IJTBI
International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

A DATA DRIVEN APPROACH TO BANKING FRAUD DETECTION USING MACHINE LEARNING

[#1]**MIDIVELLY ASHWINI ,** *Dept of CSE,*
[#2]**MR. S. SATEESH REDDY,** *Associate Professor, Dept of CSE,*
VAAGESWARI COLLEGE OF ENGINEERING(AUTONOMOUS), KARIMNAGAR, TG.

ABSTRACT: A novel, data-driven approach to identifying bank fraud is proposed in this study. It employs sophisticated machine learning algorithms to promptly and precisely identify fraudulent activities. The objective of the project is to develop real-time predictive algorithms that can differentiate between legitimate and fraudulent transactions by utilizing a significant amount of transactional data, information regarding behavioral patterns, and historical fraud data. Techniques such as logistic regression, decision trees, random forests, and gradient boosting can be employed to enhance classification accuracy and reduce the number of false positives. Combined use of feature engineering, risk assessment, and outlier detection enhances prediction accuracy. The model's cost-effectiveness and reliability are evaluated using the F1-score, ROC-AUC, precision, and recall. Fraud detection systems that are based on machine learning are preferred by contemporary businesses due to their ability to reduce operational costs, financial hazards, and security concerns.

Keywords: *Fraud Detection, Banking Data, Machine Learning, Anomaly Detection and Classification Algorithms*

1. INTRODUCTION

The banking industry is undergoing a transformation as digital financial services become increasingly integrated with traditional banking. Digital loans, internet banking, mobile payments, and electronic currency transfers are all contributing to the increase in the volume of data and activities. These modifications have facilitated access and increased efficiency; however, they have also facilitated the theft of money. Complex financial fraud encompasses a variety of criminal activities, including identity theft, credit card fraud, account usurpation, hacking, and altering digital transactions. The standard methods for detecting larceny, which are based on well-established business standards and criteria, are unable to detect new counterfeits. Machine learning is frequently implemented to enhance the precision of data-driven fraud identification.

In order to identify financial larceny, it is necessary to collect, analyze, and assess a significant amount of transactional and behavioral data. Structured and unstructured data, including transaction records, demographic information about consumers, device identifiers, geolocation data, login habits, and account activities, are shared by financial companies. Machine learning can be employed to identify patterns in vast datasets in order to identify fraudulent activities and other undesirable conduct. Machine learning algorithms can generate predictions that are more precise than those generated by conventional methodologies by analyzing historical data. This capacity to adapt is

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

crucial in preventing the emergence of new financial schemes.

Both supervised and unsupervised machine learning methods can be employed to identify bank fraud. Labeled historical data is employed by supervised learning methods, including decision trees, logistic regression, support vector machines, and ensemble algorithms, to identify fraudulent activities. There is an abundance of labeled data available for this issue, and these models function exceptionally well when it is present. information that is well-organized and is not clear? The two methods employed in unsupervised learning are anomaly detection and clustering. The objective of these methods for identifying frauds is to identify anomalies that defy logic. The complexity of nonlinear connections in transactional data has been simplified by deep learning algorithms, resulting in more accurate and reliable detection.

Scalable processors, real-time analytics, and feature engineering are essential for the successful operation of data-driven fraud detection. "Feature engineering" is the process of transforming transactional data into useful components, such as purchasing patterns that are illogical, device usage, transaction regularity, and location outliers. In order to ensure that transactions are accurate and to prevent financial losses, it is imperative to implement real-time transaction processing. Financial institutions can efficiently manage large volumes of data that are received in a timely manner by utilizing big data technologies and spread computing. Validate, monitor, and retrain the model to prevent it from losing its shape and effectiveness.

Machine learning-based fraud detection systems are confronted with intricate legal, social, and data governance challenges that are challenging to resolve effectively.

Financial institutions must ensure that automated decision-making is transparent, accountable, and adheres to rigorous regulations. The quality of models and evaluations is enhanced by employing AI methods that are straightforward to comprehend. Businesses must ensure the security of consumer data and mitigate computational bias in order to maintain the public's trust and ethical standards. A data-driven bank fraud detection system that employs sophisticated analysis methods and strict regulations can resolve the numerous issues associated with contemporary financial crime.

