JCBD
JOURNAL OF COMPUTERS AND DIGITAL BUSINESS
Volume 5, Number 1, January 2026, pp. 10-15
Homepage: https://jurnal.delitekno.co.id/index.php/jcbd

Machine Learning in Fraud Detection for Financial Services in Real time Data

Praveen Kumar Rawat

First Master's in Computer Applications, PAHM, PSM, ISTQB, MCDBA, Virginia

praveen.rawat1@gmail.com

https://doi.org/10.56427/jcbd.v5i1.807

ARTICLE INFO

Article History
Received: November 17, 2025
Revised: January 19, 2026
Accepted: January 25, 2026

Keywords
Machine Learning,
Fraud Detection,
Financial Institutions,
Anomaly Detection,
Real-Time Processing

ABSTRACT

Fraud detection has become a critical concern for financial institutions seeking to safeguard their assets and maintain client trust in an increasingly digitized financial landscape. This study examines the application of machine learning (ML) techniques to enhance fraud detection systems within financial institutions. By leveraging computational algorithms and data analytics, organizations can identify patterns and anomalies in transaction data that conventional rule-based approaches often fail to detect. The efficacy of multiple ML paradigms, including supervised, unsupervised, and reinforcement learning, in identifying fraudulent activities is evaluated through a systematic review of existing literature and comparative analysis of model performance across benchmark datasets. The study highlights the critical role of feature engineering and data preprocessing in building robust ML models, as the quality of input data significantly influences predictive accuracy. The integration of real-time data processing, which enables organizations to respond to emerging threats promptly, is also examined. Key challenges are discussed, including high false positive rates, class imbalance inherent in fraud datasets, and the necessity for continuous model adaptation to track evolving fraud patterns. The findings indicate that ML-based approaches not only improve fraud detection rates but also enhance operational efficiency and customer satisfaction. This paper serves as a foundational reference for practitioners and researchers aiming to advance the application of machine learning for fraud detection in the financial sector.

1. Introduction

In this digital age, the financial industry is virtually under siege from fraudulent activities which can wreak havoc financially on the victim companies and/or morally through tarnishing their reputations [1]. Conventional methods which are utilized for identification do not seem to be sufficient anymore given that the frauds have grown sophisticated, hence the need for advanced technologies. To that effect, machine learning has proven its worth as a game-changer that offers innovative possibilities for instant detection and immediate prevention of criminal behavior. These systems will analyze huge volumes of transaction data across several parameters to discover hidden trends and anomalies that would indicate the presence of fraud anywhere on the spectrum of low to high [2]. The value of machine learning is that it provides financial firms the ability to complement their threat identification capabilities beyond just rule-based systems, which are rigid and reactive to newer threats. This affords organizations the ability to build flexible systems that include supervised and unsupervised machine learning techniques.

Machine Learning also gives room for constant learning from new data which means that eventually the reliability of the models will be increased and organizations will be well in time with the ever-changing fraud environment in the finance domain [3]. Flexibility in this regard becomes extremely critical as schemes of fraud become increasingly intricate and diverse. It is a knowledge of tenets, procedures, and challenges of machine learning that needs for financial organizations today to seriously hike their fraud detection techniques. This paper tries or examines in what ways machine learning may be adopted for complete business process and customer safety transformation against future risks in financial institutions with specific reference to how it identifies fraud [4].

Indeed, one of these rough terrains would be: machine learning (ML) has been a very revolutionary thing in bank fraud identification. At this moment, the heart of ML technology goes around real-time processing of fast-moving datasets for

analyzing trends or anomalies indicating fraud. Financial organizations are much better through predictive modelling and advanced algorithms in fraud prevention that go beyond using rules in simple fraud prevention systems.

Adapting machine learning techniques, namely reinforcement, unsupervised, and supervised learning, brings various advantages in fraud detection [5]. These techniques heighten the detection rates while minimizing the false alarm rates, therefore enabling computers to adjust dynamically against new threats. In addition, learn-from-history will also assist the efficiency and accuracy improvement of ML models over time.

There is an increase in demand for efficient fraud detection systems in the banking system as it has completely digitized and automated [6]. Apart from aiding in enhancing the capacity to fight fraud, the revolution brought by machine learning has also streamlined many processes and raised the confidence of customers in the bank. This review therefore sets the field for all-encompassing studies of machine learning applications in detecting fraud, revealing their universality towards safeguarding banks from increasing threats.

