International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-13

https://doi.org/10.70917/ijcisim-2026-3132
Article

Developing a Strategic Crisis Management Model for Enhancing Stability of the Real Estate Sector: A Case Study of Chinese Cities

Xueying Zhang [1,*], Gerelmaa Jamsran [1], Narmandakh Dolgor [1] and Huadong Zhang [1]

[1] Global Leadership University (GLU), Ulaanbaatar, 11000, Mongolia
* Correspondence author: zhangxueying1717@163.com

Abstract: The real estate bubble and economic crisis have significant impacts on Chinese real estate enterprises. This paper constructs a strategic crisis management model for improving the stability of the real estate industry based on a machine learning model. Based on the research data of cases in Chinese cities, the PSR model is introduced to construct a strategic crisis warning index system. It is proposed to use the LASSO feature selection method to conduct refined screening of feature indicators, and use it as the input layer of XGBoost. After optimizing the hyperparameters using the genetic algorithm, the model training and prediction are carried out. The results show that the LASSO-GA-XGBoost model has a higher warning accuracy rate for the stability crisis of the real estate industry compared to other models. The average ACC value and AUC value for early three-year crisis warning are 96.39% and 90.71%, respectively. Moreover, the prediction results have smaller errors compared to the actual results. Based on SHAP, the parameter features are further analyzed, and it is found that the importance ranking is the growth rate of real estate development investment and the rolling year-on-year income of land transfer fees.

Keywords: PSR model; LASSO regression; genetic algorithm; XGBoost; real estate industry stability crisis warning

1. Introduction

Among all industries in China, the real estate industry started relatively late. It truly took shape after the government announced the implementation of the market economy system in 1992. Although its development has been rapid, it is still in the initial stage, which makes the real estate market less mature and leads to more changing factors [1-4]. In addition, the real estate project development involves numerous steps, has a long cycle, requires large single investment amounts, and is immovable [5-6], which makes strategic crisis management in the real estate industry particularly important.

Strategic crisis management is a new management category formed by enterprises based on their exploration of the patterns of crisis occurrence and summary of experiences in handling crises. It represents the deepening of crisis handling by enterprises and their proactive response to crises [7-9]. The content of strategic crisis management mainly includes: prediction and management before the crisis occurs, emergency handling during the crisis, and post-crisis work. The main purpose of strategic crisis management is to analyze and predict possible crises based on the enterprise's own situation and the external environment, and then formulate targeted measures to prevent problems before they occur and minimize the possibility of crisis occurrence [10-13]. The strategic crisis management for the stability improvement of the real estate industry can be divided into three types: rapid expansion financial strategy with the characteristics of "high debt, high expected return, and low distribution", stable development financial strategy with the characteristics of "low debt, high return, and moderate distribution", and defensive contraction financial strategy with the characteristics of "low debt, low return, and low distribution" [14-17]. Due to the complexity and variability of the social economic environment, as well as the reasons of the enterprises themselves, when constructing and applying the strategic crisis management model, all factors must be comprehensively considered.

This paper first introduces the PSR model to construct a crisis warning index system for the stability of the real estate industry from the three levels of pressure, state, and response, including 13 indicators. It uses the Pearson correlation analysis method and Lasso regression method to test the correlation between features and eliminate redundant features. Then, based on the XGBoost algorithm, a crisis warning base model for the stability of the real estate industry is constructed, and the hyperparameters of it are optimized using the genetic algorithm. Finally, the SHAP additive explanation algorithm is used to extract the key indicators that have an impact on the crisis warning results of the stability of the real estate industry, and these key indicators' values are displayed in a visual way to show the impact of these key indicators on the crisis warning results of the stability of the real estate industry.

2. Index System and Data Sources

2.1. Stability Crisis Early Warning Index System for the Real Estate Industry

This paper introduces the PSR model to construct a strategic crisis early warning indicator system for enhancing the stability of the real estate industry. In the PSR model, the elements at the pressure (P) level typically represent the damage and disturbance caused by external factors to the system, the elements at the state (S) level represent the current state of the system under pressure, and the response (R) is the countermeasure taken by the system when facing risk pressure. This paper selects 13 indicators from the pressure-state-response level to construct a crisis early warning indicator system for the stability of the real estate industry. Table 1 shows the crisis early warning indicator system for the stability of the real estate industry.

Table 1. The real estate industry stability crisis warning index system

Target layer	Index	Code
Pressure layer	Money supply and economic growth gap	X1
	The mortgage of the house is poor	X2
	The people buying home buy panic public opinion index	X3
State layer	Real estate development investment growth	X4
	The commodity housing cycle	X5
	House price income is more than the long-term deviation	X6
	The land is rolling in revenue	X7
	New start and sales area coefficient	X8
	The second-hand house is listed for the transaction ratio	X9
Response layer	Head housing companies link short debt coverage	X10
	Local regulation policy easing index	X11
	The size of the property acquisition in the local AMC	X12
	The local state-owned enterprises account for the amount of the land	X13

2.2. Data Sources and Preprocessing

The data sources of this article include macro financial data: the official website of the People's Bank of China, and the "China Statistical Yearbook" of the National Bureau of Statistics; micro enterprises: CSMAR (Guotai An) and CNRDS (China Research Data Service Platform) real estate listed company financial reports; market transactions: the China Real Estate Index System (CREI), and the CRIC database; policy texts: the local regulations database of Peking University's Baodaifa, combined with Python web scraping and sentiment analysis to generate. 35 cities were selected (4 first-tier cities, 26 second-tier cities, and 5 typical third-tier cities), with a time span from 2014 to 2025, totaling 1680 sample points.

