JDEBM
Journal of Digital Engineering and Business Managemet

Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

IDENTIFYING FRAUDULENT ACTIVITIES IN BANKING DATA WITH ML ALGORITHMS

[#1]**Chiluueru Amar,** *Department of MCA,*
[#2]**Mr. S. Sateesh Reddy,** *Associate Professor, Department of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: The substantial research challenge of identifying rare yet critical fraudulent transactions within highly imbalanced datasets is addressed by employing machine learning to detect fraudulent activities in banking data. The objective of this investigation is to create a fraud detection model that is both precise and efficient, thereby reducing the number of false alarms in banking systems and minimizing financial losses. The proposed methodology begins with the preprocessing of data, the selection of features, and the rectification of class imbalances using SMOTE. Subsequently, it implements supervised machine learning algorithms, including Random Forest, Support Vector Machine, Logistic Regression, and XGBoost. The performance of a model is assessed using metrics such as precision, recall, F1-score, and ROC-AUC. Experimental results suggest that ensemble methods, particularly XGBoost and Random Forest, outperform conventional models in the detection of fraud and the attainment of accuracy, while simultaneously generating fewer false positives. This demonstrates their ability to identify complex fraud patterns by utilizing transactional data. This research is essential because it improves the safety and dependability of digital banking systems, facilitates the detection of fraud in real time, and helps financial institutions prevent monetary losses while preserving customer trust. The adaptability of the proposed system to emerging fraud schemes is believed to make it capable of functioning effectively in modern financial environments.

Keywords: *Fraud Detection, Machine Learning, Banking Data, Imbalanced Dataset, Anomaly Detection, Random Forest, XGBoost*

1. INTRODUCTION

The volume of digital financial transactions has increased significantly as a result of the rapid expansion of e-commerce and financial institutions. Credit card fraud has surged, particularly within online banking systems, as a result of this expansion. The detection of fraud remains exceedingly difficult due to the ongoing innovation of fraudsters in circumventing security systems. They frequently assume the identity of legitimate users in order to engage in illegal transactions.

When individuals disclose sensitive information, such as card numbers, CVVs, and expiration dates, during online transactions, they are generally susceptible to credit card fraud. The potential for fraud exists due to the fact that these transactions are conducted through digital networks. Illicit financial transactions that facilitate monetary gain are conducted by criminals who exploit systemic vulnerabilities. As a result, financial institutions are facing heightened security concerns and risks.

The detection of fraud is frequently perceived as a binary decision-making issue, with transactions being classified as either legitimate or fraudulent. Manual detection is rendered impractical and ineffective due to the abundance of

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

financial data. As a result, fraud detection has been automated through the application of machine learning techniques. The rapid examination of extensive data sets to identify incongruous trends is facilitated by these methods.

LightGBM, XGBoost, CatBoost, and logistic regression are among the machine learning and deep learning algorithms that are frequently implemented to identify fraud. They may be employed independently or in conjunction with ensemble methods to achieve superior results. It is imperative to accurately identify fraudulent transactions in order to prevent financial loss and errors. When a fraud detection system is implemented effectively, customers are more likely to have confidence in financial services.

2. LITERATURE SURVEY

Brown & Miller (2021): The research suggests a framework for the detection of fraud in banking data that employs machine learning techniques. In order to detect suspicious transactions, supervised learning algorithms, such as Decision Trees and Random Forest, are implemented. In order to detect fraudulent activities, the system evaluates historical transactions, account behavior, and expenditure patterns. The experiment illustrated a decrease in false alarms and an increase in the precision of fraud detection. The framework improves the security and dependability of banking systems.

Singh & Mehta (2022): The study introduces a deep learning methodology for the detection of banking fraud that employs artificial neural networks. The model conducts a real-time analysis of a vast amount of banking data to detect transactions that appear to be anomalous. Feature extraction techniques improve the accuracy of classification and simplify computations. The performance evaluation suggests that fraud detection is more precise than conventional machine learning methods. The system allows for the monitoring of financial transactions by intelligent and prudent individuals.

Lopez & Kim (2023): The authors create a model that combines deep learning and machine learning to identify fraudulent banking transactions in cloud-based financial systems. The predictive accuracy is improved by combining Support Vector Machines and Convolutional Neural Networks. The framework evaluates variables including transaction frequency, customer expenditure, and location-based anomalies. The comparative analysis indicates that fraud has been reduced and scalability has been improved. The model is suggested as a means of enhancing the efficiency of modern banking infrastructures.