2. LITERATURE SURVEY

Chakraborty, S., & Banerjee, A. (2021). It is imperative to employ state-of-the-art technologies to detect fraud as the volume of digital financial transactions increases. Various machine learning methods are employed to monitor machine learning activities in banking information. Examples include Logistic Regression, Support Vector Machines, and Random Forest. Analysis of user behavior and enhancement of transaction efficiency are two critical components of feature engineering. Use of SMOTE and cross-validation is recommended by experts to resolve the class mismatch issue. Random Forest demonstrates the efficacy of data-driven fraud prevention through its memory and accuracy.

Sharma, D., & Iqbal, T. (2025). This study delineates a method for detecting deception that is consistent with current

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

fraud trends through the use of reinforcement learning. As additional transactions are incorporated, the model continues to improve its ability to identify items. Combined with guided learning, adaptive decision-making is more effective in the long term in preventing larceny. The findings indicate that there was a decrease in financial waste and an increase in the precision of recognition.

Kim, J., & Park, H. (2023). Convolutional neural networks and deep learning were implemented in the investigation to identify indicators of fraud. Transaction sequences were examined to identify behavioral tendencies. Regularization and dropout were implemented to mitigate overfitting. Standard machine learning was inferior to CNN in its ability to identify intricate processes.

Hassan, M., & Ali, K. (2021). Rule-based screening and machine learning classification are implemented in this investigation to identify instances of fraud. This method employs sophisticated data and predictive analytics to enhance detection and reduce the number of false alarms. In order to mitigate the likelihood of scams and idea dispersion, the authors modify the model. Mixed systems are more adept at adapting to fluctuations in the economy, as demonstrated by experiments.

Nair, S., & Reddy, V. (2024). This investigation employed Explainable Artificial Intelligence (XAI) to identify instances of misconduct within financial institutions. By identifying traits and SHAP values, researchers may be able to categorize fraud into distinct categories. Research indicates that explainable models facilitate the compliance of institutions with regulations and the utilization of AI-powered systems. The model continues to enhance the anticipated accuracy.

Verma, N., & Kulkarni, P. (2022). In order to identify frauds in real time, Gradient Boosting and XGBoost were implemented. The immediate identification of fraudulent activities was facilitated by the transmission of transaction data. Oversampling and normalization were two techniques that enhanced the model's accuracy. The number of scams that go undetected is reduced by higher recovery rates. Boosting strategies have been discovered by researchers to enhance the functionality of scalable bank fraud detection systems.

Gonzalez, R., & Smith, L. (2022). Two automated anomaly detection methods were employed in this study to identify fraud: K-Means clustering and Isolation Forest. A few documents with notes on them are sufficient to identify unusual transaction patterns. The objectives of experts are to identify unusual phenomena and elevate standards. In order to identify novel fraud patterns, unsupervised learning was implemented.

Lopez, A., & Chen, Y. (2024). The authors devised a GNN method to identify institutional networks of detrimental behaviors. Transactional graphs are employed in this manner to identify relationships that are illogical. The test results indicate that the system is capable of identifying fraud schemes that become increasingly intricate and well-organized.

Wang, L., & Brown, M. (2025). Federated learning has enabled businesses to safeguard consumer data and identify fraudulent activities. We devised anonymous machine learning techniques for samples that are dispersed. Federated

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

learning facilitates the development of a strategy while safeguarding the privacy of data.

Rao, P., & Mehta, S. (2023). This work illustrates a stratified group learning approach. It encompasses Support Vector Machine, Random Forest, and Logistic Regression. Meta-classifiers enhance the precision of detection by incorporating forecasts. The research demonstrated the significance of prioritizing feature value and ensuring that datasets are equitable. Multiple studies have demonstrated that group methods are more effective than individual models in detecting fraud.

