International Journal of Science Academic Research
Vol. 05, Issue 04, pp.7251-7258, April, 2024
Available online at http://www.scienceijsar.com
ISSN: 2582-6425

Research Article

SUSTAINABILITY MODEL ANALYSIS AND APPLICATION OF PROPERTY INSURANCE

[1]Zhan Shu, [1,] *Diping Zhang and [2]Jianya Gu

[1]School of Science, Zhejiang University of Science and Technology, Hangzhou, 310023, China
[2]School of Marxism, Zhejiang University of Science and Technology, Hangzhou, 310023, China

Received 11th February 2024; **Accepted** 15th March 2024; **Published online** 19th April 2024

Abstract

Nowadays, extreme weather events bring a great crisis to property owners and insurance companies, in response to this problem, this paper selects five indicator systems for analysis, namely economic indicators, extreme weather indicators, owner behavior indicators, real estate indicators, and historical landmark indicators, further deals with outliers, missing values, and normalizes the data. It predicts the future trend of changes based on specific numerical situations by building insurance models and protection models in order to take better measures to protect local properties.

Keywords: Logistic probability, Pearson correlation coefficient, Entropy weight method, Topsis model.

INTRODUCTION

Background

As the impact of global climate change becomes more pronounced, the frequency and intensity of extreme weather events are increasing. This undoubtedly poses a great challenge to the insurance industry, and at the same time provides opportunities for innovation in insurance business models. To meet this challenge, insurers need to reassess their insurance strategies and product designs to better meet customer needs and reduce their own risks. Insurers need to consider a number of factors when facing the risks posed by climate change. Insurers address these factors and thus reassess their claims handling and risk control strategies, and therefore need to use technology tools such as data analytics and artificial intelligence to improve the accuracy of their risk assessment and pricing and to reduce their own risks. The following questions will be addressed in this paper:

- Insurance modeling to determine the most appropriate time for an insurance company to choose to write a policy, with different regions giving clear decisions.
- Owner behavior variables were added and correlation analyses were conducted with other insurance model correlation variables as well as insurance model probability results to see the degree of relevance of owner behavior to decision making.
- The optimized insurance model is shown by adding data on extreme weather variables and selecting two regions on different continents as examples.
- By incorporating variables related to real estate development, an insurance yield assessment model was developed to determine the best locations for communities and real estate openers to build.
- Identify a historic landmark building that should be preserved and protected by building a preservation model and give relevant measures to protect the building. And as an example of a historic landmark under an extreme weather event, evaluate the results through insurance modeling and preservation modeling.

Notations

The following table lists the key symbols used in this paper:

Table 1. Notations

Symbol	Definition	Unit
σ	standard deviation	/
ω	Net premium margin	%
C_i	Proximity of indicator i to the optimal value	/
r_{ij}	Correlation coefficient between indicator i and indicator j	/
y_i	XGboost predicted results	/

DESCRIPTION AND PRE-PROCESSING OF DATA

Description of data

Sources of data: In this paper, we select indicators related to Hangzhou, Harbin, Wuhan, Kunming, and Urumqi in China to build a universal insurance model, then introduce data from San Francisco in the U.S. to explore the impact of the number of extreme weather events on the underwritten policies of the two regions on different continents, then search for data related to the buildings related to the cultural and community significance, to build a preservation model and to formulate the preservation measures. Specific data sources are listed in the table below:

Table 2. Data sources

Name	Website
Weather forecast for 24 hours	https://www.tianqi24.com/
EPS Data	https://www-epsnet-com-cn.ez.zust.edu.cn/index.html#/Index
NBS	http://www.stats.gov.cn/
U.S.Bureau of Economic Analysis	https://www.bea.gov/
WeatherSpark	https://tw.weatherspark.com/

Corresponding Author:** ***Diping Zhang,
School of Sciences, Zhejiang University of Science and Technology, Hangzhou, 310023, China.