The quality of the dataset will affect the upper limit of the model's prediction effect. Therefore, the first step in model construction is to analyze and process the dataset to prepare for the subsequent algorithms. This article first conducts a preliminary analysis of the feature data and sample distribution, then performs missing value handling, categorical variable processing, and data standardization processing, and conducts a preliminary exploration of the relationships between features through Pearson correlation coefficients to provide a reference for subsequent feature selection.

(1) Descriptive statistics. This article statistically analyzes the maximum value, minimum value, mean, standard deviation, and variance of the features, making a preliminary judgment on the features and providing a basis for data standardization processing.

(2) Sample distribution exploration. This article explores the sample distribution through visualization methods, providing a basis for the subsequent handling of sample imbalance problems.

(3) Missing value handling. Data missing situations can generally be divided into two types: one is

the missing situation of a certain feature's data, and the other is the missing situation of a certain sample's data. If the missing situation is ignored, it will affect the model's output results. If it is directly deleted, it will affect the model's generalization ability due to the reduction in data volume. Therefore, different handling methods need to be selected according to different missing situations. This article first counts the missing situations of the samples, and for samples with a missing rate greater than 20%, this article uses the deletion method to directly delete them. Then, it counts the missing situations of the features, and for features with a missing rate greater than 20%, this article uses the mean method to fill in the data, otherwise, it directly deletes.

(4) Categorical variable processing. Computers cannot recognize categorical feature data with practical significance. Therefore, we need to encode it to convert categorical data into discrete data. This article uses one-hot encoding to process the categorical variables. One-hot encoding is to represent categorical features or discrete values using a similar binary number. The number of categories determines the dimension of the encoded feature. If it belongs to this category, then the value at the corresponding position in the encoding is 1, and the other category positions are all 0. This encoding extends the discrete feature's values to the Euclidean space, and the distance calculation between features can be made more reasonable through this encoding method.

(5) Standardization processing. Due to different calculation standards and units, the quantities of feature data may have significant differences. If the values of certain features are much larger than those of other features, their impact on the model will also be greater than other features, which may lead to the almost loss of data information of other features. Therefore, it is necessary to standardize the sample data.

The standardization processing method is:

$$y = \frac{x - \mu}{\sigma} \tag{1}$$

Here, μ represents the mean of the data, and σ represents the standard deviation of the data.

(6) Exploration of relationships between features. In this paper, the Pearson correlation coefficient is used to investigate the relationships between features, and a heat map is employed for visualization. The Pearson correlation coefficient can measure the linear correlation between features. The value range of r_{xy} is $[-1,1]$. If it is greater than 0, it indicates a positive correlation between the features; if it is less than 0, it indicates a negative correlation; and if it is equal to 0, it indicates no linear correlation between the features. Its calculation formula is as shown in Equation (2):

$$r_{xy} = \frac{\sum x_i y_i - n\overline{xy}}{\sqrt{\left(\sum x_i^2 - n\overline{x}^2\right)}\sqrt{\left(\sum y_i^2 - n\overline{y}^2\right)}} \tag{2}$$

Here, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

3. Model and Method

3.1. Extreme Gradient Boosting Algorithm and Parameter Optimization

The Xgboost algorithm was proposed by Chen Tianqi as a decision tree ensemble algorithm. It is an extension of the Gradient Boosting Decision Tree (GBDT). This algorithm can generate new trees during each iteration. It can combine and enhance weak learners with low classification performance into a strong learner with high accuracy. This algorithm has the advantages of fast training speed and good learning effect, and is therefore often used in the field of machine learning [18].

(1) Xgboost Principle

(a) Model Function

Given the dataset $D = \{X_i, y_i\}$, the input X_i is predicted by a linear superposition mode to obtain y_i. Suppose k trees are learned, the models used are as follows:

$$\begin{gathered} \hat{y} = \phi(X_i) = \sum_{k=1}^{K} f_k X_i, f_k \in F \\ F = \{f(X) = \omega_q(X)\}\left(q : R^m \to T, \omega \in R^T\right) \end{gathered} \tag{3}$$

Among them, $f(X)$ represents the regression tree, F represents the regression set, $q(X)$ is the classification at the leaf node, T is the number of leaf nodes, ω is the leaf node score, and $\omega_q(X)$ represents the prediction of $f(X)$ for the sample.

(b) Objective function

The objective function of the parameter space is as follows:

$$Obj(v) = L(v) + \Omega(v) = \sum_{i}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{4}$$

With $L(v)$ and $\Omega(v)$, the probability of model overfitting decreases, where $\Omega(f) = \gamma T + \frac{1}{2}\lambda\omega^2$:

(c) Model optimization

This step mainly involves finding the optimal parameters. The specific approach is to optimize the first tree of the CART, then the second tree, and so on. Each iteration of the objective function is based on the optimized model for tree construction. After t iterations, the calculation formula becomes:

$$\mathrm{Obj}(t) = \hat{y}_i^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t\left(X_i\right)\right) + \Omega\left(f_t\right) \tag{5}$$

After further calculation, the following results can be obtained: the tree structure form, the form of cumulative addition by leaf nodes, the optimal value of each leaf node, and the value of the objective function are respectively:

$$\begin{aligned}
\widetilde{L^{(t)}} &= \sum_{i=1}^{n}\left[g_i\omega_q\left(x_i\right) + \frac{1}{2}h_i\omega_q^2\left(x_i\right) + \gamma T + \lambda\frac{1}{2}\sum_{j=1}^{T}\omega_j^2\right] \\
\widetilde{L^{(t)}} &= \sum_{j=1}^{T}\left[G_i\omega_j + \frac{1}{2}\left(H_j + \lambda\right)w_j^2\right] + \gamma T \\
W_j^* &= -\frac{G_i}{H_j + \lambda} \\
\widetilde{L^{(t)}} &= -\frac{1}{2}\sum_{j=1}^{T}\frac{\left(\sum_{i\in I_j} g_i\right)^2}{\sum_{i\in I_j} h_i + \lambda} + \gamma T
\end{aligned} \tag{6}$$

The Xgboost algorithm has a wide range of applications. It can solve both classification and regression problems. Moreover, the loss function of this algorithm is more precise. Introducing the penalty factor into it can effectively avoid the overfitting problem and improve the prediction accuracy.

