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

DOI: 10.23977/pree.2025.060106
ISSN 2616-2253 Vol. 6 Num. 1

Research on Interpretable Machine Learning Models for Identifying Corporate Bond Default Risk

Yunpeng Zhao

Treasury Department, Bank of China, New York, NY 10018, USA
JackZ20241@outlook.com

Keywords: Bond Default; Risk Identification; Machine Learning

Abstract: Against the backdrop of increasingly severe credit bond default risks in China, how to accurately identify and efficiently warn of corporate bond default risks has become a major focus of academic and practical fields. This study aims to overcome the shortcomings of traditional default risk warning models in terms of predictive ability, hyperparameter adjustment, and model interpretability. We have constructed a novel corporate bond default risk warning model, LightGBM-NSGA-II-SHAP, by organically integrating LightGBM, NSGA-II, and SHAP algorithms. Through empirical testing, the warning accuracy of this model exceeds 85%, and its performance is significantly better than traditional methods. In addition, the application of SHAP algorithm enables the visualization of the impact of warning features, and the results show that features such as coupon rate, net profit margin of fixed assets, total issuance amount, and accounts receivable turnover rate are crucial for identifying bond defaults.

1. Introduction

In recent years, bond default events have broken out frequently in China, which has become a significant financial risk problem, and has attracted extensive attention from academia and industry[1]. In this field, the main research directions include the selection of warning indicators and the optimization of warning models. Regarding warning indicators, previous studies have pointed out that micro indicators such as corporate financial status, non-financial characteristics, and bond attributes, as well as macroeconomic factors such as GDP, interest rates, and price indices, are key predictive factors[2]. The warning effect of combining financial and non-financial indicators has been proven to be more significant. Therefore, this study incorporates multiple financial indicators such as debt paying ability, profitability, operational efficiency, development potential, capital structure, and cash flow, as well as bond attributes, non-financial characteristics of enterprises, and macroeconomic indicators, into the construction of an early warning feature system[3]. Considering the high requirement for timeliness in credit risk warning, this study chose quarters as the time window to explore the impact of different time periods on the effectiveness of the warning model. Traditional default risk warning methods, such as Altman Z-Score and option theory models, although widely used, have certain limitations. In recent years, with the continuous advancement of machine learning technology, the LightGBM model has been selected as the core model for research due to its excellent predictive performance. In addition, in order to optimize the

hyperparameters of the model, this study introduced the NSGA-II algorithm, thereby improving the warning effect. In terms of interpretability of the model, this study adopted the SHAP algorithm to visualize the contribution of each feature to the prediction results, filling the gap in interpretability of existing methods[4]. Therefore, the contribution of this study mainly lies in three aspects: firstly, constructing a comprehensive warning feature system that comprehensively considers financial, bond, non-financial, and macroeconomic indicators; Secondly, by integrating multiple machine learning algorithms, the warning accuracy and explanatory power of the model have been improved; Thirdly, by analyzing the performance of the model under different time windows, a new perspective has been provided for the study of bond default warning[5].

2. Principle of Early Warning Model Algorithm

LightGBM aims to improve computational efficiency while maintaining classification accuracy through single-sided sampling (GOSS) and unique feature bundling (EFB) techniques. When applying LightGBM for bond default risk warning, the goal of the model is to fit the warning feature set by optimizing the loss function x and label vector y. The loss function is shown in formula (1).

$$L(x) = \arg\min L\left(y, \hat{f}(x)\right) \tag{1}$$

Among them, L represents the error between the model warning results and the actual default status. The LightGBM model uses k regression trees to construct the final risk warning model, as shown in formula (2).

$$f_k(x) = \sum_{i=1}^{k} f_i(x) \tag{2}$$

In each iteration of the algorithm, the risk warning objective function is optimized ithrough second-order Taylor expansion, as shown in formula (3).

$$\Gamma_t = \sum_{i=1}^{n}\left(g_i f_t(x_i) + \frac{h_i f_t^2(x_i)}{2}\right) + \sum_{i=1}^{k}\Omega\left(f_k(x)\right) \tag{3}$$

