

Financial Risk Modeling and Fraud Detection Using Machine Learning: A Method to Reduce Business Risk

Xin Jin

*Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
2698581073@qq.com*

Keywords: Machine Learning, Financial Risk Modeling, Fraud Detection, Credit Risk Assessment

Abstract: This paper discusses the application of machine learning technology in financial risk modeling and fraud detection, aiming at reducing commercial financial risks. Based on the Internet, big data and supercomputing technology, machine learning has significantly enhanced the ability of financial institutions to extract and interpret information through automatic data analysis, and thus greatly improved their risk control efficiency. The research covers the nonlinear association identification of supervised learning in credit risk assessment, the new fraud detection mode of unsupervised learning in fraud detection, and the efficient performance of deep learning in abnormal transaction identification. Empirical analysis shows that machine learning can significantly improve the accuracy of credit risk scoring model, accelerate the detection speed of fraudulent transactions, and achieve remarkable results in the verification of anti-money laundering in trade financing, effectively shortening the processing time and improving business efficiency. This study not only provides a theoretical basis for financial institutions to build an efficient internal risk management system, but also provides practical guidance for regulatory authorities to improve the external risk control environment, which is of great significance to improving the risk prevention and control ability of the financial system.

1. Introduction

With the gradual transformation from digital to intelligent in economic and social fields, traditional risk management methods have been difficult to meet the needs of financial innovation and development. In recent years, there have been frequent problems in P2P network loan platform, such as the outbreak of problems such as fund pool, self-financing and related guarantee, which indicates that the operational risk of online loan format is in a blowout trend, further confirming the urgency of coordinated development of financial risk management. Machine learning technologies such as Internet, big data and supercomputing, as the core technologies of artificial intelligence, are gradually becoming indispensable technologies in financial risk management [1]. In 2017, the document "Development Plan of New Generation Artificial Intelligence" issued by the State Council proposed to establish a financial big data system, enhance the ability of financial multimedia data processing and interpretation, and encourage the financial industry to use intelligent monitoring technology to build an intelligent alarm and monitoring mechanism for

financial risks.

Machine learning technology shows significant advantages in credit rating, fraud identification and abnormal transaction monitoring: through a large number of automatic data analysis, the ability of financial institutions to extract and interpret information can be greatly enhanced, and the time required for human processing information can be reduced, which will greatly promote the efficiency of risk control of financial institutions. Studying the application of machine learning in financial risk management will not only help financial institutions to establish a more efficient internal risk management system, but also help regulators to create a more perfect external risk control environment, which has important theoretical value and practical significance for improving the risk prevention and control ability of the entire financial system.

2. The theoretical basis of machine learning in financial risk management.

2.1 Basic concepts and framework of machine learning

As the core technology of artificial intelligence and an important infrastructure in the era of big data, machine learning has been continuously optimized through Qian Qian's ten thousand iterations, and the prediction accuracy in external data has been continuously improved (see Figure 1)[2], so that it can extract its features from a large number of sample data without coding, and then discover the potential relationship between samples and generate corresponding models. Machine learning methods are generally divided into two types: one is marked learning method, that is, samples are marked by existing sample data, so that specific inputs can be predicted; The other is the unlabeled data learning method. In this case, the samples that are not used as sample data are not labeled with any information at all, so the data pattern can be deduced by using unknown factors without estimating the variable values. The framework of machine learning includes four basic processes: data acquisition, feature extraction, model creation and reasoning. The process of model creation in machine learning is the process of constantly changing its parameters to find the best prediction effect, and finally a reliable prediction model will be obtained. This automatic learning and optimization process makes machine learning especially suitable for dealing with complex risk identification and early warning tasks in the financial field.

