Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1593–1604
Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

CREDIT RISK MEASUREMENT OF REAL ESTATE ENTERPRISES BASED ON THE RANDOM FOREST MODEL

HUI JIANG, JUNFU CUI, AND YANG LIU*

Abstract. The measurement and early warning of credit risk in real estate enterprises represents a crucial step in strengthening macroprudential supervision. This paper constructs a credit risk evaluation index system for real estate enterprises based on debt servicing ability, profitability, operating ability, growth ability, and cash flow. Subsequently, the data of 117 listed real estate companies in China's A-share market was employed to measure their risks and identify the key influencing factors using the random forest model. The AUC value on the test set was 81.25%, with an accuracy rate of 83.33%. These findings suggest that the random forest model has particular applicability in measuring credit risk in real estate enterprises. By ranking the importance of the eigenvalues of the random forest model, it can be determined that the gearing ratio, operating income growth rate, cash ratio, net profit margin on sales, and total cash-to-liabilities ratio have a significant impact on the credit risk of real estate enterprises. It is, therefore, incumbent upon real estate companies to pay close attention to the performance of debt service and profitability, including such key indicators as the asset-liability ratio and the operating income growth rate, to better cope with the inherent credit risk.

1. Introduction

The real estate industry plays a pivotal role in the national economy, and its virtuous cycle is paramount for the economy's sustained growth. For an extended period, Chinese real estate enterprises have predominantly operated with high debt, high leverage, and high turnover structure, resulting in relatively fragile balance sheets [6]. In China's economic deceleration, overcapacity in the real estate industry, subdued income expectations of residents, inadequate effective demand, and robust expectations of declining house prices, the capital chain of real estate enterprises has been disrupted, precipitating a wave of defaults. In 2023, the default scale of the real estate industry reached 23.226 billion yuan, accounting for 76.31% of the default scale, while the rollover scale reached 66.693 billion yuan, representing 84% of the rollover scale. The frequent defaults of real estate enterprises and the downturn of the real estate market have intensified the accumulation of debt risks for commercial banks and local governments [22]. This may induce systemic financial risks for the whole economy and jeopardize economic and social stability. The prevention and

---

2020 *Mathematics Subject Classification.* 91G40, 91G42.

Key words and phrases. Random forest model, machine learning, credit risk measurement, real estate firms, financial indicators.

This research was funded by Key R&D Program (Soft Science Project) of Shandong Province (No.2025RKY0301) and Philosophy and Social Sciences Planning Research Project of Jinan City (No.JNSK2025C096).

*Corresponding author.

resolution of risks in key real estate areas, along with the resolute avoidance of systemic risks, represent the most pressing priorities of China's current economic agenda. The ability to promptly and accurately identify and alert to potential risks associated with real estate enterprises is a crucial element in preventing and resolving risks in pivotal sectors and promoting a stable and prosperous real estate market.

The conventional approach to measuring real estate risk, relying on traditional statistical techniques, is inadequate for addressing the economic system's inherent nonlinearity, complexity, and dynamic nature. Attempts to apply these methods to the real estate market, with its high-dimensional and noisy data, have proven challenging. In the context of the digital economy, the accumulation of data elements, the development of machine learning models, and the advancement in computer arithmetic are transforming macro-prudential regulation. The advantages of machine learning methods in analyzing financial data are gradually becoming apparent and are widely employed in economic research [14]. In light of the considerations above, this study has selected 117 real estate enterprises as research samples, comprising 37 in the default group and 80 in the non-default group. Secondly, the credit risk indicator system is constructed from various dimensions, including debt repayment ability, profitability, operating ability, growth ability, and cash flow. Finally, based on machine learning technology, the random forest model is used to monitor and issue early warnings regarding the risk of real estate enterprises in the current situation.

This paper makes two contributions to the existing literature. One notable limitation of existing studies is their reliance on traditional risk measurement methods, such as logistic and KMV, to assess real estate corporate risk. There is a paucity of scholarship employing machine learning techniques to quantify risk. This paper uses the random forest machine learning method to measure real estate enterprises' credit risk, providing a new method for risk monitoring and early warning. Secondly, the eigenvalues of the random forest model were employed to rank the credit risk influencing factors of real estate enterprises. The results indicated that the asset-liability ratio, the operating income growth rate, the cash ratio, the net profit margin on sales, and the total cash to liabilities ratio exert the most significant influence on the credit risk of real estate enterprises.

