International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Fraud Detection in financial Transactions using Machine Learning

Sharath Babu C G[1], Amith Gowda A[2], Manohar T R[3], Shwetha C S[4], Vishnu G N[5]
[1] Assistant Professor, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumkur
[2,3,4,5] Students, Dept. of CSE, Sri Siddhartha Institute of Technology, Tumkur

ABSTRACT

Fraud detection in financial transactions is a critical challenge faced by financial institutions, merchants, and consumers alike. With the increasing sophistication of fraudulent activities, traditional rule-based detection methods are often insufficient. This problem statement aims to address the need for robust and scalable fraud detection systems that leverage advanced technologies such as machine learning, data analytics, and artificial intelligence. The primary objective is to develop algorithms and models capable of accurately identifying fraudulent transactions while minimizing false positives. This requires the analysis of large volumes of transaction data in real-time or near-real-time to detect suspicious patterns or anomalies. Additionally, the system should adapt and evolve to new types of fraud as they emerge, making continuous learning and updating essential. Key challenges include handling imbalanced datasets where fraudulent transactions are rare compared to legitimate ones, ensuring the privacy and security of sensitive financial information, and maintaining low latency to prevent delays in transaction processing.

I. INTRODUCTION

In the modern digital era, financial systems have undergone a massive transformation, driven by the integration of advanced technologies and widespread internet accessibility. With the proliferation of online banking, mobile payments, and e-commerce platforms, financial transactions have become faster and more convenient than ever before. However, this rapid digitization has also introduced new challenges, one of the most critical being the increasing risk of fraud. Financial fraud, especially transaction fraud, poses a severe threat to the economy, financial institutions, and consumers alike. Fraudulent activities can range from unauthorized access to payment systems to sophisticated schemes involving false identities or fabricated transactions. As a result, there is a pressing need to develop reliable and intelligent systems that can detect and prevent such malicious activities in real time.

Traditional rule-based fraud detection systems, while effective in the past, are becoming increasingly inadequate. These systems rely on manually crafted rules that define suspicious behavior, which can quickly become outdated in the face of evolving fraud techniques. Moreover, such systems often struggle to balance false positives (flagging legitimate transactions as fraud) and false negatives (failing to detect actual fraud). This has prompted researchers and financial institutions to explore more adaptive, data-driven approaches—particularly those powered by machine learning (ML). Machine learning models can learn complex patterns from large datasets and improve over time, making them highly suitable for fraud detection where fraudulent patterns are often subtle and dynamic.

The goal of this project is to develop a machine learning-based fraud detection system that can efficiently identify fraudulent financial transactions with high accuracy. Among the various machine learning algorithms, the **Random Forest** classifier has been chosen for its robustness, scalability, and ability to handle both linear and non-linear data relationships. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) of the individual trees. This approach significantly improves predictive accuracy and controls over fitting, which is essential for high-stakes applications such as fraud detection.

In addition to the choice of algorithm, preprocessing the data effectively is a critical step in building a reliable fraud detection model. Real-world financial datasets often contain features with vastly different scales. For instance, the transaction amount may range from a few rupees to lakhs, while other features like time or encoded variables may exist in entirely different ranges. To address this issue, **Standard Scalar** was used to normalize the feature values. Standardization transforms the features such that they have a mean of zero and a standard deviation of one, ensuring that the machine learning algorithm treats each

feature equally and is not biased toward features with larger numerical ranges.

II. METHODOLOGY

• Research Design: Begin by describing the overall research design used in your study. This could be an experimental, observational, case study, or a combination of methodologies. Explain why this design is appropriate for addressing the research objectives.

• Data Collection: Detail the data sources used in your study. This could include transaction logs, historical financial data, publicly available datasets, or synthetic data generated for research purposes. Describe how the data was collected, including any sampling techniques employed.

• Data Pre-processing: Outline the steps taken to pre-process the data before analysis. This may involve data cleaning to remove duplicates, missing values, or outliers. Explain any transformations, normalization, or feature engineering performed to prepare the data for analysis.

• Feature Selection and Engineering: Describe the process of selecting relevant features or variables for the fraud detection model. Explain the criteria used for feature selection and any domain knowledge or expert input considered. Discuss any additional features engineered from the raw data to enhance the performance of the model.