(2) Parameter Optimization of Xgboost Based on Genetic Algorithm

Since the genetic algorithm has the ability of global optimization in the problem of model parameter optimization and has a wide range of applications, this paper uses the genetic algorithm to optimize the parameters of Xgboost. The parameters that Xgboost needs to optimize include [max_depth, learning_rate, n_estimators, min_child_weight, subsample, reg_lambda]. The specific meanings of the related parameters are shown in Table 2.

Table 2. The relevant meaning of the xgboost parameter

Parameter name	Explain
Max_depth	The value is used to avoid fitting, representing the maximum depth of the tree
Learning_rate	This value is the learning rate, which reduces the weight of each step, and each round of iteration can avoid the fitting
N_estimators	This value represents the number of iterations, the number of decision trees
Min_child_weight	The effect of this parameter is to avoid excessive fitting, but the value is not suitable, otherwise it can lead to an underfitting
Subsample	The parameter represents the sampling rate, and its typical value is 0.5~ 1
Reg_lambda	This parameter controls the regularization part of the xgboost

This paper optimizes the 6 parameters of GA_XGBoost using genetic methods, which incorporate an elite retention strategy. Based on the genetic algorithm, an individual fitness calculation function $f(x)$ is established as follows:

$$\frac{1}{MSE} = n\frac{1}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \tag{7}$$

$$\frac{1}{H(p,q)} = \frac{1}{\sum_i p(i)\cdot\log\left(\frac{1}{q(i)}\right)} \tag{8}$$

Since the problems in this article are divided into regression problems (for predicting the comprehensive evaluation index) and classification problems (for predicting the evaluation grade), Equation (7) and Equation (8) are used as the fitness functions for regression problems and classification problems respectively, and relevant parameters are set. The specific steps are as follows:

(1) Set the parameters of the genetic algorithm (number of iterations T, population size M, crossover probability P_c, mutation probability P_m).

(2) Set the search range for the parameters of Xgboost [max_depth, learning_rate, n_estimators, min_child_weight, sub-sample, reg_lambda].

(3) Encode the parameters using binary coding.

(4) Randomly generate M individuals to form the initial population $pop(t), t=1$.

(5) Calculate the individual fitness of the population $pop(t)$ in the t-th generation based on the evaluation function $f(x)$.

(6) By using the selection operator in genetic operations (roulette wheel selection method), a new population $newpop(t)$ can be obtained.

(7) Based on the new population $newpop(t)$, generate a new population $crosspop(t)$ with a probability of P_c using the two-point crossover operator. Here, the elite retention strategy is adopted, and the optimal individual does not participate in the crossover operation.

(8) Based on $crosspop(t)$, generate a new population $mutpop(t)$ with a probability of using the single-point mutation operator. Here, the elite retention strategy is adopted, and the optimal individual does not participate in the crossover operation.

(9) Determine whether to stop the optimization. If $t < T$, then $t = t+1$, and the new population $pop(t) = mutpop(t-1)$ is returned to step 5. Otherwise, stop the evolution and enter step 10.

(10) Calculate the individual fitness of $mutpop(t)$ using $f(x)$, then select the individual with the best fitness, decode this individual to obtain the optimal parameters of Xgboost, and achieve automatic determination of parameters to improve the algorithm accuracy.

3.2. Lasso Regression for Variable Selection

Lasso was first introduced by Tibshirani in 1996. This method is a compression estimation that reduces the variable set, serving as an alternative algorithm to ridge regression [19]. Lasso regression penalizes the absolute values of the regression coefficients. It adds a penalty term, namely the L_1 regularization term, to the linear regression model. This penalty term compresses the regression coefficients of some irrelevant variables to 0, so these irrelevant variables will not have an impact on the model in the model, thereby achieving the purpose of variable selection. The objective function of Lasso is formula (9):

$$\hat{\beta}^{Lasso} = \left(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p\right) = \arg\min_{\beta\in R^p} \|Y - \beta X\|_2^2 + \lambda\|\beta\|_1 \tag{9}$$

When $\lambda \to \infty$ approaches infinity, the degree of compression on the estimated parameter β becomes greater, that is, it converges to 0. The number of variables retained in the model also decreases. As the value of λ gradually decreases, the number of variables in the model increases. During the calculation process, when there is multicollinearity or irrelevant items in the model, Lasso regression will eliminate these variables, achieving the effect of variable selection. In summary, to facilitate the understanding of Lasso regression, it can be regarded as a constrained optimization problem. Because the feasible region of the optimization problem is all restricted to convex sets with sharp points, it ensures the sparsity and enumerability of the optimal solution. Therefore, the constrained optimization problem can be solved.

3.3. Interpretability of the SHAP Model

For this article, establishing an effective stability crisis early warning model for the real estate industry is not enough. A good stability crisis early warning model for the real estate industry can only determine the probability of a stability crisis occurring in the real estate industry, but it cannot tell exactly which indicators caused the stability crisis. Therefore, the main purpose of this article is to study the specific impact of indicators on the final prediction results, so that listed companies can adjust their operational strategies as early as possible to avoid falling into a stability crisis.

