Population, Resources & Environmental Economics (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/pree.2025.060102
ISSN 2616-2253 Vol. 6 Num. 1

Research on Financial Performance Prediction Model of Listed Companies Based on Multi Source Data and Machine Learning Algorithms

Peng Dong

School of Business, Stevens Institute of Technology, Hoboken, New Jersey, 7030, United States
pengdong0128@gmail.com

Keywords: Based on multi-source data, financial performance prediction, support vector machine, Logit regression model, machine learning algorithm, manufacturing listed companies, risk warning, data analysis

Abstract: This study aims to develop a financial performance prediction model for listed companies based on multi-source data and machine learning algorithms to address the financial uncertainty faced by enterprises in complex economic environments. The decline in financial performance will affect the internal operations and financial stability of the company, affecting its investors, creditors, and other stakeholders. Building an efficient financial performance forecasting system can help companies identify potential operational risks early, provide data support for management to develop more accurate response strategies, and avoid or mitigate the negative impact of financial difficulties. This article proposes a prediction framework based on multi-source data, and constructs Logit regression model and support vector machine model for prediction experiments through multi-dimensional financial indicator preprocessing. Research has found that there is a certain gap in prediction accuracy between the two models, with support vector machine models performing better in identifying companies with deteriorating financial performance, especially in dealing with complex and multi-source data, demonstrating strong robustness and adaptability. This article also reveals the potential and limitations of current machine learning algorithms in financial performance prediction. Although support vector machines have better predictive performance than Logit regression, there is still room for improvement in accuracy in practical applications. Future research will consider integrating more non-financial data, such as industry trends, market sentiment, etc., and further improving the predictive performance of the model by optimizing algorithm structure and parameters. We hope to provide more scientific and forward-looking support for the financial management and decision-making of enterprises, laying a solid foundation for enhancing their ability to cope with uncertain risks.

1. Introduction

The sustained decline in a company's financial performance will have a significant impact on its daily operations and funding chain, triggering systemic risks that affect shareholders, creditors, and

other stakeholders. Accurately predicting changes in financial performance and warning potential financial crises has important strategic significance. An efficient financial performance forecasting system can provide early warning for enterprises, help management take targeted measures before problems worsen, and reduce the possibility of risk losses.

With the increasingly complex global economic environment and the wide range and complex structure of financial data sources for enterprises, traditional forecasting methods have shown many limitations in dealing with multi-source heterogeneous data. Thanks to the development of big data and artificial intelligence technology, the integration and analysis of multi-source data provide new possibilities for financial performance prediction. Machine learning algorithms have significant advantages in processing complex data, which can improve the accuracy and robustness of predictions. The research on financial forecasting models based on multi-source data and machine learning has important theoretical value and provides powerful tools for practical management and decision-making of enterprises.

This article reviews the relevant research results on financial distress prediction at home and abroad, and provides a detailed review of the application status of Logit regression models and support vector machines in financial prediction. It analyzes their advantages and disadvantages in different scenarios. Support vector machines demonstrate superiority in handling nonlinear and complex data due to their strong generalization ability, while Logit regression models still have important reference value in specific linear scenarios. This article will combine the advantages of these two methods and propose an innovative financial forecasting model framework based on multi-source data.

This article introduces research methods and innovations, including improvements in data preprocessing, indicator screening, and model optimization. This article aims to improve the accuracy and reliability of financial performance forecasting, provide more forward-looking decision support for enterprises in complex economic environments, help them cope with potential financial risks, and enhance market competitiveness.

2. Related research

Although the academic community has accumulated relatively mature research results in the field of financial crisis warning, research focused on financial performance prediction is still insufficient. The purpose of this study is to combine existing financial distress warning models with advanced machine learning techniques to construct a financial performance prediction model for listed manufacturing companies in China. By analyzing multi-source data, extracting key financial indicators, and combining machine learning algorithms, this article aims to reveal potential trends in corporate financial performance and provide reliable decision support for enterprise managers. Based on the emergency database of the General Hospital of the Chinese People’s Liberation Army and under the guidance of professional clinicians, the X Liu team designed the inclusion and exclusion criteria of the research experiment in view of the injury situation of traumatic hemorrhagic shock, extracted the medical index data of related patients, and conducted data preprocessing[1]. The X Zhang team conducted a two-year field experiment using four rice varieties under three irrigation regimes[2]. The X Chen team proposed a model fusion method based on multi-source heterogeneous data and different learning algorithms for predicting TP (MF-MSHD)[3]. V Nistane aims to propose a prediction method based on the combination of optimized health indicators (OHI) and machine learning algorithms[4].

3. Theoretical basis and methodological exploration of financial performance forecasting

3.1 Exploration of Crisis Management Theory in Financial Performance Forecasting

In the process of enterprise operation, unexpected events often trigger highly uncertain crisis situations, requiring managers to make wise decisions quickly within limited time and resources. Effective crisis management can resolve potential risks, turn crises into opportunities, and drive enterprises towards new directions of development[5].If not handled properly, the crisis may worsen, leading to serious financial losses and even bankruptcy of the enterprise. When facing internal and external challenges, enterprises must adopt proactive and flexible strategies to turn crises into opportunities.

