Indian Journal of Computer Science and Technology
Volume 5, Issue3 (September-December 2026), PP: 654-657.
www.indjcst.com
ISSN No: 2583-5300

Decision Tree-Based Fraud Detection and Financial Data Protection in Digital Banking

Shanaya Desai
Aditya Birla World Academy, Mumbai, Maharashtra, India.

To Cite this Article: *Shanaya Desai, "Decision Tree-Based Fraud Detection and Financial Data Protection in Digital Banking", Indian Journal of Computer Science and Technology, Volume 05, Issue 03 (September-December 2026), PP: 654-657.*

Copyright: ©2026 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *As banking becomes more digital to increase convenience of online transactions, the risk of financial scams increases noticeably because banks now have to store huge amounts of personal and sensitive financial data in their systems. This increases the need for effective fraud detection in order to protect user privacy. Conventional detection methods are not as effective as those that can be enforced using newer technologies. Artificial Intelligence and Machine Learning models can be helpful in stopping potential scams by identifying patterns. The aim of this research paper is to propose a decision tree based machine learning model that can help identify suspicious transactions and terminate them before any significant harm is caused, while also protecting user data. A decision tree algorithm is the most ideal for this scenario since its decision making process is the easiest to interpret compared to other algorithms. The proposed model considers data minimisation, which only provides the required data to the algorithm while maintaining complete confidentiality of sensitive information. This model has not been experimentally tested and hence the degree of its effectiveness may vary under various circumstances. This research paper wishes to evaluate how fraudulent transactions can be detected while still considering the importance of customer data privacy.*

Key Word: *Machine Learning, Artificial Intelligence, Digital Banking, Fraud Detection, Decision Tree, Cybersecurity, Financial Data Privacy.*

I.INTRODUCTION

The evolution of online banking has revolutionised how people and organisations make payments. Financial fraud can be defined as any intentional act of deception involving financial transactions for unlawful gain. (Lopez, 2025) [1] In today's time, digital banking continues to become prevalent as both individuals and organisations rely on these systems for faster, efficient and more accessible methods of payment. However, this increase in reliance towards online financial services has also raised concerns not only about user privacy, but also about the identification and prevention of fraudulent activity. Such activity leads to financial loss of customers and organisations but also reduced trust in the bank and its digital banking services. Therefore, banks need to enforce reliable methods of fraud detection. Although manual efforts to detect financial malpractice have been made, they are time consuming, expensive and impractical in this day and age where AI and machine learning algorithms are increasingly used.

Artificial Intelligence (AI) and Machine Learning (ML) provide an efficient alternative to identifying suspicious financial activity. Machine Learning models can be trained using test data to identify patterns and categorise transactions. A literature review performed by Ali et al. [2] assessed how machine learning can be used to detect financial fraud and also listed several common techniques that are used. Similarly, a review by Hernandez Aros et al. [3] also researched technical approaches to tackling such scams by highlighting various models and datasets.

While these advancements are significant, another major concern remains to be the privacy of customer data. Revealing financial records such as spending patterns and frequency to AI/ML models increases the risk of sensitive information being breached. Thus, the use of such data must be taken into consideration and restricted access should be encouraged. The effectiveness of the fraud detection system should not solely be reliant on its ability to detect fraudulent transactions but also its ability to handle sensitive data.

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. (Kavlakoglu, 2021). [4] It can classify the given data using a list of conditions provided. Therefore, this research proposes a decision tree algorithm for digital banking transactions.

II.MATERIAL AND METHODS

1. ML for financial fraud detection:

Ali et al. conducted a literature review of 93 articles and reports on ML-based financial fraud detection systems. The study examined the types of algorithms used and how its performance was measured. They concluded that machine learning can help identify patterns in transactional data which may be difficult when done manually. According to the source, Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) are the most commonly used techniques. SVMs can categorise transactions based on decision boundaries whereas ANNs are used to analyse complex data

relationships. However, when compared to a decision tree model, SVMs and ANNs are harder to interpret. While the aforementioned systems work exceptionally when high quality training data is provided, the decision tree works better under all conditions since it is based on a series of rules.

Afriyie et al. compares three supervised machine-learning algorithms for credit-card fraud detection: Logistic Regression, Random Forest and Decision Tree. They used datasets that replicated credit card transactions from 2020 to examine how effectively each algorithm could distinguish between fraudulent and legitimate transactions. Their final judgement stated that Random Forest attained a 96% accuracy, hence outperforming the others. This research compares three different algorithms using quantitative data which is useful when deciding which model is ideal, however the tests were conducted under experimental conditions hence it cannot fully represent the complexity of real banking activity.

