European Journal of Business, Economics & Management

Vol. 1 No. 2 2025

Article **Open Access**

Application of Real Time Machine Learning Models in Financial Fraud Identification

Xuanrui Zhang [1,*]

[1] College of Engineering, University of California, Berkeley, Berkeley, CA, 94720, United States
* Correspondence: Xuanrui Zhang, College of Engineering, University of California, Berkeley, Berkeley, CA, 94720, United States

Received: 12 May 2025
Revised: 17 May 2025
Accepted: 29 May 2025
Published: 05 June 2025

Abstract: With the continuous advancement of digital and intelligent technologies, financial fraud methods are becoming more diversified, making it difficult for traditional rule engines and manual auditing methods to keep up with the speed of change and achieve accurate and real-time detection standards. At this point, efficient and self-learning real-time machine learning algorithms have become important tools for identifying financial fraud. This algorithm can conduct in-depth analysis of massive amounts of data generated by financial transactions, detect suspicious fraudulent activities in real time, and respond quickly by issuing alerts. This article focuses on the characteristics of machine learning models and deeply analyzes their application in financial fraud recognition, covering multiple aspects such as data preprocessing, feature engineering, model selection and optimization. The actual test results show that the experimental results validate that the real-time machine learning model exhibits excellent performance in accuracy, processing speed, and adaptability, bringing more advanced and intelligent anti fraud technology to the financial industry.

Keywords: real-time machine learning; financial fraud; fraud identification; data stream processing; model application

1. Introduction

With the advancement of financial technology, financial fraud has become increasingly diverse and complex, which undoubtedly poses great threats and challenges to the financial industry. Traditional fraud prevention methods mainly rely on manually set rules and fixed data processing methods, making it difficult to respond in real-time to rapidly changing fraud patterns. Real-time machine learning models, relying on their cutting-edge computational methods and big data analysis techniques, can more quickly and accurately identify and prevent financial fraud. This article aims to analyze the application and advantages of real-time machine learning models in financial fraud identification, in order to provide reference for technological innovation in the industry.

2. Basic Theories Related to Financial Fraud Identification

2.1. Overview of Financial Fraud Identification

In the financial industry, the identification of fraudulent behavior is achieved through in-depth research on transaction data, user behavior characteristics, and related contextual information, aiming to effectively prevent the occurrence of fraudulent behavior. Traditional fraud detection methods mainly rely on manually set rules and historical data, which are inadequate in the face of changing and complex fraud methods. With the

advancement of big data and artificial intelligence technology, the identification methods of financial fraud have begun to integrate advanced technologies such as machine learning and deep learning. These technologies can effectively detect fraud patterns by analyzing numerous historical transaction cases and provide real-time warnings and identification when transactions occur. Machine learning algorithms can continuously update recognition strategies as the environment evolves to address emerging fraudulent methods, and therefore have been widely applied in the current financial field [1].

2.2. Characteristics of Financial Fraud Identification

In financial fraud identification, key characteristics include large amounts of data, variable behavior patterns, and high demand for response speed. Due to the extremely large and continuously growing data generated by financial transactions, detection systems must be able to effectively process large amounts of data. The diversity of fraudulent methods, with the advancement of technology, criminals constantly change their methods, posing new challenges to traditional rule-based detection techniques. Financial fraud is often covert and has a time delay, making rapid response the core of fraud identification. In order to keep up with the updates of fraudulent methods, the identification system needs to be continuously upgraded and improved. It can learn by analyzing historical data and detect abnormal activities in real transactions in real time.

3. The Current Status of Financial Fraud Identification

3.1. Frequent Occurrence of Multi-Channel Fraud

The frequent occurrence of financial fraud through multiple channels affects many industries, from banking services to online payments, e-commerce transactions, and cross-border transactions. These fraudulent methods are becoming increasingly complex, posing challenges for financial institutions to identify all risks in a timely manner. The explosion and diversification of data sources from multi-channel operations, as well as the requirement for high timeliness, pose a huge challenge to the computing power of the current system. Meanwhile, the noise and outliers in these cross platform data weaken the identification accuracy of the model and make retraining more complex and time-consuming. The frequent cross-platform fraud also amplifies the conflict between model accuracy and false alarm rate. The rapid updating of data makes it difficult for the model to fully adapt, resulting in prominent false alarm problems on some platforms, while other platforms face the risk of missed reports due to insufficient data. To quantify the distribution of multi-channel fraud, the following formula is used:

$$F = \sum_{i=1}^{n} fraud(C_i) \tag{1}$$

Among them, F is the total number of fraud events, $fraud(C_i)$ is the number of fraud events in the i-th channel, and n represents the total number of channels. This formula effectively summarizes the distribution characteristics of multi-channel fraud events, providing basic support for subsequent data integration and analysis [2].