How to make complex models interpretable, the SHAP interpretable framework is a very good tool. SHAP is an interpretive method for machine learning models proposed by Lundberg (2017) [20]. As a classic post-hoc interpretation framework, SHAP gives an importance value (Shapley Value) for each sample and each feature respectively, thereby explaining the model's prediction results. For the output of a model, each feature becomes a participant in the final result of the model. Each sample in the model generates a corresponding SHAP value, and the SHAP value is the sum of the contribution degree values of each feature in that sample.

Suppose x_{ij} represents the j -th feature of the i -th sample, SHAP represents the predicted value of the model for the i -th sample, $SHAP_{base}$ represents the baseline of the model, that is, the average of the prediction contributions of all sample variables in the model, then the SHAP value satisfies the following equation:

$$SHAP_i = +shap\left(x_{i1}\right) + shap\left(x_{i2}\right) + \cdots + shap\left(x_{im}\right) \tag{10}$$

Among them, $shap\left(x_{ij}\right)$ represents the SHAP value of x_{ij}, and m represents the dimension of the features in the model. $shap\left(x_{ij}\right)$ can be understood as the contribution value of the j -th feature of the i -th sample to the final predicted value $\hat{y}_i$; when $shap\left(x_{ij}\right) > 0$, it indicates that this feature has a positive effect on the predicted value; when $shap\left(x_{ij}\right) < 0$, it indicates that this feature has a negative effect on the predicted value. On the one hand, the SHAP value can reflect the magnitude of the influence effect of the feature on the specific sample in terms of the sample's characteristics; on the other hand, it can also reflect the specific direction of the influence effect.

The SHAP value has three characteristics: local accuracy, incompleteness, and consistency.

(1) Local accuracy means that the sum of the contributions of all features is equal to the output of the model. This property satisfies the basic requirements of the additive explanation framework, that is:

$$shap(x) = SHAP_{base} + \sum_{j=1}^{m} shap_j * I_j \tag{11}$$

Among them, $SHAP_{base}$ represents the baseline SHAP value, $shap_j$ represents the SHAP value corresponding to feature j, and I_j is an indicator function. When this feature appears, its value is 1, otherwise it is 0.

(2) Absence, that is, the contribution of the missing feature is 0. Here, the absence is not that a certain feature value in the structured data is empty, but that a certain feature is not observed in the sample, that is:

$$I_i = 0 \Rightarrow shap_i = 0 \tag{12}$$

(3) Consistency means that if the model structure changes, but the influence of a certain feature on the output result increases or remains unchanged, then the contribution of that feature to the overall result also increases or remains unchanged.

Finally, $SHAP_i$ is mapped to the model's prediction probability for the sample through the Sigmoid function:

$$\hat{p}_i = \frac{1}{1+e^{-SHAP_i}} \tag{13}$$

When $\hat{p}_i \geq 0.5$, the sample is classified as 1, which represents a positive sample; otherwise, it is 0. Based on the above SHAP framework, this paper explains the established XGBoost stability crisis early warning model.

3.4. Evaluation Indicators for Model Performance

In order to analyze the stability crisis early warning capability of the LASSO-GA-XGBoost model, this paper uses two evaluation indicators: accuracy (ACC) and the area under the receiver operating characteristic curve (AUC). AUC can be understood as the probability that, when randomly selecting a positive sample and a negative sample, the model predicts the true positive rate is greater than the false positive rate. Before defining ACC and AUC, the confusion matrix defines true positive (TP), false positive (FP), true negative (FN), false negative (FN), true positive rate (TPR), false positive rate (FPR), true negative rate (TNR), and false negative rate (FNR).

Based on this, the definitions of ACC and AUC are as shown in Equations (16) and (17):

$$TPR = \frac{TP}{TP + FN} \tag{14}$$

$$FPR = \frac{FP}{FP + TN} \tag{15}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$AUC = \int_0^1 TPR\left(t_i\right) dFPR\left(t_i\right) \tag{17}$$

4. Results and Analysis

4.1. Selection of Indicators

(1) Correlation Analysis of Characteristics

When constructing the early warning model, we first used the Pearson correlation coefficient to measure the correlation between the selected financial indicators, which served as the basis for feature selection. During the selection process, we fully considered the actual meaning of the crisis warning indicators in the real estate industry and eliminated redundant features to improve the training efficiency of the early warning model. The results of the correlation analysis of the indicators are shown in Figure 1. Most of the features with an absolute value of the Pearson correlation coefficient greater than 0.8 were defined as highly correlated. We removed the highly correlated crisis warning indicators to avoid data redundancy and improve the training efficiency of the subsequent early warning model. From the figure, while eliminating, we can select one of the two highly correlated crisis warning indicators from the perspective of their actual meanings. It can be found that the correlation coefficients of all indicators in the figure are less than 0.8, and all indicators can be retained.

