Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 8005493, 14 pages
https://doi.org/10.1155/2022/8005493

Research Article

Credit Debt Default Risk Assessment Based on the XGBoost Algorithm: An Empirical Study from China

Jun Wang,[1] Wei Rong,[1] Zhuo Zhang,[2] and Dong Mei[3]

[1]*School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, China*
[2]*School of Economics, Fudan University, Shanghai, China*
[3]*Institute of Agricultural Resources and Regional Planning (IARRP), Chinese Academy of Agricultural Sciences, Beijing, China*

Correspondence should be addressed to Wei Rong; tantorrong@smail.swufe.edu.cn and Dong Mei; meidong199312@163.com

Received 17 January 2022; Accepted 22 February 2022; Published 19 March 2022

Academic Editor: Nawab Muhammad Faseeh Qureshi

The bond market is an important part of China's capital market. However, defaults have become frequent in the bond market in recent years, and consequently, the default risk of Chinese credit bonds has become increasingly prominent. Therefore, the assessment of default risk is particularly important. In this paper, we utilize 31 indicators at the macroeconomic level and the corporate microlevel for the prediction of bond defaults, and we conduct principal component analysis to extract 10 principal components from them. We use the XGBoost algorithm to analyze the importance of variables and assess the credit debt default risk based on the XGBoost prediction model through the calculation of evaluation indicators such as the area under the ROC curve (AUC), accuracy, precision, recall, and F1-score, in order to evaluate the classification prediction effect of the model. Finally, the grid search algorithm and *k*-fold cross-validation are used to optimize the parameters of the XGBoost model and determine the final classification prediction model. Existing research has focused on the selection of bond default risk prediction indicators and the application of XGBoost algorithm in default risk prediction. After optimization of the parameters, the optimized XGBoost algorithm is found to be more accurate than the original algorithm. The grid search and *k*-fold cross-validation algorithms are used to optimize the XGBoost model for predicting the default risk of credit bonds, resulting in higher accuracy of the proposed model. Our research results demonstrate that the optimized XGBoost model has a significantly improved prediction accuracy, compared to the original model, which is beneficial to improving the prediction effect for practical applications.

1. Introduction

China's bond market has continued to develop, and the degree of marketization has continued to increase. Direct financing through the issuance of bonds by companies has become common. In order to enable companies to raise funds, the threshold for bond issuance has been lowered, and bond defaults have gradually emerged over the years. On March 5, 2014, the "11 Chaori Bond" failed to pay the required interest on time, which constituted a default, setting a precedent for the default of listed credit bonds in China. A total of six bonds defaulted during this year. In 2015, "11 Tianwei MTN2" failed to redeem interest on time and constituted a substantial default, which opened the curtain of default by state-owned enterprises. In 2018, "17 Shanghai Huaxin SCP002" failed to repay the principal and interest on schedule, constituting another substantial default and breaking zero defaults on AAA-rated bonds. With the further popularization and development of bond financing, the volume of credit bond issuance in China is on the rise, and the bond industry is growing larger; however, the number of credit debt defaults has also increased. In 2018, China's credit debt default amounted to 120.77 billion, exceeding the total amount of credit debt defaults since 2014. The default amount of credit bonds in 2019 and 2020 was expected to be 150.12 billion and 169.7 billion, respectively. Therefore, the default risk of Chinese credit bonds is worthy of attention.

From the perspective of the ownership of the issuers of default credit bonds, most are private enterprises, which account for a higher proportion than state-owned

enterprises. However, the proportion of state-owned enterprises in default has increased in recent years. From an industry perspective, the industries in which default credit bond issuers are located are mainly concentrated in traditional industries, such as coal, characterized by overcapacity, and industrial machinery, characterized by strong cyclicality. From the perspective of the issuer's ratings and debt ratings, most of the default credit bonds are at a relatively high rating level at the time of issuance. However, some problems gradually appear in the process of enterprise development while, at the same time, the supervision of bond-issuing enterprises is not strict, leading to a high number of bond defaults. Therefore, it is of great significance to accurately predict and measure the default risk of credit bonds. The bond market inherently has information asymmetry. Although there are already open bond risks, such as credit ratings, which reduce information asymmetry between investors and issuers, many bonds still maintain relatively low credit ratings in the event of default. As such, a high level is unable to provide early warning of the risk of default in a timely manner. Therefore, at this stage, information asymmetry reduction through the credit rating mechanism still has certain limitations. Considering the results of relevant domestic and foreign research, the current use of bond issuance data, the issuer's financial indicators and the macroenvironment, and the risk of credit bonds can yield a good forecasting effect. Therefore, in this article, we use the above indicators to predict the default risk of credit bonds.

The aims and objectives of the research study include the following:

(i) We use the principal component analysis method to reduce the dimensions of the selected macro- and microindicators to obtain representative factors, then use the XGBoost algorithm to extract variables that have an important impact on the default risk of credit bonds. The proposed XGBoost model is able to predict the default risk of credit bonds

(ii) Secondly, we use grid search algorithm and k-fold cross-validation to optimize the parameters of the XGBoost model, in order to construct an optimized XGBoost model

(iii) Finally, the original model and the optimized model are compared, through the use of indicators such as the AUC. The optimal default risk prediction model has improved accuracy, compared to the original default risk prediction model

(iv) The improved model can help investors to avoid investment risks, as well as allowing regulators to strengthen credit bond risk management, improve bond market institutions and systems, and provide an important basis for the future development of the bond market. Therefore, the study of more effective credit bond default risk prediction methods not only may further stimulate the vitality of China's bond market but can also help investors to achieve capital appreciation

2. Literature Review

Deng et al. [1] have stated that credit risk can be identified as a default risk. There is a large body of classic literature on credit risk research, both at home and abroad. Kanapickiene and Spicas [2] have constructed a statistical model based on logistic regression for enterprise trade credit risk assessment for small and microenterprises. Coşer et al. [3] have developed several predictive models for loan default risk assessment, using different classifiers including LightGBM, XGBoost, logistic regression, and random forest (RF). Ma et al. [4] have analyzed the development of China's peer-to-peer online lending and the credit risks of borrowers, developed a credit risk assessment indicator system, and used the indicator system to develop a back-propagation neural network to complete the risk assessment model. Chen [5] has stated that, in order to prevent and control government debt risk, we need early warning and early prevention. They also developed a machine-learning-based model for risk assessment consisting of a back-propagation neural network. Combined with the problems studied in this article, we mainly focus on the selection of indicators for predicting bond default risk and research on the XGBoost algorithm.

2.1. Selection of Indicators for Predicting Bond Default Risk. Research into the selection of indicators for predicting bond default risk, both at home and abroad, has mainly been conducted from two aspects: at the macrolevel (i.e., represented by macroeconomic indicators such as GDP and CPI) and the microlevel (i.e., represented by corporate financial data indicators).

First, we consider the selection of macrolevel indicators. Many scholars generally believe that the quality of the macroeconomic environment directly affects the default risk of enterprises. For this reason, scholars have conducted relevant research. Collin-Dufresne et al. [6] have stated that the determinants of credit spreads include corporate factors and market factors, where market factors include such factors as the consumer price index (CPI), producer price index (PPI), and interest rates. Giesecke et al. [7] have found that GDP change is a strong predictor of the default rate, by studying the degree of financial and macroeconomic variables predicting the default rate. Dai and Sun [8] have found that risk-free interest rates are negatively correlated with bond credit spreads, while GDP and M1 are positively correlated with credit spreads. Mili et al. [9] have used the financial data of bond issuers, macroeconomic data, and bond market variables as the basic attributes of primary bonds as explanatory variables, while the bond recovery rate was the explained variable. They found that the recovery rate of domestic bonds in developing countries is more affected by corporate fundamentals, while developed countries are more affected by macroeconomics and bond markets. Hu [10] has pointed out that China's GDP growth rate, industry prosperity, and reverse repurchase interest rates are particularly critical for default predictions, in a study on bond default analysis based on random forests.