Among them, g_i and h_i are the first-order and second-order degrees of the loss function, respectively, and $\sum_{i=1}^{k}\Omega\left(f_k(x)\right)$ is the regularization term aimed at preventing overfitting. When selecting the optimal splitting feature for each tree, calculate the splitting benefit to maximize the information gain (formula (4).

$$G = \frac{1}{2}\left(\frac{\left(\sum_{i\in I_L} g_i\right)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{\left(\sum_{i\in I_R} g_i\right)^2}{\sum_{i\in I_R} h_i + \lambda} + \frac{\left(\sum_{i\in I} g_i\right)^2}{\sum_{i\in I} h_i + \lambda}\right) \tag{4}$$

LightGBM adopts a Leaf wise growth strategy and histogram algorithm to optimize model performance. Compared with traditional strategies, the Leaf wise strategy selects the leaf with the highest gain for splitting in each iteration, effectively improving the warning effect of the model while suppressing overfitting. The histogram algorithm improves training efficiency by discretizing continuous features into integers to construct histograms.

In order to improve the performance of the LightGBM model in bond default risk warning, this paper applies the Fast Non Dominated Sorting Genetic Algorithm II (NSGA-II) to optimize its

hyperparameters. NSGA-II is an algorithm that solves multi-objective optimization problems by simulating natural selection and evolution processes to find the optimal solution set. We defined hyperparameters to be optimized, such as learning rate, number of leaves, maximum depth, and sampling ratio, and randomly generated an initial population, with each individual representing a combination of hyperparameters. Subsequently, each individual was trained using LightGBM and their fitness was evaluated based on the model's warning accuracy and efficiency. Next, the population is non dominated sorted, individuals are classified according to their strengths and weaknesses, and the crowding distance of each individual is calculated to reflect their density distribution in each level. Based on these rankings and crowding levels, select individuals suitable for reproduction. Generate a new generation of individuals through crossover and mutation operations, and merge them with the original population to form a new population. The new population will undergo sorting and crowding calculation again to determine the next generation of individuals. Output the optimal hyperparameter combination when the predetermined termination conditions are met. This process not only optimizes model performance, but also maintains diversity in understanding.

In order to improve the interpretability of the warning model and identify key warning features, this study adopted the SHapley Additive exPlans (SHAP) algorithm. The SHAP algorithm was proposed by Lundberg in 2017 based on the Shapley value in game theory, aiming to quantify the marginal contribution of each warning feature to the model's prediction results. The risk warning result of the model consists of the SHAP value of each feature and the mean of all sample target variables, as shown in formula (5):

$$g(z) = \phi_0 + \sum_{i=1}^{M} \phi_i z_i \tag{5}$$

Among them, ϕ_i represents the SHAP value of the i-th feature, ϕ_0 is the mean of the target variable, M is the total number of features, and z_i indicates whether the feature exists (0 or 1). The calculation formula for SHAP value is:

$$\phi_i = \sum_{S \subseteq N \setminus \{X_i\}} \frac{|S|!(p-|S|-1)!}{p!} \left(f(S \cup \{X_i\}) - f(S) \right) \tag{6}$$

Among them, p is the total number of features, S is a subset containing other features, f(S) is the model output of subset S, and $f(S \cup \{X_i\})$ is the model output after adding feature i. By calculating these SHAP values, we can gain a detailed understanding of the specific impact of each warning feature on model decision-making.

3. Experimental Design

This study used the LightGBM-NSGA-II model to identify bond default risk and conducted interpretability analysis on the identification results using the SHAP algorithm. The research process is divided into four main steps: (1) data collection and preprocessing. Extract raw data from the database, perform feature filtering, and then divide the dataset into training and testing sets; (2) Training and optimization of the model. Train the LightGBM model based on the training set and optimize the hyperparameters of the model using NSGA-II; (3) Testing and performance evaluation of the model. Verify the recognition performance of the optimized model using a test set and evaluate it from both accuracy and efficiency perspectives; (4) Explanation of the model. Identify

key features that have a significant impact on bond default prediction results through SHAP algorithm, and analyze the specific effects of these features on the model's prediction results[6].