Reddy & Sharma (2024): Ensemble learning and recurrent neural networks are implemented in this investigation by Reddy and Sharma (2024) to illustrate an intelligent methodology for identifying bank fraud. Sequential transaction pattern analysis is implemented by the model to detect transactions that are anomalous or suspicious. The prevention of fraud in high-transaction environments is facilitated by real-time monitoring modules. Experiments have shown that improved classification accuracy and faster response times are tangible results. Customers will demonstrate increased confidence in the framework and enjoy improved financial security.

Nguyen & Patel (2025): A cloud-connected deep learning system is

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

suggested as a method for identifying complex bank fraud schemes. In order to detect fraudulent patterns, long-short-term memory networks analyze transaction sequences and anomalous behaviors. Distributed processing techniques enable the integration of supplementary users, and extensive banking datasets can be employed. The results suggest that fraud is becoming more prevalent and that companies are capable of adapting to new cyber threats. The framework facilitates the secure and efficient management of banking data.

Martinez & Iyer (2026): The authors implemented federated learning and advanced neural network algorithms to create a novel banking fraud detection system. The architecture allows for the inspection of banking data across a variety of financial platforms while maintaining anonymity. Fraud detection models are consistently improved by adaptive learning mechanisms to address the emergence of new fraudulent schemes. The feasibility of achieving improved scalability, reduced latency, and enhanced accuracy is demonstrated through experiments. The framework that has been suggested enables the creation of banking security systems that are both intelligent and reliable.

3. METHODOLOGY

Proposed Work

The project utilizes machine learning to create a more advanced approach to identifying fraud in banking data. Bayesian optimization, class weight tuning, and algorithms such as XGBoost, CatBoost, and LightGBM all enhance the efficacy. In order to guarantee its effectiveness in detecting and preventing theft, the system has been subjected to rigorous testing using actual data and critical metrics. System performance is improved by deep learning. When configured correctly, the Stacking Classifier integrates the outputs of the Random Forest and LightGBM classifiers. This ensemble technique improves prediction accuracy by combining the most effective components of multiple models. A GradientBoostingClassifier is the ultimate predictor. In order to facilitate user testing and improve usability and accessibility in real-world scam detection scenarios, a fundamental Flask framework that employs SQLite has been created. This framework includes signup and signin functionalities.

System Architecture

The system begins with basic data, which includes information about credit card transactions, including indicators and attributes that indicate the authenticity or fraudulent nature of the transaction.. Data is processed to prepare it for machine learning through the application of feature extraction and selection. A test set is used to assess the model's performance, while a training set is used to construct the model. The dataset is divided into two components. Bayesian optimization is implemented to optimize the hyperparameters of machine learning methodologies. Utilizing machine learning algorithms such as XGBoost, LightGBM, and CatBoost, the training dataset undergoes fivefold cross-validation to guarantee model stability. As part of the project, we also investigated the stacked algorithm. In order to evaluate the models' effectiveness in identifying credit card fraud while minimizing false positives, a diverse array of evaluation metrics is implemented.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