Second, we consider the selection of microlevel indicators. Scholars believe that the financial data of the issuer

can truly reflect the company's own operating conditions and profitability, such that the probability of default by the issuer can be judged in the future. Mizen and Tsoukas [11] have used profitability, cash flow, liquidity, financial leverage, solvency, and company size as explanatory variables in an empirical model to predict bond default ratings. Deng et al. [1] have selected 13 financial indicators, such as return on total assets, asset–liability ratio, current ratio, and quick ratio to construct a logistics model for the credit risk prediction of listed companies. Jones et al. [12] have used the data of companies whose credit ratings had changed during 1983–2013 as a sample, using financial indicators (e.g., current ratio, ROE, and asset–liability ratio), macroeconomic indicators (e.g., GDP and inflation rate), company governance proxy variables, and other variables (e.g., company size and year of establishment) as explanatory variables to test the predictive performance of a binary classifier. Shin and Kim [13] have focused on the impacts of liquidity and credit risk on yield spreads. The proxy variables of credit risk are bond rating, coupon rate, default distance, and stock volatility. Jia and Wang [14] have used the KMV (Kealhofer, McQuown, and Vasicek) model to measure the default risk of local government bonds. From the results of their empirical research, it can be seen that the larger the scale of bond issuance, the greater the default risk of debt. Yan and Xu [15], using the Logit model to study the default probability of China's credit bond market, have shown that the scale of corporate assets, profit before interest and tax/total assets, cash ratio, multiple of interest earned, turnover rate of current assets, payable account turnover days, asset–liability ratio, industry to which the company belongs, and the type of issuer company are all related to the company's probability of default.

2.2. Research on XGBoost Algorithm. Modern credit risk measurement tools are more stable and accurate in measuring credit risk. With the development of computers and various disciplines, machine learning algorithms have gained great attention [16, 17]. Among them, the decision-tree-based XGBoost algorithm is a boosting algorithm, proposed by Chen and Guestrin [18], which can perform multithreaded parallel computing. Iteratively generating new trees involves combining many weak learners with low classification accuracy into a strong learner with high classification accuracy, in order to ultimately obtain more accurate predictions. This algorithm has been rapidly spread and widely used since its inception and has achieved satisfactory prediction results in many high-dimensional and unbalanced data analysis and prediction tasks.

XGBoost has been widely used in the field of risk prediction in recent years, due to its significantly better classification and forecasting effect than traditional models. Tao [19] has used the XGBoost algorithm to build a predictive model. The results showed that, based on existing data, the model can effectively predict the business risk of an enterprise and has higher accuracy than traditional methods. Chang et al. [20] have used the XGBoost classifier to construct a credit risk assessment model for financial institutions. Their results showed that, with AUC as the evaluation index, the XGBoost classifier performs better than classifiers such as logistic regression and support vector machines. Zhao et al. [21] have analyzed credit risk based on the XGBoost algorithm. Their experiments showed that the XGBoost algorithm has obvious advantages in accuracy and classification effect, compared with the commonly used decision tree, GBDT (Gradient-Boosting Decision Tree), and support vector machine algorithms, thus verifying the effectiveness and accuracy of the XGBoost model. Huang and Yen [22] selected 16 financial variables from the financial statements of listed companies in Taiwan as the input for six models including the XGBoost model. Their empirical results showed that XGBoost provides the most accurate forecast of financial distress. Xia et al. [23] have used the XGBoost algorithm to assess feature importance, proposed a comprehensive credit rating model based on the XGBoost algorithm, and found that the classification effect of the comprehensive credit rating method is better than that of the conventional method.

In order to further improve the predictive ability of the model, scholars have optimized the model's parameters. Xia et al. [24] have used the Bayesian hyperparameter optimization method to optimize the parameters of the XGBoost algorithm. Their results showed that the Bayesian hyperparameter optimization algorithm is superior to random search, grid search, and manual search methods. Zhao et al. [21] have used the raster search method to optimize parameters and used cross-validation to evaluate model performance. The accuracy of the model and the classification effect were both improved. Guo et al. [25] have used Bayesian optimization to adjust the hyperparameters, and the optimized model had a higher accuracy rate than the existing model. Liu et al. [26] have used Bayesian parameter tuning methods to construct an interpretable credit scoring model based on the XGBoost integration, which has both good performance and interpretability; its AUC, accuracy, and F value were also superior to those of other algorithms. Shen [27] has derived the mathematical principles of XGBoost in detail and used grid search to find optimal parameters for the model. Li et al. [28] have used cross-validation, grid search, and early stopping methods to determine the hyperparameters of the model in the article "Complex network link prediction based on integrated model," then proposed a new prediction method based on the integrated algorithm. Similarly, [29–32] have addressed the predictive modeling issue from different perspectives regarding dealing with metacharacteristics of the data and then modeling it to select the best predictive model for a given problem.

In summary, domestic and foreign scholars have carried out a significant amount of research on the selection of bond default risk prediction indicators and the application of the XGBoost algorithm in default risk prediction. Furthermore, the XGBoost algorithm can be expected to be more accurate than the original algorithm after parameter optimization. Therefore, in this paper, we use the grid search algorithm and k-fold cross-validation to optimize the XGBoost model when predicting the default risk of credit bonds, in order to obtain more accurate prediction results.

3. Research Design

3.1. Selection of Sample Data and Indicators. For the selection and determination of sample data, we extracted the relevant data of credit bond issuances from 2014 to 2020 and the financial data of the relevant issuing companies. Considering the availability of data, we took the credit bonds of listed companies as the research object. Among them, 195 listed credit bonds defaulted from 2014 to 2020. After excluding related bonds with incomplete data, there were 152 bonds remaining. In addition, according to the maturity of default bonds, the scale of issuance, and the credit rating of the main body at the time of issuance, similar bonds are matched among normal bonds. After matching, 3465 normal bonds were obtained, and 1726 normal bonds were obtained after removing a large number having missing values. Therefore, the final sample consisted of 152 default credit bonds and 1726 normal credit bonds.

Considering the selection and determination of indicators, based on relevant domestic and foreign research results, we selected a total of 31 predictive indicators at the micro- and macrolevel of enterprises and bonds. Among them, there were 20 indicators at the microlevel of enterprises and bonds and 11 at the macrolevel. The microlevel predictive indicators were selected from two perspectives: basic bond information and corporate financial information. The basic bond information included bond issuance scale, coupon rate at the time of issuance, issuance period, and whether there was a guarantor. Whether there was a guarantor is a dummy variable, which was assigned a value of 1 if there was a guarantor; otherwise, it was assigned a value of 0. It is generally believed that the larger the scale of bond issuance, the higher the coupon rate, the longer the repayment period, and with the lack of a guarantor, the higher the risk of default of a bond. Considering corporate financial data, we selected 16 indicators from the aspects of profitability, cash flow, solvency, capital structure, income quality, and operating capacity of a company. Generally speaking, the better the financial status of the issuer, the smaller the default risk of the bond. At the macrolevel, 11 indicators were selected, including GDP (gross domestic product), CPI (consumer price index), producer price index (PPI), M2, M1, M0, consumer confidence index, consumer expectations index, business climate index, entrepreneur confidence index, and interbank lending rate (1 day). Generally speaking, the macroenvironment is negatively related to the default risk of credit bonds [32].