2. Literature review

2.1. Credit Risk of Real Estate Enterprises.

2.1.1. *Causes of Credit Risk in Real Estate Enterprises.* The research on the mechanism of credit risk generation in the context of real estate enterprises can be divided into three categories. The initial category of research concerns the decline in housing prices, which may disrupt capital chains and, subsequently, the emergence of credit risk for real estate enterprises. An increase in house prices will result in a corresponding expansion of the scale of real estate enterprises, leading to a financial structure characterized by high debt, high leverage, and high turnover [11]. A decline in housing prices can readily result in the disruption of the capital chain of real estate enterprises and a reduction in liquidity, thereby increasing the risk of default on real estate enterprise credit [7, 16]. Secondly, a decline in consumers' psychological expectations of the real estate market will also exacerbate risks in real estate enterprises. In an economic downturn and when residents' income expectations are low, consumers' psychological expectations will decline, reducing consumer demand in the real estate market [9]. A decline in demand in the real estate market will hurt the operating income of real estate enterprises, reducing the liquidity of enterprise funds and subsequently increasing the probability of credit default for real estate enterprises. Thirdly, the real estate enterprise's mismanagement constituted the proximate cause of the credit default. The inadequate management of real estate enterprises, including the indiscriminate borrowing of funds to expand operations, the misappropriation of capital for disparate projects, and the inability to diversify investments, has resulted in significant challenges regarding operating income, leverage, and liquidity. This, in turn, has led to the credit default of real estate enterprises [1].

2.1.2. *Credit Risk Measurement for Real Estate Enterprises.* Most existing studies focus on measuring real estate enterprises' credit risk through conventional models, such as the KMV and logistic models. On the one hand, scholars employ financial index data to construct a credit risk evaluation system, utilizing the logistic model to assess the credit risk of real estate enterprises [10]. For example, Zhang et al. [20] put forth a hybrid neural network comprising an attention-based convolutional neural network (CNN) and a bidirectional long short-term memory (BiLSTM) unit, which they employed to predict the credit risk of listed real estate enterprises by integrating the characteristics of logistic regression outputs. Conversely, scholars use the KMV model to assess the credit risk of real estate companies, utilizing capital market data [19].

2.2. **Research related to the application of machine learning models in economics.** The advantages of automated pattern recognition, continuous self-optimization, and effective handling of nonlinear, high-granularity, and large-sample data have enabled intelligent technologies such as big data and machine learning to significantly improve the accuracy of data prediction and the efficiency of decision support. These technologies have played an essential role in economics, particularly in risk monitoring and early warning [4].

The first category of literature pertains to the utilization of machine learning models in the domain of economic growth forecasting and causality identification. Firstly, machine learning models were applied in economic growth forecasting. Chen et al. [5] used a random forest algorithm to predict Treasury bond futures index changes in their study. Moreover, scholars have employed machine learning to predict asset prices, agricultural price fluctuations, natural gas uranium prices, agrarian supply chain finance [12], and other related variables. Secondly, the application of machine learning models in identifying causal relationships. Zhou et al. [21] employed a dual machine learning methodology to assess the influence of the digital economy on urban ESG development. Finally, machine learning models are utilized to monitor individual economic behavior. Huang et al. [8] employed a machine learning model to monitor the trading behavior of the cryptocurrency market, effectively identifying instances of price manipulation and shipping fraud on online social platforms. The second literature category is applying machine learning models to risk metrics. Scholars have employed machine learning algorithms to assess corporate credit risk, demonstrating that the model exhibits superior classification and identification capabilities and enhanced stability compared to alternative models, including logistics and KMV [2,18]. Peng et al. [15] utilized a machine learning approach to assess farmers' credit risk and validated this method's efficacy against a traditional model. The third category of literature is the application of random forest models in risk management. Scholars have found that random forest models can effectively identify key credit risk factors and predict default probabilities. Furthermore, the random forest model has been shown to possess notable advantages and practicality in model prediction and risk early warning when compared to other algorithms, including boosting, bagging, decision tree, and support vector machine [13].