Indicator system

a. Indicators of independent variables

In order to establish the insurance model, this paper measures the specific situation of underwriting policies in a certain region by collecting data on premium income and insurance claims expenditure, through the difference between the two as a profit, and then refer to a number of indicator systems to establish logistic regression models to determine the probability of meeting the conditions of underwriting policies, and the explanations of the indicators of the independent variables are shown in Table 3 below:

Table 3. Indicators of independent variables

Symbol	Definition	Unit
N_{BF}	Premium income	$
Nzc	Insurance payout	$

b. System of economic indicators

In the research on insurance demand by foreign scholars, Mantis and Farmer (1968) found that population size has a positive relationship with the demand for life insurance; Lewis and Camphell (1980) found that income and insurance demand have a positive relationship; Burett and Palmer (1984) found that educational attainment also has a positive relationship with insurance demand. Considering the structure of the insurance market and product structure as well as the availability and accessibility of data, therefore, this paper adopts the Gross Domestic Product (GDP) of each region, the number of population (QOP), the education level (EDU) and the Per Capita Disposable Income (PCDI) as the explanatory variables[1] , to study the influence of insurance claims and economic factors, the system is as follows:

Table 4. System of economic indicators

Symbol	Definition	Unit
GDP	gross value of production	$
QOP	population	/
EDU	Level of education	/
PCDI	per capita disposable income	$

c. System of Extreme Climate Indicators

Since the increase in the number of extreme weather events may have an impact on the underwriting policies in the region, this paper selects five relevant variables to establish an indicator system [2] to show the impact of extreme weather on insurance claims as follows:

Table 5. Extreme weather indicator system

Symbol	Definition	Unit
EHT	The daily maximum temperature ≥ 35℃	d
ELT	The daily minimum temperature ≤0℃	d
N_D	Precipitation months less than 10 mm	m
N_{Ep}	Months of rainfall of more than 100 mm	m
N_{Uw}	Extreme Weather	d

d. Owner Behavioral Indicator System

Since this paper also needs to explore the impact of the decision of the owner's behavior, the following four indicators are selected to explore the relationship between them:

Table 6. Owner Behavioral Indicator System

Symbol	Definition	Unit
N_{De}	Death toll of various production and operation safety accidents	/
N_Z	Total number of patents granted	/
N_C	Savings deposits of urban and rural residents at the end of the year	$

e. Real estate indicator system

For the purpose of later exploring how and where communities and real estate developers plan to build and develop, this paper introduces the following four metrics to adjust the insurance model to evaluate optimal building scenarios:

Table 7. Real estate indicator system

Symbol	Definition	Unit
N_{Dz}	Completed investment this year	$
S_{SG}	Building area under construction	m^2
S_{XS}	Commercial housing sales area	m^2
S_{GZ}	Area of land purchased this year	m^2

f. Architectural landmark indicator system

Table 8 Architectural landmark indicator system

Level 1 ndicators	Level 2 indicators	Level 3 indicators	Definition	Unit
Historical building protection level	History and culture	H_Y	Historical year	y
		G_C	Grade of cultural relics protection	/
		T_R	Total revenue of tourist attractions	/
	Social economy	E_O	Employment of working residents	$
		D_I	Disposable income of regional residents	/
		N_H	Number of architectural heritage protection policies	$
	Community service	S_C	Surrounding infrastructure conditions	/
		P_S	Public service level	/

Data pre-processing

In order to eliminate the influence of the scale on the data for the calculation, we delete the outliers in the previous basis, the table data and then data standardization, so that the data can be calculated more accurately. Each sample data in the table is processed according to the formula: $\varrho_i = \frac{x_i - \overline{x}}{\alpha_x}$,where ϱ_i denotes the standardized variable, $\overline{x}$ denotes the mean of the group, and α_x denotes the standard deviation of the data, some of which are as follows:

Table 9. Data standardization

City	Year	N_{BF}	N_{zc}	GDP	…	N_{De}	N_Z	N_C
Hangzhou	2020	2.03	2.07	1.34	…	-0.66	0.00	2.14
Hangzhou	2021	1.67	1.59	1.24	…	-0.36	-0.29	1.30
Hangzhou	2022	1.23	1.17	0.92	…	-0.27	-0.51	1.00
Harbin	2020	-0.70	-0.06	-0.75	…	-0.17	-0.28	0.04
…	…	…	…	…	…	…	…	…
Urumqi	2021	-0.95	-0.90	-1.03	…	-1.14	-0.81	-1.13
Urumqi	2022	-0.94	-1.04	-1.09	…	-1.12	-0.88	-1.12

Outlier handling: In this paper, the 3σ principle is used to determine outliers, i.e., if the data follow a normal distribution, an outlier is defined as a value in a set of result values that deviates more than three times the standard deviation from the mean. That is, under the assumption of normal distribution, the probability of the occurrence of values beyond three times σ (standard deviation) from the mean is very small (as in the following equation), with a probability of less than 0.3%, and therefore can be considered outliers.

$$P(|x-\mu|>3\sigma)\leqslant 0.003$$

where is the data point in the dataset, is the mean value corresponding to that dataset, and is the standard deviation corresponding to that dataset.