3.2. Descriptive Statistical Analysis of Sample Data. Through the statistical analysis of the microindicators of default bonds and normal bonds in the sample, we can intuitively understand the difference between the two types of bonds from two aspects: basic bond information and issuer financial information. In addition to calculating the mean value of each variable, we also conducted a mean difference test for each index. The specific results are shown in Table 1.

First, judging from the bond issuance data, there were significant differences between normal bonds and default bonds, in terms of total issuance, coupon rate, maturity, and whether there was a guarantor. From the perspective of total bond issuance, the average total issuance of default bonds was about 8.774 billion which was higher than the average total issuance of normal bonds (681 million). Considering the coupon rate of bonds, the average issuance interest rate of default bonds was 6.772%. The average coupon rate was 5.618%, such that the difference between the two was 1.154%. From the perspective of the bond issuance period, the average issuance period of default bonds was 3.415 years, while the average issuance period of normal bonds was 3.673 years. From the perspective of whether there was a guarantor, the proportion of guarantors in default bonds was higher than that of normal bonds. From the above statistical results, those related to the total amount of bond issuance and the bond coupon rate were in line with the expected analysis results, while those related to the bond maturity and whether there was a guarantor were contrary to the expected results. We believe that the main reason for this is that the selected default bonds and normal bond data have a large difference. Generally speaking, the default risk of a bond with longer issuance maturity is greater than that of a bond with a shorter maturity, and the default risk of a bond without a guarantor at the time of issuance is greater than that of a bond with a guarantor.

Secondly, considering the financial data of the issuing company, from the perspective of profitability, we selected three indicators—return on assets (ROA), profit before interest and taxes/total operating income, and return on total assets (TTM)—to evaluate the profitability of companies. According to the statistical results, the three indicators were greater for normal bonds than those for default bonds. Among them, the return on assets (ROA) and the return on total assets (TTM) showed significant differences, indicating that the financial situation of normal bond issuers is better than that of default bond issuers. At the same time, it was found that, regardless of whether default bonds or normal bonds were issued, the profitability index of the issuer was positive, indicating that their profitability in the reporting period before bond issuance was stable. From the perspective of cash flow, we selected the cash recovery rate of all assets and the net cash flow from operating activities/operating profit (TTM) to evaluate the cash flow of companies. The cash recovery rate of all assets reflects the ability of a company's assets to generate cash. The larger the value of this indicator, the better their ability. The cash recovery rate of all defaulted bonds was 0.650 higher than that of normal bonds, indicating that the issuers of defaulted bonds had a stronger ability to generate cash before issuing assets than the issuers of normal bonds. Subsequent defaults on bonds may have been caused by problems in the company's operations and poor cash flow. The difference in net cash flow/operating profit (TTM) between the two groups was large, but not significant. From the perspective of debt solvency, two indicators—current ratio and cash ratio—were selected to evaluate the debt solvency of companies. The larger these two indicators, the stronger a company's solvency. The statistical analysis results indicated that the solvency of normal bond issuers was higher than that of default bond issuers. From the perspective of capital structure, we selected three

Table 1: Variable descriptive statistics.

	Normal credit bond		Default credit		Difference test
	Number of samples	Mean	Number of samples	Mean	Mean difference
Total issuance (100 million yuan)	1726	6.81	152	8.774	-1.964***
Coupon interest rate (% at the time of issuance)	1726	5.618	152	6.772	-1.154***
Term (years)	1726	3.673	152	3.415	0.258**
Guarantee	1726	0.118	152	0.211	-0.092***
Return on total assets (ROA) (%)	1726	3.081	152	1.163	1.919***
Ebit/gross operating income (%)	1726	34.127	152	19.272	14.854
Return on total assets (TTM)	1726	3.285	152	1.954	1.331***
Cash recovery rate of all assets	1726	-0.686	152	0.65	-1.336***
Net cash flow from operating activities/operating profit (TTM) (%)	1726	-95.521	152	-424.956	329.435
Current ratio	1726	3.331	152	1.56	1.772***
Cash ratio	1726	0.62	152	0.573	0.048**
Assets and liabilities (%)	1726	60.485	152	63.012	-2.527**
Tangible assets/total assets (%)	1726	30.292	152	16.031	14.261***
Current assets/total assets (%)	1726	63.291	152	48.442	14.848***
Net income from operating activities/total profit (%)	1726	61.737	152	-49.293	111.03***
Operating profit/total profit (%)	1726	88.308	152	80.834	7.474
Accounts payable turnover days (day)	1726	92.314	152	94.51	-2.196
Accounts receivable turnover rate	1726	21.643	152	9.922	11.721***
Total asset turnover rate (TTM) times	1726	0.216	152	0.412	-0.197
Working capital (100 million yuan)	1726	53.477	152	-18.49	71.968***

*, **, and *** represent the significance levels of 10%, 5%, and 1%, respectively.

indicators—asset–liability ratio, tangible assets/total assets, and current assets/total assets—to evaluate the capital structure of companies. The lower the asset–liability ratio, the higher the tangible assets/total assets, and the current assets/total assets, the less debt a company has or the stronger their ability to repay the debt. The results demonstrated that the proportion of tangible assets and the proportion of current assets of normal bond issuers were significantly higher than those of default bond issuers. The debt-to-asset ratios of the two were equivalent, at the 60% level. Among them, default bond subjects were slightly higher than normal bond subjects. The capital structure data were in line with expectations: the asset–liability ratio was high, and the proportions of tangible assets and current assets were relatively low, which all increase the risk of future defaults. From the perspective of income quality, we selected two indicators—net income from operating activities/total profit and operating profit/total profit—to evaluate the income quality of companies. The net income from operating activities reflects the profitability of a business entity. Earnings quality can be used to study the sustainability of a company's future earnings. From the statistical results, the average percentage of operating profits of normal bonds was larger than that of default bonds, indicating that the profits of the issuers of normal bonds rely more on the company's main business. From the perspective of operating capability, we selected four indicators—accounts payable turnover days, accounts receivable turnover rate, total asset turnover rate TTM, and working capital—to evaluate the operating capability of companies. The results indicated that there were significant differences between the mean values of the indicators accounts receivable turnover rate and working capital. The turnover rate and working capital of normal bond issuers were higher than those of default bond issuers. This shows that the purpose of issuing bonds by default bond issuers is likely to raise working capital and make up for the company's funding gap.

3.3. Principal Component Analysis Method

3.3.1. KMO and Bartlett's Test. The Kaiser–Meyer–Olkin (KMO) test is mainly used to determine the suitability of extraction of principal components from data, and the coefficients of the test are distributed in the range (0,1). It is generally believed that if the coefficient value is greater than 0.6, the sample set is considered to meet the requirements of principal component analysis. The null hypothesis of Bartlett's test is that there is no correlation between the variables in the experimental sample set. Each variable has its own meaning, and a certain factor cannot be used to replace several variables to simplify the number of variables; that is, there is no need for principal component extraction.

According to the output of SPSS, the KMO test coefficient in our experiment was 0.682 (see Table 2). According to the coefficient correspondence table, we believe that the data structure used in this experiment was general and had correlation. The P value of the Bartlett test was less than 0.01, and the null hypothesis could be rejected. According to the output results of these two experiments, we considered that the experimental data set in this paper could be analyzed by principal component analysis, in order to reduce the dimensionality of the original data set and extract common factors.