2. Cybersecurity in Indian banking:

Chhabra Roy and Prabhakaran published a study that focuses on cyber fraud within Indian banks. The authors used focused group discussions with risk officers and cybercrime experts in order to identify the types of internal financial scams and what factors contribute to it. They also investigated a K-NN based machine learning approach to help classify and predict risks and its severity.

3. Data privacy in ai based fraud detection:

Ab et al. uses federated learning to detect potential fraudulent transactions across Japanese banks. The model entails a collaborative system where all customer data is not combined in a central location, which ensures absolute confidentiality of data. This theory is tested practically, rather than only relying on hypotheses and hence it is reliable. Contradictingly, such a system requires intensive coordination and technical infrastructure to implement. Furthermore, the model's effectiveness cannot be judged based on its performance in Japan since Indian frameworks differ.

Proposed Model

The proposed model aims to detect and prevent potentially fraudulent transactions using a decision tree machine learning algorithm. This algorithm is selected because it uses a series of rules and conditions to classify given data, which makes its decision making process easy to interpret. The model will analyse features of a particular transaction and categorise it as either legitimate or potentially fraudulent. The latter will be flagged and will undergo further investigation.

Although the motive of the model is to aid fraud detection and prevention by identifying suspicious payment activity, the decision tree algorithm itself does not inherently prevent any form of fraud, hence manual banking support would be required to respond to such activity.

The objectives of the proposed model are as follows:

1. To identify potentially fraudulent transactions
2. To incorporate basic data shadowing measures to control unnecessary exposure of sensitive financial information.

The model will be trained using a variety of labelled data values containing sample values of legitimate and fraudulent transactions such as the amount and frequency. This will help the machine learning model learn the patterns that help distinguish between the two categories. It can then follow a series of decision branches until it reaches a final classification.

While the decision tree algorithm is relatively easy to interpret, it also has limitations. Imbalance in data provided to the model might cause fluctuations in its performance since in reality, fraudulent transactions only represent a small fraction of total transactions.

The proposed system will follow the steps of data processing, classification and response, as demonstrated below.

Proposed System Workflow

- Transaction data input: obtain a suitable public or authorised dataset containing labelled examples of legitimate and fraudulent transactions. Private customer records are not required for the prototype.
- Data processing: check the dataset for missing, duplicate or inconsistent values. Select only features relevant to fraud classification and exclude unnecessary personal identifiers.
- Model training and classification: train the Decision Tree model using labelled transaction examples. The model predicts whether each transaction is legitimate or potentially fraudulent.
- Flagging and review: transactions classified as potentially fraudulent are flagged for further review. The model provides a prediction; it does not independently investigate or resolve a case.
- Evaluation: assess the model using accuracy and precision, and where possible also recall, F1-score and a confusion matrix. These measures should be calculated on a held-out test set

Privacy safeguards

- Data minimisation: provide only the transaction features needed for classification and remove unnecessary identifiers.
- Dataset selection: use public, synthetic or otherwise authorised data for prototype development instead of private customer banking records.
- Restricted access: limit access to the dataset, trained model and model outputs to authorised users.
- Human review: require further review of flagged transactions instead of treating every model prediction as a final decision.

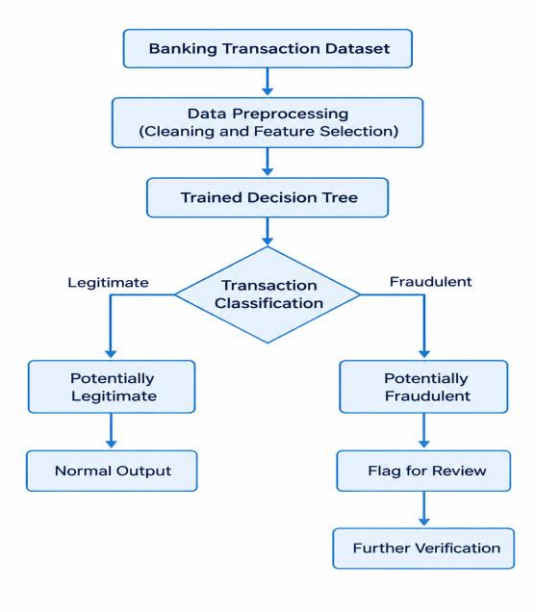


Diagram 1. Flowchart of the proposed Model

III. RESULTS

The working comparison supplied with the research report lists the following accuracy and precision values. It also includes projected privacy-protection scores to represent the relative interpretability of each model. The figures are reproduced as supplied and should be treated as indicative comparison values until they are verified through a consistent experiment.

Performance metric	Decision Tree	Random Forest	Logistic Regression
Accuracy	82%	95%	85%
Precision	80%	92%	84%
Privacy Protection	85%	40%	75%

Table 1. Model comparison values reproduced from the supplied research report.