In conclusion, the extant literature has predominantly employed conventional metrics, such as Logistic and KMV, to quantify real estate corporate credit risk. However, when many economic variables are used as risk predictors, the limited dimensions of the variables that traditional metric models can accommodate make it challenging to integrate them with macro-micro-mixed-frequency big data [17]. This ultimately results in an inability to capture risk's complex and dynamic nature fully, limiting the effectiveness of risk analysis and judgment. The application of machine learning models, exemplified by random forests, has led to notable advancements in the accuracy of data prediction and the efficiency of decision support. These models possess the capabilities of automated pattern recognition, continuous self-optimization, and robust generalization and prediction, which have become instrumental in risk monitoring and early warning. Nevertheless, a paucity of studies have employed the random forest model in the domain of credit risk assessment for real estate enterprises. Consequently, this study has adopted the random forest model to quantify the credit risk of real estate enterprises and ascertain the pivotal risk factors that significantly influence the risk.

3. Methods and data

3.1. Random forest model.

3.1.1. *Method theory.* The Random Forest algorithm was initially proposed by Breiman [3] in 2001. It is an integrated learning method, integrated by multiple decision trees, which can handle large amounts of data and does not require censoring or variable filtering on large data sets. This makes it possible to comprehensively and accurately capture the credit risk characteristics of enterprises. By Breiman's definition, a random forest can be conceptualized as an integrated prediction model, R, which can be expressed as follows: $R = \{h(x,\ \theta_k),\ k = 1, 2, 3, \ldots, K\}$, where θ, k represents a random vector that obeys an independent homogeneous distribution, and k denotes the number of decision trees in the random forest. Given the independent variable x, each decision tree model decides the optimal classification result by voting. A random forest is a prediction model comprising multiple decision trees. Let's consider a decision tree to be an expert in classification. A random forest can be viewed as a collective of experts working together to classify a given task.

Given the original training dataset $\gamma = (X, Y)$ the number of samples is N, $(x_1, y_1), \ldots, (x_n, y_n), \ldots, (x_N, y_N)$ each sample contains D features. The main steps of the Random Forest algorithm include drawing M training samples: the Bagging technique is used to select N samples in the dataset (X, Y) in a releasing manner, constituting a new set of samples (X^m, Y^m), (x_1^m, y_1^m),..., (x_n^m, y_n^m),..., (x_N^m, y_N^m). The samples that were not taken form the OOB set.

Train M CART trees: construct CART trees on each set of drawn training samples (X^m, Y^m). D_m features $D_m < D$ are randomly sampled from the original D features for each tree node. Then, in the feature subspace composed of D_m features, split features are selected according to the maximum impurity descent method. Finally, the splitting yields the next level of child nodes. Keep looping the above process until the stopping condition of decision tree construction is satisfied (the number of samples in leaf nodes is less than a threshold, etc.).

The classification results are determined by majority voting. This entails using each CART tree in the forest to predict the out-of-bag (OOB) samples independently. Subsequently, a vote is taken based on the results of each tree, and the final predicted category is output using majority voting. Additionally, the number of votes cast for each category in the voting result can be counted, and the category with the most significant number of votes can be identified as the predicted category. The corresponding number of votes can then be taken as the expected probability.

3.1.2. *Feature Importance Score.* Estimating the importance of features to a classification problem by calculating their respective importance scores represents a fundamental aspect of the random forests methodology. Random forests typically employ the average precision descent method to measure the importance of features. The algorithmic process comprises eight principal stages, as follows:

Step 1: Bag samples the original training dataset $\gamma = (X, Y)$ to obtain a set of sample subsets $\gamma_1, \ldots, \gamma_m, \ldots, \gamma_M$.

Step 2: Use the sample subset $\gamma_1$to train the decision tree α_1, the out-of-bag samples of the current decision tree are δ_1^{oob}.

Step 3: Apply the decision tree α_1 to predict the out-of-bag sample of δ_1^{oob}, and the number of correctly categorized samples in the prediction result is noted as R_1^{oob}.

Step 4: For features $d = 1, 2, \ldots, D$, randomly disrupt the order of the dth feature in sequence on the out-of-bag sample as δ_1^{oob} to form D new out-of-bag samples $\delta_{1,d}^{oob}$.

Step 5: On the D new out-of-bag samples, apply the decision tree α_1 to make a prediction. The number of correctly categorized samples in the prediction result is denoted as $R_{1,1}^{oob}, \ldots, R_{1,d}^{oob}, \ldots, R_{1,D}^{oob}$.

