

The use and risks of machine learning in finance

Wang Xinyue^{1,*}

¹*School of Mathematics and Statistics, Beijing Technology and Business University, Beijing, China*

**Corresponding author*

Abstract: *Under the influence of the rapid development of the market economy, the financial market also shows a good development trend, and statistics plays a very important role in the financial market, which can promote its long-term stable development. Under the background of big data, the integration of machine technology and statistical technology and the use of machine learning can better utilize the advantages of the two to identify the risks and problems in the financial market, which can help to identify the potential risks and challenges in the financial market, formulate scientific prevention and response strategies, and further lay a solid foundation for the future development of the financial market.*

Keywords: *machine learning, Risk prediction, Credit risk assessment*

1. Introduction

With the advent of the digital era, the amount of data in various industries has shown explosive growth. How to mine valuable information from massive data has become one of the hot issues that academics and industries are currently focusing on together. The financial industry, as a data-concentrated field, has a huge amount of data and a variety of data structures, which brings unprecedented challenges to financial researchers and professionals in related fields. The development of artificial intelligence has been included in China's long-term strategic planning[1], and the financial sector has become one of the application areas of artificial intelligence due to its highly informatized and digitized nature. Many financial institutions in China have realized strategic upgrading of financial technology through AI technology, and realized product innovation, process reengineering and service upgrading in a variety of business scenarios[2]. With the improvement of the entire social credit system, the growth of external data companies, and the opening and sharing of data by government agencies, finance-specific big data can provide the basic nutrients for the continued development of artificial intelligence, transforming from purely in-line data analysis to the combination of social data, forming a smarter means of technological empowerment.

2. Importance of Machine Learning in Finance

2.1. Reliability of machine learning

An individual credit rating system is essential when dealing with financial issues. Banks, investment firms, and stock markets transact up to billions of dollars every day. Due to the bias and selfishness that may exist in human nature, some people tend to commit fraud during money transactions. Even though banks exercise extreme caution and carefully verify the creditworthiness of companies, defaults on bank debts by individuals and even multinational corporations still seem to be a common phenomenon in the financial sector. Several financial institutions utilize scoring models to reduce credit risk in credit assessment, disbursement and monitoring. Credit scoring models based on classical statistical theory are widely used. However, these models perform poorly in terms of resilience when a large number of data inputs are involved. As a result, some of the assumptions in classical statistical analysis fail to hold, which in turn affects the accuracy of predictions. To address such issues, machines embedded with machine learning can further improve the accuracy of predictions when processing requests.

2.2. Fastness of machine learning

Trading stocks in the stock market is very difficult. People usually do a lot of analysis in historical

data, charts and formulas to predict the future of a stock, while others simply place random bets. All these actions sound very hectic and time-consuming. Machine learning algorithms are capable of performing precise in-depth analysis of thousands of data sets and can give succinct and accurate predictions in a short period of time, helping to alleviate the hassle of organizing and analyzing big data.

2.3. Accuracy of machine learning

People are not capable of or do not like to do repetitive monotonous tasks, and this repetitive labor tends to produce many errors, whereas machines can perform repetitive tasks for an unlimited amount of time. Machine learning algorithms will do the hard work of analyzing data and recommending new strategies when humans need them, and can also detect subtle or non-intuitive patterns more effectively than humans can to identify fraudulent transactions. In addition, unsupervised machine learning models can analyze and process new data without interruption and then automatically update their models to reflect the latest trends.

2.4. Risk prediction for machine learning

With the rise of the big data era, a large amount of financial data has brought great challenges to the traditional financial risk management system, the problem of financial risk management has become more and more prominent, and the traditional analysis methods can no longer meet the current financial risk management needs. Artificial intelligence technology is an emerging and promising cross-discipline, which simulates human thinking mode and behavior through the computer to carry out risk analysis and decision-making. With the rapid development of various computer technologies and the common progress of the economy and society, machine learning algorithms are gradually introduced into the field of financial risk prediction. In the face of the sample complexity of financial risk data, the diversity of indicators and other characteristics, the choice of machine learning models for its analysis and prediction, the study of machine learning in the field of financial risk prediction can further strengthen the efficiency of financial risk management and promote the good development of society.

3. Purpose and significance of the study

3.1. Application of machine learning algorithms in finance

3.1.1. Early warning of risks

With the arrival of the information age, the financial field has a huge volume and more types of data, which need to be analyzed and processed with the help of computer software. Under the perspective of big data, the type of financial business in China has become more diversified, and the financial instruments are also more abundant. Statistics and financial management are closely linked, and can statistically manage all kinds of information in economic activities, realize the sharing of data and information, and effectively process the data to promote the successful completion of various tasks. In recent years, China's financial market has developed rapidly on the basis of the popularization of computer technology and the transformation and development of the real economy, and has made certain achievements. Supported by computer technology and network technology, the scale of the financial market is rapidly expanding, but because of the virtual nature of network financial products, experts in the field of finance are not able to discover their potential risks in time, nor can they formulate scientific risk prevention and response programs in conjunction with the specific development situation, which is obviously not conducive to the healthy development of the financial market [3].