3.3.2. Crushed Stone Diagram of Principal Component Analysis. We conducted principal component analysis on 31 variables to generate a scree plot (see Figure 1). The variance contribution rate of each factor measures the ratio of the variance explained by the factor to the total variance of the original variable, and the characteristic value (or eigenvalue) is an indicator which measures the importance of a factor. The data in the graph show that the variance contribution of the first principal component is 17.8%, and its eigenvalue is 5.5, while the variance contribution of the second principal component is 12.8%, and its eigenvalue is 4.0. Compared with the first principal component, the variance contribution of the second component showed a significant decrease. In this paper, the principal component factors were extracted according to their feature value being greater than 1, and the variance contribution of the selected top 10 principal components was 70.4%; that is, the sum of these 10 principal components could explain 70.4% of the information in all variables.

3.3.3. Principal Component Structure Table. The principal component score coefficient matrix indicates the explanatory strength of the principal component factors with respect to the individual variables. According to the results shown in Table 3, it can be seen that the first principal component was greatly affected by macrofactor variables, such as the corporate climate index, the entrepreneur confidence index, and the interbank lending rate (1 day). The second component was greatly influenced by variables such as M1, consumer confidence index, and consumer expectation index. Therefore, the first and second components better measure the impacts of macrofactors on default bonds. The third component was greatly affected by variables such as current ratio, tangible assets/total assets, and current assets/total assets. Therefore, the third component can measure the solvency and capital structure of the bond issuer. The fourth principal component was greatly affected by variables such as ROA, TTM, and asset–liability ratio. The fifth principal component was greatly affected by variables such as the number of days of accounts payable turnover and profit before interest and taxes/total operating income. The sixth principal component was greatly affected by variables such as current assets/total assets and working capital. The seventh principal component was greatly affected by variables such as whether there was a guarantor, the bond issuance period, and the total issuance and, as such, can be used to measure the basic information of the bond itself.

Table 2: KMO and Bartlett's test results.

KMO metric		0.682
Bartlett's sphericity test	Approximate chi-square	37,845.748
	Degree of freedom	465
	Significance	0

The eighth principal component was greatly affected by variables such as account receivable turnover rate and net cash flow/operating profit (TTM) generated from operating activities. The ninth principal component was greatly affected by variables such as the total issuance and the turnover rate of accounts receivable. Finally, the tenth principal component was greatly affected by variables such as net cash flow generated by operating activities/operating profit (TTM) and the coupon rate at the time of bond issuance.

Overall, from a macro perspective, the macroeconomic background has a strong influence on credit bond defaults. From a microlevel perspective, the basic information of the bond itself, such as the total issuance, coupon rate, issuance period, and whether there is a guarantor, has a strong relationship with the probability of the bond defaulting. At the same time, the issuer's financial data has a strong relationship with whether the bond defaults. Therefore, the principal components extracted from the original variables by principal component analysis were considered appropriate for the following analysis.

3.3.4. Analysis of the Importance of Variables. We used the XGBoost algorithm to perform programming calculations in the Python software to obtain the feature importance scores of all variables, then analyze the importance of features for the 10 principal component factors obtained by principal component analysis. Figure 2 shows a histogram of the feature importance scores for all variables, and Figure 3 shows a histogram of feature importance scores for the principal component factors. From Figure 2, it can be seen that the characteristics of maturity, total issuance, M1, coupon rate, CPI, and PPI scores were relatively high. From the XGBoost model, these indicators had a greater impact on the accurate prediction of bond default risk. These indicators are mainly the basic information and macroeconomic indicators of bond issuance. Combined with the previous difference test, there were significant differences in the coupon rate and total issuance indicators between default bonds and normal bonds. It is generally believed that the higher the bond coupon rate, the more urgent a company's need to raise funds, the higher the financing cost the company will bear, and the greater the company's risk of default in the future. In addition, the tangible assets/total assets and the TTM feature importance scores were also high. From the perspective of the difference analysis, these two indicators were significantly higher for normal bond issuers than for default bond issuers, indicating that normal bond issuers are significantly better than default bond issuers, in terms of capital structure and profitability.

It can be seen, from Figure 3, that principal components 7, 6, 8, 9, and 4 had higher feature importance scores,

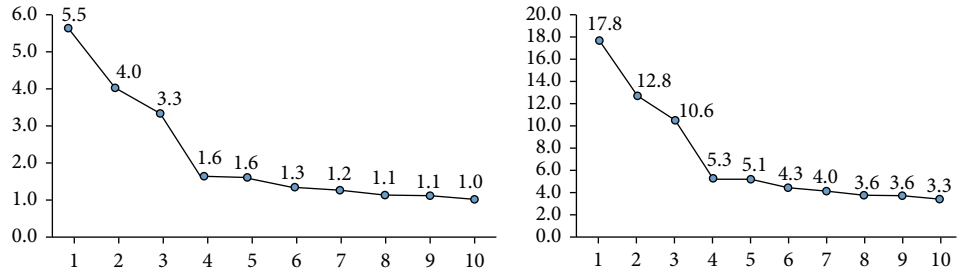

Figure 1: Principal component analysis scree plot.

Table 3: Principal component structure table.

	Principal components									
	1	2	3	4	5	6	7	8	9	10
Total issuance (100 million yuan)	-0.082	0.146	-0.036	-0.022	-0.009	0.121	-0.399	0.032	0.678	0.012
Coupon interest rate (%)	0.488	0.257	0.088	-0.018	-0.173	0.099	0.246	0.124	-0.122	0.313
Term (years)	-0.053	0.458	0.421	-0.098	-0.056	-0.171	0.331	0.076	0.178	0.056
Guarantor	-0.002	0.047	-0.084	-0.106	-0.086	-0.114	0.581	0.382	0.234	0.266
Return on total assets (ROA) (%)	0.229	0.441	-0.399	0.473	0.189	0.16	0.093	-0.015	0.133	-0.05
Ebit/gross operating income (%)	-0.017	0.066	0.139	-0.132	0.731	0.202	0.116	0.127	0.07	0.123
Return on total assets (TTM)	0.166	0.398	-0.44	0.537	0.202	0.205	0.052	0.029	0.094	-0.103
Net income from operating activities/total profit (%)	0.114	0.164	-0.009	0.251	-0.022	0.216	0.157	-0.072	0.208	0.049
Operating profit/total profit (%)	0.034	-0.052	-0.094	0.302	-0.318	-0.071	0.013	0.155	0.175	0.151
Cash recovery rate of all assets	0.12	-0.008	-0.437	0.113	0.196	-0.234	-0.141	-0.283	0.006	0.3
Net cash flow from operating activities/operating profit (TTM) (%)	0.004	0.025	-0.134	-0.017	0.115	-0.241	-0.157	-0.318	0.235	0.596
Assets and liabilities (%)	0.044	-0.124	-0.399	-0.423	-0.262	0.559	-0.115	0.132	0.062	0.145
Tangible assets/total assets (%)	-0.139	0.004	0.688	0.36	0.163	-0.236	0.179	-0.223	0.016	-0.146
Current assets/total assets (%)	-0.16	-0.061	0.627	0.067	-0.175	0.498	0.214	-0.145	0.102	0.005
Current ratio	-0.074	0.012	0.783	0.268	0.036	0.053	-0.115	0.193	-0.029	0.16
Cash ratio	-0.039	0.214	0.471	0.348	0.06	-0.016	-0.393	0.307	-0.021	0.179
Accounts payable turnover days (day)	-0.038	0.011	0.121	-0.351	0.76	0.224	0.063	0.047	-0.034	0.04
Accounts receivable turnover rate	0.042	0.03	-0.02	0.148	-0.026	0.252	0.135	-0.483	-0.351	0.263
Total asset turnover rate (TTM) times	0.123	0.175	-0.474	0.33	-0.008	0.3	0.132	0.022	-0.076	-0.126
Working capital (100 million yuan)	-0.194	-0.29	0.524	-0.01	-0.128	0.388	-0.155	-0.297	0.161	0.013
GDP	0.778	0.436	0.116	-0.176	0.021	-0.049	-0.062	-0.104	0.109	-0.187
CPI	-0.207	-0.679	-0.129	0.126	0.133	-0.082	0.034	0.001	0.182	-0.25
PPI	0.661	0.184	0.143	0.086	-0.031	0.088	-0.251	0.207	-0.289	0.148
M0	-0.464	-0.282	-0.1	0.157	0.132	0.076	-0.233	0.328	-0.188	0.049
M1	-0.283	0.812	0.038	-0.077	-0.049	0.004	-0.171	0.052	-0.174	-0.051
M2	-0.792	0.547	-0.069	-0.088	-0.039	0.001	0.018	-0.049	0.018	-0.006
Consumer confidence index	0.531	-0.787	0.006	0.144	0.039	0.024	0.047	0.08	-0.005	0.082
Consumer expectation index	0.544	-0.772	0.011	0.144	0.038	0.019	0.058	0.075	0.007	0.08
Business climate index	0.937	0.028	0.116	-0.09	0.03	-0.031	-0.077	-0.039	0.043	-0.141
Entrepreneur confidence index	0.927	0.084	0.115	-0.123	0.026	-0.039	-0.041	-0.069	0.11	-0.136
Interbank offered rate (1 day)	0.856	0.375	0.151	-0.08	-0.027	-0.004	-0.046	0.003	-0.007	0.005