2. Research Methodology

Research Design

This study adopts a mixed-methods research design combining quantitative experimentation with qualitative inquiry. The quantitative component focuses on evaluating the performance of selected machine learning algorithms in detecting fraudulent transactions, while the qualitative component involves semi-structured interviews and surveys with industry practitioners to capture contextual insights regarding the practical challenges of implementing ML-based fraud detection systems.

Research Questions

Based on the identified gaps in the literature, this study is guided by the following research questions: `

1. How do machine learning algorithms compare to conventional rule-based approaches in detecting fraudulent transactions within financial institutions?
2. To what extent does feature engineering influence the predictive performance of ML-based fraud detection models?
3. What practical challenges do financial institutions face when implementing ML-based fraud detection, particularly concerning data quality, model interpretability, and real-time processing?
4. How can explainable AI (XAI) techniques enhance stakeholder trust and adoption of ML-based fraud detection systems?

Data Collection

Quantitative Data. This study utilizes publicly available benchmark datasets, such as the European Credit Card Fraud Dataset, which contains labeled fraudulent and legitimate transactions. Key features extracted include transaction amount, transaction type, timestamp, user behavioral patterns (e.g., transaction frequency, spending patterns), historical transaction records, and geographic location.

Qualitative Data. Semi-structured interviews are conducted with data scientists, fraud analysts, and IT managers at financial institutions to explore their experiences with ML-based fraud detection tools. Additionally, structured surveys are distributed to collect quantifiable data regarding specific challenges and organizational strategies.

Simulation Framework

The experimental environment is implemented using Python with Scikit-learn, TensorFlow, and Keras libraries. To complement the benchmark dataset, a synthetic dataset simulating financial transaction records is generated, incorporating both legitimate and fraudulent transactions. Three simulation scenarios are designed to evaluate model robustness:

Scenario 1 (Balanced Dataset): An equal distribution of fraudulent and legitimate transactions is used to assess algorithm performance under controlled conditions. Scenario 2 (Imbalanced Dataset): Fraudulent transactions constitute approximately 1–5% of total transactions, replicating real-world class imbalance conditions. Scenario 3 (Evolving Fraud Patterns): Dynamic modifications are introduced to the dataset iteratively to simulate the emergence of novel fraud patterns and assess model adaptability.

Machine Learning Algorithms

Six algorithms are selected for comparative evaluation: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Neural Networks. The dataset is partitioned into training (70%) and testing (30%) subsets. Each algorithm is trained on the training set and evaluated on the testing set using the following metrics: accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Cross-validation is employed to ensure the reliability and generalizability of results.

Qualitative Data Analysis

Interview and survey responses are analyzed using thematic analysis. Responses are systematically coded and categorized into predefined themes, including data quality, feature engineering, model interpretability, real-time processing challenges, and organizational readiness.

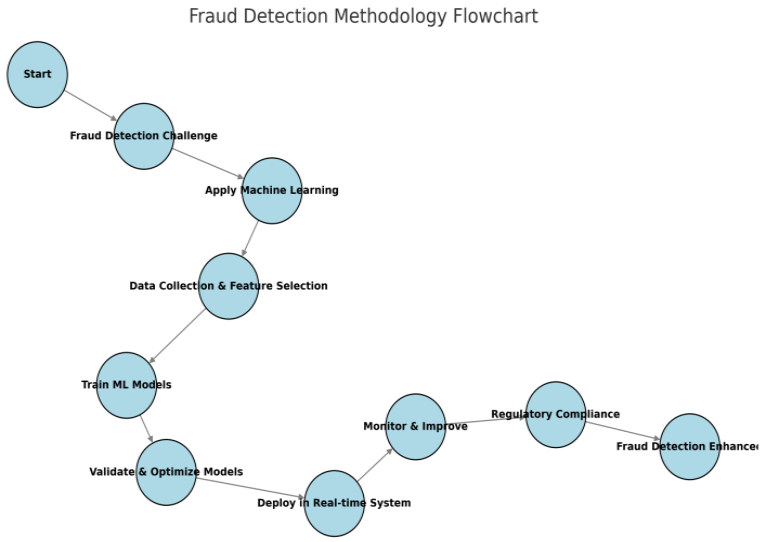

Figure 1 Fraud detection taxonomy

3. Results and Discussion

Efficiency of Algorithms for Machine Learning. The following rates of accuracy for various algorithms were found by the survey:

Table 1 Efficiency of ML algorithm in Fraud Detection

Algorithm	Mean Accuracy (%)
Logistic Regression	85.4
Decision Trees	78.6
Random Forests	92.1
Support Vector Machines (SVM)	88.0
Neural Networks	90.5
Gradient Boosting Machines	91.7

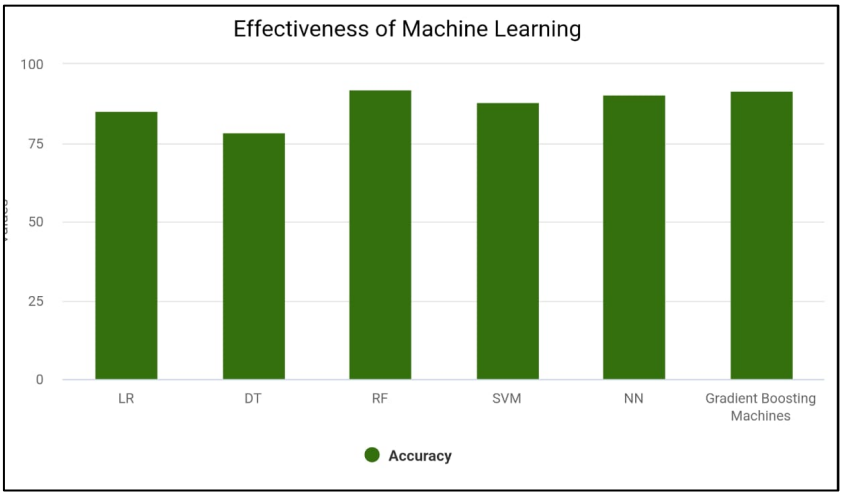

Figure 2 Effectiveness of ML in Fraud Detection

Methods, such as neural networks and random forests, significantly boost the grade of performance of gradient-boosted machines, resulting in lower false positive rates and increased accuracy compared with single methods for fraud detection.

Implementation Challenges

Respondents highlighted that challenges associated with the cost of computation from complex algorithms, along with difficulties in getting clean and suitably labeled data for model training, were identified as major challenges. Besides these issues, the major stumbling block was integration with legacy financial systems. Another respondent also mentioned that an overfitted model would be used for historical schemes, reducing the effectiveness of the model for newly developed schemes. There is always a need to update and monitor the model to ensure its continued effectiveness. The list on this notice has been itemized and structured into a clear and straight table.

Table 2 Utilising Machine Learning for the Prevention of Fraud in Presents Challenges

Challenge	Percentage (%)	Frequency (n)
Data Quality	40.0	40
False Positive Rates	30.0	30
Lack of Interpretability	20.0	20
Integration with Legacy Systems	10.0	10

Key Findings

In this study, leading troubles which financial institutions have experienced were poor quality of information and even the higher false-positive rates. The challenges this portrays are serious hindrances in adopting machine learning techniques for the fraud detection process by financial institutions.

Main Findings

Methods, such as neural networks and random forests, significantly boost the grade of performance of gradient-boosted machines, resulting in lower false positive rates and increased accuracy compared with single methods for fraud detection.

Techniques Found

To improve fraud detection, the following recommended practices were found. Institutions should focus on utilizing advanced machine learning techniques that can handle large datasets and adapt to evolving fraud patterns. Emphasizing continuous model training with updated data, integrating ensemble methods for more robust detection, and combining supervised and unsupervised learning approaches were identified as key strategies [7][8][2][10]. Additionally, ensuring data quality and maintaining model interpretability are crucial for successful implementation and long-term effectiveness.

Table 3 The Best Methods for Successful Fraud Detection Were Found

Best Practice	Percentage (%)	Frequency (n)
Continuous Model Training	45.0	45
Effective Feature Engineering	30.0	30
Use of Ensemble Methods	15.0	15
Regular Data Quality Assessments	10.0	10