This study was conducted from the Wind database(https://www.wind.com.cn/)We extracted bond default data of Chinese listed companies from 2014 to 2023. Due to the possibility of multiple defaults by companies in the same quarter, this study selected 220 samples of bond default events that occurred for the first time in each quarter. In order to ensure the rationality of the sample, this study adopted a method similar to that of Pang Chunchao et al., matching each default sample with two non default samples at a matching ratio of 1:2. The specific operation is to use the CSMAR database(https://data.csmar.com/)Select companies that have similar characteristics to the default sample in terms of their main business and current total asset size as the control group. The final dataset contains 660 samples, with defaulting companies marked as 1 (label=1) and non defaulting companies marked as 0 (label=0). In data processing, the quarter in which the default occurred was set as Q, and warning features were selected from the quarter before the default (Q1), the first two quarters (Q2), the first three quarters (Q3), and the first four quarters (Q4). The warning features include a total of 36 items, covering financial indicators, bond attributes, macroeconomic indicators, and corporate characteristics. Among them, financial indicators involve six aspects including debt paying ability, development ability, cash flow ability, profitability, operating ability, and ratio structure, totaling 24 indicators, while non-financial indicators have 12 items. According to the research methods of Xu Shuyue and Cao Yanhua, missing value processing was applied to the data, resulting in the construction of four datasets.

Figure 1 shows the Pareto front of dataset 1 optimized by NSGA-II, where y1 represents the error omission rate (FOR) and y2 represents the recall rate (TPR). In the figure, green dots mark the optimal solutions selected in the study. The hyperparameter combinations of these optimal solutions are used to construct a default risk warning model. Table 1 presents the specific results of hyperparameter optimization on four datasets.

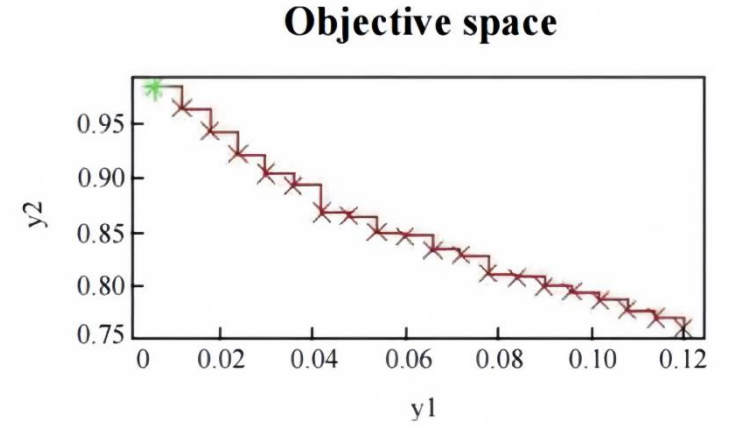

Figure 1. The Pareto front of $Dataset_{Q1}$ hyperparameter optimization results for $Dataset_{Q1}$

Table 1. Hyperparameter optimization results of four datasets

	Learing-rate	Num-leaves	Max-depth	subsample
$Dataset_{Q1}$	0.013	29	9	0.54
$Dataset_{Q2}$	0.179	31	9	0.70
$Dataset_{Q3}$	0.050	51	4	0.58
$Dataset_{Q4}$	0.176	47	3	0.53

4. Design of Model Effect Evaluation Indicators

This study designed six classic machine learning models for comparison to evaluate the effectiveness of the proposed method. Specifically, LightGBM-GA was used for bond default identification, and its performance was compared with the hyperparameter optimization algorithm NSGA-II to determine its advantages over GA. The hyperparameter ranges of methods 1 and 2 are the same, while method 3 uses the unoptimized LightGBM model with hyperparameters. Methods 4, 5, 6, and 7 respectively use classic models such as XGBoost, Adaboost, ANN, and SVM. The performance of all models on the test set will be evaluated by accuracy (ACC), error omission rate (FOR), recall rate (TPR), and false positive rate (FPR) to compare their effectiveness in identifying bond defaults. The confusion matrix is used to calculate these evaluation metrics, where ACC and FOR measure the predictive accuracy of the model, while TPR and FPR reflect the predictive efficiency of the model. Specifically, True Positive (TP) refers to the number of actual non default samples correctly identified as non default, True Negative (TN) refers to the number of actual default samples correctly identified as default, False Positive (FP) represents the number of actual non default samples incorrectly identified as default, and False Negative (FN) represents the number of actual default samples incorrectly identified as non default. The detailed description of the above four evaluation indicators is shown in Table 2.