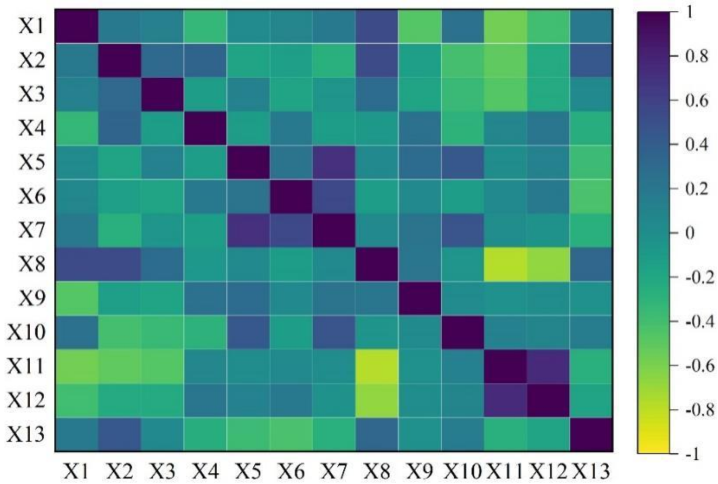


Figure 1. The correlation analysis results of the index

(2) LASSO Regression

To reduce the complexity of the predictive model structure and improve the accuracy of the prediction results, this paper constructs a LASSO regression to select the indicators that have a significant impact on the stability crisis of the real estate industry from the early warning indicator system. In the LASSO regression, the indicators with non-zero influence coefficients are the ones that have a significant impact on the stability crisis of the real estate industry. By setting different penalty parameters λ, the LASSO regression model obtains different optimal solutions to minimize the prediction error. The goal of this paper using LASSO regression to perform dimensionality reduction on the indicators is to control the number of indicators to be around 10. Therefore, when $\lambda = 0.009$, a set of optimal solutions containing 10 indicators that have a significant impact on the stability crisis of the real estate industry is obtained. The results of the influence coefficients are shown in Figure 2.

From the figure, it can be seen that there are 10 indicators that have significant early warning capabilities for the stability crisis of the real estate industry. Among them, X3 the resident housing purchase panic public opinion index, X5 the commodity housing devaluation cycle, X9 the second-hand house listing and transaction ratio, X12 the local AMC non-performing asset acquisition scale, and X13 the local state-owned enterprise platform land acquisition amount proportion have a positive impact on the stability crisis of the real estate industry, which means that the increase of these 5 indicators will make real estate enterprises more likely to have a stability crisis. X1 the money supply and economic growth gap, X4 the real estate development investment growth rate, X7 the rolling year-on-year increase of land revenue, X8 the ratio of new construction and sales area, and X11 the local regulatory policy tightness index have a negative impact on the stability crisis of the real estate industry, which means that the increase of these 5 indicators will make the real estate industry reduce the stability crisis. The other indicators have no significant impact on the stability crisis of the real estate industry, so they are not selected into the real estate industry stability crisis early warning model. The LASSO regression results show that the 10 selected indicators can be used as leading indicators for the stability crisis of the real estate industry.

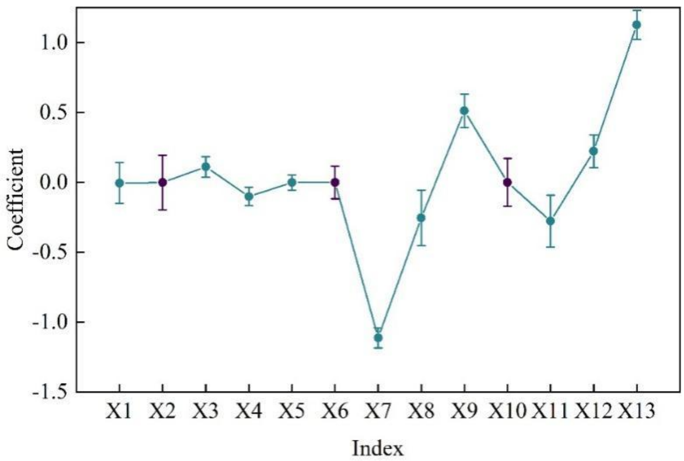


Figure 2. Lasso regression impact coefficient

4.2. Comparison of Model Validity

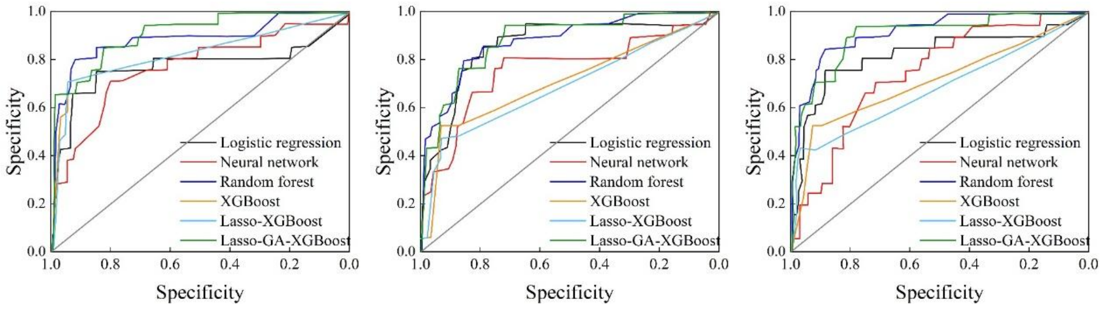
In order to comprehensively evaluate the effectiveness of the LASSO-GA-XGBoost model in predicting the stability crisis of the real estate industry, this paper uses two evaluation indicators to assess the model's predictive performance: accuracy (ACC) and the area under the receiver operating characteristic curve (AUC). This paper also selects logistic regression, artificial neural network, random forest, XGBoost, and LASSO-XGBoost to predict the stability crisis of the real estate industry, and conducts a comparative test of the predictive performance with the LASSO-GA-XGBoost model. Table 3 shows the ACC and AUC values of the crisis prediction models.