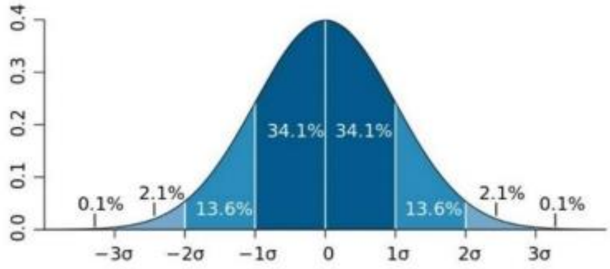

Figure 1. Outlier Handling

The outliers in the data are filtered and removed based on the 3σ principle and with the help of Matlab code.

INSURANCE PROBABILITY MODEL BASED ON LOGISTIC REGRESSION

Establishment of Insurance Probability Model

This article uses logistic regression [3] to conduct probability analysis on the level of underwriting policy conditions. We use a linear predictor and additional normalization factors to model the logarithm of the probability of classification results.

$$\ln\left(P(y=K|x)\right)=k_{n,K}x_n-\ln Z$$

We use an additional term - ln (Z) to ensure that all probabilities can form a probability distribution, so that the sum of these probabilities is equal to 1.

$$\sum_{k=1}^{K}P(y=k|x)=1$$

Exponentializing the equation yields the following formula:

$$P(y=K|x)=\frac{1}{Z}e^{k_{n,K}x_n}$$

The sum of all probabilities is equal to 1, the formula for Z can be obtained:

$$1=\sum_{k=1}^{K}P(y=k|x)=\sum_{k=1}^{K}\frac{1}{Z}e^{k_{n,k}x_n}=\frac{1}{Z}\sum_{k=1}^{K}e^{k_{n,k}x_n}$$

By using the formula above for calculation, it can be concluded that：

$$Z=\sum_{k=1}^{K}e^{k_{n,k}x_n}$$

Combine the above formulas to obtain the final probability formula：

$$\omega=\frac{N_{BF}-N_{ZC}}{N_{BF}}\times 100\%$$

$$P(y=K|x)=\frac{e^{k_{n,K}x_n}}{\sum_{k=1}^{K}e^{k_{n,k}x_n}}$$

Meanwhile, selecting the net premium margin as the probability condition, if the net premium margin is greater than or equal to 35%, it indicates that the region is suitable for underwriting policies. The calculation formula is as follows：
Among them, ω represents the net premium margin, N_{BF} represents premium income (\$), N_{ZC} represents insurance payout expenses (ω).

Solution of Insurance Probability Model

Based on the insurance probability model mentioned earlier, this study used the economic indicators and owner behavior indicators mentioned earlier to solve the insurance probability model for five cities in China - Hangzhou, Harbin, Wuhan, Kunming, and Urumqi. The results of the model are as follows:

Table 10. Insurance Probability Model Results

City	Year	Predicting the probability of training outcomes_0.0	Predicting the probability of training outcomes_1.0
Hangzhou	2020	0.005011	0.994989
Hangzhou	2021	0.003532	0.996468
Hangzhou	2022	0.001533	0.998467
Harbin	2020	0.188588	0.811412
Harbin	2021	0.907846	0.092154
Harbin	2022	0.987955	0.012045
Wuhan	2020	0.000667	0.999333
Wuhan	2021	0.802402	0.197598
Wuhan	2022	0.181945	0.818055
Kunming	2020	0.486104	0.951390
Kunming	2021	0.673013	0.326987
Kunming	2022	0.264306	0.736694
Urumqi	2020	0.016790	0.983210
Urumqi	2021	0.014367	0.985633
Urumqi	2022	0.003825	0.996175

The table of evaluation indicators through cross validation set is as follows：

Table 11. Evaluation indicator table

	Accuracy	Recall rate	Precision	F1
training set	0.946	0.946	0.946	0.946
test set	0.848	0.848	0.648	0.741

It can be found that the training and testing sets of the model have good test results, both above 80%, indicating that the data selected for establishing the model is reasonable and can represent most of the characteristics of insurance risk. At the same time, according to the specific probability model results, it can also be seen that developed cities in China have lower insurance risks and higher returns, such as Hangzhou, where insurance has been mostly profitable in recent years. Secondly, if the regional environment is relatively closed and there is less communication with the outside world, such as Urumqi, the premium risk will also be relatively small, which may be caused by a comfortable environment.