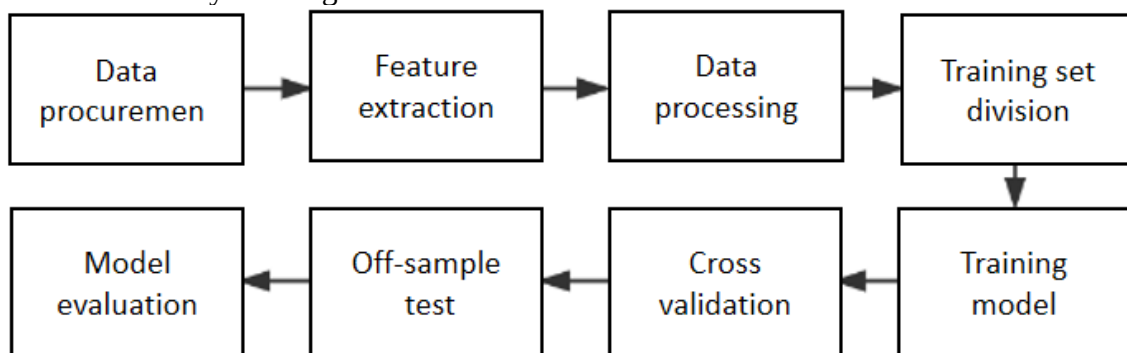


Figure 1 Basic framework of machine learning

2.2 The connotation and characteristics of financial risk management

Financial risk control is an effective way of financial risk management. From the perspective of ensuring the smooth development of finance, all kinds of risks existing in the financial field can be avoided or solved [3]. From the perspective of characteristics, first, financial risks are complex and systematic, and there is a direct correlation between financial risks. For example, credit risk will

induce difficulties in capital flow and turn into market risk. Therefore, in the face of the current situation, we should implement the whole process of risk management; Second, the changing speed of the financial market and the development of new technologies not only bring new risk characteristics, but also change the traditional risk forms, which requires adjusting the coping strategies with the times; Third, science and technology are developed, and the use of high-tech tools is becoming more and more critical. For example, based on big data technology, risk identification, risk assessment and risk early warning can be realized.

2.3 Application advantages of machine learning in the financial field

Machine learning technology is fully applied in the application of financial risk management, which has the following five advantages:

(1) The comprehensive data processing ability is strong, and machine learning can comprehensively apply structured and unstructured financial data to comprehensively evaluate possible risks;

(2) Self-adjusting learning function, machine learning can automatically adjust when obtaining new data, optimize the model and improve the accuracy of prediction;

(3) Extensible, which can expand new data sources and risk factors, and with the changes of financial markets and the emergence of new risks, the system can quickly respond and improve actions, making risk control more effective;

(4) Strong anti-noise or error input ability, even if there is more noise and error input, the machine still ensures the system performance.

Finally, from the perspective of economic benefits, the application mode of artificial intelligence has saved a lot of labor costs. For example, the anti-money laundering inspection in China Bank requires two hours of manual inspection, while the use of machine learning only takes two minutes, which can further improve the accuracy of risk judgment, so that the work efficiency and accuracy are improved together. This advantage makes machine learning a necessary technical function in the current financial risk control construction [4].

3. Key technologies of machine learning financial risk modeling

3.1 The principle of supervised learning and its application mechanism in credit risk

Supervised learning is a core aspect of machine learning and plays a key role in evaluating credit risk. The general idea is to learn a mapping relationship between mapping features and risks by using historical data with labels [5]. The supervised learning of credit risk should first model the predictive value of measurement continuity variables, and select the linear regression basic model in the modeling process:

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon$$

The parameter β is usually estimated by least square method. When there is multiple correlation between independent variables, more penalty factors can be introduced by means of more complex ridge regression and Lasso regression to increase the model's ability to interpret information. In practical application, supervised learning method can deal with quantity and quality information at one time, such as consumer's financial, trading and social data. Through the standardization process, the model can learn the causes of overdue repayment and its various combinations [6]. For example, in credit rating, the model not only needs to consider traditional factors such as income and financial situation, but also needs to mine non-traditional factors such as users' consumption

behavior and social behavior (such as Weibo, WeChat, QQ, etc.), and then output them as non-performing loan ratio between 0 and 1 through logistic regression. If the influence of multiple factors is dealt with, supervised learning can also use support vector machine to project the low-dimensional features into the high-dimensional space to achieve better risk segmentation. This past data-oriented learning process enables the model to continuously improve its forecasting ability, and provides a stronger numerical basis for financial institutions such as banks to make risk decisions.

3.2 Unsupervised learning algorithm and its technical characteristics in fraud detection

Unsupervised learning has its own technical characteristics in the application of fraud detection, which mainly shows that unsupervised learning does not need to use pre-marked data for the training process, but directly identifies the internal distribution law of data through clustering, thus realizing the discovery of data category centers by self-organizing method in preventing financial fraud, and aggregating transaction information into several non-overlapping Clusters, and each cluster represents a possible transaction mode. Its formula can be expressed as:

$$P(\text{Fraud}|X) = f(X_1, X_2 \cdots, X_n)$$

Where X is the transaction attribute vector. Compared with the traditional rule-based fraud detection method, unsupervised learning has three advantages: first, it can adapt to the changing fraud means through self-adjustment, and use the outlier detection algorithm to actively find those abnormal behaviors that deviate from normal trading activities without making any rules in advance. For example, in the practice of combating money laundering, criminals can distinguish abnormal trading activities through the aggregation nature of trading behavior even if they adjust the specific means of money laundering.