Based on the results of the two evaluation indicators, the warning framework based on the LASSO-GA-XGBoost model constructed in this paper performs better than other models in predicting the stability crisis of the real estate industry. The LASSO-GA-XGBoost model, using the financial indicators of the previous year, has the maximum ACC and AUC for predicting the stability crisis of the real estate industry, which are 97.49% and 91.68% respectively. In addition, the random forest model, as a Bagging-type ensemble algorithm, also performs well in predicting the stability crisis of the real estate industry. The ACC and AUC values are second only to the LASSO-GA-XGBoost model, at 96.65% and 89.95% respectively. Based on the results of the two evaluation indicators, this paper can draw the conclusion that the LASSO-GA-XGBoost model has a higher warning accuracy, more reasonable feature classification and sorting ability, and smaller error between the prediction results and the actual results compared to other models. This means that the warning framework based on the LASSO-GA-XGBoost model can provide more effective forward-looking warnings for the stability of the real estate industry in the event of a crisis. In contrast, the accuracy, classification ability, and error of the artificial intelligence models in predicting the stability crisis of the real estate industry are almost superior to those of the traditional logistic regression model, which once again verifies the importance of introducing artificial intelligence models into the stability crisis warning framework of the real estate industry.

Table 3. The crisis alert model ACC and AUC values

Early period	Index	LR	ANN	RF	XGBoost	Lasso-XGBoost	Lasso-GA-XGBoost	Mean
1	ACC	0.9209	0.9414	0.9665	0.9444	0.9382	0.9749	0.9477
	AUC	0.7799	0.7743	0.8985	0.835	0.8296	0.9168	0.839
2	ACC	0.9221	0.9442	0.9528	0.9137	0.9251	0.9531	0.9352
	AUC	0.8655	0.7631	0.8846	0.7084	0.6912	0.8893	0.8004
3	ACC	0.9361	0.9228	0.9468	0.9189	0.947	0.9637	0.9392
	AUC	0.829	0.754	0.9219	0.7185	0.6875	0.9151	0.8043
Mean	ACC	0.9264	0.9361	0.9554	0.9257	0.9368	0.9639	
	AUC	0.8248	0.7638	0.9017	0.754	0.7361	0.9071	

Additionally, the results of the true positive rate (also known as sensitivity) and the true negative rate (also known as specificity) obtained by setting different classification thresholds in this study were plotted as an ROC curve (Receiver Operating Characteristic curve). Figure 3 shows the ROC curves of the LASSO-GA-XGBoost model and other risk warning models when the early warning lead time is 1, 2, and 3 years. From the ROC curves, it can be seen that when the early warning lead time is 1, 2, and 3 years respectively, the ROC curves of the LASSO-GA-XGBoost model and the random forest model are closer to the upper left corner of the graph than other models. This indicates that the LASSO-GA-XGBoost model and the random forest model have more prominent warning capabilities for the stability crisis in the real estate industry when setting different threshold values for feature selection. Moreover, the logistic regression, neural network, XGBoost, and LASSO-XGBoost models show varying degrees of decline in prediction performance when the early warning lead years increase, especially XGBoost and LASSO-XGBoost. The prediction performance of XGBoost and LASSO-XGBoost drops sharply when the early warning lead time is 2 years and 3 years respectively. The reason for this is that the hyperparameters of XGBoost and LASSO-XGBoost models are set to default values, which have a significant impact on the classification results and thereby limit the warning ability of the XGBoost model for the stability crisis in the real estate industry. This further demonstrates the significance of using genetic algorithms to optimize the hyperparameters of the real estate industry stability crisis warning model in this study.



(a) t-1 (b) t-2 (c) t-3

Figure 3. The early warning period is the roc curve of the 1-3 year

4.3. Analysis of the Importance of Crisis Warning Indicators

In the SHAP method, the Shapley value can measure the marginal contribution of the input features of the model to the prediction result, and can be used as a metric to measure the influence of the features on the prediction result. Therefore, the average value of the absolute values of the Shapley values of each financial indicator is calculated as the measurement indicator of importance, and the selected indicators in this study are sorted, and the key features for crisis warning are extracted. The results of the feature importance of the indicators are shown in Figure 4. Each sub-figure represents the ranking results of feature importance in three different time windows of t-1 year, t-2 year, and t-3 year. The horizontal axis of the sub-figure represents the importance result of each financial indicator, and the vertical axis represents each feature, and it is sorted from top to bottom.

One year before the crisis (t-1), the key indicators that have an impact on the stability crisis warning of the real estate industry are the resident housing purchase panic public opinion index (X3), the growth rate of real estate development investment (X4), and the ratio of listing and transaction of second-hand houses (X9), which reflect the self-fulfillment of market expectations and the depletion of enterprise liquidity as the final trigger before the crisis outbreak. And from the ranking results of feature importance, in the two years before the crisis (t-2) and the three years before the crisis (t-3), the key indicators for the stability crisis warning of the real estate industry are the rolling year-on-year income of land transfer fees (X7), and compared with the important results of other features, the importance of this indicator is far greater than the important results of other indicators. We can find that the feature importance of some indicators will change as the time approaches the stability crisis of the real estate industry. For example, the ratio of listing and transaction of second-hand houses (X4), in the three years before the stability crisis of the real estate industry, it was only ranked eighth, but as time approaches, the ranking result of this indicator gradually moves forward.