Step 6: Repeat steps 2 to 5 for the subset of samples $\gamma_1, \ldots, \gamma_m, \ldots, \gamma_M$, to obtain the number of correctly classified samples before and after the perturbation $\{R_2^{oob}, R_{2,d}^{oob}, \ldots, R_{2,D}^{oob}\}$,..., $\{R_M^{oob}, R_{M,1}^{oob}, \ldots, R_{M,D}^{oob}\}$.

Step 7: Calculate the importance score of the dth feature $P_d = \frac{1}{M} \sum_{m=1}^{M} (R_m^{oob} - R_{m,d}^{oob})$.

Step 8: Collect the importance scores of all D features.

The fundamental premise of MDA is that if feature d substantially influences the predictive outcome, the number of accurate samples will be markedly reduced when the perturbation is introduced. In the majority of cases, the importance score will be positive. However, as illustrated in the importance formula in step 7, the score for the feature may be close to zero or negative. A score approaching zero indicates that adding a perturbation to the data for feature d does not affect the prediction results. This suggests that the feature is irrelevant to the prediction results and may be considered irrelevant.

3.1.3. *Model Effectiveness Evaluation.* The receiver operating characteristic (ROC) curve is a graphical representation that demonstrates the efficacy of a classification model in forecasting defaults on real estate loans. The horizontal coordinate of the ROC curve represents the false positive rate (FPR) in the prediction of defaults on real estate loans, while the vertical coordinate represents the actual positive rate (TPR) in the prediction of defaults on real estate loans. The closer the distance between the model's ROC curve and the upper-left corner of the axis, the more effective the model is at classifying and predicting defaults on real estate loans.

AUC represents the area under the ROC curve, which is a quantitative indicator with a value between 0 and 1. It can be used to evaluate the overall prediction performance of a classification model. In comparison to accuracy, AUC has the advantage of being less affected by data imbalance, thus is more commonly used as an evaluation index to comprehensively evaluate the prediction performance of the model in the context of loan default prediction for real estate enterprises. Table 1 below presents a summary of the default prediction performance of models with varying AUC values.

Table 1. Evaluation of default prediction performance of models with different AUC values

AUC value	Evaluation of the model's default prediction performance
0-0.5	Average prediction performance, even weaker than random prediction
0.5-0.7	Better predictive performance
0.7-0.85	Predicts performance very well
0.85-0.95	Excellent predictive performance
0.95-1	Unlikely to be realized in practice when AUC=1 represents perfect prediction

3.2. **Data.** The present study employed a sample comprising 117 real estate listed companies. Of the aforementioned companies, 37 were classified as the default group, having either defaulted on their bonds or having a credit rating below B or Special Treatment (ST). The remaining 80 companies that did not default on their bonds and had a non-ST rating were designated as the non-default group. The data presented in this paper were sourced from the Wind database and the CSMAR database. The empirical analysis employed an 8:2 ratio for the training and test sets, respectively.

4. Empirical analyses

4.1. **Construction of credit risk indicator system.** This paper proposes a novel credit risk measurement index system for real estate enterprises, which is constructed through an in-depth analysis of the causes of credit risk in this sector. The index system is derived from 19 secondary indicators selected from five key dimensions of solvency, growth ability, profitability, operating ability and cash flow.(as illustrated in Table 2). This paper presents an overview of the descriptive statistics associated with the eigenvalues of the samples. Regarding solvency (X1-X4), the mean value of the gearing ratio (X1) is 65.9%, while the standard deviation reaches 22.1%. This indicates that the overall solvency of the firms within the sample is relatively stable, although there may be stratification and some degree of variability across the different categories of the sample. In terms of growth capacity (X5-X8), the operating revenue growth rate (X5) and net profit growth rate (X6) are 5.010 and 9.509, respectively. This suggests that the majority of enterprises are operating within normal parameters. With regard to profitability (X9-X13), the standard deviation of the indicator indicates that the sample is generally stable, and the medians of the firms within the sample are all positive, which suggests that the firms in the sample are still performing well in terms of profitability. With regard to operational capacity (X14-X17), the standard deviation of accounts receivable turnover and inventory turnover reached 144.588 and 96.109, respectively. This suggests that there is a significant disparity between the samples in terms of operational capacity. Nevertheless, the standard deviations of the cash flow indicators (X18-X19) are relatively minimal, suggesting a relatively balanced cash flow position between the samples.