3.2. Prominent Issues of False Positive Rate and False Negative Rate

In financial fraud identification systems, false positive rate (FPR) and false negative rate (FNR) are key parameters for measuring the effectiveness of the mechanism. The false positive rate reflects the frequency at which the system incorrectly identifies legitimate transactions as fraudulent, while the false negative rate describes the frequency at which the system fails to detect actual fraudulent behavior. In practical applications, many financial institutions face the challenge of high false positive and false negative rates, especially when handling complex transaction patterns and catering to diverse user needs. A higher false alarm rate can lead to customer complaints and may also result in unnecessary resource consumption, such as excessive auditing and manual intervention [3]. A higher rate of underreporting may result in fraudulent activities not being dealt with in a

timely manner, causing significant economic losses to financial institutions. To measure the performance of financial fraud identification systems, the false positive rate and false negative rate can be calculated using the following formula:

$$FPR = \frac{FP}{FP+TN}, FNR = \frac{FN}{FN+TP} \quad (2)$$

Among them, FP represents the number of false positives, TN represents the number of correctly identified non fraudulent behaviors, FN represents the number of missed detections, and TP represents the number of correctly identified fraudulent behaviors. This formula provides a quantitative analysis of the two main challenges encountered by the system in specific applications, thereby indicating an adjustment path for improving model performance.

3.3. Uneven Data Quality

The problem of uneven data quality is particularly prominent in the identification of financial fraud. The data from different channels exhibit significant differences in completeness, accuracy, and timeliness, especially in the presence of a large amount of interference and duplicate information in numerous unstructured data streams, which undoubtedly exacerbates the demand for computing power in system processing. The uneven distribution and unstable quality of data make it difficult for the model to fit the actual application scenario during training, which in turn affects the detection efficiency. The negative impact of low-quality data on model accuracy is significant, as it not only amplifies the conflict between false positives and false negatives, but also challenges the stability of the model's performance when dealing with different datasets. The frequent changes in data further increase the difficulty of model adjustment and place higher demands on the system's data processing and integration capabilities.

3.4. High System Construction and Maintenance Costs

The construction and maintenance costs of financial fraud identification systems continue to rise, posing a major challenge. Faced with the complexity and variability of fraud methods, the system needs to integrate multiple real-time monitoring and analysis techniques. This places extremely high demands on the system's hardware and computing capabilities. The model development and deployment cycle is relatively long, and during the retraining stage, it is necessary to process massive real-time data and continuously adjust algorithms to cope with evolving fraud methods. The pursuit of efficient performance by the system has led to a continuous increase in operating costs, such as the replacement of data storage and processing equipment, as well as the sustained demand for computing resources for algorithm improvement. In the process of fraud identification, the system must also balance the control of accuracy and false alarm rate, making the conflict between resource investment and performance improvement increasingly apparent, becoming a long-term challenge that financial institutions must address [4].

4. Application Path of Three Real Time Machine Learning Models in Financial Fraud Identification

4.1. Data Source Integration and Real-Time Stream Processing

The two key factors in ensuring the effective application of real-time machine learning models in financial fraud identification are data source integration and real-time stream processing. Financial institutions need to collect information from various sources, including transaction records, payment system operation logs, user geographic location information, and user behavior data. These pieces of information often present diverse formats, strong timeliness, and uneven quality. By utilizing real-time data processing technology, it is possible to continuously integrate, clean, and standardize this information, providing a solid foundation for the efficient operation of real-time analysis models.

Taking the anti-fraud system of a large bank as an example, the system achieves deep integration of massive amounts of information from different channels through real-time data stream processing technology, including transaction records from terminals, past account information, and user network behavior data. In the actual application process of the system, this real-time data processing mechanism can continuously aggregate and analyze data streams from multiple sources, filter out interference factors, fill in missing data, and transmit the processed high-quality data to intelligent algorithm models for in-depth analysis. After obtaining these optimized data, the model can quickly assess the risk value of each transaction and detect potential fraudulent behavior in a timely manner. The calculation of the comprehensive risk score in the system can be represented by the following formula:

$$R = \frac{\sum_{i=1}^{n} F_i \cdot W_i}{\sum_{i=1}^{n} W_i} \tag{3}$$

Among them, R is the comprehensive fraud risk score, F_i is the risk score of the i-th channel, W_i is the weight of that channel, and n is the number of channels. This formula provides scientific decision support for data source integration and real-time stream processing by quantifying the risk contribution of multi-channel data, significantly improving the application efficiency and accuracy of real-time machine learning models in complex financial fraud scenarios.

