern recognition by machine learning models.

The preprocessed data is then passed to the model application module. Here, it is structured into training and testing datasets, ensuring that the model can be trained and evaluated effectively. Dimensionality reduction techniques like PCA (Principal Component Analysis) may also be applied at this stage to improve efficiency. Subsequently, in the machine learning phase, the system uses algorithms such as XGBoost and Decision Tree to analyze the data. These models are trained to distinguish between fraudulent and non-fraudulent transactions based on patterns identified in historical data. Finally, the system produces a result, which is presented to the user. This result includes the fraud prediction and the confidence score, allowing real-time fraud detection and decision-making.

The activity diagram visually outlines the step-by-step flow of actions involved in the fraud detection process using machine learning techniques. It begins with the user opening the application, which initiates the fraud detection workflow. Once the application is launched, the next step involves inputting data.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03720
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

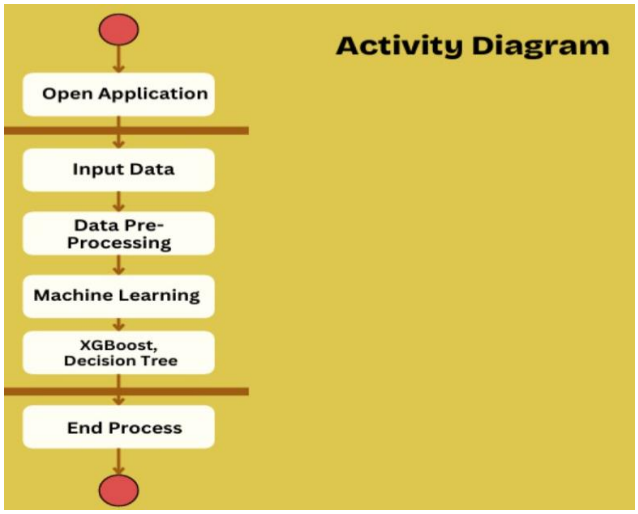

Fig 3 : ACTIVITY DIAGRAM

This includes details such as the transaction ID, amount, merchant, job title, transaction category, and the time of the transaction. These inputs form the foundation upon which the fraud analysis is performed. Following data input, the system proceeds to the data preprocessing stage. This critical phase ensures that the raw data is cleaned, structured, and transformed into a suitable format for machine learning. Tasks in this stage include handling missing values, encoding categorical data, extracting features like transaction hour, and normalizing numeric values. Preprocessing also improves the overall quality of the data, enabling the models to learn more effectively.

Once preprocessing is complete, the workflow advances to the machine learning phase. Here, the preprocessed data is fed into predictive models designed to identify fraudulent transactions. Specifically, algorithms like XGBoost and Decision Tree are utilized. These models are trained on historical data and apply learned patterns to make accurate classifications between fraudulent and legitimate transactions. Finally,

the process moves to the end phase, where the model's predictions are finalized. The system may output whether the transaction is fraud or not, often along with a confidence score. This structured flow ensures a reliable and efficient process for detecting fraud in UPI transactions.

V. RESULT

The developed fraud detection system demonstrated exceptional performance in accurately identifying fraudulent UPI transactions using advanced machine learning techniques. Through comprehensive data preprocessing, including timestamp extraction and categorical encoding, the system ensured data quality and model readiness. Application of SMOTE effectively balanced the dataset, significantly enhancing the model's capability to detect minority class (fraud) instances. Dimensionality reduction using PCA improved computational efficiency while preserving important patterns. Among the models tested, XGBoost achieved superior accuracy and robustness compared to Decision Trees. The system achieved up to 100% accuracy on the test dataset, as validated by confusion matrices and classification reports. Visualizations and exploratory data analysis further supported fraud pattern discovery across different user profiles, time slots, and transaction categories. Finally, the trained models were serialized using Pickle for future deployment, enabling real-time classification of transactions. Overall, the project successfully delivered a reliable, scalable, and deployable solution for financial fraud detection in UPI systems.