Table 2. The evaluation indicators for bond default identification results

Evaluating indicator	Indicator Description	Calculation formula
Accuracy rate (ACC)	Actual default samples and non default samples are Identify the correct proportion.	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$
Error omission rate (FOR)	The proportion of samples incorrectly identified as non defaulting in actual default samples.	$FOR = \frac{FN}{FN + TN}$
recall (TPR)	The proportion of samples identified as non default that are actually non default samples.	$TPR = \frac{TP}{FN + TP}$
False positive rate (FPR)	The proportion of non default samples misjudged as default to the actual non default samples.	$FPR = \frac{FP}{FP + TN}$

5. Experimental results and analysis

In this study, we plotted SHAP dependency graphs for important features such as bond interest rate, fixed asset net profit margin (y4), total issuance amount (total), and accounts receivable turnover rate (j1) (see Figure 2). These charts are used to demonstrate how these features affect the warning results. The vertical axis in each graph represents the SHAP value, while the horizontal axis displays the actual value of the feature. Each data point represents a sample, and smooth curves are used to depict the relationship between feature values and SHAP values. This visualization method helps us intuitively understand the specific contributions of each key feature to the prediction results.

In Figure 2, we analyzed the impact of key features on SHAP values to better understand their contribution to bond default prediction. When the coupon rate is in the range of (0.005, 0.034), the

SHAP value shows a decreasing trend, indicating a reduction in default risk within this range; When the coupon rate is in the range of (0.034, 0.085), the SHAP value increases, indicating an increase in the probability of default; In the range of (0.085, 0.088), the SHAP value decreases again. The net profit margin of fixed assets (y4) is within the range of (-2.338, -1.009), and an increase in SHAP value indicates an increase in the likelihood of default; In the range of (-1.009, 1.377), the SHAP value decreases with the increase of eigenvalues, indicating a decrease in risk; In the interval of (1.377, 3.076), the SHAP value rises again. The total issuance amount (total) is in the range of (1 * 10 ^ 7, 1.5 * 10 ^ 8), and the SHAP value decreases, indicating a decrease in default probability; In the range of (1.5 * 10 ^ 8, 1 * 10 ^ 9), as the SHAP value increases, the default risk also increases; In the intervals of (1 * 10 ^ 9, 1.5 * 10 ^ 9) and (1.5 * 10 ^ 9, 6 * 10 ^ 9), the change in SHAP value shows an initial decrease followed by an upward trend. The accounts receivable turnover ratio (j1) is in the range of (0.089, 2.503), and the SHAP value increases with the increase of characteristic values, indicating an increase in default risk; In the range of (2.503, 31.672), the SHAP value decreases and the likelihood of default decreases; In the interval of (31.672, 59.327), the SHAP value rises again. Figure 3 provides a visualization of the feature influence of a single sample through SHAP analysis, where red represents positive influence and blue represents negative influence. The numerical values in the arrow boxes show the SHAP values of each feature and their specific contributions to the prediction results.