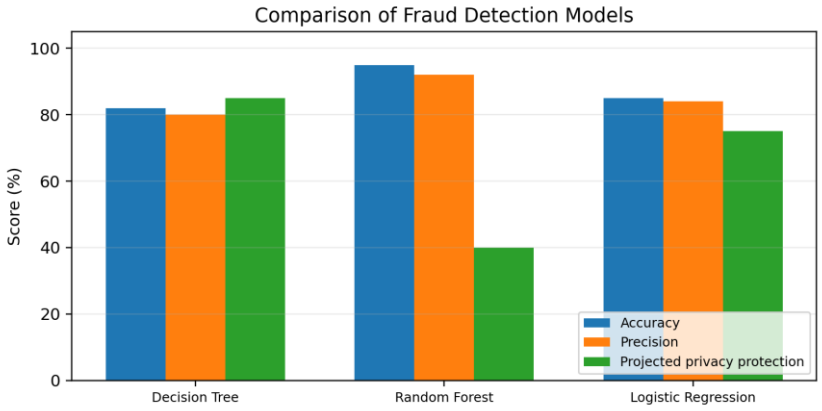


Diagram 2. Accuracy, precision and projected privacy-protection scores. The privacy values are estimates, not experimentally measured outcomes.

As presented in the graph and table above, the Random Forest model gives a higher accuracy and precision rate than the decision tree and logistic regression, however when compared with the privacy protection aspect it would not be able to provide as much security as the decision tree model. This is because a random forest combines multiple trees, making its predictions more complex to explain and interpret, whereas logistic regression is not ideal for complex data and non-linear patterns. A decision tree uses smaller datasets and hence is easier to interpret and limits customer data revelation, thus

providing the highest privacy protection metrics. Overall, the decision tree is the better choice because confidentiality of sensitive financial information is equally as important as the fraud detection system.

IV. CONCLUSION

In conclusion, the proposed Decision-Tree based model provides an efficient alternative to identifying suspicious transactions using pattern recognition. This model classifies banking transactions based on its features due to its easily interpretable decision-making process. Unfortunately, the actual performance of this model has not been tested under experimental conditions and hence any conclusion made through this research about its accuracy or ability is not reliable. Future scope will focus on implementing the proposed model and evaluating its performance on real data. Overall, the proposed work provides a starting point for studying the relationship between machine learning-based fraud detection and customer-data privacy. Its practical value will depend on the results of implementation, the reliability of its predictions and the effectiveness of the safeguards used to handle financial data.

REFERENCES

1. Ali, Abdulalem, et al. "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review." Applied Sciences, vol. 12, no. 19, 2022, article 9637. https://doi.org/10.3390/app12199637.
2. Afriyie, Jonathan Kwaku, et al. "A Supervised Machine Learning Algorithm for Detecting and Predicting Fraud in Credit Card Transactions." Decision Analytics Journal, vol. 6, 2023, article 100163. https://doi.org/10.1016/j.dajour.2023.100163.
3. Cherif, Asma, et al. "Credit Card Fraud Detection in the Era of Disruptive Technologies: A Systematic Review." Journal of King Saud University—Computer and Information Sciences, vol. 35, no. 1, 2023, pp. 145–174. https://doi.org/10.1016/j.jksuci.2022.11.008.
4. Hernandez Aros, Ludivia, et al. "Financial Fraud Detection through the Application of Machine Learning Techniques: A Literature Review." Humanities and Social Sciences Communications, vol. 11, 2024, article 1130. https://doi.org/10.1057/s41599-024-03606-0.
5. Kavlakoglu, Eda. "What Is a Decision Tree?" IBM Think, https://www.ibm.com/think/topics/decision-trees.
6. Kanamori, Sachiko, et al. "Privacy-Preserving Federated Learning for Detecting Fraudulent Financial Transactions in Japanese Banks." Journal of Information Processing, vol. 30, 2022, pp. 789–795. https://doi.org/10.2197/ipsjjip.30.789.
7. Lopez, Jude Gerald. "What Is Financial Fraud? & Its Types." Razorpay Learn, 21 July 2025, https://razorpay.com/learn/what-is-financial-fraud/.
8. Roy, Neha Chhabra, and Sreeleakha Prabhakaran. "Internal-Led Cyber Frauds in Indian Banks: An Effective Machine Learning-Based Defense System to Fraud Detection, Prioritization and Prevention." Aslib Journal of Information Management, vol. 75, no. 2, 2023, pp. 246–296. https://doi.org/10.1108/AJIM-11-2021-0339.
9. Wang, Shuang, et al. "Data Privacy and Cybersecurity Challenges in the Digital Transformation of the Banking Sector." Computers & Security, vol. 147, 2024, article 104051. https://doi.org/10.1016/j.cose.2024.104051.