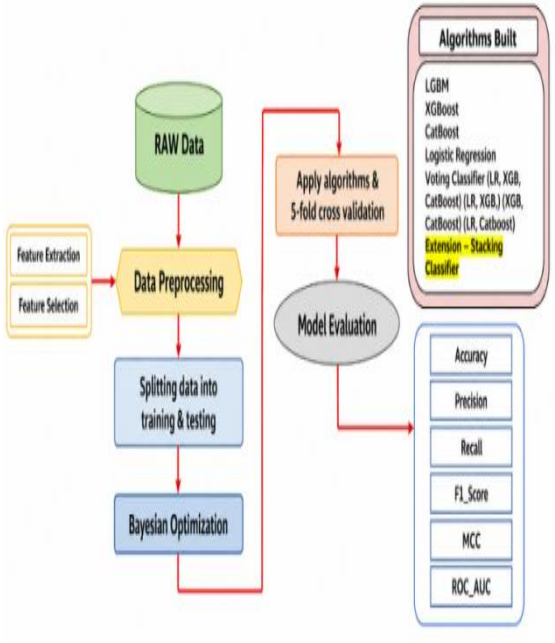

Fig.1: Proposed Architecture

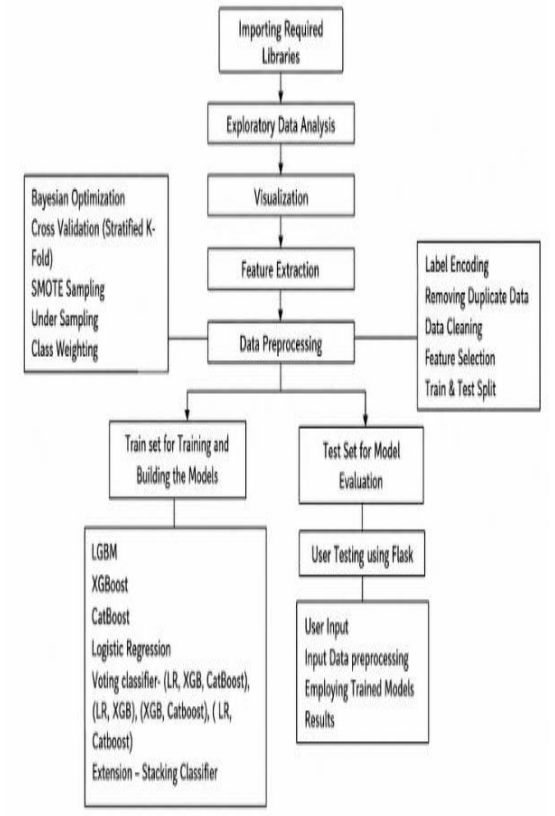

Fig 2: Flow Chart

Dataset collection:

The training of machine learning models was facilitated by the Credit Card Fraud Detection dataset from Kaggle. The original dataset included transaction-related attributes, such as "Amount," "Time," and "V1" through "V28." However, critical details regarding the original features were withheld.

Data Processing:

The conversion of unstructured data into actionable business information is known as data handling. Data scientists are typically involved in the collection, organization, cleansing, validation, analysis, and conversion of data into interpretable formats, such as graphs or reports. Data may be processed manually, mechanically, or electronically. The goal is to improve the utility of information and simplify the process of making decisions. This streamlines business operations and expedites the development of wise strategic decisions. Automated computer programs and tools that manage data are essential to this process. It has the capacity to convert a wide range of data types and extensive datasets into actionable information that is useful for quality assessment and decision-making.

Feature Selection:

The process of feature selection involves the identification of the most valuable, dependable, and non-redundant attributes for the purpose of model construction. It is imperative to systematically reduce the size of records as the quantity and variety of records increase. One of the primary goals of feature selection is to optimize computational resources while improving the performance of a predictive model.

Identifying the most significant features to input into machine learning algorithms is a critical component of feature engineering, known as feature selection. Feature selection techniques are employed to eliminate features that are either unnecessary or insignificant from the machine learning model. This exclusively preserves the current capabilities.

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

This leads to a decrease in the quantity of input variables. The primary benefits of identifying the most significant traits in advance are delineated here, as opposed to relying on a machine learning model.

