Machine Learning Based Financial Applications of Data

Hangyi Li

Department of Industrial Engineering, Tongji University, Shanghai 200092, China

perling@tongji.edu.cn

Abstract. As a matter of fact, with the rapid development of big data and machine learning technology, the field of data finance is facing huge opportunities and challenges especially in recent years. With this in mind, based on machine learning model, this study makes a deep analysis and research on data finance. Through the application and comparison of multiple machine learning models, this paper finds that different models have different predictive effects on data finance. At the same time, this paper also discusses some frontier problems as well as challenges in the field of data finance in detail. According to the analysis, it provides ideas and references for future research. In the meantime, some practical application cases are added in order to more intuitively demonstrate the application of machine learning models in data finance. Overall, these results shed light on guiding further exploration of data finance in terms of machine learning scenarios.

Keywords: Machine Learning, Data Finance, Predictive Model, Risk Assessment, Investment Strategy.

1 INTRODUCTION

With the increasing complexity and uncertainty of the financial market, the traditional financial analysis methods have been difficult to meet the needs of the market. The continuous development of big data and machine learning technology provides new solutions for the financial field [1-3]. Data finance is an important branch of the financial field, which involves a lot of data analysis and processing. Through the analysis and mining of historical data, the law and trend of the market can be found, so as to provide strong support for investment decisions. This paper chooses the application of machine learning model in data finance as the research topic, aiming to explore the prediction effect and advantages and disadvantages of different machine learning models in data finance. Some practical application cases are added in order to more intuitively demonstrate the application of machine learning models in data finance [4].

Data finance is a rapidly growing field that covers a large amount of data analysis and processing. With the increasing complexity and uncertainty of the financial market, traditional financial analysis methods have been difficult to meet the needs of the market. Therefore, the development of the field of data finance provides new solutions for the financial field [5-7].

R. Magdalena et al. (eds.), *Proceedings of the 2024 9th International Conference on Social Sciences and Economic Development (ICSSED 2024)*, Advances in Economics, Business and Management Research 289,
https://doi.org/10.2991/978-94-6463-459-4_17

The development of data finance goes back decades. Initially, data analysis in the financial sector relied on manual operations and simple statistical methods. However, with the continuous increase of data and complexity, traditional analysis methods have been unable to meet the needs. Therefore, with the development of computer technology and big data technology, machine learning models have been widely used in the financial field [7].

The research in the field of data finance is of great significance. First of all, through the analysis and mining of a large number of data, you can find the rules and trends of the market, so as to provide strong support for investment decisions. Secondly, research in the field of data finance can help financial institutions better understand the needs and behaviors of customers, so as to formulate more accurate marketing strategies and risk control strategies. In addition, research in the field of data finance can also promote the transparency and fairness of financial markets, and improve the efficiency of financial markets [7].

The field of data finance is closely linked to other fields. First, the field of data finance is closely related to computer science. Computer science provides powerful data processing and analysis capabilities to the field of data finance, making machine learning models widely used. In addition, the field of data finance is also closely related to economics, finance, statistics and other disciplines. These disciplines provide the theoretical foundation and analytical tools for the field of data finance, supporting research and development. At the same time, the development of the field of data finance has also promoted the innovation and development of these disciplines.

2 APPLICATIONS

With the increasing complexity and uncertainty of the financial market, the traditional financial analysis methods have been difficult to meet the needs of the market. The continuous development of big data and machine learning technology provides new solutions for the financial field. Data finance is an important branch of the financial field, which involves a lot of data analysis and processing. Through the analysis and mining of historical data, the law and trend of the market can be found, so as to provide strong support for investment decisions.

A support vector machine (SVM) is a supervised learning model that can make classification or regression predictions based on training data sets. In the aspect of stock price prediction, some studies show that SVM can achieve better prediction results. For example, Li et al. used SVM model to predict stock prices and found that the model can effectively predict the volatility trend of stock prices. In addition, SVM can also be applied to credit risk assessment and market volatility prediction [8].

A random forest (RF) is an ensemble learning model that makes predictions by building multiple decision trees and taking the average of their outputs. In terms of credit risk assessment, Zhou et al. used the RF model to perform credit scores on customers and found that the model was effective in identifying high-risk customers. In addition, RF can also be applied to market forecasting and stock price forecasting [9].

Neural network (NN) is a computational model that simulates the neural network structure of human brain, which can map the input data in a complex non-linear way. In terms of market volatility prediction, Zhang et al. used NN model to forecast market volatility and found that the model could effectively predict the market volatility trend [10]. In addition, NN can also be applied to stock price prediction and credit risk assessment.

Different machine learning models have their advantages and disadvantages in the prediction of data finance. For example, SVM model has high accuracy, but it has some limitations when dealing with multiple classification problems. RF model has good stability and generalization ability, but it may be troubled by overfitting when dealing with high dimensional data. NN model has strong nonlinear mapping ability, but it is easy to be affected by local minima during training. Therefore, it is necessary to choose the appropriate model according to different needs in practical application.

3 METHODOLOGY AND EXPERIMENTAL DESIGN

This paper uses the method based on machine learning model to analyze and study data finance. Specifically, this paper selects three models of support vector machine (SVM), random forest (RF) and neural network (NN) for experiment and analysis. Through the training and testing of these three models, one can find their advantages and disadvantages in the forecast of data finance and the scope of application.

Before model training, the data need to be preprocessed. First of all, it is necessary to obtain historical trading data from relevant financial institutions, including stock price, volume and other information. The data then needs to be cleaned and collated to remove outliers and missing values. In addition, it is also necessary to normalize the data to eliminate the impact of dimensions between different features on model training.

