Forecasting Financial Performance: A Comparative Study of Machine Learning Models in Accounting

Ze Tang

Renmin University of China, Beijing, 100872, China

Email: 3800679351@qq.com

Abstract. This paper explores the application of different machine learning models in financial performance prediction and makes a comparative study. Machine learning models play an important role in financial performance prediction, and the commonly used models include linear regression, decision tree, random forest, support vector machines, and neural networks. Linear regression is suitable for handling linear relationships, but has limited effect for predicting complex nonlinear patterns. Decision trees and random forests are able to handle nonlinear relationships and are robust to feature selection and data incompleteness. Support vector machines perform well in handling high-dimensional data and non-linear patterns, while neural networks have the ability to process complex patterns and large-scale data. The choice of suitable model depends on the dataset, prediction target and business scenario, and to evaluate the model performance, accuracy, recall and F1 score. With the development of technology and the enrichment of data, we can expect more advanced models to apply in the field of accounting and finance in the future. To sum up, machine learning models provide a powerful tool for financial performance prediction, but in practical application, appropriate models need to be selected according to the requirements and data characteristics.

Keywords: Financial performance; accounting; machine learning model

1 Introduction

With the rapid development of science and technology, machine learning technology has entered every corner of the society, among which, the field of accounting is no exception. Accounting, as a knowledge of managing enterprise finance, its importance is self-evident. In accounting, financial performance prediction occupies a pivotal position. It can help the enterprise executives to develop long-term strategic planning and provide the basis for the enterprise's decision-making. Therefore, it is of great significance to improve the accuracy and efficiency of financial performance prediction for modern enterprises.

In recent years, with the rise of big data and artificial intelligence, the application of machine learning models in financial performance prediction has become increasingly extensive. Compared with traditional statistical methods, machine learning

Z. Wang et al. (eds.), *Proceedings of the 4th International Conference on Economic Development and Business Culture (ICEDBC 2024)*, Advances in Economics, Business and Management Research 299,
https://doi.org/10.2991/978-94-6463-538-6_33

models can handle more complex and changeable data relationships, and automatically extract useful information, so as to provide enterprises with more accurate and timely prediction results[1].

Among the numerous machine learning models, models such as linear regression, decision tree, random forest, support vector machine and neural network are commonly used choices. These models have their own characteristics and are suitable for different scenarios and data types. For example, linear regression models perform well with data with distinct linear relationships, while decision tree and random forest models are better at handling data with complex nonlinear relationships. Support vector machine and neural network models have stronger generalization ability and can deal with more complex data patterns.

In financial performance prediction, we can choose suitable machine learning models for modeling according to different needs and data characteristics. For example, we can use the linear regression model or the neural network model, and the decision tree and the random forest model may be more applicable.

But it's important to note that machine learning models are not everything. In the application process, we also need to combine the theoretical knowledge and practical experience of accounting to adjust and optimize the model reasonably. At the same time, we also need to interpret and apply the prediction results of the model prudently, so as to avoid blind relying on the model.

In addition, with the continuous development of technology, new machine learning models and methods continue to emerge. We should keep learning and exploring new technologies, constantly updating and perfecting our prediction models to adapt to the increasingly complex and changeable financial environment.

In conclusion, the application of machine learning models in financial performance prediction has become a trend. We should give full play to its advantages to provide more solid and efficient theoretical support for accounting research. At the same time, we should also maintain a prudent and critical attitude towards the model to ensure that it can maximize its effectiveness in practice.

2 Overview of the Machine-learning Model

Machine learning plays an increasingly important role in financial performance prediction. According to the different learning methods, machine learning models can be roughly divided into three categories: supervised learning, unsupervised learning and semi-supervised learning[2]. In the practical application of financial performance prediction, supervised learning models such as linear regression, decision tree, random forest can predict future financial performance by training data with known labels. Unsupervised learning models, such as cluster analysis, can be used to discover potential patterns and structures in the data, thereby assisting in the analysis of the financial situation. The semi-supervised learning combines the characteristics of supervised learning and unsupervised learning, using both known labeled data and unlabeled data for training. In addition, complex models such as support vector machine (SVM) and neural network are also widely used in financial performance prediction,

which can deal with more complex and non-linear data relationships and improve the accuracy of prediction[2].