4.2. Adaptive Learning Mechanism

In the process of identifying financial fraud, the adaptive learning mechanism of real-time machine learning models is a key strategy to address the complexity and dynamic changes of fraudulent behavior. Faced with the constantly changing methods of fraud, traditional static models often fail to effectively grasp the characteristics of new types of fraud. The adaptive learning mechanism can adapt to changes in fraud scenarios in real time by changing model parameters and classification thresholds, ensuring recognition efficiency while reducing the possibility of misjudgments and missed detections. This technology relies on a real-time data feedback loop, identifying newly emerging fraud patterns and flexibly updating the decision boundaries of the model, thereby ensuring the continuous improvement of recognition performance [5].

Taking a certain payment platform as an example, the platform adopts advanced real-time machine learning models to achieve self-optimized learning when processing large-scale online transactions. In practical applications, when the system detects an increase in the frequency of a new type of fraud pattern, it will automatically adjust the discrimination criteria of the model, thereby enhancing awareness of potential risky transactions. To avoid the impact of false positives on normal transactions, the system will conduct in-depth analysis of real-time feedback information, optimize classification boundaries, and reduce misclassifications of legitimate transactions. By adopting this flexible adjustment strategy, the platform is able to quickly adapt to updates in fraudulent methods and improve overall accuracy while reducing false negatives, greatly enhancing the system's response speed and stability performance. This dynamic adjustment process can be represented by the following formula:

$$T = T_0 + \gamma \cdot \frac{FP - FN}{FP + FN} \tag{4}$$

Among them, T is the dynamically adjusted threshold, T_0 is the initial threshold, γ is the adjustment coefficient, FP and FN represent the number of false positives and false negatives, respectively. This formula balances false positives and false negatives, dynamically adjusts classification thresholds, and enhances the adaptability of real-time machine learning models in dealing with complex fraudulent behavior, providing theoretical support and operational methods for the practical application of adaptive learning mechanisms in financial fraud recognition.

4.3. Real Time Fraud Scoring and Risk Assessment

When executing real-time fraud scoring and risk assessment processes, the system gathers transaction information streams from different sources, involving account details, transaction amounts, time stamps, and geographic locations, among many other factors. These pieces of information will undergo screening by preliminary processing units to remove redundant, inaccurate, and interfering data, while performing normalization operations on data features to maintain data consistency standards. The preliminary processed data is then sent to the feature extraction stage to construct a feature set related to fraud risk, including factors such as transaction frequency, device category, and past transaction behavior. These data sets are then input into machine learning algorithms for fraud scoring. The model assigns risk values based on past fraud cases and real-time data features, and divides the scores into risk levels by continuously adjusting threshold values, dividing transactions into three categories: high, medium, and low risk. For transactions identified as high-risk, the system will immediately raise an alert and proceed to the manual review process. Medium risk transactions need to go through a reanalysis module to further refine model predictions. Low risk transactions can be automatically processed. The risk rating results will be stored in real-time in the database for subsequent data analysis and model optimization. The model performance monitoring module will continuously evaluate the accuracy of the scoring results and optimize the model parameters based on the latest data to adapt to constantly changing fraud patterns. Figure 1 shows the key steps of real-time fraud scoring and risk assessment, which provides an intuitive understanding of how the system quickly processes transaction data and dynamically evaluates risks, thereby improving the efficiency and accuracy of fraud detection.

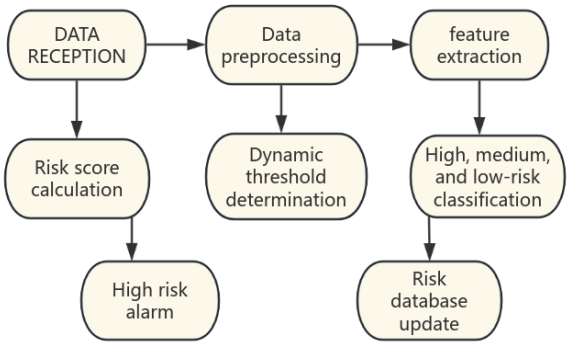

Figure 1. Schematic Diagram of Real-Time Fraud Scoring and Risk Assessment Process.