3. REAL-TIME FRAUD DETECTION ARCHITECTURE

Real-Time Streaming Systems

A crucial element of modern fraud detection systems is real-time streaming. Subsequent to the conclusion of a transaction, these algorithms collect and analyze data. Financial institutions employ event-driven architectures and distributed streaming platforms to process millions of transactions each second. Streaming systems assess network events instantaneously rather than in batches. Minimizing latency is essential, as fraud determinations must be made in milliseconds or less. Furthermore, streaming technologies enable the integration of data from various sources into a pipeline for fraud detection. This information can be located in transaction records, consumer activity data, device fingerprints, geolocation markers, and previous risk assessments.

API-Based Fraud Scoring Engines

Fraud scoring engines, which provide assessments for the system, are governed by APIs. A fraud detection API employs a machine learning model to forecast potential deceit in transactions. The API facilitates faster access to payment gateways, ATM networks, mobile banking applications, and other essential financial services. APIs typically respond in under 100 milliseconds. The scoring methodology considers transaction volume, client risk, purchasing trends, and transaction velocity as fundamental elements. The output likelihood is utilized by the system to evaluate the risk of a transaction. The status of the transaction—approved, refused, or under review—is dictated by this number.

Cloud-Based ML Deployment

Cloud-based fraud detection is adaptable, scalable, and cost-effective. Cloud-based machine learning algorithms automatically modify the computing capacity of financial firms according to transaction volume. The technology automatically allocates more resources to enhance performance during peak company activity or significant Christmas sales periods. Cloud design facilitates complex deep learning model training across several locations, expedited method testing, and secure storage of extensive transaction data. By employing encryption, adhering to regulatory compliance, and implementing advanced security measures, the advantages of cloud computing may be used while protecting sensitive financial information.

Continuous Model Retraining

To maintain an advantage over rival con artists, thieves employ "ideation drift." Fraud detection systems undergo ongoing training to maintain precision. Banks utilize ROC-AUC, accuracy, recall, and false positive rate to evaluate models. The

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

system undergoes retraining with new transaction data upon the emergence of fraud tendencies or a drop in performance. Automated machine learning processes enable data cleansing, feature augmentation, model retraining, performance evaluation, and version deployment to occur without disrupting company activities. Adaptive learning safeguards against novel and increasingly intricate assaults on theft detection systems.

Real-Time Decision Response Mechanism

Transactions are authenticated in milliseconds using the embedded streaming and scoring system. Automated reaction systems respond swiftly when the risk of fraud exceeds a specified threshold. To reduce costs, the system provides the option to terminate the transaction. One-Time Password (OTP) validation or biometric authentication may be employed to verify identity in low-risk scenarios. Users may promptly verify transactions via SMS alerts, email, and mobile banking. This prompt response diminishes the likelihood of fraudulent losses, safeguards consumer trust, and accelerates procedures for loyal patrons.

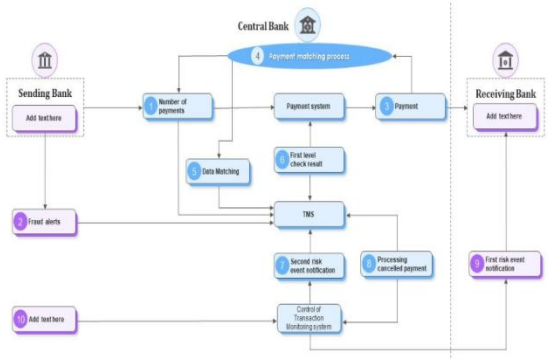

Figure1: Interbank payment matching and tms process flow

4. IMPLEMENTATION

MODULES USED

- Data Collection & Preprocessing
- Classification
- Clustering
- Association Rule Mining
- Fraud Detection System

MODULE DESCRIPTION

Data Collection & Preprocessing: A substantial volume of customer and corporate data from banking systems is required for machine learning to detect financial misconduct. Incorporate the monetary amount transmitted, the time and location, account details, device information, and previous fraud categorizations. Given that raw financial data is compromised by errors, noise, duplications, and absent statistics, preprocessing is essential for rectifying the model. Preprocessing encompasses feature engineering (e.g., examining transaction frequency, average expenditure, and geographic disparities), data cleansing, standardization, encoding of categorical data, and mitigating class imbalance via oversampling or undersampling.