4.2. Parameterization of the Random Forest Model.

4.2.1. *A sample equalization treatment.* Given the relatively limited number of default group samples included in this study, the resulting data set was inherently imbalanced. To mitigate the risk of overfitting the model during training, the SMOTENC method was employed to oversample the data. In accordance with the tenets of the SMOTENC methodology, an exceedingly large numerical value will exert an influence on the novel samples generated through the Euclidean distance. Consequently, this study initially employs the min-max approach to normalize the data samples. As illustrated in Table 3, the expansion of the SMOTENC method permitted the original 117 data points to be augmented to 160, thereby enabling the default group samples within the dataset to be more comprehensively leveraged.

4.2.2. *Parameter Optimization for Random Forest Models.* Modifying the pivotal parameters enables the model to be more accurately aligned with the attributes of the data, thereby mitigating the impact of generalization error and enhancing the efficacy of the risk assessment. As illustrated in Table 4, the grid search and cross-validation techniques are employed for the purpose of optimizing the model. The resulting values for the random_state and n_estimators parameters are 1 and 100, respectively, while the max_depth parameter is set to 4.

TABLE 2. Credit risk indicator system

The type of indicator	Indicator Symbols	Define	Average	Std
Solvency	X1	Gearing ratio	0.659	0.221
	X2	Cash ratio	0.469	0.815
	X3	Current ratio	1.925	1.860
	X4	Quick ratio	0.767	0.889
Growth Capacity	X5	Operating income growth rate	-0.463	5.010
	X6	Net profit growth rate	-1.901	9.509
	X7	Total assets growth rate	-0.087	0.504
	X8	Total liabilities growth rate	-0.04	0.326
Profitability	X9	Return on equity	-0.695	6.650
	X10	Return on total assets	-0.002	0.056
	X11	Return on assets	0.023	0.107
	X12	Gross profit margin	0.388	1.451
	X13	Net profit margin on sales	-0.022	0.889
Operating Capacity	X14	Total asset turnover	0.240	0.248
	X15	Accounts receivable turnover ratio	69.092	144.588
	X16	Inventory turnover ratio	18.154	96.109
	X17	Fixed assets turnover ratio	59.076	173.892
Cash Flow	X18	Ratio of cash to current liabilities	-0.038	0.965
	X19	Total cash to liabilities ratio	-0.058	0.916

TABLE 3. Results of the sample equalization processing

	Sample size of the default group	Sample size of non-default group
Pre-Expansion	37	80
post-expansion	80	80

TABLE 4. Important parameters of Random Forest

Parameters	Hidden meaning
N_estimators	Number of trees in forests
Max_features	Limit the number of features considered when branching
Max_depth	Maximum depth of the tree

4.3. **An examination of the efficacy of the random forest model in measuring credit risk for real estate enterprises.** In this paper, the credit risk of real estate enterprises is quantified and evaluated through the construction of an ROC curve, and the results show that the ROC curve of the random forest model demonstrates an upward trajectory. The area under the curve (AUC) value for the credit risk prediction model is 81.25%, with an accuracy rate of 83.33%, a recall rate of 75%, and a precision rate of 75%. This indicates that the random forest model extended by combining SMOTENC technology is capable of more accurately measuring credit risk data and demonstrating an excellent fitting effect. In light of these findings, it can be concluded that the random forest model demonstrates excellent applicability to the measurement of credit risk in real estate enterprises.