Before model training, it is necessary to extract features from the data. According to the different research problems, different features can be selected for extraction. For example, in the stock price forecast, one can extract the historical stock price, trading volume and other features; In the credit risk assessment, the customer's age, income, credit history and other characteristics can be extracted; In the prediction of market volatility, one can extract the characteristics of market volatility and trading volume.

After data preprocessing and feature extraction, the model needs to be trained. Specifically, you need to divide the data set into a training set and a test set, then use the training set to train the model, and use the test set to evaluate the predictive effect of the model. In the training process, it is necessary to select appropriate parameters and optimization algorithms to obtain the best model performance.

In order to evaluate the prediction effect of the model objectively, it is necessary to select the appropriate evaluation index. Commonly used evaluation indicators include accuracy, recall rate, F1 value, MSE (mean square error), MAE (mean absolute error), etc. According to different research questions, different evaluation indicators can be selected for evaluation. For example, in stock price forecasting, MSE and MAE can be used to measure the prediction error of the model; In credit risk assessment, accuracy and recall rate can be used to measure the classification effect of the model.

4 EMPIRICAL ANALYSIS

Through experiments on three machine learning models (SVM, RF and NN), one gets the following experimental results. In terms of stock price prediction, SVM model has the best prediction effect, and its MSE and MAE values are the smallest, which indicates that this model can predict the volatility trend of stock price well. The RF model and NN model had the second best predictive effect, but their MSE and MAE values were slightly higher than those of the SVM model. In terms of credit risk assessment, the RF model has the best classification effect and the highest accuracy and recall rate. SVM model and NN model have the second best classification effect, but their accuracy and recall rate are slightly lower than RF model. In terms of market volatility prediction, NN model has the best performance and the smallest prediction error. SVM model and RF model performed second, but their prediction error was slightly higher than NN model.

According to the experimental results, one can draw the following conclusions. Different machine learning models perform differently in the prediction of data finance. SVM models perform best in stock price prediction, RF models perform best in credit risk assessment, and NN models perform best in market volatility prediction. Different problems require the selection of appropriate machine learning models. For classification problems, such as credit risk assessment and market volatility prediction, RF and NN models can be selected. For regression problems, such as stock price prediction, SVM model can be chosen. The effectiveness of machine learning models is influenced by a variety of factors. Including the quality and scale of data, the selection of features, the parameters of the model and the optimization algorithm. Therefore, in practical application, it is necessary to select and adjust according to the specific situation. Machine learning model has good prediction effect and generalization ability. Through training and testing, it can be found that machine learning models can effectively discover the laws and trends of the market, and provide strong support for investment decisions and market risk management.

5 CONCLUSION

Through the research on the application of machine learning model in data finance, the following conclusions can be drawn. Machine learning model has wide application value in the field of data finance. Through the application and comparison of several machine learning models, it is found that different models have different forecasting effects and advantages in stock price prediction, credit risk assessment and market volatility prediction. Different problems require the selection of appropriate machine learning models. For classification problems, such as credit risk assessment and market volatility prediction, RF and NN models can be selected; For regression problems, such as stock price prediction, SVM model can be chosen.

The prediction effect of machine learning models is affected by many factors. Including the quality and scale of data, the selection of features, the parameters of the model and the optimization algorithm. Therefore, in practical application, it is

necessary to select and adjust according to the specific situation. Machine learning model has good prediction effect and generalization ability. Through training and testing, it can be found that machine learning models can effectively discover the laws and trends of the market, and provide strong support for investment decisions and market risk management.

There are still some shortcomings in this study. This paper only selects three machine learning models for experiment and analysis, and fails to comprehensively discuss the application of other models in data finance.

In the future, the scope of research can be further expanded to explore the application of more machine learning models in data finance. The data source of this paper is limited to the historical transaction data of a financial institution, and it fails to cover more market data. In the future, data sources can be further expanded to obtain more comprehensive market data and improve the forecasting effect of the model. The research of this paper mainly focuses on the application of machine learning model in data finance, and fails to involve more problems in the financial field. In the future, the application of machine learning model in financial risk management and financial market prediction can be further discussed. This paper mainly adopts traditional machine learning methods for experiments and analysis, and fails to involve cutting-edge technologies such as deep learning. In the future, the application of deep learning in data finance can be further explored to improve the predictive performance of the model.

REFERENCES

1. Guida, T.: Big data and machine learning in quantitative investment. John Wiley Sons (2019).
2. Mądra-Sawicka, M.: Financial management in the big data era. In Management in the Era of Big Data (pp. 71-81). Auerbach Publications (2020).
3. De Prado, M. L.: Advances in financial machine learning. John Wiley Sons (2018).
4. Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., Herrera Viedma, E.: Machine learning methods for systemic risk analysis in financial sectors (2019).
5. Pérez-Juárez, M. A., Aguiar-Pérez, J. M., Alonso-Felipe, M., Del-Pozo-Velázquez, J., Rozada-Raneros, S., Barrio-Conde, M.: Exploring the Possibilities of Artificial Intelligence and Big Data Techniques to Enhance Gamified Financial Services. In Next-Generation Applications and Implementations of Gamification Systems (pp. 187-204). IGI Global (2022).
6. Damrongsakmethee, T., Neagoe, V. E.: Data mining and machine learning for financial analysis. Indian Journal of Science and Technology, 10(39), 1-7 (2017).
7. Tatsat, H., Puri, S., Lookabaugh, B.: Machine Learning and Data Science Blueprints for Finance. O'Reilly Media (2020).
8. Li, X., Wu, P., Wang, W.: Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. Information Processing Management, 57(5), 102212 (2020).
9. Zhou, J., Wang, C., Ren, F., Chen, G.: Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. Decision Support Systems, 149, 113611 (2021).
10. Zhang, P., Ci, B.: Deep belief network for gold price forecasting. Resources Policy, 69, 101806 (2020).