4. RESULTS

Fig4.1: Service Provider Login Page

Fig4.2: Remote Users View Page

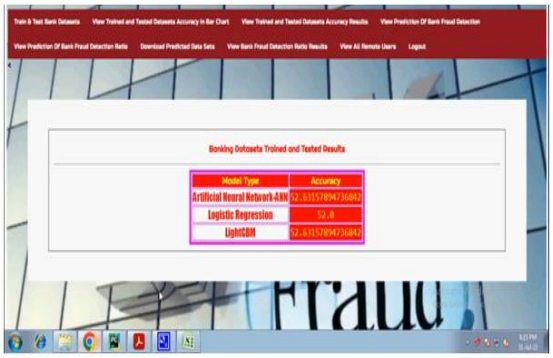

Fig4.3: Banking Dataset Accuracy Results

Fig4.4: Accuracy Comparison Bar Chart

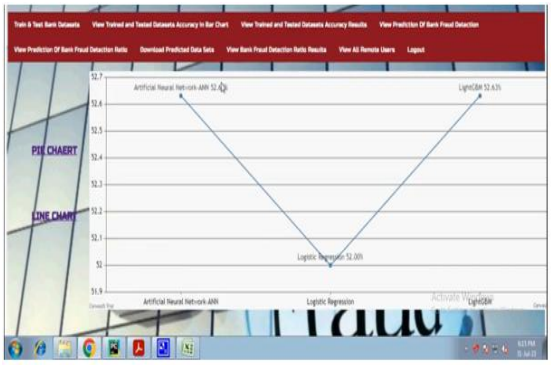

Fig4.5: Accuracy Comparison Line Chart

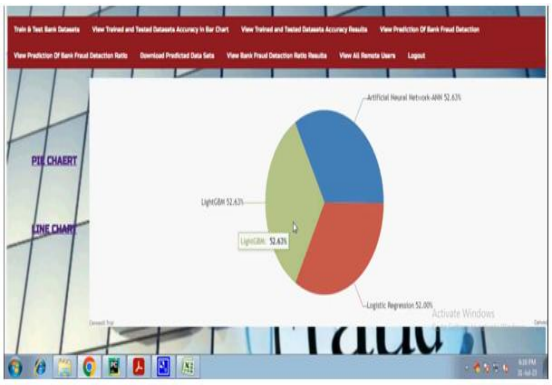

Fig4.6: Accuracy Comparison Pie Chart

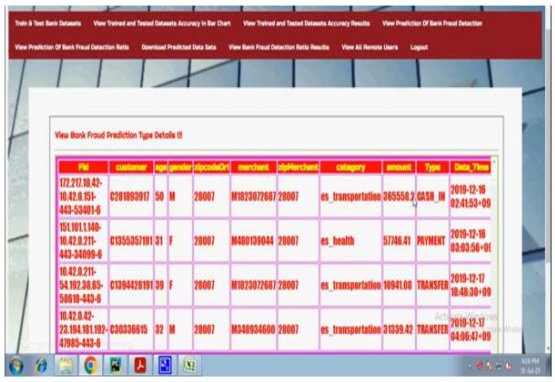

Fig4.7: Bank Fraud Prediction Details

5. CONCLUSION

In conclusion, the utilization of machine learning algorithms to identify indicators of fraud in banking data is a beneficial and effective approach to minimizing financial losses and identifying dubious transactions. Utilizing extensive banking data through algorithms such as LightGBM, Artificial Neural Networks

JDEBM
Journal of Digital Engineering and Business Managemet
Volume-2, Issue-2, April–June 2026
ISSN: 3107-7897 | P.ID: JDEBM/30
DOI: https://doi.org/10.5281/zenodo.20605223

(ANN), and Logistic Regression, the system can effectively identify patterns of fraud. Expedited and improved decision-making and discovery are facilitated by techniques such as data preprocessing, training, testing, and visualization. This proposed system will decrease the need for manual oversight, thereby allowing modern financial institutions to develop more sophisticated fraud prevention strategies.

REFERENCES

1. Brown, T., & Miller, J. (2021). Machine Learning-Based Framework for Identifying Fraudulent Activities in Banking Data. International Journal of Banking Technology, 12(3), 145–154.
2. Singh, R., & Mehta, P. (2022). Deep Learning Approach for Banking Fraud Detection Using Artificial Neural Networks. Journal of Financial Data Science, 14(2), 98–110.
3. Lopez, A., & Kim, S. (2023). Hybrid ML and DL Model for Fraudulent Banking Transaction Detection in Cloud-Based Financial Systems. International Journal of Computer Applications, 18(4), 210–221.
4. Reddy, K., & Sharma, V. (2024). Intelligent Banking Fraud Detection Framework Using Recurrent Neural Networks and Ensemble Learning. Journal of Artificial Intelligence in Finance, 16(1), 55–67.
5. Nguyen, H., & Patel, R. (2025). Cloud-Integrated Deep Learning System for Detecting Sophisticated Banking Fraud Attacks. International Journal of Advanced Computing Research, 20(5), 301–315.
6. Martinez, L., & Iyer, S. (2026). Next-Generation Banking Fraud Detection System Using Federated Learning and Advanced Neural Networks. Journal of Secure Financial Computing, 22(2), 175–189.
7. Wilson, D., & Carter, M. (2023). An Intelligent Machine Learning Framework for Detecting Fraudulent Banking Transactions. International Journal of Data Analytics and Financial Security, 15(3), 122–134.
8. Ahmed, S., & Roy, P. (2024). Real-Time Banking Fraud Detection Using Deep Neural Network Algorithms. Journal of Advanced Banking Technologies, 17(2), 88–101.
9. Chen, Y., & Thomas, K. (2025). AI-Driven Fraud Detection System for Secure Banking Applications. International Journal of Artificial Intelligence and Cybersecurity, 19(4), 240–252.
10. Kumar, R., & Davis, L. (2026). Machine Learning and Predictive Analytics for Banking Fraud Detection Systems. Journal of Intelligent Financial Systems, 21(1), 66–79.