3 Linear Regression

Linear regression, as a classical statistical method, is widely used in data analysis and prediction in various fields. In the context of financial performance prediction, the linear regression model plays a pivotal role. The model is based on the underlying assumption that there is a linear relationship between the dependent variable (usually what we want to predict, such as the company's future financial performance) and the independent variable (usually historical financial data or other relevant indicators).

In the financial field, the application of linear regression model is mainly reflected in the analysis of corporate financial data. For example, by collecting the financial data of the company in the past few years, such as operating income, net profit, total assets, etc., we can use the linear regression model to predict the company's future financial performance. The advantage of this approach is that it is simple and intuitive, and easy to understand and operate.

However, the linear regression models also have some limitations. The most obvious limitation is that it can only deal with linear relationships. In the real world, the relationships between many economic phenomena and financial data are often nonlinear. For example, when the size of a company reaches a certain level, its growth rate may slow, and this nonlinear relationship is difficult to accurately describe by linear regression models.

Moreover, the linear regression model also assumes that the data are independent of each other and that the error terms follow a normal distribution. If these assumptions do not hold, the predictive effects of the model may be severely affected. For example, if there is an autocorrelation between the data (i. e., the data of a period are related to the data of the previous period), then the predictions of the linear regression model may be biased.

To overcome these limitations of the linear regression models, the researchers propose many improvements. For example, polynomial regression can capture the nonlinear relationship by introducing high order terms[3]; ridge regression and principal component regression can reduce the multicollinearity problem of the model[4]; the generalized linear model can relax the assumption that the error terms obey normal distribution and make the model more flexible and applicable.

In conclusion, linear regression, as a simple and commonly used prediction model, has some application value in financial performance prediction. However, due to its limitations, we need to select and adjust the specific situation in the practical application. Through continuous exploration and innovation, we believe that there will be more complete and efficient models in the future, to help us better predict and analyze financial performance[5].

4 Decision Tree with a Random Forest

In this data-driven era, the financial sector is increasingly dependent on data. In order to more accurately predict the financial performance of enterprises, financial analysts and data scientists constantly explore various algorithms and technologies. Decision trees and random forests, as two powerful machine learning algorithms, are playing an increasingly important role in financial performance prediction.

The decision tree, as the name suggests, is a tree-like structure model that simulates the human decision-making process. It gradually builds the model that can predict the target variable by recursively partitioning the data. Each node of the decision tree represents a judgment condition on a feature attribute, each branch represents a possible attribute value, and each leaf node represents a prediction result. This intuitive structure makes the decision tree easy to understand and interpret, especially in the financial sector, where it can help analysts quickly identify the key factors that affect financial performance[6].

However, a single decision tree model may be affected by overfitting, where the model performs well on training data but poorly on new data. To solve this problem, the random forest algorithm has emerged. Random forests are actually an ensemble learning model composed of multiple decision trees. In random forests, each decision tree is trained on randomly selected samples and feature subsets, and this randomness increases the diversity of the model and thus reduces the risk of overfitting.

The prediction results of random forests are usually obtained by averaging or voting on the predictions of multiple decision trees, which further improves the robustness and accuracy of the model. In the financial sector, random forest can be used to predict the profitability, credit risk, stock price and many other aspects. By building a random forest model, analysts can have a more comprehensive understanding of the company's financial position and provide a more reliable basis for investment decisions.

In addition to prediction accuracy, decision trees and random forests have an important advantage: their ability to handle nonlinear relationships. In the financial data, the relationship between many variables is often not linear, which makes it difficult to capture accurately in the traditional linear regression models. Decision trees and random forests are able to handle this nonlinear relationship well to predict financial performance more accurately.

