Research Article

The Construction of Corporate Financial Management Risk Model Based on XGBoost Algorithm

Rongyuan Qin

School of Economics and Management, Hanjiang Normal University, Hubei Shiyan 442500, China

Correspondence should be addressed to Rongyuan Qin; qinrongyuan@hjnu.edu.cn

Received 22 February 2022; Revised 16 March 2022; Accepted 18 March 2022; Published 13 April 2022

Academic Editor: Miaochao Chen

Corporate financial management is a tedious task, and it is a complicated thing to rely solely on the human resources of financial personnel to manage. With the continuous development of intelligent algorithms and machine learning algorithms, new ideas have been brought to enterprise financial risk assessment. This method will not only save a lot of financial and material resources but also improve the accuracy of enterprise financial risk assessment. Compared with machine learning algorithms such as random forests and support vector machines, the extreme gradient boosting (XGBoost) algorithm is more widely used, and it has unique advantages in terms of speed and accuracy. This study selects the XGBoost learning algorithm to predict the risk assessment in corporate finance. In this study, the data preprocessing method is used to preprocess and classify the enterprise financial data source effectively, and then the XGBoost algorithm is used to assess the risk of enterprise financial data, and finally a set of enterprise financial risk assessment model is established. The research results show that the XGBoost model selected in this paper has high reliability in predicting the financial risk assessment of enterprises, and the prediction errors are all within 3%. The largest forecast error is only 2.68%, which comes from the profit and loss of the enterprise’s financial situation. The smallest error is only 0.56%, which is a trustworthy enough error for corporate financial forecasting. There is a high correlation between the type of enterprise financial risk assessment and the actual type of risk. At the same time, this paper also has a good dependence on the preprocessing method of enterprise financial data.

1. Introduction

The operating status and development trend of the company's financial affairs are important steps in the company development. With the development of economic globalization and the development of e-commerce, the company's financial status is particularly important, which is related to the company development and future [1–3]. Traditional corporate finance only relies on financial managers to use professional knowledge to deal with tedious corporate finance, which is not only inefficient but also has a high error rate. At the same time, traditional corporate financial management methods cannot effectively predict the future financial development of the company, which limits the formulation of the company business strategy. Similarly, with the continuous development of random economy, a company business is not limited to its own financial affairs, and it may face international business or even e-commerce business, which will generate all kinds of cumbersome financial data. These company financial data are important guarantees for the company risk assessment [4, 5]. A company's financial risk assessment is extremely important to a company's long-term development. If a reasonable forecast and risk assessment can be made based on the company's financial development, it will be extremely beneficial to the company's financial and long-term operations. Therefore, using the company daily financial information to reasonably assess the company development risk is extremely meaningful work and research. In recent years, with the continuous development of machine learning algorithms and the continuous advancement of computer hardware equipment, algorithms capable of processing highly nonlinear and high-dimensional data have emerged, which is extremely beneficial for the intelligent risk
Many economic researchers have done a good job of researching and evaluating corporate finance and risk levels, and have produced many valuable studies. These studies will have a good guiding significance for the operation of the company finance and the assessment of the risk level. For the credit risk in supply chain finance business, Zhang et al. [12] established a supply chain risk assessment model using the KMV model and copula function and effectively quantified credit risk. The conclusion shows that this risk assessment model can well predict the risk pollution in the supply chain, which is a meaningful study for commercial banks. Aisaiti et al. [13] conducted effective assessments of financial operations and risks for Chinese rural farmers. They used the knowledge of dependent variables and intermediary variables to effectively evaluate the profit perception and risk perception of financial business. The results of this study suggest that the reinforcement factor of the model affects the positive correlation between perceived benefit and perceived risk assessment. This is valuable for risk assessment of rural financial business. Xu et al. [14] established a complex grid evaluation model with the risk evaluation task of Internet finance as the research object. They divided the Internet financial network into subsystem models such as sub-supervision financial network and supervision subnetwork and studied the propagation relationship between different risk factors. They came to the conclusion that the central point of the entire Internet is the main factor of risk propagation, which is valuable for the development of Internet finance. [15] conducted effective analysis and forecasting of risk assessment tasks in trade credit financing (TCF) and credit guarantee (PCG) financing. They proposed a conditional value-at-risk system and a least squares method to assess the risk attitude of decision makers. The results show that the equilibrium assessment risk model and the initial investment cost can affect the risk attitude of retailers. Wang et al. [16] mainly conducted sufficient research on the problems of model lag and high information risk level between banks and enterprises in supply chain finance business. They used the Internet of Things technology and deposit and loan financing model to establish a supply chain risk assessment model. The results show that the risk assessment model established in this paper based on the Internet of Things technology is beneficial to reduce the operational risk of enterprises. At the same time, this model has certain reference significance for enterprises to reduce financial crisis. Zhu and Hua [17] believe that the current Internet financial industry has had a certain impact on the development of banks and the macroeconomy, and the Internet financial industry has certain risks. They used the systemic contingent claims analysis (SCCA) model to assess and forecast financial risks in the banking industry. Through this model, they assessed the impact of the Internet on the banking industry and concluded that the risk of the banking industry will continue to rise in the future. Wulandari et al. [18] assessed production risks based on the correlation between financial production risks and financing providers in horticultural farms. They introduced parameters such as coefficient of variation, skewness, and kurtosis to measure the risk relationship between bank commercial loans and production. Their findings could help farmers fully understand the risk relationship between financing and production. Qi et al. [19] used the random forest (RF) method to analyze and classify the financial risks in Internet finance and used the BP neural network method to effectively predict the risks. Wang and Liu [20] proposed an innovative financial risk prediction model based on the Internet of Things technology and the method of backpropagation neural network (BPNN). The conclusion shows that the risk assessment model based on IoT technology and neural network technology has high accuracy.