The second is the hidden association between criminals excavated by correlation analysis method and the organizational form found. The calculation formula of correlation degree is:

$$R = \text{Support}(X \rightarrow Y) / \text{Support}(X)$$

Among them, Support indicates the degree of support.

Third, organizations can be predicted, and criminal behaviors can be classified and analyzed in real time, so as to predict behavior patterns that can cause economic losses. Taking the anti-money laundering monitoring system based on unsupervised learning developed by Bank of China as an example, the system can shorten the processing time from 2 hours to 2 minutes, and improve the accuracy of identifying suspicious transactions, which undoubtedly becomes an effective risk prevention and control tool in the financial field.

3.3 Deep learning network structure and its technical advantages in abnormal transaction identification

Deep learning is the latest technology to deal with abnormal transaction detection, which makes full use of the advantages of multi-layer neural network architecture and can achieve good detection results in the process of abnormal transaction detection. The typical structure system of deep learning consists of three parts (see Figure 2): the input layer receives the original transaction data such as transaction amount, transaction time and transaction frequency, and then the hidden layer ($h_1, \dots, h_m$) extracts features according to a series of formulas. The calculation formula of each hidden layer is:

$$h(l) = f(W(l)h(l-1) + b(l))$$

l represents the number of layers of the network in the structure, W is the weight matrix of the hidden layer, b is the offset vector, and f is the activation function of the output layer. The output result of the output layer finally determines the abnormal risk score, and its formula is:

$$y = \text{softmax}(W(L)h(L-1) + b(L))$$

For trader behavior monitoring, under the framework of deep learning, the model can input all kinds of information such as transaction data (such as the quantity and price of transactions), communication data (such as telephone recording/email content) and behavior data (such as login schedule/usage path) from the data source, and output abnormal transaction warnings through multi-level feature extraction and fusion. For example, the voice converter with deep learning as the core developed by SmartLanguage Company, a British company, automatically detects abnormal events discussed by traders on the phone, thus effectively avoiding internal transactions. In addition, the deep learning model can automatically adjust the network parameters with market changes and new abnormal trading methods.

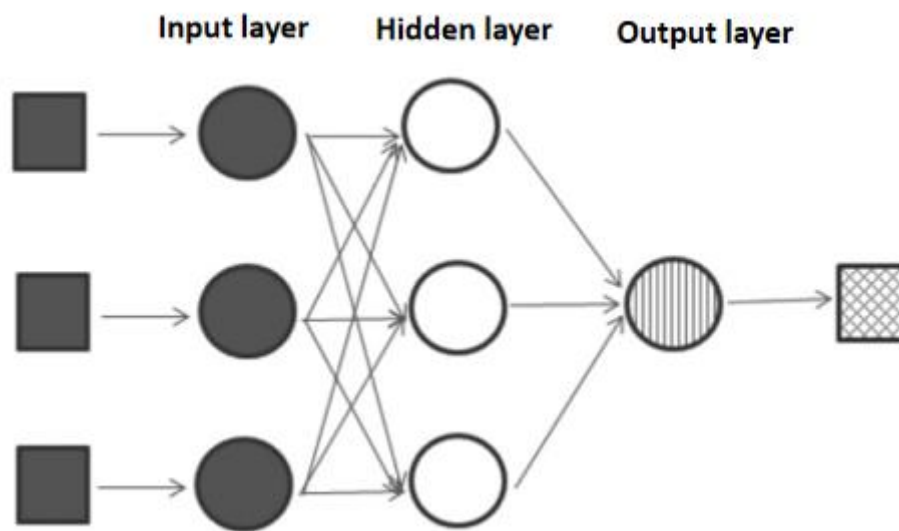


Figure 2 Deep learning analysis

4. Empirical analysis of financial risks based on machine learning

4.1 Construction and application of credit risk scoring model

In the financial risk control industry, big data and big data hybrid have gradually eliminated a single linear risk control mode. The hybrid model of big data is applied to credit risk prediction because machine learning can play a good solution to credit risk recognition rate and nonlinear correlation. For example, ZestFinance Company has designed a new model which is higher than the original rating mechanism and used for technological innovation strategy. The formula of this model is:

$$\text{CreditScore} = \sum_{i=1}^n w_i X_i$$

X_i is a set of variables, including the original traditional credit indicators (X_1 is income level, X_2 is financial stability), new variables (X_3 is social media activity level, X_4 is the stability of social network) and the characteristic attributes of consumers' personal preferences (X_5 is purchase frequency, X_6 is the distribution of purchase categories). Using deep learning, the characteristic weight parameters of these variables can be automatically changed, so as to adjust with the

combination of characteristics, and finally get the borrower's credit value with more accurate prediction results.

The most important innovation of this model is that it can find the potential correlation between borrowers' behavior attributes. For example, by analyzing the correlation between online communication and consumer behavior, we can explore the reasons that may have an impact on loan repayment (see Figure 3). The empirical research proves that the new model has a good effect in the application and operation. Compared with the traditional loan audit method, the processing speed is increased by about 90%, and the risk management and control ability is improved by 40%[7]. This successful attempt also proves the correctness and effectiveness of machine learning in credit rating, and it is also a new risk management solution for financial institutions.

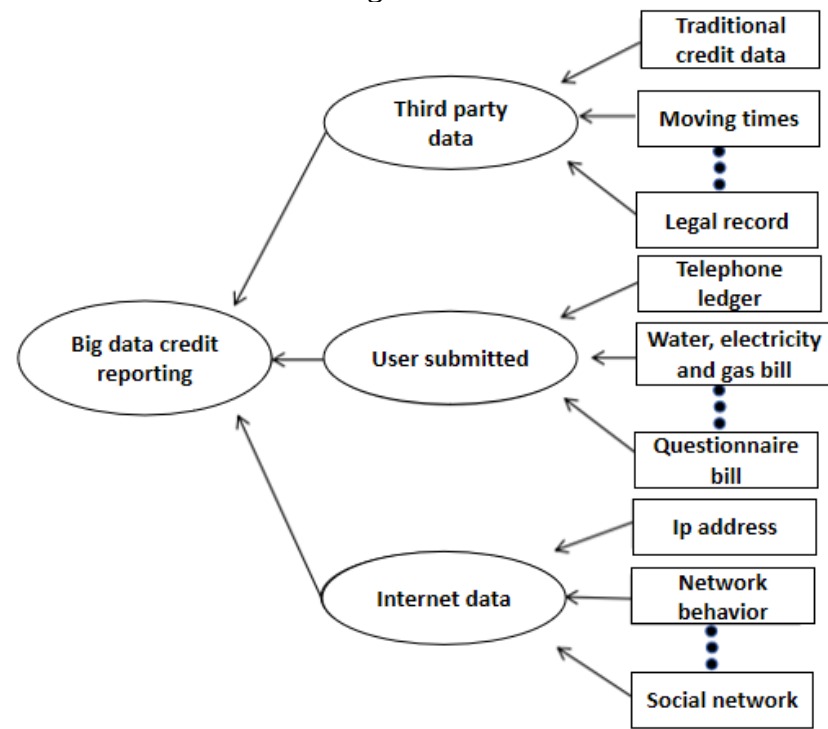


Figure 3 Big data sources of Zest Finance

4.2 Implementation and evaluation of fraud transaction detection system

Under the background of the rapid evolution of financial fraud means, Monzo built a real-time risk assessment system based on unsupervised learning, which broke through the traditional method of relying on historical labels and realized dynamic defense. In terms of system architecture, the system includes four information dimensions, namely: time attribute (including short-term and long-term characteristics such as minute/hour frequency and day/week pattern), monetary attribute (abnormal amount distribution and type), geographical attribute (cross-domain behavior, regional risk heat map) and merchant characteristics (risk score, merchant-user correlation map). Through these four characteristics, the system is constructed. Through online learning, the system uses unsupervised expectation maximization algorithm to learn conditional probability table to realize real-time updating of feature weights. In terms of real-time, in order to achieve millisecond-level decision-making, the system uses Kafka to process the big data of high-throughput transaction information, and then Flink completes feature calculation and risk rating, which can decide to make decisions every second. For example, when a cross-border theft case occurs, three high-risk cross-border transactions within five minutes will trigger the system to stop all subsequent

transactions and freeze accounts.