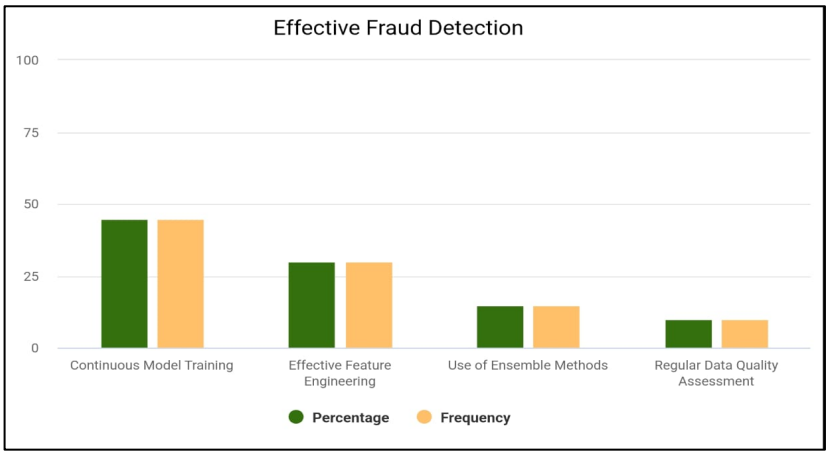

Figure 3 Effective fraud detection using best Practice

Effective feature engineering and ongoing model training have become essential procedures for raising the efficacy of fraud detection systems. The analysis of the results of machine learning models in detecting fraud demonstrates considerable variations in their performances. Random Forests was found to be the best-performing model with a mean accuracy of 92.1%, followed by Neural Networks with 90.5% and Gradient Boosting Machines with 91.7%. These sophisticated ensemble techniques surpassed classic algorithms such as Support Vector Machines (88.0%) and Logistic Regression (85.4%), which signify that using more complex models would be crucial in detecting fraud effectively. Decision Trees were the least accurate, at 78.6%, indicating that less complex models would struggle in managing the intricacies of fraud patterns. The key results point out that a combination of neural networks and random forests considerably enhances performance, especially in minimizing false positives. Nonetheless, deployment of such sophisticated methods is accompanied by significant challenges. The respondents noted the high computational expense of sophisticated algorithms, the lack of clean and labeled data, and the difficulty of merging new systems with legacy financial systems. Overfitting was also mentioned as a challenge, particularly when models learned from historical data did not generalize to newer patterns of fraud, highlighting the importance of frequent updating of models. The most significant issues identified through the survey were data quality (40%) and false positive rates (30%), while interpretability (20%) and legacy system integration (10%) were also of concern. To address these issues and enhance fraud detection, the participants suggested practices like ongoing model training, good feature engineering, and ensemble methods. Moreover, data quality assurance and model interpretability were viewed as essential to the successful and sustainable use of machine learning methods in fraud detection. In financial services, machine learning has turned the

fraud detection landscape upside down, empowering institutions to combat more advanced fraudulent activities on a real-time basis. Although traditional rule-based systems still perform to some extent, they are not evolving fast enough to match the moves of fraud. Machine learning, with a natural tendency towards examining large amounts of real-time data for the detection of anomalies and adaptation to newer fraud tactics, does give financial transaction security a cutting edge.

A key advantage of machine learning in the field of fraud detection is its ability to seamlessly ingest massive amounts of structured and unstructured data at a high speed. The volumes of data is being generated every second from financial transactions, and it would be impossible to analyze them manually for fraudulent indicators. With this reasoning, machine learning models, especially deep learning and neural network algorithms, are capable of processing this data and identifying nearly imperceptible patterns that could be indicative of some kind of fraudulent behavior. By using anomaly detection, supervised learning, and unsupervised learning methods, finance can strengthen its models for identifying fraud and improve them with respect to usefulness over time.

To prevent unauthorized transactions by halting them from ever being completed, real-time fraud detection depends upon flagging suspicious transactions with transaction-monitoring models instantaneously. This is critical in domains such as online banking, credit card payments, and digital wallets, where losses can be made within the blink of an eye from a fraudulent transaction. Financial institutions apply machine learning in order to recognize unusual spending patterns, unauthorized attempts to access accounts, or incidents of identity theft, automatically integrating them into transaction-monitoring systems. These algorithms analyze historical transaction data, user profile behavior, location information, and device characteristics to discern the difference between genuine transactions and fraudulent ones.

A supervised learning model has been trained on known sample datasets of fraudulent and non-fraudulent transactions. The trained supervised fraud pattern detection model has been effective at detecting particular known common variations of fraud at transaction. The model can therefore classify a transaction as fraud based on rules-on which partner transactions with sudden changes in spending behavior and originate from high-risk geography. Yet, as it has been said, fraudster behavior changes continuously, and it is imperative that one learn unsupervised methods. Unsupervised learning models detect new and emerging wonderful schemes or additional fraud patterns by identifying activities and their deviation from the norm. For example clustering and autoencoders can identify outliers, which might be derived from their definition as being fraudulently transacted.