In today's rising artificial intelligence, the machine learning method shows good arithmetic effect and strong scenario applicability in risk prediction [4]. According to the literature in the field of risk early warning, the traditional early warning models mainly cover several kinds of models such as Logit model, multivariate cumulative model, and Markov zone system transfer model. However, traditional econometric models have two obvious shortcomings in their predictive ability: first, real-life data often cannot meet the strict criteria set by the models; from another perspective, econometric models require highly consistent coefficient estimation, and the over-pursuit of "unbiasedness" may generate large variance, which is detrimental to accurate prediction. This is detrimental to accurate prediction. The

weaknesses mentioned above are the strengths of machine learning models. In recent years, new machine learning models have gradually gained attention, and their scope of application has been expanding, especially in the two fields of micro enterprise risk warning and macro financial risk warning. The use of machine learning methods as a means of early warning has successfully solved the problems of traditional models, enhanced the accuracy and practicality of early warning, and provided an effective complement to econometric methods. The financial risk prediction model based on machine learning has more outstanding prediction performance than the financial prediction model based on traditional statistical methods, and can provide theoretical support for financial decision-making in reality.

3.1.2. Forecast of financial market trends

The Bank of China Research Institute released the 2024 Economic and Financial Outlook Report, which concluded that the global economic growth momentum in 2023 will be insufficient, with the growth rate of developed economies slowing down and the overall performance of emerging economies stabilizing. The growth of global trade is slightly weak, the production boom of each country is gradually declining, and the pulling effect of domestic demand on the economy is weakening. In the financial industry, the prediction of the market direction has always been a major problem in the financial industry. Alternatively, fundamental factors such as macroeconomic data and company financial statements are studied to assess the intrinsic value of an asset and predict market direction accordingly. Traditional methods usually rely on historical data and models to make predictions, and are unable to fully consider new factors and unexpected events that may occur in the future, resulting in predictions that may not be accurate enough; and also ignore the non-linear relationships and complexity that exist in the financial market, oversimplify the complexity and uncertainty of the market, and ignore the impact of the emotions and behaviors of market participants on the direction of the market. It is unable to fully reflect the real situation of the market.

With the deep development of artificial intelligence technology, fintech is in a period of unprecedented transformation [5]. Therefore emerging a process of extracting valuable information and knowledge from large, complex data sets and conducting deep mining and analysis through artificial intelligence technology combined with big data analysis methods can help financial institutions analyze and mine massive financial data to discover potential trends and patterns. By analyzing historical data, machine learning can quickly capture hidden patterns in the market and predict future trends. Algorithms can also be used for high-speed trading decisions and execution, as well as automated trading operations. Real-time monitoring of market conditions and a large amount of historical trading data can identify short-term changes in market trends and respond quickly to achieve efficient high-frequency trading.

3.1.3. Credit risk assessment

Financial institutions need to assess the credit risk of borrowers when lending or offering credit products to determine the likelihood of default. Credit risk assessment involves a comprehensive analysis of the borrower's credit history, financial situation, behavioral history and other aspects, so as to help financial institutions formulate appropriate loan terms and interest rates and effectively manage risks. In the field of credit risk assessment, the traditional assessment method is the expert scoring method, i.e., credit industry experts subjectively assess credit risk elements through their own professional knowledge and industry experience, such as the 5C, 5P and 5W element methods [6]. With the development of financial globalization, the financial sector's attention to credit risk has gradually increased, and the credit risk assessment methods have been constantly innovated, and both quantitative hard information and stereotyped soft information have been widely used in the field of credit risk assessment. Therefore, the choice of machine learning can help financial institutions to build more accurate credit scoring models to better predict the probability of default by analyzing a large amount of data and identifying the key factors that affect the credit risk of borrowers. The model built by machine learning can automatically select the most relevant features to help financial institutions identify the most influential factors on credit risk assessment and improve the accuracy of the assessment model.

At the data level, machine learning usually employs different forms of resampling methods to balance the distribution of classes in a dataset, either by increasing the number of minority classes or decreasing the number of majority classes in order to eliminate the class imbalance problem. These include 3 types of methods: undersampling, oversampling, and hybrid sampling. At the algorithmic level, the training dataset is usually not overly pre-processed, but rather the traditional classifiers are improved and redesigned to improve their applicability to the special classification needs of the

unbalanced dataset by taking into account the cost differences of various misclassification cases.