Moreover, these two models are also relatively robust to feature selection and the incompleteness of the data. In the financial sector, the quality of data is often uneven, sometimes even with a lot of missing values. Decision trees and random forests are able to deal with these issues to some extent, allowing analysts to make effective analysis and predictions on incomplete datasets.

In conclusion, the decision tree and the random forest play an important role in predicting financial performance in the financial field. They not only deal with nonlinear relationships and data completeness, but also provide intuitive and easily understood prediction results. With the continuous progress of technology and the increasing enrichment of data, the application prospect of these two algorithms in the financial field will be even broader.

5 Support Vector Machine SVM

Support vector machine (SVM) is a classifier based on statistical learning theory. Its core idea is to find a hyperplane that can maximize the classification interval, so as to classify the data. This approach is unique in that it does not only focus on current classification results, but also strives to ensure that the boundaries between classifications are as broad as possible, thus effectively avoiding the problems of overfitting and underfitting[7].

In the financial field, financial performance prediction is a crucial task. By analyzing historical data, companies can predict their future financial performance and thus make more informed decisions. And SVM is a great approach for this task. It can handle high-dimensional data and maintain good classification effect even in the case of high data dimension. In addition, the SVM is able to handle non-linear patterns, which means that it can capture the complex relationships in the data to give more accurate predictions.

Of course, any method has its applicable scope and limitations. The SVM, though powerful, is not a panacea either. In practice, appropriate models and methods need to be selected according to specific data features and prediction targets. Meanwhile, proper data preprocessing and feature selection are also needed to improve the prediction accuracy and stability of the model[8].

However, with the continuous progress of technology and the deepening of research, the application prospect of machine learning methods such as SVM will be broader in the financial field. With the continuous development of big data and artificial intelligence technologies, we have reason to believe that future financial performance forecasting will be more accurate, efficient and intelligent.

In conclusion, the support vector machine (SVM), as an advanced machine learning method, has an important application value in financial performance prediction. Through its strong classification and forecasting ability, we can more accurately predict the future financial performance, so as to provide strong support for the development of the enterprise. At the same time, we also need to continue to explore and study to explore more possibilities and potentials and promote the sustainable development of the financial sector.

6 Nerve Net: Neural Net

With the rapid development of science and technology, artificial intelligence and machine learning technology have gradually penetrated into all aspects of our lives. Among them, neural network, as the core component of artificial intelligence, has also made remarkable achievements in many fields. In the financial field, especially in the financial performance prediction, the application of neural network technology is gradually showing its strong potential and value.

Neural networks, the term is derived from the simulation of the neuronal structure in the human brain. Just as neurons in the human brain transmit and process information through electrochemical signals, neural networks also transmit and learn data

through layers of neuronal nodes. This unique structure allows neural networks to handle very complex and nonlinear data relationships, which are difficult areas for conventional methods.

When we talk about financial performance forecasts, we are actually talking about how to predict our future financial results based on historical data, market trends, corporate strategies, and many other factors. This prediction involves not only a large amount of data, but also the complex relationships and non-linear patterns of the data. Neural networks are a powerful tool that can help us identify hidden patterns and trends from these data[9].

In practical applications, neural networks can automatically generate prediction models by learning from large amounts of data[10]. This model can not only process the data quickly, but also constantly adjust itself according to the new data, so as to improve the accuracy of the prediction. In addition, neural networks can also process a variety of data types, such as text, images, etc., which allows us to incorporate more information into the prediction model and further improve the accuracy of prediction.

However, neural networks are not a panacea. The prediction results are still influenced by various factors such as data quality, model selection and training methods. Therefore, in practical applications, we need to carefully choose the data, rationally design the model, and adopt scientific training methods to ensure the reliability of the prediction results.