In the context of increasingly globalized economy, enterprises realize that it is difficult to solve deep-seated problems solely by improving the external environment. The insufficient and erroneous internal management level has become an important factor restricting the sustainable development of enterprises. Traditional management theories have shown certain limitations in dealing with complex and changing crisis situations, and there is an urgent need for new theories and methods to deeply analyze the causes of crises and develop effective response measures. Taking the research team from Wuhan University of Technology as an example, they pointed out that crises have a certain degree of reproducibility, and by establishing a systematic warning mechanism, enterprises can effectively reduce the probability of crisis occurrence. They combined qualitative and quantitative analysis, comprehensively considered internal and external factors, constructed a comprehensive evaluation index system, and determined the weights of each index through the Analytic Hierarchy Process and Fuzzy Set Theory, achieving accurate risk monitoring and early warning.

This article aims to establish a financial indicator prediction model for Chinese manufacturing listed companies. This model helps to identify and reflect potential risks of deteriorating financial performance in a timely manner, providing scientific basis for management decisions, improving the company's ability to respond to crises, and protecting the legitimate rights and interests of relevant parties.

3.2 Support Vector Machine Theory and Practice in Financial Performance Forecasting

Vector Machine, developed by Cortes and Vapnik in the mid-1990s, is a complex machine learning technique based on optimization algorithms. Over time, their applications in data analysis and machine learning have significantly increased, mainly due to successful solutions to complex tasks such as regression and image recognition. Vector support machines have been proven to be highly effective in responding to financial crises and evaluating efficiency, and are therefore widely used in fields such as scientific research, engineering, and economic management.

The basic theory of vector machines is based on statistical learning principles, with a focus on minimizing structural risk to improve the model's generalization ability. In the application stage, SVM uses nonlinear mapping techniques to transform input data into an arrogant feature space, where it searches for the most suitable hyperplane for classification. Usually, this nonlinear transformation is achieved through kernel functions, which effectively solve the problem of arrogant datasets and significantly improve the model's ability to handle complex data structures. By defining the optimal classification boundary in this arrogant space, SVM not only improves classification accuracy, but also transforms the learning process into a second-order programming problem, ensuring the stability and uniqueness of the obtained model.

Vector support is highly valued by the ability to resolve classification and regression problems, and significantly improved the efficiency of predicting the performance of listed companies. One of

the main advantages of SVM is to reveal complex relationships that may not happen immediately. This ability makes the SVM an indispensable resource for officers seeking to identify potential financial vulnerabilities and improve management strategies. By analyzing historical financial data, SVM can provide reliable forecasts for future financial performance and build a solid foundation for wise decision making. In this regard, the administrator can quickly adjust the strategy based on the dynamics of the market and ultimately increase the risk burden of the enterprise.

4. Selection of predictive indicators and data preprocessing methods

The construction of financial forecasting models for listed companies depends on careful selection and full analysis of financial indicators, which are key criteria for measuring corporate financial conditions and provide decision-making information on the behavior of operational indicators and cash flows to decision makers. Using a single metric also has no overall impact on the overall performance of the firm. In order to better reflect the financial situation, it is necessary to diversify several indicators. Examining different indicators and their relationships can greatly enhance the depth and accuracy of financial analysis.

The ratio between gross interest rates and income increases is emphasizing corporate earnings and competitive advantage, as indicators that measure liquidity and efficiency of the company's short-term debt repayment capabilities. This analysis helps to accurately identify potential financial risks in a changing environment, provide a more reliable, data driven perspective, and help guide management decisions.

Integrating different data sources and machine learning algorithms can increase the efficiency and accuracy of analysis and processing of multiple financial indicators. Machine learning is particularly excellent in identifying complex models of data concentration, and these models can predict financial fluctuations more accurately. Using historical data, the algorithm identifies the nonlinear relationship between different financial indices and determines the important factors that affect the financial results.

Create a predictive model that combines the set of financial indicators and their possible interactions, allowing organizations to quickly identify potential financial problems and report risks to the management immediately. This comprehensive decision support approach allows companies to gain competitive advantage in competitive markets, promote sustainable development, and eventually build a solid foundation for long-term success.

5. Empirical analysis and verification of financial performance prediction model

Since it is the most important condition variable, it is important to carefully analyze the transition of net income. This financial index is the basis for the DuPont's analytical framework and provides a view of the company's profitability. The calculation method divides the net net income of a company by an average value of a pure asset, and the molecule shows the shareholders profit, and the denominator represents the total investment of shareholders. This indicator clearly shows the net profit of the stock, which is an important measure to measure the overall profitability and warehouse management efficiency.