Correlation analysis of owner behavior

This study detects whether there is a high correlation between independent variables by calculating the correlation coefficient matrix, using three indicators of owner behavior, namely the number of deaths from various production and operation safety accidents, the total number of patent authorizations, and the year-end balance of savings deposits of urban and rural residents. Perform Pearson correlation coefficient analysis with four economic indicators and insurance net profit margin, and the results are shown in the following figure:

	ω	GDP	EDU	PCDI	QOP	N_{De}	N_Z	N_C
ω	1	0.27673	-0.0387	0.3113	0.02934	-0.0316	-0.0051	0.17593
GDP	0.27673	1	0.53778	0.79503	0.87958	-0.088	0.69148	0.89345
EDU	-0.0387	0.53778	1	0.00747	0.71894	0.42655	0.86916	0.35512
PCDI	0.3113	0.79503	0.00747	1	0.50201	-0.2423	0.23417	0.76878
QOP	0.02934	0.87958	0.71894	0.50201	1	0.09372	0.7284	0.85493
N_{De}	-0.0316	-0.088	0.42655	-0.2423	0.09372	1	0.06768	-0.1096
N_Z	-0.0051	0.69148	0.86916	0.23417	0.7284	0.06768	1	0.44628
N_C	0.17593	0.89345	0.35512	0.76878	0.85493	-0.1096	0.44628	1

Figure 2. Pearson correlation coefficient

Based on the explanation of the correlation coefficient, refer to the table below:

Table 12. Explanation of correlation coefficient size

Relevance	Minus	Positive
irrelevant	[-0.09,0]	[0,0.09]
weak correlation	[-0.3,-0.1]	[0.1,0.3]
medium relevance	[-0.5,-0.3]	[0.3,0.5]
strong correlation	[-1.0,-0.5]	[0.5,1.0]

Conclusion: For indicators related to owner behavior, there is a strong correlation with many economic indicators, and there is also a partial impact on underwriting policies. Among them, the year-end balance of savings deposits of urban and rural residents has the greatest impact. According to the above correlation coefficients, it can be found that homeowners can improve their economic level, enhance their safety awareness, increase the benefits of policies, and also increase the probability of underwriting policies.

Analysis of extreme weather insurance models in San Francisco and Hangzhou

The aim of this study is to obtain the eligible contracting insurance probabilities for San Francisco and Hangzhou from 2016 to 2022 without extreme weather indicators using the insurance probability model established in the previous text. The results are as follows:

Table 13. Probability of Contracting Insurance (before)

City	Year	0 probability	1 probability
San Francisco	2016	0.010439	0.989561
San Francisco	2017	0.008995	0.991005
San Francisco	2018	0.007054	0.992946
San Francisco	2019	0.005125	0.994875
San Francisco	2020	0.002355	0.997645
San Francisco	2021	0.004353	0.995647
San Francisco	2022	0.001044	0.998956
Hangzhou	2016	0.016311	0.983689
Hangzhou	2017	0.014536	0.985464
Hangzhou	2018	0.009636	0.990364
Hangzhou	2019	0.010436	0.989564
Hangzhou	2020	0.005011	0.994988
Hangzhou	2021	0.003531	0.996468
Hangzhou	2022	0.001533	0.998466

Introducing five indicators from the extreme weather system mentioned earlier, namely extreme high temperature, extreme low temperature, dry months, extreme precipitation months, and number of abnormal weather, to improve the insurance probability model, the results are as follows:

Table 14. Probability of Contracting Insurance (after)

City	Year	0 probability	1 probability
San Francisco	2016	0.005943	0.994057
San Francisco	2017	0.004422	0.995578
San Francisco	2018	0.002933	0.997067
San Francisco	2019	0.002012	0.997988
San Francisco	2020	0.000474	0.999526
San Francisco	2021	0.001454	0.998546
San Francisco	2022	0.000366	0.999634
Hangzhou	2016	0.012816	0.987184
Hangzhou	2017	0.010436	0.989564
Hangzhou	2018	0.007321	0.992679
Hangzhou	2019	0.009343	0.990657
Hangzhou	2020	0.003946	0.996054
Hangzhou	2021	0.002932	0.997068
Hangzhou	2022	0.001241	0.998759