indicating that these principal components have a high contribution to accurately predicting the default risk of bonds. At the same time, it can be seen that the feature importance scores of other principal components were not low, indicating that these principal components also contribute, to a certain extent, when predicting the default risk of bonds. Therefore, in subsequent experiments, we included all of the ten factors extracted by principal component analysis into the model, with the hope that the XGBoost model could obtain more effective prediction results.

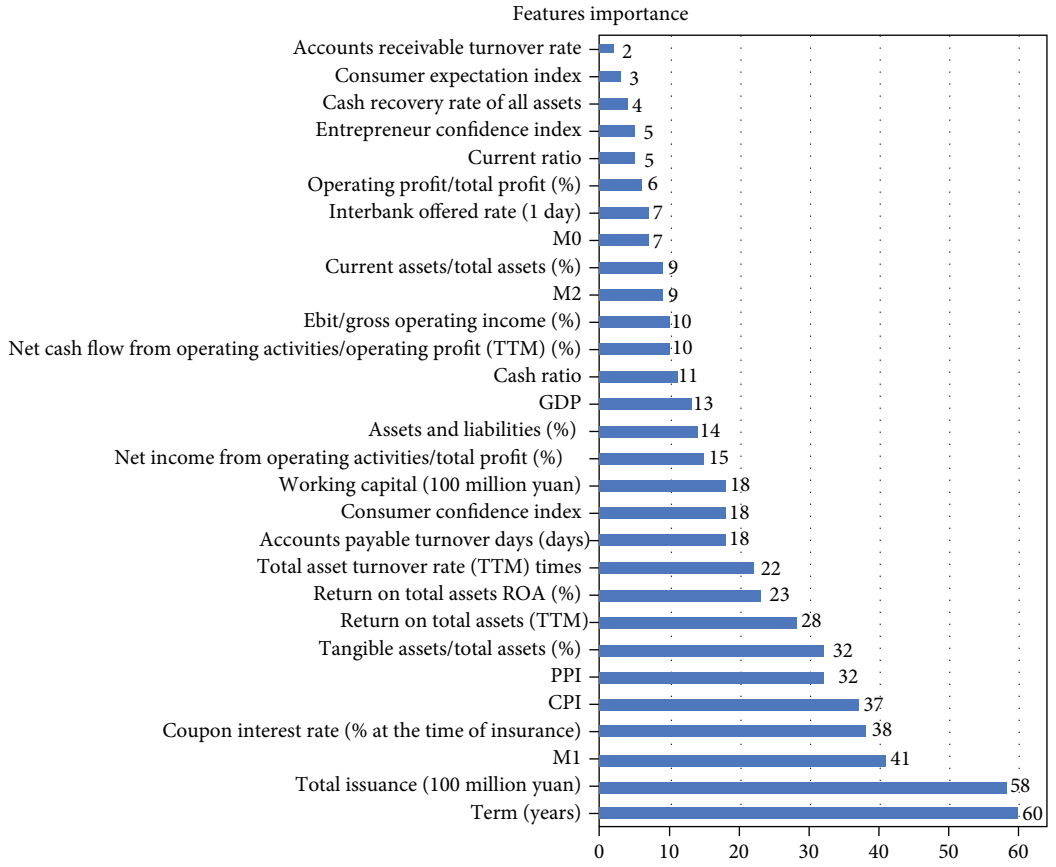

Figure 2: Feature importance scores of all variables.

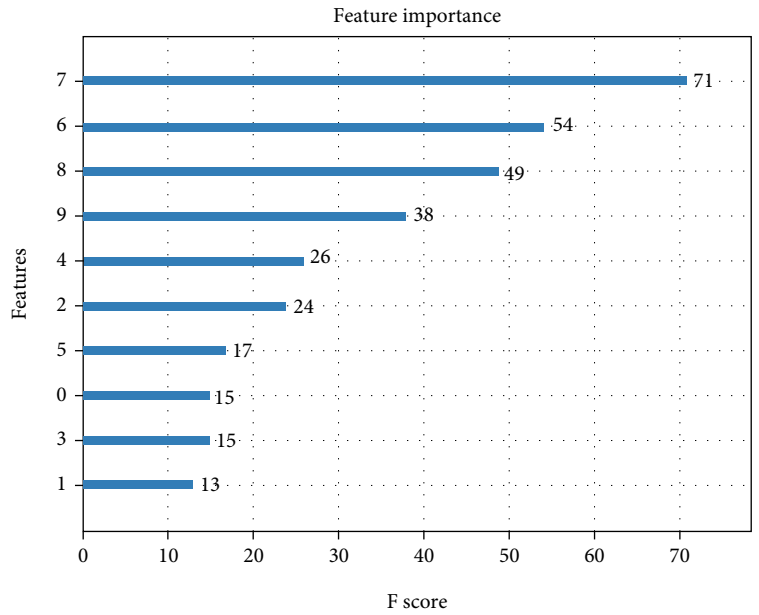

Figure 3: Feature importance scores of principal component factors.

3.3.5. Model Design. The model used in this article is the XGBoost model, which adopts an integrated idea and, so, can be used to solve both classification and regression problems. The algorithm mainly applies integrated ideas, solves the minimum loss function through second-order Taylor expansion, determines the split node, and builds the final

model. XGBoost comes from the additive model in boosting thought, namely,

$$\hat{y}_i = \sum_{k}^{K} f_k(x_i), \tag{1}$$

where f_k denotes a tree and the model has K trees in total. The objective function of the model is

$$L(\varnothing) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \tag{2}$$

where the first item is the loss function, in which $\hat{y}_i$ is the predicted value, y_i is the true value, and i is the number of samples. The second term is the regular term, where $\Omega(f_k) = \gamma T + (1/2)\lambda\|\omega\|^2$, also called the penalty term, in which T represents the number of leaf nodes and ω is the value on the leaf nodes. The regular term is added to control the complexity of the model, that is, to strike a balance between the complexity of the model and the effect of the model. According to the objective function, the forward step algorithm is used to solve the decision tree f_t in the current state:

$$L^t = \sum_i l\left(\hat{y}_i^{t-1} + f_t(x_i), y_i\right) + \Omega(f_t). \tag{3}$$

In the tth round of the current state, as the result of round $t-1$ is known, we can obtain f_t by optimizing the above formula. In the above, the regular term $\sum_k \Omega(f_k)$ in the first $t-1$ round is a constant term, which has no effect on the optimization result, so it can be removed. After that, Taylor expansion is performed on L^t to solve the model. Due to spatial constraints, we do not expand upon the model-solving process.