The risk assessment of corporate finance is very important to the long-term development of a company. At present, a large number of researchers have carried out many studies on the risk level of corporate finance and the prediction of corporate finance [21]. From the above literature review, it can be seen that there has been a lot of research on the use of mathematical models for corporate financial data, but there are relatively few studies related to risk assessment, and there are even fewer models involving machine learning algorithms [22, 23]. Based on the XGBoost model, this paper performs an efficient classification task for corporate financial data and a risk assessment for corporate financial data. From the perspective of time and cost, it has certain value and significance for the company’s financial forecast and risk assessment.

The goal of this study is to effectively classify the company financial data, and this paper needs to complete the company risk level assessment based on the company’s financial classification information. This paper is organized as follows. Section 1 mainly introduces the development status of company finance and company risk assessment. Section 2 introduces the connection and necessity of company finance, risk assessment, and machine learning algorithms. Section 3 mainly introduces the XGBoost algorithm and classification method. Section 4 illustrates the feasibility and accuracy of the XGBoost algorithm in the classification and risk prediction of corporate financial data. This research mainly uses some statistical parameters to predict the factors of XGBoost algorithm in predicting enterprise financial management, such as enterprise profit and loss, enterprise loan, and...
other factors. These statistical parameters are mainly prediction error, classification ratio, and prediction hotspot distribution map. The last part is the summary of the article.

2. The Importance of XGBoost Algorithm for Company’s Financial Risk Assessment and Financial Risk Prediction Data Source

2.1. The Significance of XGBoost Algorithm for Financial Risk Prediction of Companies. Machine learning algorithms can well map the nonlinear relationship between company finances and financial risks, and it is a way that traditional financial methods cannot realize [24, 25]. The company financial data include company sales performance, employee performance, bank loans, and other data. There is a complex relationship between these data, but these data will also affect the company development, and it can also provide a good risk assessment for the company [26]. If the company’s financial risk assessment can be done well, it can help the company avoid some financial risks, thereby ensuring the company’s long-term development and normal operation [27–29]. The traditional financial method is that the financial personnel summarize and analyze the company’s financial affairs through bookkeeping and other methods and then make a rough prediction of the company risk level according to the development trend of financial data. Such a forecast method is difficult to guarantee the accuracy of the company risk level. Financial personnel often rely on financial analysis software such as Excel and financial expertise to analyze the company’s financial integration, which is a very difficult task for the discovery of nonlinear and complex relationships between company financials. A machine learning algorithm is an algorithm that specializes in dealing with the relationship between complex data [19]. It can not only complete the classification of financial data but also complete the regression prediction of the company financial data. Also, these machine learning algorithms save time and financial resources. XGBoost is a mature algorithm with relatively high prediction accuracy in machine learning algorithms. Compared with decision trees and support vector machines, it is also a good embodiment of the idea of integrated learning [30]. The XGBoost algorithm can not only effectively classify the company financial data but also map the relationship between the company financial data and the company risk level, so as to complete the risk assessment task that cannot be completed by financial personnel.