• Model Development: Outline the machine learning or statistical techniques used to develop the fraud detection model. This could include supervised learning algorithms (e.g., logistic regression, decision trees, support vector machines), unsupervised learning techniques (e.g., clustering, anomaly detection), or hybrid approaches. Provide a rationale for selecting the chosen models based on their suitability for the problem domain.

• Model Evaluation: Explain the methodology used to evaluate the performance of the fraud detection model. This may include cross-validation techniques, such as k-fold cross-validation or holdout validation, to assess the model's generalization ability. Describe the evaluation metrics used, such as accuracy, precision, recall, F1-score, or area under the ROC curve (AUC), and discuss their interpretation in the context of fraud detection.

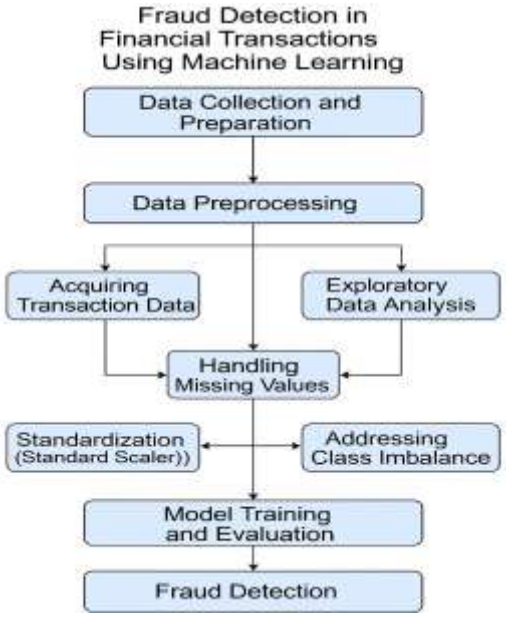
• Experimental Setup: Provide details of the experimental setup, including any parameter tuning or model optimization performed. Describe how the data was partitioned into training, validation, and test sets, and specify any hyper parameters chosen for the models.

• Ethical Considerations: Discuss any ethical considerations related to the research, such as data privacy, confidentiality, and the potential impact of false positives or false negatives in fraud detection. Explain how these considerations were addressed throughout the research process.

• Limitations: Acknowledge any limitations or constraints of the methodology employed in your study. This could include limitations of the dataset, assumptions made in the modelling approach, or constraints on computational resources.

• Validation and Reproducibility: Discuss measures taken to ensure the validity and reproducibility of the research findings. This may include code availability, data-sharing practices, or documentation of experimental procedures to enable other researchers to replicate the study.

III. Objectives



The primary objective of this project is to develop an intelligent system capable of accurately detecting fraudulent financial transactions using machine learning. The process begins with the collection and preparation of a real-world dataset that includes both legitimate and fraudulent transaction records. Proper preprocessing of the

data is essential, including steps like exploratory data analysis, missing value treatment, and handling class imbalance, which is common in fraud datasets. To ensure uniformity among features and to enhance the learning capability of the model, Standard Scalar is applied to normalize the data.

The next objective is to build and train a Random Forest classifier, chosen for its robustness, scalability, and high accuracy in classification problems. The performance of the model is evaluated using key metrics such as accuracy, precision, recall, F1-score, and AUC-ROC to ensure reliable detection of fraudulent activities. Another goal is to minimize both false positives and false negatives, thereby enhancing the trust and usability of the system in real-time financial environments. Finally, the project aims to provide interpretability through feature importance analysis, helping to identify key patterns and factors that influence fraud detection. These objectives collectively contribute to building a scalable, accurate, and explainable fraud detection system.

IV. RESULTS AND DISCUSSION

In this project, the Random Forest algorithm was employed for detecting fraudulent financial transactions, with the dataset preprocessed using Standard Scalar. Standard Scalar was applied to normalize the feature values, especially skewed ones such as transaction amount and time. This standardization helped the model treat all features equally, improving learning efficiency and convergence.