The XGBoost model mainly includes three types of parameters: general parameters, boosting parameters, and learning parameters. The general parameters are mainly used to set the overall functions of the model, which include booster, silent, nthread, num_pbuffer, and num_feature; the promotion parameters mainly include learning_rate, gamma, max_depth, min_child_weight, subsample, colsample_bytree, lambda, alpha, and scale_pos_weight; and the learning parameters are mainly used to guide the execution of task optimization, including objective, eval_metric, base_score, and seed.

Based on these three types of parameters, we optimized the important parameters and built an optimized XGBoost model to improve the model's predictive effect on bond default risk.

4. Analysis of Model Results

4.1. Evaluation Index. A variety of indicators can be used to evaluate the performance of a model, such as the F1 value, recall rate, and precision rate. We mainly use the AUC value, F1 value, precision, accuracy, and recall rate as evaluation criteria.

First of all, we define TP as the number of positive samples that are correctly predicted by the classification model, FN as the number of positive samples that are incorrectly predicted by the classification model as a negative class, FP as the number of negative samples that are incorrectly predicted by the classification model as a positive class, and TN as the number of negative samples correctly predicted by the model.

The AUC is defined as the area under the ROC curve.

The true positive rate (TPR) is defined as the proportion of positive samples that are correctly predicted by the model, while the false positive rate (FPR) is defined as the proportion of negative samples that are incorrectly predicted by the model as positive:

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{FN}}. \end{aligned} \tag{4}$$

For a specific classifier, the points (FPR, TPR) are connected to form the ROC curve. The ROC curve is a commonly used two-class model evaluation standard. This curve is a graph which shows whether the effect of the classifier in an experiment is good. However, as the ROC curve cannot quantitatively evaluate the classifier, the AUC—which is the value of the area under the ROC curve—is generally used to measure the effect of the model. The AUC value can well-describe the overall performance of the model. The larger the FPR, the more actual negative classes in the predicted positive classes; the larger the TPR, the more actual positive classes in the predicted positive classes. Therefore, it is generally believed that the smaller the FPR, the larger the TPR, the closer the ROC curve is to the upper left corner, and the larger the AUC value, indicating a better classification effect of the model.

Accuracy is defined as the proportion of correctly classified samples to the total number of samples:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{5}$$

Precision is defined as the ratio of the number of positive samples correctly predicted by the classification model to the samples that are predicted as positive by the classification model:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{6}$$

Recall is defined as the proportion of positive samples correctly predicted by the classifier:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

The F1-score (F1-score), accuracy, and recall rate indicators sometimes have contradictions. At this time, one can

use the weighted indicator F1-score to comprehensively consider them:

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (8)$$

According to the mainstream methods in the literature at present, the AUC value evaluation model is usually used, as other evaluation indicators require that a threshold be manually set to convert the obtained probability into the corresponding label category; as such, this threshold will greatly affect the accuracy of the classification model and other indicators. The AUC value takes into account the change of the threshold, and so, the evaluation of the model will be more accurate. Therefore, in this article, we mainly take the AUC as the main evaluation index and other evaluation indices as auxiliary indices.

4.2. Optimizing the XGBoost Model. Grid search is an algorithm which can be used to search for optimal parameters, mainly to optimize model performance, by traversing a given set of parameter combinations. Cross-validation (cross-validation) involves the reuse of data. First, the original data are divided into a training set and a test set, according to a certain ratio, where the training set is used to train the model and the test set is used to evaluate the built model.

In k-fold cross-validation, all of the data in the training set need to be used. All the data of the training set are divided into k parts equally; then, the kth part is taken as the validation set, and the remaining $k-1$ parts are used as the cross-validation training set. Each model parameter that needs to be optimized must undergo a full round of cross-validation, and the average score is obtained through the cross-validation scoring method. The parameter with the highest score is the value of the optimal parameter, and all the optimal values for the parameters are obtained. The corresponding model is considered the optimal model.

The grid search algorithm and 5-fold cross-validation were used to optimize some important parameters of XGBoost, and the AUC area was used as the standard criterion to measure the prediction effect of the model.

4.2.1. learning_rate (Learning Rate) Optimization. The value range of learning_rate is [0,1]. The smaller the value, the slower the calculation speed; however, with a large value of this parameter, convergence may not be possible. The typical value range is 0.01–0.2. As shown in Figure 4, according to the score graph of learning_rate, it can be seen that, when the learning rate was 0.15, the test set had the highest score (0.9677), slightly higher than the scores when the learning rate was 0.10 and 0.20. Therefore, 0.15 was taken as the optimal learning rate.

4.2.2. max_depth (Maximum Depth) and min_child_weight (Minimum Weight Sum) Optimization. The value range of max_depth is [0, +∞). The larger the parameter, the easier it is for overfitting to occur. The typical value range is [3,10]. The value range of min_child_weight is [0, +∞). For this parameter, the larger the value, the more conservative the algorithm, and the less likely it is to overfit. As shown in Figure 5, according to the score map for max_depth and min_child_weight, it can be seen that, when the combination point (maximum depth, minimum weight sum) was (8, 0), the score on the test set was higher (0.9780). So, the optimal maximum depth was 8, and the optimal minimum weight sum was 0.

4.2.3. Gamma (Loss Threshold) Optimization. Gamma is a parameter that controls the number of leaves, with a value range of [0, +∞). The larger the parameter, the more conservative the algorithm and the less likely it is for overfitting to occur. As shown in Figure 6, according to the gamma score chart, when the loss threshold was 0.1, the test set had the highest score (0.9676). Therefore, the optimal loss threshold of 0.1 was selected.

4.2.4. colsample_bytree (Column Sampling Rate) Optimization. colsample_bytree uses a feature column sampling similar to a random forest, with a value range of (0, 1]. As shown in Figure 7, when the column sampling rate was 0.3, the average test set score was the highest (at 0.9707). Thus, the optimal column sampling rate was 0.3.

4.2.5. reg_lambda (L2 Regularization Parameter) Optimization. The value range of reg_lambda is [0, +∞). The larger the parameter, the less likely it is for overfitting to occur. According to Figure 8, when the L2 regularization parameter was set to 6, the highest score value of the test set was obtained (0.9719). Therefore, 6 was selected as the optimal L2 regularization parameter.

4.3. Comparative Analysis of Optimized XGBoost Model and Default XGBoost Prediction Results. It can be seen, from the process of parameter optimization, that the optimization of learning rate, maximum depth and minimum weight sum, loss threshold, column sampling rate, and L2 regularization parameters improved the prediction effect of the model.