4.4. **An analysis of the importance of credit risk characteristics of real estate enterprises based on the random forest model.** This paper presents a ranking of the importance of credit risk characteristics of real estate companies, the five most important indicators, as determined by the ranking system, are the gearing ratio, operating income growth rate, cash ratio, net profit margin on sales, and total cash to liabilities ratio. A high gearing ratio (X1) in solvency indicates that in the event of poor operational performance, the company may encounter challenges in meeting debt obligations in a timely manner, which increases credit risk. Consequently, as a significant factor influencing the credit risk of real estate enterprises, the gearing ratio, whether high or low, will directly impact the enterprise's solvency and financial stability. A low operating income growth rate (X5) in growth capacity indicates that the enterprise is subjected to considerable market competition and exhibits suboptimal operating efficiency. It is, therefore, incumbent upon real estate companies to direct their attention to the improvement of the operating income growth rate in order to achieve long-term sustainable development. A low cash ratio (X2) in solvency indicates a deficiency in cash reserves, which may impede the ability of real estate companies to meet maturing debt obligations in a timely manner. Consequently, it is imperative for real estate companies to adopt a prudent approach to cash flow management, ensuring the availability of sufficient cash reserves to navigate potential debt repayment challenges and operational risks. A low net profit margin on sales (X13) profitability may be perceived as indicative of suboptimal business operations or the presence of potential risks, which may ultimately reduce investors' willingness to invest in the company. It is therefore recommended that real estate companies adopt measures to improve the net profit margin on sales in order to achieve long-term sound development. A low total cash to liabilities ratio (X19) in cash flow is indicative of a lack of solvency and a poor financial condition, which, in turn, increases credit risk. It is therefore recommended that real estate companies implement effective measures to manage accounts receivable to reduce credit risk.

5. Conclusions

The present study utilises the random forest model to assess the credit risk of Chinese real estate firms. The results demonstrate that the credit risk measurement model of real estate enterprises based on random forest exhibits superior performance. The model demonstrates a predictive accuracy of 83.33% for bond default in the subsequent year, with an AUC value of 81.25% and a precision rate of 75%. These findings suggest that the random forest model has considerable applicability to the credit risk measurement of real estate enterprises. Secondly, the importance ranking of the eigenvalues of the random forest reveals that the asset-liability ratio, operating income growth rate, cash ratio, net profit margin on sales, and total cash to liabilities ratio exert a significant influence on the credit risk of real estate enterprises.

The findings of this study corroborate those of Zhang et al. [20]. Both studies emphasize the potential of machine learning techniques in improving the accuracy of credit risk prediction. The distinction between the two studies lies in the fact that this one not only considers the predictive capacity of the model but also places particular emphasis on the assessment of feature importance. This allows financial institutions to gain a deeper understanding of the key factors affecting the credit risk of real estate firms, thus enabling them to formulate more accurate risk management strategies. Secondly, the five key risk factors identified in this study (asset-liability ratio, operating income growth rate, cash ratio, net sales interest rate, and total cash-to-debt ratio) are consistent with those identified by Li et al. [11] through the Random Forest Fusion methodology. The distinction is that this study places greater emphasis on risk factors that are specific to the real estate industry, as opposed to merely generic financial indicators. Finally, the AUC value of the random forest model in this study (81.25%) is comparable to the performance of the credit risk model assessed by Peng et al. [15] using a machine learning approach. The model in this study exhibits superior performance in terms of accuracy (83.33%) and precision (75%), which may be attributed to the SMOTENC technique employed during the model training process. This technique helps to balance the dataset and enhance the model's generalization capabilities.

It is recommended that real estate companies enhance their sensitivity to the asset-liability ratio and operating income growth rate and reinforce their focus on solvency to more effectively mitigate credit risk. Firstly, in terms of solvency, it is recommended that the management of funds be strengthened and the liability structure be optimized in order to ensure the timely repayment of mature debts, a rational arrangement of debt repayment plans, and the avoidance of short-term debt concentration. Secondly, with regard to profitability, it is recommended that a diversified business strategy be implemented with a view to reducing the enterprise's dependence on a single market or business, thus enhancing the stability of profitability. In terms of operational capacity, supply chain management can be optimized to enhance capital utilization efficiency and productivity, thereby strengthening the enterprise's operational capacity. Concurrently, real estate enterprises may enhance their capital turnover by accelerating the pace of project promotion and optimizing inventory management. With regard to the cash flow position, it is recommended that cash flow management be strengthened and that a sound cash flow forecasting and monitoring mechanism be established in order to ensure the stability and adequacy of cash flow.

References

[1] E. I. Altman, X. Hu, and J. Yu, *Has the Evergrande debt crisis rattled Chinese capital markets? A series of event studies and their implications*, Finance Research Letters **50** (2022): 103247.

[2] O. A. Bello, *Machine learning algorithms for credit risk assessment: an economic and financial analysis*, International Journal of Management **10** (2023), 109–133.

[3] L. Breiman, *Random forests*, Machine Learning **45** (2001), 5–32.