By adding extreme weather indicators, new regional insurance probabilities were obtained. Compared with the original probability, it can be found that by adding extreme weather indicators, the insurance model has been further improved, and the predicted results are closer to the actual situation. Compared with the original probability, there has been a slight increase, indicating that extreme weather will affect the insurance probability. By adding this indicator, the model can be improved. At the same time, by observing the probability model of the two regions, it can be found that extreme weather has a greater impact on the insurance probability. The overall probability curve of the two regions shows an upward trend, but for the data of Hangzhou, China, in 2019 and the data of San Francisco, America, in 2021, there is an obvious downward trend in these two years. The premium income in Hangzhou is reduced due to the environmental degradation caused by the COVID-19 epidemic, In 2021, San Francisco experienced extreme drought and early invasion of the western United States, leading to extreme water scarcity in the San Francisco Bay Area, which affected the premium environment. At the same time, in 2022, some parts of San Francisco reached the highest temperature in history, reducing the probability.

COMMUNITY REAL ESTATE DECISION MODEL

KMO test as well as sphericity test

Firstly, KMO test and spherical test are carried out. Here we mainly use principal component analysis to comprehensively evaluate the multidimensional indicators, for principal component analysis, in general, the greater the connection between the indicators, the better the effect of dimensionality reduction. Therefore, we use KMO test and Bartlett's spherical test to verify the indicators in advance and judge the relationship between the indicators. For the indicators without KMO test and Bartlett spherical test, we use the entropy weight method to calculate the comprehensive score, and utilize the entropy weight method to calculate the weight. Among them, KMO test and Bartlett's spherical test to determine the covariance or correlation between the collected indicators.

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}}{\sum\sum_{i \neq j} r_{ij}^2 + \sum\sum_{i \neq j} a_{ij}^2}$$

where r_{ij} denotes the simple correlation coefficient. α_{ij} 1,2,3, $\cdots$, k denotes the partial correlation coefficient.

Table 15. Results of KMO test and Bartlett's test of sphericity

name of index		personal basic information
KMO value		0.522
Bartlett sphericity test	Approximate chi-square	35.737
	df	3
	P	0.000**

Note: ***, **, * represent 1%, 5%, and 10% significance levels, respectively.

Therefore, for this test result, it fails to pass the test, so entropy weight method is chosen for comprehensive evaluation.

Entropy weight method + Topsis modeling

Entropy weighting method for indicator weights: This paper uses the capital city of each province in China in recent years, based on the relevant insurance probability model indicator data of these cities, based on which the indicator weights of each indicator system are derived, and the results are shown in the table below:

Table 16. Indicator weights

Itemized dimension	Basic index	Index weight	comentropy	Dimension index weight
economy	GDP	0.06812	0.93405	0.14376
	QOP	0.02701	0.97385	
	EDU	0.03783	0.96337	
	PCDI	0.01081	0.98953	
Extreme weather	EHT	0.05100	0.95062	0.36452
	ELT	0.12163	0.88223	
	N_D	0.06397	0.93806	
	N_{Ep}	0.06305	0.93895	
	N_{Uw}	0.06488	0.93718	
Real estate	N_{Dz}	0.10085	0.90235	0.28773
	S_{SG}	0.03360	0.96746	
	S_{XS}	0.06165	0.94031	
	S_{GZ}	0.09164	0.91127	
Owner's behavior	N_{De}	0.05158	0.95006	0.20398
	N_Z	0.10802	0.89541	
	N_C	0.04439	0.95702	

We can find out: extreme weather-related indicators have the greatest impact on the insurance probability model, with indicator weights as high as 0.36, followed by the real estate industry, which also has a close relationship with the insurance industry, with indicator weights as high as 0.287.For the indicator weights of the individual indicators, the per capita disposable income, the extreme low temperatures, and the volume of real estate investment in the current year are higher, with indicator weights greater than 1, which suggests that the three factors have an important impact on the insurance investment.

Topsis modeling: Topsis method is an effective multi-indicator evaluation method. This method selects the optimal solution by constructing positive and negative ideal solutions to the evaluation problem and scoring each solution by calculating its relative closeness to the ideal solution. We perform Topsis scoring on the insurance benefit level of each city by first obtaining the matrix after normalization and then recalculating it by using the weights of the indicators calculated by the entropy weighting method above to obtain the insurance benefit level score of each city, and then finally normalize the data, and the normalization steps are as follows:

Suppose there are n objects to be evaluated and m normalized evaluation metrics form the following normalization matrix X:

$$\begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

The matrix for normalization is denoted Z,each element in Z is:

$$Z_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{m} X_{ij}^2}} (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$$

After data normalization matrix X becomes matrix Z.