4.4. Model Performance Monitoring

In the field of financial fraud identification, continuous monitoring of model performance is crucial to ensure the stable and efficient operation of real-time machine learning models. Due to the processing of massive real-time information in fraud detection, high standards for accuracy and response speed are required, making continuous tracking of key performance indicators of the model particularly crucial. Performance monitoring covers multiple aspects such as data transmission speed, response time, accuracy, false positive rate (FPR), false negative rate (FNR), and performance comparison among different model versions. These parameters can reveal the stability and credibility of the model in practical applications, providing reference data for model optimization.

Taking a cross-border payment platform as an example, the real-time antifraud system relies on an efficient monitoring platform to conduct in-depth monitoring of algorithm execution status. The system continuously captures the input frequency of transaction data to ensure smooth operation even in high concurrency situations. If the system detects that the prediction processing time exceeds the preset threshold or the false alarm rate increases, it will automatically activate the alarm mechanism and conduct in-depth investigations, including the model's reduced adaptability to data changes or abnormal

changes in data flow. The payment platform regularly compares and analyzes the performance of different model versions, using historical datasets to evaluate each version, ensuring that each update enhances the accuracy and efficiency of the model. By continuously monitoring the distribution characteristics of predicted results, the platform can identify potential data quality issues or channel failures, thereby ensuring the long-term stable operation of the anti fraud system. The core indicators and standards for model performance monitoring in Table 1 are as follows.

Table 1. Core Indicators and Standards for Model Performance Monitoring.

Indicator Name	Describe	Monitoring Frequency	Threshold Standard
Data input rate	The speed at which the system receives data (transactions per second)	real time	≥1000 transactions /second
Prediction delay	The time difference between the model receiving data and outputting results	real time	≤100 milliseconds
accuracy	Proportion of correctly predicted results by the model	Every hour	≥95%
False alarm rate (FPR)	The proportion of non fraudulent transactions misjudged as fraudulent	Every day	≤3%
Missed Report Rate (FNR)	Proportion of unidentified fraudulent transactions	Every day	≤2%
response time	The time from the system receiving the alarm to triggering the action	real time	≤1 second
Comparison of Model Versions and Differences	Performance differences of different versions of models on the same dataset	Each iteration	≥0.5% increase

This table comprehensively displays the key dimensions of model performance monitoring, helping the system to continuously track and optimize model performance, thereby improving the efficiency and accuracy of financial fraud identification.

5. Conclusion

In the financial field, dynamically adjusted machine learning models are widely used to identify fraudulent behavior. Through strategies such as data fusion, adaptive learning mechanisms, real-time scoring, and performance monitoring, they provide efficient technical support for dealing with complex fraud strategies. These advanced technologies greatly improve the accuracy of fraud identification and effectively reduce the occurrence of false alarms and missed alarms, while meeting the business's demand for instant feedback. With the rapid advancement of financial technology, fraud methods are also constantly evolving. Therefore, real-time machine learning models need to integrate more cutting-edge algorithms and big data technologies to continuously improve the efficiency of the models and ensure the high adaptability and stability of the system. In the future, integrating emerging technologies such as blockchain and privacy-preserving computing, real-time fraud detection systems are expected to promote the intelligence and reliability of the financial industry while ensuring security and privacy.

References

1. B. Borketey, "Real-time fraud detection using machine learning," *J. Data Anal. Inf. Process.*, vol. 12, pp. 189–209, 2024, doi: 10.4236/jdaip.2024.122011.
2. A. Immadisetty, "Real-time fraud detection using streaming data in financial transactions," *J. Recent Trends Comput. Sci. Eng. (JRTCSE)*, vol. 13, no. 1, pp. 66–76, 2025, doi: 10.70589/JRTCSE.2025.13.1.9.
3. H. O. Bello, A. B. Ige, and M. N. Ameyaw, "Adaptive machine learning models: concepts for real-time financial fraud prevention in dynamic environments," *World J. Adv. Eng. Technol. Sci.*, vol. 12, no. 2, pp. 21–34, 2024, doi: 10.30574/wjaets.2024.12.2.0266.

4. G. Manoharan, A. Dharmaraj, S. C. Sheela, K. Naidu, M. Chavva, J. K. Chaudhary et al., "Machine learning-based real-time fraud detection in financial transactions," in *Proc. 2024 Int. Conf. Adv. Comput., Commun. Appl. Informatics (ACCAI)*, May 2024, pp. 1–6, doi: 10.1109/ACCAI61061.2024.10602350.
5. M. Malempati, "A data-driven framework for real-time fraud detection in financial transactions using machine learning and big data analytics," *SSRN*, 2023, doi: 10.2139/ssrn.5230220.