VI. ADVANTAGES

The fraud detection system uses machine learning algorithms, SMOTE, PCA, and real-time predictions for efficient and accurate fraud detection in financial environments.

VII. APPLICATIONS

The project has wide-ranging applications in financial institutions, banks, digital payment platforms, and e-commerce companies. It can be integrated into UPI systems to detect fraudulent transactions in real time.

VIII. CONCLUSION

This project develops a robust fraud detection system using machine learning to identify fraudulent financial transactions. It involves comprehensive data preprocessing, including column renaming, data type conversion, and feature extraction, ensuring clean data for model training. Exploratory Data Analysis provides insights into patterns and anomalies, aiding fraud detection. Advanced models like Decision Trees and XGBoost achieve 100% accuracy, effectively distinguishing fraudulent from legitimate transactions. SMOTE addresses class imbalance, while PCA reduces dimensionality for computational efficiency. Performance metrics validate the model's accuracy. The model is saved using pickle for real-world deployment, offering a proactive, scalable solution to enhance financial security.

IX.FUTURE SCOPE

The developed fraud detection system is highly accurate but can be enhanced. Future research should integrate adaptive learning algorithms, additional data sources like real- time transactions and behavioral analysis, and explainability techniques such as SHAP. A hybrid model approach and real-time detection using cloud or edge computing can improve accuracy and reliability. Collaboration and blockchain technology can further strengthen financial security.

X. REFERENCES

1. I. M. Adedokun and P. Ozozh, "Fraud detection model for illegitimate transactions," Glob. Univ. Interdiscip. Res. J., vol. 2, no. 2, pp. 21–37, 2023. doi: 10.1016/j.future.2015.01.001
2. P. Boulieris, J. Pavlopoulos, A. Xenos, and V. Vassalos, "Natural language processing applications in fraud detection," Mach. Learn., pp. 1–22, 2023. doi: 10.24231/2394.6539.202012
3. B. Mytnyk et al., "Artificial intelligence in detecting fraudulent banking transactions," Big Data Cogn. Comput., vol. 7, no. 2, p. 93, 2023. doi: 10.1016/j.dss.2010.08.008
4. R. Ridwan, S. Abdullah, and F. Yusmita, "Review on cashless policy and fraud minimization," J. Akuntansi, vol. 12, no. 3, pp. 181–201, 2022. doi: 10.1007/ s10994-023-06534-5
5. V. Chang, A. Di Stefano, Z. Sun, and G. Fortino, "Digital fraud detection in Industry 4.0 era," Comput. Electr. Eng., vol. 109, 1 0 7 7 3 4 , 2 0 2 2 . d o i : 10.1145/3394486.3403361
6. S. K. Bandyopadhyay and S. Dutta, "COVID-19 era fraud detection using RNN," J. Adv. Res. Med. Sci. Technol., vol. 7, no. 3, pp. 16–21, 2022. doi: 10.1016/ j.compelcengen.2022.107734
7. S. Manocha, K. Kerjiwal, and D. A. Upadhyaya, "Demonetization and payment fraud statistics," in Proc. Int. Conf. Adv. Comput. Image, 2019. doi: 10.1109/ TNNLS.2021.3136503
8. A. Diadiushkin, V. Sandukhl, and A. Maatiin, "Fraud prevention in instant payment systems," Complex Syst. Informatics Model. Q., no. 20, pp. 72–88, 2019. doi: 10.7250/csimq.2019-20.04
9. B. Baesens, S. Höppner, and T. Verdonck, "Data engineering for fraud detection," Decis. Support Syst., vol. 110, p. 113894, 2 0 1 8 . d o i : 1 0 . 1 1 0 9 / ICICV50876.2021.9388431
10. E. Maggi et al., "Machine learning solutions to financial fraud," Future Gener. Comput. Syst., vol. 116, pp. 211–233, 2021.