As a relative number, this indicator has good comparability, enabling effective comparison of financial performance of different companies and helping investors and managers analyze the competitive situation within the industry. This characteristic makes the return on equity a key tool for cross industry comparisons, helping to identify potential investment opportunities and strategic directions.

The comprehensive nature of return on equity is strong, and it is a direct reflection of a company's profitability. Through the DuPont analysis method, it is decomposed into sales net profit

margin, total asset turnover rate, and equity multiplier. The interaction between these three factors more comprehensively reveals the company's operating efficiency and financial health. By delving into these factors, we can gain a clearer understanding of the strengths and weaknesses of enterprises in resource allocation, market response, and operational management, providing guidance for further optimization.

The return on equity has a direct impact on the fundamental interests of all stakeholders, especially on shareholder returns. The changes reflect the financial situation of the enterprise and convey the management's emphasis on shareholder value. Conducting in-depth research on it will contribute to corporate governance and provide important references for the sustainable development of enterprises. A comprehensive analysis of the external economic environment, industry characteristics, and internal management of the enterprise can more accurately assess the company's development potential and risks.

Analyzing changes in the return on equity can provide profound insights into the operating performance of listed companies, helping stakeholders to have a more comprehensive understanding of the company's financial situation. This comprehensive understanding will achieve effective risk control and return optimization, providing data support and theoretical basis for investor decision-making, management strategy formulation, and policy making.

The dynamic analysis of financial performance will promote the flexible response of enterprises in complex market environments and enhance their market competitiveness. With the continuous changes in the economic environment, enterprises need to respond quickly based on real-time data, and the continuous monitoring of the return on equity provides real-time feedback to enterprises, helping them find the best adjustment strategy in the market. Strengthen its leading position in the industry and promote its long-term stable development. Building a diversified financial performance prediction model with return on equity as the core will provide enterprises with more comprehensive financial analysis tools, helping them to stand undefeated in the increasingly fierce market competition.

6. Conclusion

The enterprise financial performance prediction model constructed in this article demonstrates significant practicality, although it mainly focuses on predicting the trend of financial performance changes without delving into specific numerical values. Enterprises and their related stakeholders can identify potential future profit or loss situations in advance. This forward-looking approach is particularly crucial for companies facing financial pressure, prompting management to take appropriate measures before problems become apparent, adjust operational strategies in a timely manner, and effectively curb the decline in performance. This helps to reduce potential losses, enhance the adaptability of enterprises in rapidly changing market environments, and highlight the value of this model in practical applications.

It is worth noting that the accuracy of financial performance forecasts is still insufficient. The research results show that both Logit regression models and support vector machine models have a prediction misjudgment rate of about 40%. In contrast, the accuracy of traditional statistical prediction models and modern artificial intelligence models in predicting financial distress is usually as high as 80% to 90%, which clearly indicates that the accuracy of financial performance prediction is much lower than that of financial distress prediction.

The reasons for the high misjudgment rate can be attributed to two main aspects. The difference in financial indicators between companies with good financial performance and those with poor performance is relatively small, and this significant difference may lead to difficulties in pattern recognition and classification for the model. When using factor analysis to eliminate

multicollinearity between indicators, the extracted three common factors can only explain 67.65% of the original variable information, and about 35% of the information is missing, which affects the accurate description of the dependent variable and thus affects the predictive ability of the model. Future research should focus on optimizing the construction methods of models, ensuring the scientific and comprehensive selection of indicators, exploring richer datasets, and improving the accuracy and reliability of financial performance forecasting. We hope to provide more practical financial performance analysis tools for enterprises.

References

[1] Liu X, Li J, Jia R.A Study on the Prediction Model of Traumatic Hemorrhagic Shock Based on Machine Learning Algorithm[C]//International Conference on Logistics, Informatics and Service Sciences.Springer, Singapore, 2024.DOI:10.1007/978-981-97-4045-1_18.

[2] Zhang X, Xu H, She Y,et al.Improving the prediction performance of leaf water content by coupling multi-source data with machine learning in rice (Oryza sativaL.)[J].Plant Methods, 2024, 20(1).DOI:10.1186/s13007-024-01168-5.

[3] Xi C, Kaoru H, Zhiyang Y J.A model fusion method based on multi-source heterogeneous data for stock trading signal prediction[J].Soft computing: A fusion of foundations, methodologies and applications, 2023, 27(10):6587-6611.

[4] Nistane V.Optimum prediction model of remaining useful life for rolling element bearing based on integrating optimize health indicator (OHI) and machine learning algorithm[J].World journal of engineering, 2024, 21(1): 170-185.

[5] Imran M, Kuznetsov V, Paparrigopoulos P,et al. Evaluation and Implementation of Various Persistent Storage Options for CMSWEB Services in Kubernetes Infrastructure at CERN[J].Journal of Physics: Conference Series, 2023, 2438.DOI:10.1088/1742-6596/2438/1/012035.