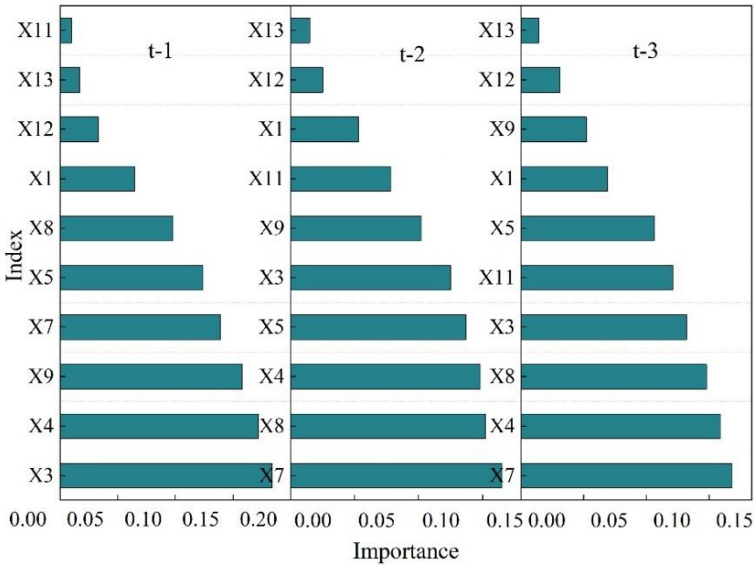


Figure 4. The characteristics of the index are the result

Based on the above analysis, we can summarize the key indicators for crisis warning of the real estate industry at different times. Moreover, by comparing and analyzing the changes in the importance of the same indicators under different times, it can provide key indicators for focus attention for market regulators, management personnel of enterprises, and market investors when conducting real estate industry stability crisis warning in practice. In different periods, by observing the changes of these key indicators, the possibility of enterprises experiencing stability crisis in the real estate industry can be judged, and thus corresponding prevention and management measures can be formulated in advance.

4.4. Analysis of the Attribution of Warning Results

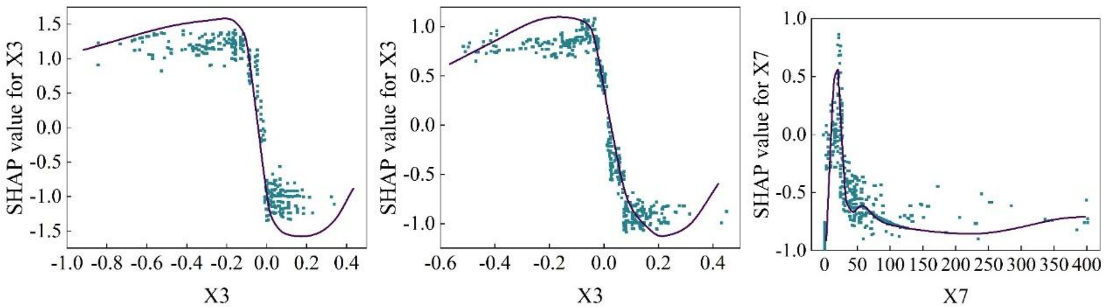
Based on the SHAP explanation algorithm, these key indicators' specific impact on the stability crisis in the real estate industry was further demonstrated through visualization. The SHAP dependency graph results of the indicators are shown in Figure 5. We sorted the feature importance in three different time windows (t-1, t-2, t-3) and took the most important features in each time window as examples to create their SHAP dependency graphs, showing the influence of each feature's value on the warning and discrimination results of the stability crisis in the real estate industry. As shown in the three subgraphs, the horizontal axis in the figure represents the feature's value, and the vertical axis represents the SHAP value of each feature. When the SHAP value is negative, it indicates that the feature makes a negative contribution to the warning result, and when the SHAP value is positive, it indicates that the feature makes a positive contribution to the warning result. The larger the absolute value of the SHAP value, the greater the marginal contribution. From the figure, we can find:

(1) For the resident housing panic public opinion index (X3) one year before the occurrence of the stability crisis in the real estate industry (t-1), when X3 gradually decreases within the range of -0.026 to -0.171, the SHAP value is positive and gradually increases. As X3's value continues to decrease, the

positive contribution to the stability crisis warning and discrimination result gradually increases, and the possibility of the stability crisis in the real estate industry gradually increases.

(2) Similarly, for the resident housing panic public opinion index (X3) two years before the occurrence of the stability crisis in the real estate industry (t-2), when the resident housing panic public opinion index (X3) gradually decreases within the range of 0 to approximately -0.1835, the SHAP value is positive and gradually increases, indicating that the possibility of the stability crisis in the real estate industry is gradually increasing.

(3) Based on the feature importance results of the previous section, we can find that the most important distinguishing feature one year before the stability crisis in the real estate industry is the rolling year-on-year land revenue (X7). Based on the SHAP dependency graph results of this feature, when the value of X7 gradually decreases within the range of 6.26 to -2.05, the SHAP value is positive and gradually increases, indicating that the possibility of the stability crisis in the real estate industry gradually increases.



(a) t-1 (b) t-2 (c) t-3

Figure 5. Index SHAP depends on the result

Based on the analysis results of the SHAP dependency graph with the above-mentioned characteristics, we can understand the specific impact of key indicators on the stability crisis in the real estate industry, and present it in a visual form, which can effectively enhance the interpretability of the discrimination model for its discrimination results.

5. Conclusion

In the process of strategic crisis management modeling for enhancing the stability of machine learning models in the real estate industry, there is a problem of excessive parameters. This paper proposes a real estate industry stability strategic crisis modeling method based on LASSO-GA-XGBoost, which reduces the parameter input in the machine learning modeling process and improves the accuracy and efficiency of the calculation. The conclusions are as follows:

(1) Feature selection was conducted on the parameters in the database. The results showed that the credit bond spreads of real estate enterprises, the long-term deviation of housing income ratio, and the cash-to-short-term debt coverage ratio of leading real estate enterprises were identified as redundant features for the stability crisis of the real estate industry.

(2) Compared with the results obtained by the XGBoost model from the database calculation, the LASSO-GA-XGBoost model obtained a higher accuracy of ACC (96.39%) and AUC (90.71%) for the average warning in the previous three years, which can provide more effective forward-looking warnings for the stability crisis of the real estate industry that is about to occur.