Based on the above analysis, we could conclude that the classification prediction results of the optimized model were intuitively better than those of the default model. In order to further clarify the improvement range, we conducted repeated experiments on the data set to measure the classification prediction effect of the model, the results of which are provided in Table 4. The results show that the mean AUC, mean accuracy, mean precision, mean recall, and mean F1-score of the optimized XGBoost model were better than those of the default XGBoost model, indicating that the model achieved a certain degree of optimization for application in predicting credit debt default risk. After optimization, the AUC increased by 0.0992, ACC increased by 0.0213, precision increased by 0.0378, recall increased by 0.2857, and F1-score increased by 0.2687. In summary, the classification prediction effect of the optimized XGBoost model was significantly improved, and the prediction accuracy was better. Therefore, we believe that the optimized XGBoost model is more suitable for default risk prediction.

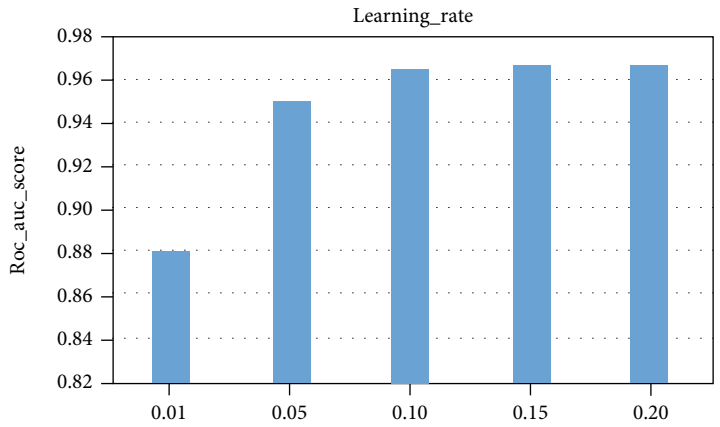

Figure 4: Learning_rate score chart.

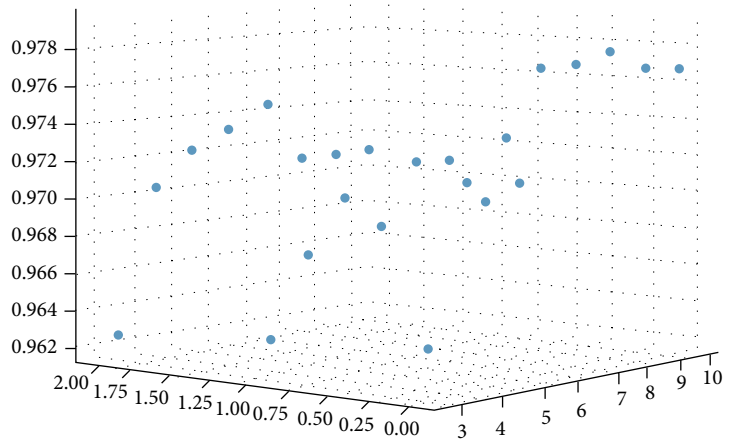

Figure 5: max_depth and min_child_weight score chart.

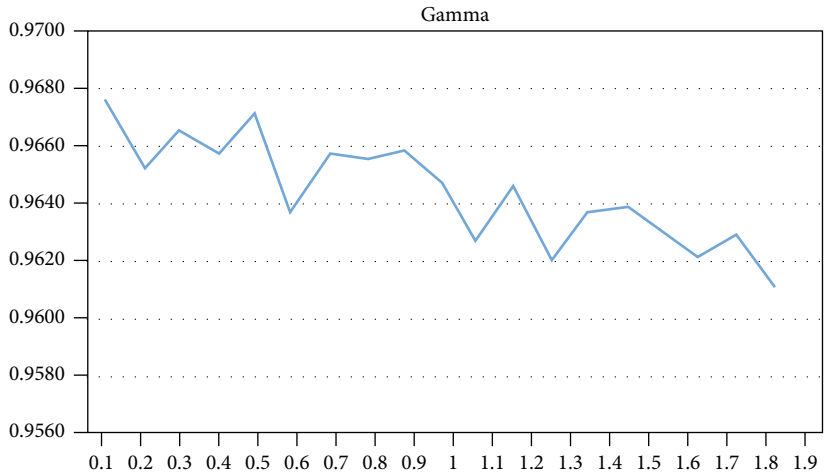

Figure 6: Gamma score graph.

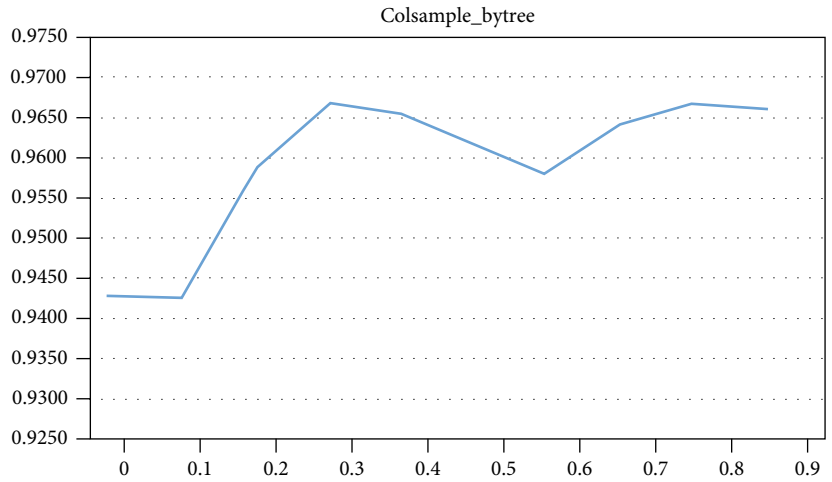

Figure 7: colsample_bytree score chart.

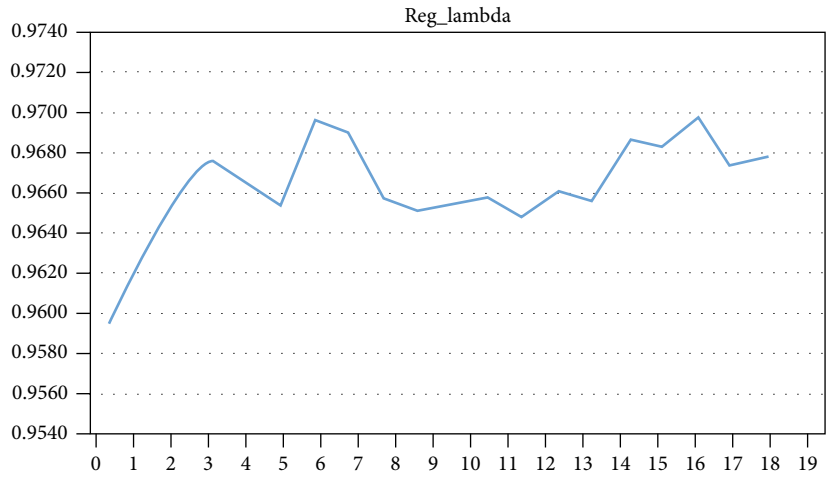

Figure 8: reg_lambda score chart.

Table 4: Comparison of prediction effects.

	AUC	ACC	Precision	Recall	F1-score
XGBoost	0.8807	0.9468	0.9167	0.3143	0.4681
optimized_XGBoost	0.9799	0.9681	0.9545	0.6000	0.7368

5. Conclusions and Prospects

Focusing the problem of credit bond default risk prediction, in this paper, we compared the prediction results of different classification models and, finally, selected the XGBoost model after parameter optimization using grid search and k-fold cross-validation as the best prediction classification model.