2.2. The Datasets Required for a Company’s Financial Risk Assessment. The main purpose of this study is to classify the types of company financial data, and it will use the XGBoost algorithm to complete the company risk level assessment based on these company financial data. This study selects the financial data of a large company to conduct research and analysis on the risk assessment model studied in this paper. According to the research purpose of this paper, first of all, five financial data of the company profit and loss data, bank loans, employee performance, e-commerce profit and loss, and cross-border business are used as the data source of the research [31]. There are also big differences in the magnitude and type of these data. It is difficult to assess the company risk level only by relying on the professional knowledge of financial personnel based on these five types of financial data. These five types of company financial data need to be preprocessed and classified before they can be used as input data for the company risk level assessment model. The XGBoost algorithm is a processing algorithm that uses a multipath decision tree to classify and regress data sources. In this study, the company financial data are input into the XGBoost algorithm as a whole data source, and the classification of five types of company finance is completed through this algorithm. The algorithm then completes the task of mapping the company financial data to the company risk level. Finally, this paper will predict the company risk level based on these five types of company financial data. Once the XGBoost model is trained, only five types of corporate financial data are provided to predict corporate financial risk levels and risk trends.

3. The Introduction to Company Financial Data Classification and Risk Assessment Prediction Algorithm and Data Preprocessing

3.1. The Introduction to XGBoost Algorithm. The full name of XGBoost algorithm is extreme gradient boosting, it is often used in some competitions and practical engineering applications, and its classification and regression effects are remarkable. It is one of the tools for massively parallel boosted trees, and it is mainly one of the schemes of multipath decision trees. It can be used for classification tasks with complex data, and it can also be used for regression tasks with complex data. Also, it has a higher level of accuracy and robustness compared to other machine learning algorithms. Figure 1 shows the workflow of the XGBoost model adopted in this study, which is mainly divided into two workflows: classification and prediction. In the first step, this research needs to preprocess the five types of data sources collected in the company’s finances in terms of magnitude and type. In the second step, these five types of company financial data need to be classified by the XGBoost algorithm. After these company financial data are accurately classified by the XGBoost model, the classified company financial data will be used to predict the company’s risk level and development trend through the regression performance of XGBoost. In this regression prediction process, the input and output label data of the given prediction model are required. This paper will select five types of company financial data as the input of the XGBoost prediction model and use the company’s financial risk trend and level as the output of the model. Once this model is trained, some optimal weights and biases suitable for predicting the company’s financial risk will be obtained. When the XGBoost model is trained, this paper will select some financial data of companies that did not participate in the training for prediction to test the accuracy and feasibility of the model. In general, the XGBoost on the left
is used for the classification task of enterprise management data, which is the first step of this model. The XGBoost on the right is used for the prediction task of corporate financial data. The input of this XGBoost comes from the classified data.

3.2. The Description of the XGBoost Algorithm and Introduction to Regression Prediction. The development of machine learning algorithms is limited by the development of computer hardware. In the early stage of machine learning algorithms, some excellent algorithms such as decision trees, random forests, and support vector machines emerged. These algorithms each have their own strengths and weaknesses, and they also have their own strengths and weaknesses for different problems. Ensemble learning is a new learning idea that fully combines the advantages of these algorithms and discards their shortcomings. Likewise, these algorithms are more suitable for classification and regression tasks. The XGBoost selected in this paper is also a kind of machine learning algorithm, which is mainly used in various large-scale competition projects and practical engineering applications, which shows that this model has better accuracy than other machine learning algorithms. The GBDT algorithm only uses the first-order derivative calculation in the loss function, while the loss function of the XGBoost algorithm uses the second-order derivative operation, and a regularization term is added to the objective solution function, which can avoid overfitting. The combined phenomenon is compared to the GBDT algorithm. Compared with the GBDT algorithm, the XGBoost algorithm well matches the computational complexity of the algorithm and the occurrence of overfitting. The decision tree method is one of the basic algorithms of XGBoost. The decision tree continuously divides the data to be classified or returned according to the number of nodes. It is only divided in the case of a tree. The XGBoost method is a bunch of decision trees, and the comprehensive results of these decision trees are output as the final predicted value.

Figure 2 shows the application process of the XGBoost method in the company’s financial risk assessment. Figure 2 is only to illustrate the application process of XGBoost in the classification of corporate financial data, so only 3 types of corporate financial data are listed here, and the branch of the decision tree used in this paper is 5. Similar to the decision number method, it first divides the original data into different nodes according to certain weights. Different from the decision tree method, the decision tree only relies on a classification or regression principle to divide nodes, and it is only divided in the case of a tree. But XGBoost is an integrated idea, which divides the decision tree into multiple trees according to different division principles and finally combines these divided nodes to output the predicted value. This XGBoost method can improve the prediction accuracy compared to a single machine learning algorithm, but the computational complexity of this method will be greatly improved. There is a complex relationship between company financial data and company risk, and only relying on one classification principle will result in inaccurate results. Therefore, this paper chooses the classification and risk prediction method of XGBoost. Corporate financial data are often huge, so deep learning methods are not suitable for the task of predicting corporate financial data risks. Therefore, the deep learning strategy requires the training and learning of large computer systems. Considering the time cost and material cost as well as the accuracy of the model, XGBoost is a more suitable task for the company’s financial risk prediction.