(3) The SHAP values of the stability crisis of the real estate industry were calculated based on LASSO-GA-XGBoost. In practical applications, it can provide decision-making basis and references for regulatory authorities, company management, and market investors, helping them better grasp the key financial indicators for the identification of the stability crisis of the real estate industry, and knowing the specific impact effects of these key indicators on the identification results, so as to be able to formulate targeted and preventive and management measures in advance, and accurately and efficiently warn and identify the stability crisis of the real estate industry, helping to maintain the stability and healthy development of the capital market.

References

1. Han, Y., Zhang, H., & Zhao, Y. (2021). Structural evolution of real estate industry in China: 2002-2017. Structural Change and Economic Dynamics, 57, 45-56.

2. Dong, X. (2017). Reform of China's housing and land systems: the development process and outlook of the real estate industry in China. Chinese Journal of Urban and Environmental Studies, 5(04), 1750027.
3. Glaeser, E., Huang, W., Ma, Y., & Shleifer, A. (2017). A real estate boom with Chinese characteristics. Journal of Economic Perspectives, 31(1), 93-116.
4. Wang, B. (2021). The evolving real estate market structure in China. In Understanding China’s real estate markets: Development, finance, and investment (pp. 9-19). Cham: Springer International Publishing.
5. Zhao, S. X., Zhan, H., Jiang, Y., & Pan, W. (2017). How big is China’s real estate bubble and why hasn’t it burst yet?. Land use policy, 64, 153-162.
6. Rogoff, K., & Yang, Y. (2021). Has China's housing production peaked?. China & World Economy, 29(1), 1-31.
7. Volodymyr, D., Olga, K., Zamiar, Z., & Lesia, K. (2021). Strategic management of the enterprise in the conditions in the instability of the market environment. Intellectualization of Logistics and Supply Chain Management, (7-8), 57-63.
8. John-Eke, E. C., & Eke, J. K. (2020). Strategic planning and crisis management styles in organizations: A review of related literature. Journal of Strategic Management, 5(1), 36-46.
9. GRYLITSKA, A. (2025). STRATEGIC RISK MANAGEMENT AND CRISIS PREDICTION TOOLS IN CONTEMPORARY ENTERPRISE MANAGEMENT. Bulletin of the Cherkasy Bohdan Khmelnytsky National University. Economic Sciences, 29(1).
10. Sobolieva, T., & Kravchenko, D. (2025). Crisis strategic analysis in enterprise management. Herald of Economics, (4), 101-108.
11. Thakur, R., & Hale, D. (2022). Strategic crisis response: managerial implications and direction for recovery and survival. Journal of Business & Industrial Marketing, 37(10), 1959-1973.
12. Stroiko, T., & Kharus, H. (2022). Anti-crisis management as the basis of ensuring the economic stability of enterprises. Three Seas Economic Journal, 3(4), 44-51.
13. Mikušová, M., & Horváthová, P. (2019). Prepared for a crisis? Basic elements of crisis management in an organisation. Economic research-Ekonomska istraživanja, 32(1), 1844-1868.
14. Toivonen, S., Heinonen, S., Verma, I., Castaño-Rosa, R., & Wilkinson, S. (Eds.). (2024). Real estate and sustainable crisis management in urban environments: Challenges and solutions for resilient cities. Taylor & Francis.
15. Ali, B. J., & Anwar, G. (2021). Real estates strategies: Analysis of strategic management practices in real estate companies. Ali, BJ, & Anwar, G.(2021). Real Estates Strategies: Analysis of Strategic Management Practices in Real Estate Companies. International Journal of Rural Development, Environment and Health Research, 5(3), 35-48.
16. Kaklauskas, A., Kelpsiene, L., Zavadskas, E. K., Bardauskiene, D., Kaklauskas, G., Urbonas, M., & Sorakas, V. (2011). Crisis management in construction and real estate: Conceptual modeling at the micro-, meso-and macro-levels. Land Use Policy, 28(1), 280-293.
17. Tien, N. H., & Thuan, T. T. H. (2019). Analysis of Strategic Risk of Domestic and Foreign Real Estate Enterprises Operating in Vietnam's Market. International journal of commerce and management research, 5(5), 36-43.
18. Ya ping Huang, Xue mei Qi, Yan Cheng & Ling ling Zhou. (2026). Prediction method of deep CBM reservoir thickness based on improved XGBoost. Applied Geophysics,23(2),507-516.
19. Yao Xiao, Anne Gelb & Aditya Viswanathan. (2026). A New Sparsity Promoting Residual Transform Operator for Lasso Regression. SIAM Journal on Numerical Analysis,64(4),1217-1238.

20. Zhe Zhang & Shengzhen Ding. (2026). Spatio-temporal dynamics and drivers of climate change vulnerability in China: an entropy weight and XGBoost-SHAP analysis. Mitigation and Adaptation Strategies for Global Change,31(6),71-71.

About the Author

Xueying Zhang studied at Chungwoon University, South Korea, from 2017 to 2019, receiving her Master's degree in 2019. Since 2019, she has worked as a lecturer at Ulanqab Vocational College in Inner Mongolia, China, for seven years. In 2025, she began doctoral studies in Business Administration at World Leaders University, Mongolia, focusing on real estate strategic management and crisis management.

Gerelmaa Jamsran, Doctor (PhD), Global Leadership University (GLU).

Narmandakh Dolgor, Doctor (Ph.D), professor, Dean of Graduate school of Global Leadership University (GLU).

Huadong Zhang graduated with a Master's degree in Economics from Vitebsk State Technological University, Belarus, in April 2023. He is a Partner at Beijing Longan Law Firm, specializing in compliance management, intellectual property, and legal risk prevention and control.