[4] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, *Explainable machine learning in credit risk management*, Computational Economics **57** (2021), 203–216.

[5] B. J. Chen, and F. Wang, *Macroeconomic Indicators, technical indicators and treasury bond futures price forecasting - An empirical test based on random forest machine learning*, Statistics and Information Forum **34** (2019), 29–35.

[6] P. Eichholtz, M. Korevaar, T. Lindenthal, and R. Tallec, *The total return and risk to residential real estate*, The Review of Financial Studies **34** (2021), 3608–3646.

[7] C. Hochstenbach and M. B. Aalbers, *The uncoupling of house prices and mortgage debt: Towards wealth-driven housing market dynamics*, International Journal of Housing Policy **24** (2024), 642–670.

[8] J. M. Huang, H. F. Pan and T. Hu, *A machine learning model for monitoring crypto digital currency manipulation*, Scientific Decision Making No. 01 (2023), 42–55.

[9] T. Kuchler, M. Piazzesi and J. Stroebel, *Housing Market Expectations*, Handbook of Economic Expectations, Academic Press, 2023.

[10] A. Levy and R. Baha, *Credit risk assessment: a comparison of the performances of the linear discriminant analysis and the logistic regression*, International Journal of Entrepreneurship and Small Business **42** (2021), 169–186.

[11] Z. H. Li and M. H. Zhu, *Pre-sale of commercial properties and credit risk taking of commercial banks-Empirical evidence from local banks in China*, International Financial Studies No. 07 (2024), 62–72.

[12] Y. Liu and H. Yan, *A credit risk evaluation on supply chain financing for farmers' cooperatives*, Journal of Nonlinear and Convex Analysis **21** (2020), 1813–1828.

[13] M. Madaan, A. Kumar, C. Keshri, R. Jain and P. Nagrath, *Loan default prediction using decision trees and random forest: A comparative study*, IOP Conference Series: Materials Science and Engineering, IOP Publishing **1022** (2021): 012042.

[14] V. Moscato, A. Picariello, and G. Sperlí, *A benchmark of machine learning approaches for credit score prediction*, Expert Systems with Applications **165** (2021): 113986.

[15] Y. L. Peng, Y. J. Peng, H. L. Zhou, S. Y. Wang and Y. S. Jiang, *Machine learning-based credit risk identification of farmers' farmland management right mortgage and its loss metric*, Systems Engineering Theory and Practice No. 02 (2025), 448–462.

[16] E. J. Ross and L. Shibut, *Loss given default, loan seasoning and financial fragility: evidence from commercial real estate loans at failed banks*, The Journal of Real Estate Finance and Economics **63** (2021), 630–661.

[17] C. H. Shen, and Y. Liu, *A verification model to capture option risk and hedging based on a modified underlying beta*, Journal of Risk Model Validation **15** (2021), 49–68.

[18] G. Teles, J. J. P. C. Rodrigues, R. A. L. Rabelo and S. A. Kozlov, *Comparative study of support vector machines and random forests machine learning algorithms on credit operation*, Software: Practice and Experience **51** (2021), 2492–2500.

[19] H. W. Wang, and J. W. Xu, *Credit risk changes of listed real estate companies before and after the new crown epidemic–an analysis based on KMV model*, Modern Business No. 04 (2021), 19–23.

[20] X. S. Zhang, Y. L. Ma and M. H. Wang, *An attention-based Logistic-CNN-BiLSTM hybrid neural network for credit risk prediction of listed real estate enterprises*, Expert Systems **41** (2024): e13299.

[21] J. P. Zhou, W. X. Xu, Z. F. Mi and C. J. Liu, *The impact of the digital economy on urban ESG development - a test based on dual machine learning approach*, Geography Research **43** (2024), 1407–1424.

[22] M. Zurek, *Real estate markets and lending: does local growth fuel risk?*, Journal of Financial Services Research **62** (2022), 27–59.

Manuscript received October 7, 2024
revised January 20, 2025

H. Jiang
School of Economics and Trade, Shandong Management University, Jinan 250357, China
E-mail address: 14438120210412@sdmu.edu.cn

J. Cui
School of Economy, Shandong Women's University, Jinan 250300, China
E-mail address: 30087@sdwu.edu.cn

Y. Liu
School of Economy, Shandong Women's University, Jinan 250300, China
E-mail address: lyang@sdwu.edu.cn