XGBoost is similar to the neural network method, which also requires nonlinear operations of weights and biases, and these calculated values need to be summed. Equation (1) represents the predicted value of a company’s financial risk assessment.

\[
\hat{y}_j = \sum \omega_j x_{jk}.
\]  

Equation (2) represents the loss function of the XGBoost model, which can be selected according to different research objects. This paper chooses the form of the mean squared error.
Financial data
source

Final-Class

Figure 2: The prediction process of XGBoost algorithm in company’s financial risk assessment.

difference shown in the equation (2) as the loss function for financial risk forecasting.

\[ l(y, \hat{y}) = (y - \hat{y})^2. \]  (2)

The solution of XGBoost is actually a process of minimizing the loss function, that is, the process of finding the minimum error between the true value and the predicted average value. Equation (3) shows the process of minimizing the loss function.

\[ F^*(\mathbf{x}) = \arg \min_{F(x,y)} [L(y, F(\mathbf{x}))]. \]  (3)

From the previous description, we can understand that the core of the XGBoost algorithm is the idea of an integrated algorithm. Equation (4) shows the process of the decision tree branch ensemble operation. It solves and averages the predicted values of multiple decision trees.

\[ \hat{y}_j = \sum_{j=1}^{K} f_k(x_j), \quad f_k \in \zeta. \]  (4)

The core of XGBoost is the integration idea, that is, each tree is continuously added to it to obtain the best boosting effect. Equation (5) shows the operation method of the first decision tree.

\[ \hat{y}_j = f_1(x_j) = \hat{y}_j^{(0)} + f_1(x_j). \]  (5)

Equation (6) shows the integration process of the second tree, and it can be seen that it is integrated on the result of the first decision tree operation.

\[ \hat{y}_j = f_1(x_j) + f_2(x_j) = \hat{y}_j^{(1)} + f_2(x_j). \]  (6)

Through equations (5) and (6), the operation of XGBoost in the integration process can be understood in detail, and \( n \) decision trees are involved in the XGBoost operation process. Equation (7) shows the integration process in the K-tree case.

\[ \hat{y}_j = \sum_{j=1}^{K} f_k(x_j) = \hat{y}_j^{(t-1)} + f_t(x_j). \]  (7)

It is similar to the neural network method, and the XGBoost method is that with the increasing number of learning samples, the model is prone to overfitting, which can easily lead to poor accuracy in the test set. Equations (8) and (9) add a penalty function to the original basis to limit the overfitting of the loss function.

\[ \Omega(f_t) = \frac{1}{2} \sum_{j=1}^{T} \omega_j^2, \]  (8)

\[ \text{Obj} = -\frac{1}{2} \sum_{j=1}^{T} G_{j}^2 + \lambda + \zeta T. \]  (9)

The process of minimizing the loss function requires derivation of the loss function, which involves many derivative operations. Equations (10) and (11) show the derivative rules in minimizing the loss function.

\[ G_j = \sum_{j \in I_j} g_j, \]  (10)

\[ H_j = \sum_{k \in I_j} h_k. \]  (11)

After the derivative is solved by equations (10) and (11), it needs to be brought back to the loss function to solve the target value. Equation (12) shows this process.

\[ h_k = \partial_{\hat{y}}^{2} l(y_k, \hat{y}^{(t-1)}). \]  (12)

3.3. The Preprocessing Operations on Company Financial Data. Through Section 2.2, it can be clearly understood that this paper selects the company’s financial data for the company’s risk assessment and prediction. At the same time, data such as the company’s profit and loss status, employee performance, and bank loans were selected as the learning data for this study. It can be seen that these data have certain differences in magnitude and type, and these data may also have vacancies or outliers, which requires preprocessing of these data. This article will use standard processing methods to normalize the financial data of these companies, and the financial data will be processed into normally distributed data. After the collected financial data are preprocessed, they will be entered into the XGBoost model in matrix form for classification and prediction tasks. This research mainly predicts five factors, such as profit and loss, loan situation, and employee performance in enterprise financial management. Therefore, the number of classifications (5) and the learning factor (0.001) were chosen in this study. The decision tree branch is selected as 3.
4. The Accuracy and Feasibility Analysis of Corporate Financial Data Classification and Risk Assessment Forecasts