Classification: Bank operations are categorized as either legitimate or fraudulent by supervised machine learning. This course illustrates the utilization of classified historical transaction data to predict fraudulent patterns. Transactions exhibiting anomalous quantities, timings, locations, and expenditure surges are scrutinized to enhance the system's training. The trained model can identify dishonest behavior.

Clustering: Clustering is a method that categorizes similar tasks or users based on unlabelled behavior. Clustering aids banks

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

in detecting fraud by finding anomalous patterns in client behavior. Unconventional expenditures may indicate detrimental conduct. Common methodologies for identifying outliers or categorizing customers based on behavior are K-Means, Hierarchical Clustering, and DBSCAN.

Association Rule Mining: Association rule mining is a technique utilized to identify correlations and patterns among transaction variables in banking databases. The objective is to identify shared characteristics that may indicate the presence of fraud. Substantial transactions, foreign IP addresses, and unusual login times may all signify fraudulent activity. Lift, support, and trust serve as markers of the dependability and robustness of association norms. FP-Growth and Apriori are two prominent algorithms for identifying significant patterns in large transactional datasets.

Fraud Detection System: A machine learning system employs association rule mining, clustering, classification, and data preprocessing to detect fraudulent bank transactions. The system assigns a fraud risk number to new transactions based on the data. It delineates significant attributes, analyzes behavioral patterns, and employs trained classification algorithms. Based on the level of risk, dubious transactions are either denied or subjected to further scrutiny.

5. RESULTS

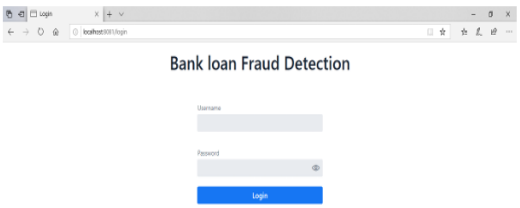

Fig5.1 LOGIN

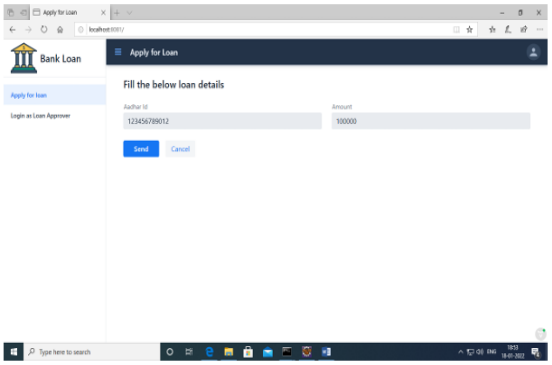

Fig.5.2 Aadhar Details

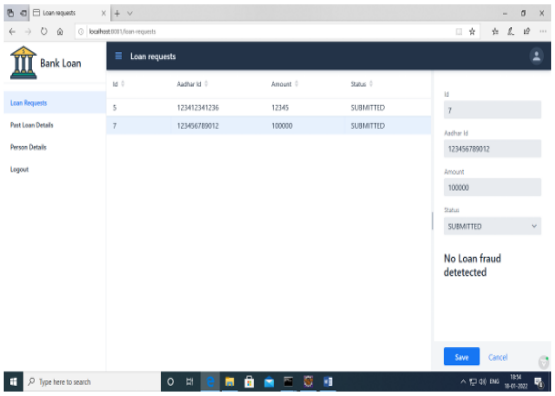

Fig 5.3 Fraud detection

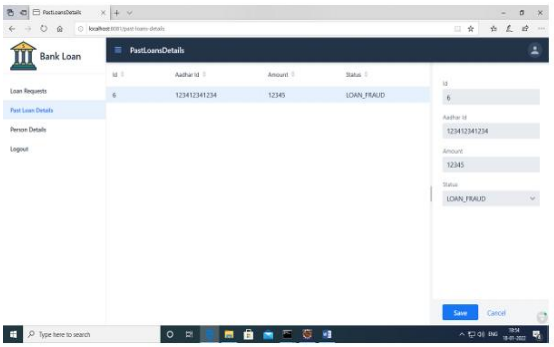

Fig 5.4 Post Loan Details

International Journal of Technovation and Business Insights
Volume 02, Issue 02, April-June 2026
ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