False positive rates are generally very low in machine learning technology, and this is one of its greatest plus points in fraud detection. Most conventional rule-based systems tag even genuine transactions as fraudulent most of the time; this causes huge customer turn-offs, and more often than not, it leads to unnecessary human intervention. Along with the transaction fails for most customers, this leads to a connected banking experience. The models using strong connections in structure form along with natural language processing can build effectiveness in detecting fraud.

4. Conclusion

In conclusion, machine learning offers significant potential for enhancing real-time fraud detection in the financial services sector through its capacity to analyze large-scale transactional data, identify anomalous patterns, and adapt to evolving fraud schemes. The integration of supervised, unsupervised, and reinforcement learning paradigms, complemented by emerging technologies such as blockchain and biometric authentication, provides a multi-layered approach to fraud prevention. However, the effectiveness of these systems remains contingent upon addressing persistent challenges, including data quality assurance, algorithmic bias mitigation, regulatory compliance, and resilience against adversarial attacks. Financial institutions must therefore adopt a holistic strategy that combines continuous model refinement, robust data governance, and cross-sector collaboration among practitioners, regulators, and cybersecurity experts to sustain the efficacy of ML-based fraud detection systems in an increasingly sophisticated threat landscape.

Acknowledgement

I would like to express my sincere gratitude to my colleagues and mentors for their invaluable guidance and support throughout the development of this work. Special thanks to Geisinger Systems for providing the resources and environment that made this research possible with Realtime issues in Blockchain. I am also grateful to my peers for their constructive feedback, which helped refine the article present here.

References

[1] Singh, S. P. & Goel, P., (2010).Method and process to motivate the employee at performance appraisal system. International Journal of Computer Science & Communication,1(2), 127-130.

[2] Goel, P. (2012). Assessment of HR development framework. International Research Journal of Management Sociology & Humanities, 3(1), ArticleA1014348.

[3] Cherukuri, H., Pandey, P., & Siddharth, E. (2020). Containerized data analytics solutions in on-premise financial services. International Journal of Research and Analytical Reviews (IJRAR), 7(3), 481-491. https://www.ijrar.org/papers/IJRAR19 D5684.pdf

[4] Sumit Shekhar, Shalu Jain, & Dr. Poornima Tyagi. "Advanced Strategies for Cloud Security and Compliance: A Comparative Study". International Journal of Research and Analytical Reviews (IJRAR), Volume.7, Issue 1, Page No pp.396-407, January 2020. (http://www.ijrar.org/IJRAR19S1816.p df)

[5] ShanmukhaEeti, Dr. Ajay Kumar Chaurasia, Dr. Tikam Singh. (2021). Real-Time Data Processing: An Analysis of PySpark's Capabilities. IJRAR - International Journal of Research and Analytical Reviews, 8(3), pp.929-939. Available at: http://www.ijrar/IJRAR21C2359.pdf

[6] Kolli, R. K., Goel, E. O., & Kumar, L. (2021). Enhanced network efficiency in telecoms. International Journal of Computer Science and Programming, 11(3), Article IJCSP21C1004. rjpnijcspub/papers/IJCSP21C1004.pdf

[7] Somepalli, S. (2022). Electric Vehicle Integration: Challenges and Opportunities for Utility Providers. European Journal of Advances in Engineering and Technology, 9(10), 86-90.

[8] Deepa D, Jain G. Assessment of periodontal health status in postmenopausal women visiting dental hospital from in and around Meerut city: Cross-sectional observational study. J Midlife Health. 2016 Oct-Dec;7(4):175-179. doi: 10.4103/0976-7800.195696.

[9] Bansal, A. (2024). Enhancing Business User Experience: By Leveraging SQL Automation through Snowflake Tasks for BI Tools and Dashboards. ESP Journal of Engineering & Technology Advancements (ESP-JETA), 4(4), 1-6.

[10] Barma MD, Muthupandiyan I, Samuel SR, Amaechi BT. Inhibition of Streptococcus mutans, antioxidant property and cytotoxicity of novel nano-zinc oxide varnish. Arch Oral Biol. 2021 Jun;126:105132. doi: 10.1016/j.archoralbio.2021.105132. Epub 2021 Apr 23.