As the data set used in this article involved 31 variables, the principal component analysis method was first used to reduce the number of variables, thus reducing the dimensionality. A total of 10 principal components are obtained, which explained 70.4% of the information of all variables. Then, we used the XGBoost model to analyze the importance of variables and output the most influential components. The XGBoost model has many parameters; therefore, we focused on optimizing the parameters of the XGBoost model, using the grid search and k-fold cross-validation algorithms for optimization of the parameters. After optimization of the parameters, the AUC value of the optimized model is 0.9799. Therefore, we finally chose to use grid search algorithm and 5-fold cross-validation to obtain the final optimal parameters of the XGBoost model. Our experimental results demonstrated that the optimized XGBoost model can yield better results, compared to the original algorithm.

Although the model discussed in this paper performed well, in terms of both accuracy and economy, there are still many shortcomings that deserve further improvement and prospects. The index body of the model needs further discussion, and the selected index has certain subjective factors. Although the selection of indicators has been referred to in the existing literature, whether it is comprehensive is still worth further discussion. In terms of sample selection, we considered the availability of data and used bonds issued by listed companies as the research sample. However, listed companies are only a small part of a large number of bond-issuing companies. Therefore, expansion of the research sample should be carried out in future research. Moreover, in terms of model selection, only the XGBoost model was selected as a classification model in this article. The model still has limitations in economic interpretation and cannot provide various index coefficients, in order to quantitatively analyze the impact of a single factor on the risk of default.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (71873108, 62072379), the Fundamental Research Funds for the Central Universities (kjcx20210103, JBK2103016), the Financial Intelligence and Financial Engineering Key Lab of Sichuan Province, the Jiaozi Institute of Financial Technology Innovation, the Southwest University of Finance and Economics (cgzh20210204), the Research Program of Science and Technology at Universities of Inner Mongolia Autonomous Region (2021GG0164), and the Financial Innovation Center of the Southwestern University of Finance and Economics.

References

[1] J. Deng, T. Qin, and S. Huang, "Research on credit risk early warning of listed companies in China based on logistic model," *Financial Theory and Practice*, vol. 2, pp. 22–26, 2013.

[2] R. Kanapickiene and R. Spicas, "Credit risk assessment model for small and micro-enterprises: the case of Lithuania," *Risks*, vol. 7, no. 2, p. 67, 2019.

[3] A. Coşer, M. M. Maer-matei, and C. Albu, "Predictive models for loan default risk assessment," *Economic Computation & Economic Cybernetics Studies & Research*, vol. 53, no. 2/2019, pp. 149–165, 2019.

[4] Z. Ma, W. Hou, and D. Zhang, "A credit risk assessment model of borrowers in P2P lending based on BP neural network," *PLoS One*, vol. 16, no. 8, article e0255216, 2021.

[5] D. Chen, "Risk assessment of government debt based on machine learning algorithm," *Complexity*, vol. 2021, 12 pages, 2021.

[6] P. Collin-Dufresne, R. S. Goldstein, and J. S. Martin, "The determinants of credit spread changes," *Journal of Finance*, vol. 56, no. 6, pp. 2177–2207, 2001.

[7] K. Giesecke, F. A. Longstaff, S. Schaefer, and I. Strebulaev, "Corporate bond default risk: a 150-year perspective," *Journal of Financial Economics*, vol. 102, no. 2, pp. 233–250, 2011.

[8] G. Q. Dai and X. B. Sun, "Research on the macro determinants of China's corporate bond credit spreads," *Financial Research*, vol. 12, pp. 61–71, 2011.

[9] M. Mili, J. M. Sahut, and F. Teulon, "Modeling recovery rates of corporate defaulted bonds in developed and developing countries," *Emerging Markets Review*, vol. 36, pp. 28–44, 2018.

[10] H. Die, "Bond default analysis based on random forest," *Contemporary Economy*, vol. 3, pp. 28–30, 2018.

[11] P. Mizen and S. Tsoukas, "Forecasting US bond default ratings allowing for previous and initial state dependence in an ordered probit model," *International Journal of Forecasting*, vol. 28, no. 1, pp. 273–287, 2012.

[12] S. Jones, D. Johnstone, and R. Wilson, "An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes," *Journal of Banking & Finance*, vol. 56, pp. 72–85, 2015.

[13] D. Shin and B. Kim, "Liquidity and credit risk before and after the global financial crisis: evidence from the Korean corporate bond market," *Pacific-Basin Finance Journal*, vol. 33, pp. 38–61, 2015.

[14] Z. Jia and Z. Wang, "Local government bond default risk measurement based on KMV model," *Financial Theory Research*, vol. 5, pp. 72–82, 2015.

[15] C. Yan and J. Xu, "How to dynamically evaluate the default probability of a company? ——based on the logit model of China's credit bond market," *Financial Market Research*, vol. 1, pp. 124–133, 2018.

[16] K. Y. Zhong, C. L. Li, and Q. Wang, "Evaluation of bank innovation efficiency with data envelopment analysis: from the perspective of uncovering the black box between input and output," *Mathematics*, vol. 9, no. 24, p. 3318, 2021.

[17] K. Y. Zhong, Y. F. Wang, J. M. Pei, S. M. Tang, and Z. L. Han, "Super efficiency SBM-DEA and neural network for performance evaluation," *Information Processing & Management*, vol. 58, no. 6, article 102728, 2021.

[18] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, 2016.

[19] S. Tao, "XGBoost-based corporate failure risk prediction," *Wireless Internet Technology*, vol. 15, no. 8, pp. 102–104, 2018.

[20] Y. C. Chang, K. H. Chang, and G. J. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing*, vol. 73, pp. 914–920, 2018.

[21] T. Zhao, S. Zheng, W. Li, and K. Liu, "Research on credit risk analysis based on XGBoost," *Software Engineering*, vol. 21, no. 6, pp. 29–32, 2018.

[22] Y. P. Huang and M. F. Yen, "A new perspective of performance comparison among machine learning algorithms for financial distress prediction," *Applied Soft Computing*, vol. 83, article 105663, 2019.

[23] L. Xia, Y. Zhang, Q. Lu, and G. Tang, "Credit rating method combining XGBoost algorithm and logistic regression," *Credit Investigation*, vol. 11, pp. 56–59, 2019.

[24] Y. Xia, C. Liu, and N. Liu, "Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending," *Electronic Commerce Research and Applications*, vol. 24, pp. 30–49, 2017.

[25] J. Guo, L. Yang, R. Bie et al., "An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring," *Computer Networks*, vol. 151, pp. 166–180, 2019.

[26] Y. Liu, Z. Zhang, H. Chen et al., "An interpretable credit scoring model based on XGBoost integration," *Data Communication*, vol. 3, pp. 27–32, 2019.

[27] C. Shen, "Principle and application of XGBoost," *Computer Products and Circulation*, vol. 3, p. 90, 2019.

[28] K. Li, L. Tu, and L. Chai, "Ensemble-model-based link prediction of complex networks," *Computer Networks*, vol. 166, article 106978, 2020.

[29] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.

[30] R. Ali, A. M. Khatak, F. Chow, and S. Lee, "A case-based meta-learning and reasoning framework for classifiers selection," in *Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication*, pp. 1–6, Langkawi, Malaysia, 2018.

[31] R. Ali, M. H. Siddiqi, M. Idris, B. H. Kang, and S. Lee, "Prediction of diabetes mellitus based on boosting ensemble modeling," in *International conference on ubiquitous computing and ambient intelligence*, pp. 25–28, Cham, 2014.

[32] K. Giesecke, F. A. Longstaff, S. Schaefer, and I. Strebulaev, "Macroeconomic effects of corporate default crisis: a long-term perspective," *Journal of Financial Economics*, vol. 111, no. 2, pp. 297–310, 2014.