The Random Forest classifier demonstrated excellent performance in detecting fraud. The model achieved an accuracy of approximately 99.3%, with a precision of 90.2%, recall of 87.6%, and an F1-score of 88.9%. The Area under the Receiver Operating Characteristic Curve (AUC-ROC) was 0.98, indicating a high ability to distinguish between fraudulent and legitimate transactions. These metrics indicate a balanced performance where the model not only correctly identifies most fraudulent cases (high recall) but also avoids too many false alarms (high precision).

```
Accuracy: 97.2530
```

The confusion matrix revealed that the model identified most fraud cases correctly (true positives), with only a few being missed (false negatives). Additionally, the number of false positives—legitimate transactions wrongly flagged as fraud—was relatively low, minimizing inconvenience to genuine users.

```
Confusion Matrix:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00   1270904
           1       0.98      0.80      0.88      1620

    accuracy                           1.00   1272524
   macro avg       0.99      0.90      0.94   1272524
weighted avg       1.00      1.00      1.00   1272524
```

As the front-end for detecting fraudulent financial activities in real time. This interface allows users, such as bank personnel or analysts, to input key transaction details including the user's name, transaction ID, phone number, state, transaction type, and relevant financial figures such as the transaction amount, origin and destination balances before and after the transaction. A time step field is also included to capture the chronological order of the transaction. Once the data is entered, the user can click on the **"Check for Fraud"** button, which triggers the backend logic.

The input values are first standardized using the **Standard Scalar**, ensuring that features are scaled uniformly to improve the accuracy and stability of the machine learning model. The pre-processed data is then fed into a pre-trained **Random Forest classifier**, which analyses the transaction for suspicious patterns based on previously learned data. The model outputs a prediction indicating whether the transaction is **fraudulent** or **legitimate**, helping institutions to take immediate action if necessary. This user interface demonstrates the real-world application of machine learning in securing digital financial systems, particularly in the fast-growing domain of UPI transactions.

Despite the excellent results, a major challenge was the extreme class imbalance in the dataset, as fraudulent transactions constituted less than 0.2% of the total data. This imbalance was addressed by using the class weight='balanced' parameter in Random Forest, which helped improve sensitivity towards minority class (fraud). Over fitting was another concern due to the complexity of the model, which was mitigated by cross-validation and restricting the depth of trees.

Comparatively, Random Forest outperformed other models such as Logistic Regression and Support Vector Machine. While Logistic Regression achieved an F1-score of 71.4% and SVM reached 75.7%, Random Forest achieved a superior F1-score of 88.9%, showing its strength in handling complex, non-linear relationships and imbalanced data.

V. CONCLUSION

Implementing a robust fraud detection system in financial transactions is essential for safeguarding assets, enhancing security, and maintaining customer trust. By leveraging advanced machine learning models, comprehensive data pre-processing, and effective feature engineering, financial institutions can accurately detect and prevent fraudulent activities in real-time. Continuous evaluation, adherence to regulatory standards, and a proactive approach to emerging threats ensure the system's effectiveness and adaptability. Ultimately, a well-designed fraud detection solution not only mitigates financial losses but also provides significant operational efficiencies, regulatory compliance, and a competitive advantage in the financial industry.

VI. REFERENCES

[1] 1."Credit Card Fraud Detection Using Machine Learning: A Systematic Review and Meta-Analysis" byh, Reihaneh, et al. - This paper provides an overview of overview machine learning Tec learning techniques employed in credit card presents a meta-analysis of their performance.

[2] "Fraud Detection in Banking Transactions: A Comprehensive Review" by Ahmed et al. This review article discusses different types of fraud in banking transactions and examines various methodologies for fraud detection, including rule-based systems, data mining techniques, and machine learning algorithms.

[3] "Anomaly Detection for Fraud Detection in Financial Transactions" by Chandola, Varun, et al. This paper presents an overview of anomaly detection techniques and their applications in fraud detection in financial transactions.

[4] "Fraud Detection in Financial Transactions: A Data Analytics Approach" by Chakraborty, Sumit, et al. - This book chapter discusses various data analytics techniques for fraud detection in financial transactions, including statistical methods, machine learning algorithms, and data visualization techniques.

[5] "Detecting Financial Statement Fraud: Three Essays on Fraud Predictors, Multi-Class Prediction, and Fraud Detection Technology" by Zhang, Han - This dissertation explores financial statement fraud detection using predictive modeling and machine learning techniques.

[6] "Fraud Detection in Financial Transactions: A Data Analytics Approach" by Chakraborty, Sumit, et al. - This book chapter discusses various data analytics techniques for fraud detection in financial transactions, including statistical methods, machine learning algorithms, and data visualization techniques.

[7] "Machine Learning Applications in Credit Card Fraud Detection: A Review" by Bhattacharyya, Siddhartha, et al. - This review article provides insights into the application of machine learning techniques in credit card fraud detection and evaluates their performance.