This paper includes two processes for the risk assessment and prediction of the company’s financial data: the classification task of financial data types and the prediction of risk assessment levels. Figure 3 shows the errors in the classification of company financial data by the XGBoost method. In general, the XGBoost classification algorithm is suitable for classification tasks of corporate financial data types. It can be clearly seen that the classification errors are all within 3%, which is an acceptable error range for corporate financial data classification tasks. The largest error is 2.48%, and the smallest error is 0.56%. The classification errors of the other three financial data types are within these two error ranges. The biggest error comes from the company’s profit and loss status. The reason for the relatively large error may be that the changes in the company’s products are changeable, and the complexity between them is more complicated than the other four types of financial data. The smallest error mainly comes from the employee’s performance factor because this part of the financial data is relatively stable and the changes are small. We can improve the classification error of this part by increasing the number of samples of this part of financial data through the XGBoost algorithm. Figure 4 more intuitively shows the error heat distribution map of the predicted and actual values of the financial profit and loss status data of the enterprise. It can be seen from Figure 4 that the classification error distribution is relatively uniform and the errors are all within 2.5%. At the same time, it is clear from Figure 4 that the difference in error distribution is relatively small for different types of corporate financial data. From the above description, we can see that XGBoost can classify the company financial data effectively, which is helpful to evaluate the performance of the company’s financial risk level.

After the company’s financial data are classified and processed by the XGBoost algorithm, these data will be used again in the form of the input layer to use the XGBoost algorithm to predict the company’s financial risk level. Figure 5 shows the predicted value of the company’s financial risk level over time. It can be clearly seen from Figure 5 that the predicted value of the risk level is in good agreement with the actual risk level, and the trend of financial data risk level changes over time is also consistent with the actual change trend. The actual risk level changes are in good agreement. Through these two points, it can be shown that the XGBoost model is suitable for the prediction of the company’s financial risk level, and the accuracy rate is relatively high. Figure 6 shows how the predicted value of the company’s financial risk level deviates from the mean and minimum value. It can be seen that the deviation distance between the predicted value of risk level and the mean value at each moment is relatively close, which shows that the weights of XGBoost are relatively similar, and there is no serious deviation, which proves that XGBoost is suitable for the evaluation task of the company’s financial risk level.

![Figure 3: The classification results of company financial data through the XGBoost method.](image1)

![Figure 4: The heat map of classification errors in corporate financial data.](image2)

![Figure 5: The predicted value of the company’s financial risk level over time by XGBoost method.](image3)
level prediction, this paper selects box plots and fan-shaped pie charts to analyze the results. It can be seen from Figure 7 that the predicted value of the risk level over time is in good agreement with the actual risk level value, but the predicted value of the risk level is larger than the actual risk level value, but the overall deviation range is relatively small. Moreover, the predicted risk level values are in good agreement with the actual risk level values in terms of the distribution trend. Figure 8 shows the proportion distribution of the company’s financial data after being classified by the XGBoost model. It can be seen that the distribution of the five types of financial data is relatively uniform, which indicates that the proportion of weights is also relatively uniform. This further illustrates the feasibility of the XGBoost model in the classification of corporate financial data, which also provides a guarantee for the risk assessment of corporate financial data.

5. Summary Part of the Company’s Financial Risk Level Forecast

With the continuous development of economic globalization and the development of various e-commerce businesses, the financial status of a company has become more and more complicated. It is difficult to complete the risk level assessment of the company’s financial affairs by relying solely on financial personnel, and it is difficult to guarantee the accuracy of the risk level prediction. The continuous advancement of machine learning algorithms and computer hardware equipment has made it possible to use intelligent algorithms to assess the company’s financial risk level, which will not only improve the accuracy of risk forecasting but also greatly improve the efficiency of forecasting.

This paper mainly uses the XGBoost algorithm in the machine learning algorithm to effectively classify the company financial data and accurately predict the risk level of these financial data. From the perspective of the company’s financial classification error, the classification errors are all within 3%, and the maximum error is only 2.48%. This shows that the XGBoost model has good accuracy in the classification task of company financial data. This part of the error is mainly from the company. The profit and loss status of this part of the dataset needs to be increased to improve the accuracy. From the perspective of the company’s financial risk assessment, both the change trend of the risk level over time and the value of the risk level are in good agreement with the actual risk level. In terms of time and accuracy, XGBoost has obvious advantages over traditional financial processing methods, and this model is reliable in the task of corporate financial risk assessment.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.
Acknowledgments
This study was supported by 2021 Annual Science Research Foundation Project of Department of Education of Hubei Province of China (Application Research on Identification of Financial Fraud of Listed Companies Based on AI) (project no. Q20213102).

References