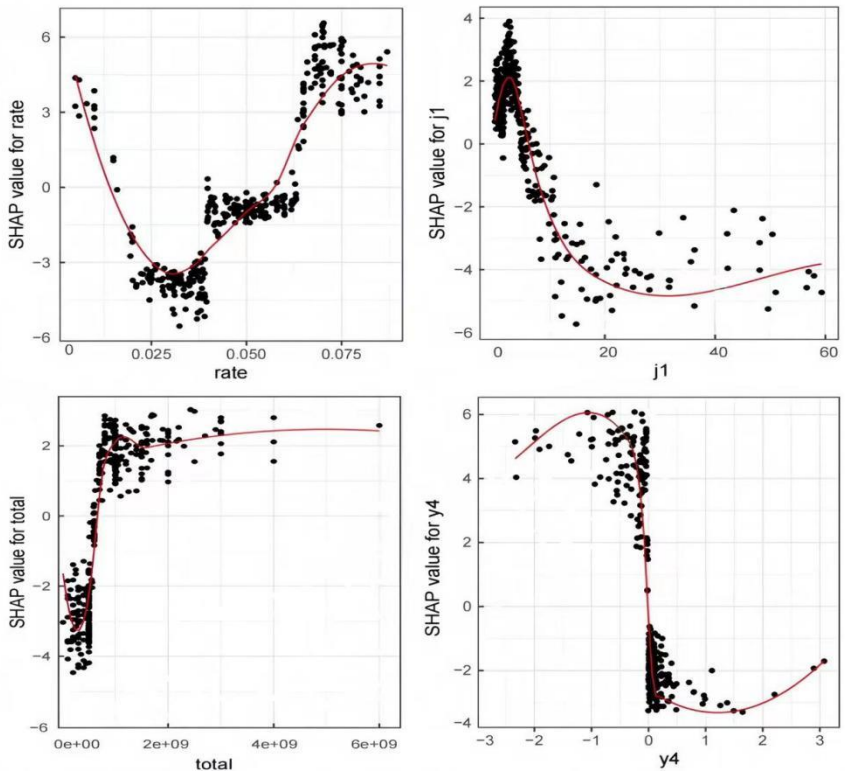

Figure 2. The SHAP dependence plots of the top four features for default identification

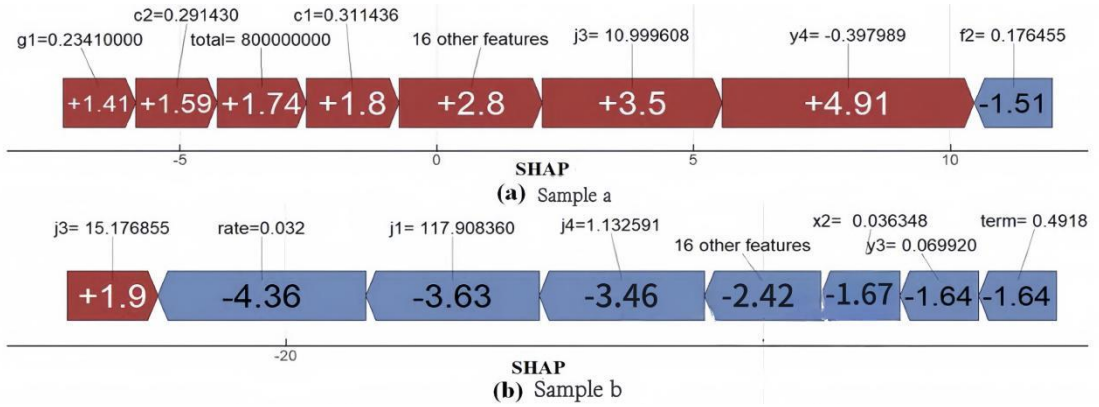

Figure 3. The SHAP force plots in testing set

The subgraph (a) in Figure 3 shows the SHAP plot of the samples predicted by the model as potential bond defaults. When analyzing sample A, it was found that 22 characteristics such as fixed asset net profit margin (y4) and cash and cash equivalents turnover rate (j3) have a positive impact on default risk, indicating that the increase of these characteristics may lead to an increase in default risk. The revenue growth rate (f2) has been identified as a negative factor for risk prediction, indicating that its contribution to default prediction is relatively small and not considered a key feature. In subgraph (b) of Figure 3, the SHAP plot of samples predicted to not default on bonds is shown. For sample b, 22 features such as coupon rate and accounts receivable turnover rate (j1) have a negative impact on the model's prediction results, indicating that an increase in these features helps to reduce default risk. Through these single sample SHAP diagrams, users can have a clearer understanding of the specific predictive contributions of each feature to a single sample, thus enabling more accurate personalized risk assessment. This detailed visual analysis helps reveal the underlying mechanisms of model predictions, supporting more targeted decision-making and risk management.