3.1.4. Trading strategies

One type of application of machine learning in finance is to improve trading strategies. In the machine learning context, supply chain financial collusion behavior is taken as the research content, and financial institutions, small and medium-sized enterprises (SMEs), core enterprises or supportive enterprises are taken as the research objects, to analyze the interaction between financial institutions' science and technology investment strategies and colluders' behavioral strategy choices. In recent years, machine learning technology has been used to improve the trading strategy from reducing trading noise, introducing new factors, and optimizing the asset portfolio, respectively.

3.2. Advantages of machine learning algorithms

With the continuous development of financial technology, the financial industry has experienced or is undergoing significant changes at many levels. Due to the increasing complexity of the modern financial system, the limitations of traditional risk modeling tools have become increasingly prominent, while machine learning methods are good at capturing the complex nonlinear relationships between variables, and have many inherent advantages over traditional economic analysis and forecasting techniques, which can better meet the modeling requirements and analytical predictions of the economy and finance, which is a typical complex and open system [6].

Mainstream machine learning models include penalized factorial regression, support vector machines, decision tree-based models (e.g., single decision trees, random forests, Boosting models, etc.), and neural networks. There is no such thing as a certain machine learning model that outperforms all other models for all prediction problems, so the most appropriate model for a particular prediction problem needs to be selected based on the actual situation.

A single model may have excellent performance in one aspect, but an integrated model is better in terms of robustness and model reliability. The assignment method can be utilized, i.e., different weights are assigned to the results of different machine learning models. The equal weight method is a special case of the assignment method, and its basic idea is simple and easy to understand, i.e., a single algorithm will not always be effective, and taking the mean of the predictions of a series of algorithms can smooth out the errors of different algorithms, and thus obtain more effective predictions. Or it can be used to identify or use integrated methods to build optimal models for systemic financial risk prediction and early warning through further model building, such as modeling the regression results of different machine learning models again for the actual values, thus obtaining a two-layer model and other machine learning modeling tools. Different models can be built to help financial institutions improve efficiency, reduce risk, enhance customer experience, and promote the innovative development of the financial industry.

3.3. Challenges of Machine Learning Algorithms

In the field of financial risk prediction, although machine learning algorithms can provide more accurate prediction results compared to traditional statistical methods, the increase in algorithmic complexity will result in a decrease in the interpretability of the model and the inability to discern the reasonableness of the model's results, leading to difficulties in giving explanatory and persuasive conclusions about wind control and dramatically decreasing the model's practical benefits.

In the future, financial innovation is likely to rely more on AI technology. However, apart from the technology developers who understand the thought processes of the technology, financial regulators, financial institutions and the public have no way of understanding how machine-learning-based AI makes investment decisions. The black box problem of training data and algorithms may hide "illusory information" and "algorithmic bias". At the data collection stage, machine learning is susceptible to the "garbage input, garbage output" problem. If the data is problematic, the decisions made by the trained algorithms will also be flawed. A variety of data issues can affect the accuracy of machine learning algorithms in assessing the risk of potential investment losses.

Currently, several mainstream machine learning algorithms based on artificial intelligence techniques may reduce the number of decision participants and standardize risk assessment among the remaining decision makers. The data used to train financial risk management algorithms is limited, and limited data sources can reduce the accuracy of the algorithms, making their application to similar assets more consistent and the results more convergent.

4. Conclusion

And the application of machine learning covers a wide range of aspects such as risk early warning, prediction of financial market trends, credit risk assessment, and optimization of trading strategies. Through machine learning algorithms, financial institutions can more accurately assess risks, improve efficiency, reduce costs, and innovate business models.

References

- [1] Liu Hongbo, Lin Bin. *Value Orientation, Issue Construction and Path Selection of China's Artificial Intelligence Development - A Quantitative Study Based on Policy Texts* [J]. *E-Government*, 2018(11): 47-58.
- [2] Shi Guang, Song Fangxiu. *The main features, risks and development countermeasures of the new round of financial technology innovation* [J]. *Economy*, 2020(12): 100-108.
- [3] Liu Y. *Research on economic forecasting and financial statistics of banks based on big data* [J]. *Investment and Cooperation*, 2023(10): 4-6.
- [4] Cui Zhenguan, Jin Yide. *Application of machine learning in early warning of China's overseas investment efficiency: Evidence from countries along the "Belt and Road"* [J]. *Finance Theory and Practice*, 2023(05): 13-25.
- [5] He Tao. *The direction and path of fintech innovation driven by AI* [J]. *China Collective Economy*, 2024(03): 103-106.
- [6] Li Hongquan, Zhou Liang. *Systemic financial risk monitoring and early warning based on machine learning technology* [J]. *Operations Research and Management*, 2023, 32(11): 212-219.