Compared with traditional methods, unsupervised learning has the advantages of no label, real-time decision-making, ability to identify unknown threats, and low maintenance cost. This model has been used for reference by Visa and others. In the future, it is necessary to combine confrontation training to prevent gan forgery attack, and combine federated learning to realize cross-institutional data sharing, so as to build a smarter and safer financial ecosystem.

4.3 Case analysis and effect evaluation

The innovative practice of China Bank in the fields of trade financing and anti-money laundering shows the comprehensive application value of machine learning in financial risk management. Through machine learning algorithm, the Bank intelligently analyzes and optimizes the weights of the four core dimensions of transaction content authenticity, capital flow, transaction behavior and historical records, thus constructing a scientific multi-dimensional evaluation system (see Table 1).

Table 1 Risk assessment index system of anti-money laundering verification system of China Bank [8]

Monitoring dimension	Specific indicators	Weight
Authenticity of transaction content	Document consistency	30%
Capital flow direction	Correspondence between receipt and payment	25%
Trading behavior	Abnormal degree of transaction frequency	25%
History	Historical violation record	20%

The biggest feature of the system is to upgrade the traditional manual file review to system automation, and analyze different types of information by using deep learning algorithms. In the authenticity of transaction content, optical character recognition (OCR) and natural language (NLP) are applied to automatically judge the consistency of business information. In the analysis of capital flow, the graphic neural network algorithm is applied to construct the graphic mode of capital flow, and the relationship between abnormal payment and collection can be identified in time. In the evaluation of trading behavior, time series analysis method is applied to find abnormal trading behavior patterns; In the playback of historical records, the algorithm of knowledge chart is applied to collect historical violations as a whole. With such a comprehensive and intelligent analysis framework, the work efficiency of the system can be greatly improved on the basis of ensuring high accuracy, that is, it has been shortened from two hours of manual judgment to two minutes of system identification now, which has been improved by 60 times.

Thanks to the self-learning characteristics of the system, we constantly improve our own judgment model, and then promote the steady improvement of anti-money laundering identification accuracy. This successful practice provides an example for the digital risk control transformation of financial institutions, and shows the potential of machine learning to improve the efficiency of financial risk control in the future.

5. Conclusion

This paper deeply discusses the application of machine learning in financial risk modeling and fraud detection, and highlights its core role in reducing financial risks of enterprises. With the transformation of financial industry from tradition to intelligence, under the background that the existing risk control methods can no longer meet the requirements, machine learning has become the core weapon of financial technology risk control with its efficient data processing ability and self-learning.

This paper introduces in detail the innovative application of supervised learning in credit risk

scoring, the unique advantages of unsupervised learning in detecting fraud and the high efficiency of deep learning in detecting abnormal transactions. Through empirical cases, this paper analyzes the effectiveness of machine learning in the construction of credit risk scoring model, fraud transaction detection, international trade loan business and anti-money laundering audit, which can significantly improve the speed and accuracy of risk detection and reduce the processing time.

This study not only provides a feasible path for financial enterprises to build efficient internal risk management, but also provides support for regulators to improve the external risk control environment, which is of great significance for improving the risk prevention level of the whole financial industry.

References

- [1] Li Cangshu. *Risk prediction and prevention of China's new financial format based on machine learning theory* [J]. *Academic Forum*, 2019,42(04):114-121.
- [2] Cui Xinning. *Application of machine learning algorithm in financial market risk analysis and prediction* [N]. *Henan Economic News*, 2025-02-06(009).
- [3] Zhou Yingxue. *Cross-domain spillover of systemic financial risks and policy responses* [D]. Jilin University, 2025.
- [4] Xu Yi. *Analysis of the application of machine learning in financial risk management* [J]. *Journal of anshun university*, 2019,21(05):110-114.
- [5] Long Huihui. *Research and prototype realization of credit risk control early warning method based on machine learning* [D]. University of Electronic Science and Technology of China, 2020.
- [6] Li Zimu. *Research on financial risk control model and algorithm based on machine learning* [D]. north china university of technology, 2019.
- [7] Li Jingzhao, Zhang Yuxin, Wang Aojun. *Research on Internet Financial Risk Management* [J]. *Cooperative Economy and Technology*, 2024,(13):52-54.
- [8] Zhang Hua. *Research on User Credit Evaluation Model Based on Telecommunication Data* [D]. Beijing University of Posts and Telecommunications, 2018.