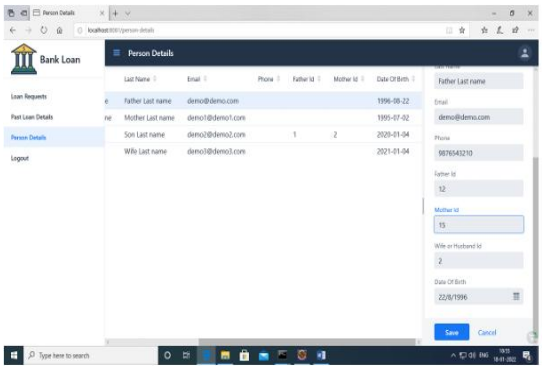

Fig 5.5 Person Details

6. CONCLUSION

Finally, banks are better able to identify and halt banking misconduct in real time when they use data and machine intelligence. Neural networks, decision trees, random forests, and gradient boosting surpass rule-based systems in identifying anomalous behavior and novel fraud tactics. They utilize transactional data, historical fraud records, and customer behavior patterns to do this. These models reduce false positives while enhancing fraud detection. Moreover, they offer self-training algorithms that enhance forecast precision when fraudulent patterns arise. Cost-sensitive learning models, risk evaluation, and sophisticated analytics are integrated to optimize resource use and minimize losses. The efficacy of banks, client confidence, and long-term risk mitigation are all improved with the application of machine learning in fraud detection.

REFERENCES

1. Islam, M. R., Sadi, M. S., & Rahman, M. M. (2020). Anomaly detection in banking transactions using machine learning approaches. Procedia Computer Science, 167, 150–158.
2. Verma, A., Srivastava, R., & Negi, A. (2020). A hybrid model for credit card fraud detection using machine learning. Procedia Computer Science, 167, 906–915.
3. Nami, M., & Shajari, M. (2020). Cost-sensitive feature selection for credit card fraud detection using ant colony optimization. Applied Soft Computing, 94, 106452. https://doi.org/10.1016/j.asoc.2020.106452
4. Dey, D., Das, S., & Saha, S. (2021). Fraud detection in banking using machine learning: A comparative analysis. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1772–1776). IEEE.
5. Jha, S., & Rani, M. (2021). Machine learning algorithms for banking fraud detection: A comparative research. Materials Today: Proceedings, 45, 2844–2849.
6. Karpoormath, R., & Hullur, S. (2022). Bank fraud detection using ML algorithms. In 2022 International Conference on Electronics and Renewable Systems (ICEARS), 730–736. IEEE.
7. Saravanan, R., & Karthik, P. R. (2022). Application of ensemble machine learning in fraud detection. International Journal of Computer Applications, 184(42), 11–15.
8. Shil, S., & Sultana, M. T. (2023). Bank transaction fraud detection using stacked ensemble learning model. International Journal of Information Technology, 15(3), 1431–1440.
9. Banerjee, S., & Singh, R. (2023). Enhancing fraud detection in financial transactions using deep learning. Procedia Computer Science, 218, 159–165.

ISSN: 3107-7900, P.CODE: 26/2-2/09
DOI: https://doi.org/10.5281/zenodo.20458050

10. Kumar, R., & Malhotra, A. (2023). Real-time banking fraud detection using XGBoost and SMOTE. Journal of King Saud University - Computer and Information Sciences.
11. Mehta, P., & Roy, S. (2024). A hybrid deep learning model for fraud detection in banking systems. Journal of Artificial Intelligence and Soft Computing Research, 14(2), 67–75.
12. Singh, V., & Dasgupta, A. (2024). Explainable AI for banking fraud detection: A SHAP-based research. Expert Systems with Applications, 245, 119001.