By applying the SHAP interpretation method, this study can reveal the impact of warning features on model prediction results and gain a deeper understanding of the complex interactive relationship between feature values and prediction results. This method not only displays the marginal contribution of each feature to the model output, but also visualizes the warning results of individual samples, thereby enhancing the interpretability of the model. This detailed explanation helps investors, regulatory agencies, and issuing companies better grasp risks, providing a scientific basis for the management and decision-making of bond default risks. By increasing the transparency of the model, this analysis method makes risk assessment more accurate and reliable, providing strong support for the risk warning process in practical applications.

6. Conclusion

This study developed a comprehensive model based on LightGBM and NSGA-II to effectively identify the corporate default risk in the Chinese bond market. By integrating financial data, bond characteristics, non-financial information of enterprises, and macroeconomic indicators, we use LightGBM as the core classifier and apply NSGA-II algorithm to optimize its hyperparameters to improve the warning accuracy and efficiency of the model. In addition, this study conducted interpretability analysis on the model using SHAP method, demonstrating in detail the specific contributions of each feature to the model's prediction results, thereby enhancing the transparency and comprehensibility of the model. The research results indicate that in multiple experiments, the model proposed in this study outperforms other comparative methods in terms of recognition accuracy and efficiency. Especially in the quarter before the default occurred (DatasetQ1), the

model showed the best warning effect, indicating that the model can provide more accurate warnings when the default is approaching. The interpretation results of SHAP method reveal the complex influence of features on the model output. For example, the net profit margin of fixed assets (y4), the shareholding ratio of the largest shareholder (g1), and the M2 growth rate (h4) are negatively correlated with the risk of bond default, while the coupon rate and current liability ratio (b4) are positively correlated. These key features should be the key focus indicators for monitoring default risk.

Based on the actual situation of China's bond market, the findings of this study have important practical significance:

Bond investors should use advanced machine learning technology to conduct in-depth financial and non-financial data analysis of bond issuing companies, pay attention to their operational capabilities and asset performance, and comprehensively consider the impact of macroeconomic environment on bond default risk.

Financial regulatory agencies should focus on monitoring the profitability, development potential, and bond issuance related indicators of bond issuing companies, identify high default risk companies using quantitative analysis tools, and strengthen information disclosure systems to ensure transparency of financial data and bond characteristics information, in order to prevent potential systemic risks.

Enterprise management should make full use of these advanced warning signals, clarify key influencing factors and their mechanisms of impact on default risk, adjust business and investment strategies in a timely manner, thereby effectively reducing the risk of bond default and minimizing potential economic losses. These measures not only help improve the financial health of enterprises, but also contribute to maintaining the stability of the entire financial market.

References

[1] Liu P, Li Y. Comparative analysis of machine learning models for bond default forecasting based on financial data of Chinese listed companies [J]. BCP Business & Management, 2022, 34: 1151-1158.

[2] Huan Z. On the effectiveness of graph statistics of shareholder relation network in predicting bond default risk [J]. Journal of Control Science and Engineering, 2022, 2022: 8401354.

[3] Lei S, Liang X, Wang X, et al. A short-term net load forecasting method based on two-stage feature selection and LightGBM with hyperparameter auto-tuning[C]. IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference, 2023: 1-6.

[4] Tang M, Zeng W, Zhao R. Corporate credit risk rating model based on financial big data [J]. BCP Business & Management, 2023, 48: 33-42

[5] Chen, H., Ma, K., & Shen, J. Interpretable Machine Learning Facilitates Disease Prognosis: Applicationson COVID-19 and Onward [J]. International Journal of Computer Science and Information Technology, 2024, 3(3), 428-436.

[6] Chen, H., Yang, Y., & Shao, C. Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding [J]. Journal of Manufacturing Systems, 2021, 58, 306-315.