In conclusion, neural networks have great potential and value in financial performance prediction. By simulating the neuronal structure of the human brain, the neural networks are able to handle complex and nonlinear data relationships, which can provide more accurate prediction results. But at the same time, we also need to recognize its limitations and maximize its advantages through scientific methods and technologies. With the continuous progress of technology, it is believed that the application of neural network in the financial field will be more extensive and deep.

7 Conclusions

In the field of financial performance prediction, the application of machine learning models has become increasingly popular, and the logic and principles behind it have gradually become well known in the industry. Different machine learning models, such as linear regression, decision tree, support vector machine, neural network, etc., all show their own characteristics and advantages in predicting the financial performance of companies.

The linear regression model is simple and intuitive, and can give a clear explanation for the relationship between continuous variables and dependent variables. However, it has limited ability to handle nonlinear relationships, which may become a problem in financial performance prediction. Decision tree models are good at handling classification problems and are able to generate easily understood decision rules, but their prediction accuracy may be affected by data noise and overfitting.

SVMs perform well with high-dimensional data and can handle nonlinear relationships. But it may not be efficient when handling large-scale datasets, and parameter

tuning is also a challenge. Neural networks, especially deep learning models, have powerful capabilities in handling complex nonlinear relationships, but their "black box" nature makes the interpretation of results difficult.

In practical application, we need to select the appropriate model according to the specific requirements and data characteristics. For example, if we aim to generate understandable prediction rules, decision trees may be a good choice, and if we pursue higher prediction accuracy, we may need to consider using neural networks.

In the future, with the development of technology and the enrichment of data, we can expect more advanced machine learning models in the field of accounting and finance. These models may include complex network structures based on deep learning, integrated learning methods, etc. At the same time, as data quality and quantity improve, there is reason to believe that the accuracy of machine learning in financial performance prediction will further improve.

In conclusion, the application of machine learning models in financial performance prediction has broad prospects and potential. We need to constantly explore and practice to maximize the strengths of these models and overcome their limitations.

References

1. Wu Wenrui, Wang Jinlong, Sun Lei. Research on Resource Sharing Pathways of Accounting Professional Practice Course [J]. Foreign trade, 2024, (02): 158-160.
2. Huang Xiaobing, Chen Lijuan. Accounting teaching reform of "four in one and three paragraphs of integration" [J]. Metallurgical Education of China, 2023, (06): 30-33. DOI:10.16312/j.cnki.cn11-3775/g4.2023.06.027.
3. Zhang Jing, Liu Lina, Zheng Luhang, et al. —— Take the "accounting" course of economics and management major as an example [J]. Education and Teaching Forum, 2023, (51): 141-144.
4. Huang Hongying. Research on the talent training ability of "cross-border integration" of accounting major under the background of four-chain integration [J]. Time-honored Brand Marketing, 2023, (20): 154-157.
5. Feng Xiaoqing, Xu Wenjing. Research on the Integration of Green and low-carbon Development and Accounting major under the background of "Double First-class" construction [J]. Business Accounting, 2023, (20): 114-118.
6. Zhang Yue. Mixed teaching research and practice of Financial Accounting based on the 345 model [J]. Accounting communication, 2023, (09): 164-171.
7. Liu Ziyi, Dong Birong, Wang Rong. Evaluation on the effect of ideological and political Construction —— Take the "Accounting" course of Nanjing Audit University as an example [J]. Accounting communication, 2022, (22): 42-46 + 71.
8. Wen Subin, Dong Birong, Zhang Xingliang. —— Take Nanjing Audit University as an example [J]. Accounting communication, 2022, (20): 27-31 + 43.
9. Zhang Yinghui, Xiao Wenjuan. Equity incentive, contract heterogeneity and financial performance [J]. Journal of Xi'an Petroleum University (Social Science Edition), 2024,33 (01): 65-73 + 88.
10. Zhang Hao, Xiong Lin. Research on the financial performance and influencing factors of listed companies in the pharmaceutical manufacturing industry [J]. Friends of the Accounting, 2024, (03): 66-73.