Subsequently, the optimal and worst values of the indicators are determined, forming the optimal and worst value vectors, respectively

$$Z_j^+ = \max\left\{Z_{1j}, Z_{2j}, \cdots, Z_{mj}\right\} (j = 1, 2, \cdots n)$$

$$Z_j^- = \min\left\{Z_{1j}, Z_{2j}, \cdots, Z_{mj}\right\} (j = 1, 2, \cdots n)$$

Next, the distance of each unit indicator value from the optimal and worst values is calculated.

$$D_i^+ = \sqrt{\sum_{i=1}^{n} \left(Z_{ij} - Z_i^+\right)^2} (i = 1, 2, \cdots, m)$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \left(Z_{ij} - Z_i^-\right)^2} (i = 1, 2, 3, \ldots m)$$

Finally, the proximity of each indicator to the optimal value is calculated:

$$C_i = \frac{D_i^-}{D_i^- + D_i^+} (i = 1, 2, 3, \ldots, m)$$

It is then ranked and evaluated. The evaluation units are ranked in order of proximity, with a larger C indicating a closer proximity to the optimal level.

The entropy weighting method and Topsis are used to find the Topsis scoring of the 28 capital cities, which is used to indicate the good or bad returns of insurance policies, and the results are shown in the heat map below:

Figure 3. Heat map of China

Solving Community Real Estate Decision Models

Analysis of Topsis Score Results: According to the scoring of the 28 provincial capital cities sought above, that is, the final comprehensive evaluation of the value of the threshold delimitation process, identified the top 25% of the rankings identified as excellent; 25%-50% identified as good, 50%-75% identified as medium, and 75%-100% as poor.

Based on the results sought, the specific thresholds for each classification are shown in the table below:

Table 17. Table of Score Thresholds

Comprehensive evaluation	Lower threshold	Upper threshold
poor	0.013131	0.025312
middle	0.025312	0.039819
good	0.039819	0.040674
optimal	0.040911	0.050176

Based on the above grades, combined with the results of the city scoring, it is possible to determine the level of insurance returns for each city in China.

Analysis of indicator weighting results: Based on the above table of indicator weights, a weight visualization image is drawn based on the four dimensions, as shown below:

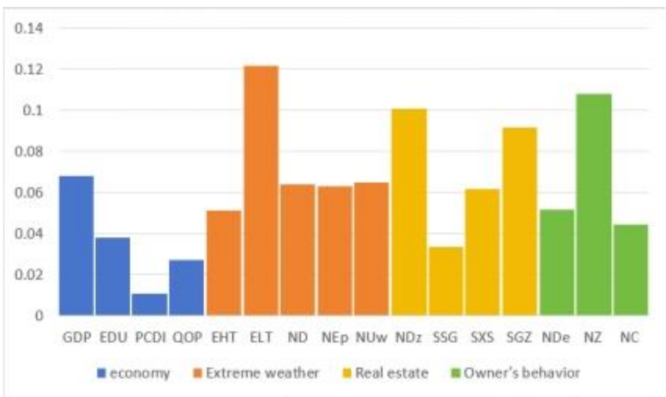

Figure 4. Bar chart of indicator weights

Based on the above weighting results, we can find that the most important thing in choosing whether an area is suitable for construction is to look at its weather conditions and the circumstances related to real estate investment. Combined with the above results, we can find that suitable locations for investment need to meet:

1. A high level of economy and better infrastructure.
2. Not much seasonal weather change, less affected by extreme weather
3. Good prospects for real estate investment, a broad market for the sale of commercial properties, abundant land resources for property construction, and high economic strength for investment.
4. The owner has a high quality of life, and has a strong sense of self-protection and high cultural literacy.

Locations that meet the above requirements are relatively more suitable for investment and construction.

PROTECTION MODEL BASED ON XGBOOST REGRESSION

Selection of distribution of historical buildings

For the establishment of the conservation model, we selected 44 typical historical buildings from different cities in 88 countries in the Americas and Asia. The distribution of the selected historical buildings is shown in the following figure:

(a)Map of Asia

(b)Map of America

Figure 5. Selection of historical buildings

AHP evaluation of historical building protection level

This article quotes AHP [4] to evaluate the protection level of historical buildings, and selects eight indicators to obtain the protection level scores of each historical building from the aspects of historical culture, social economy, and community services. The hierarchical structure diagram is as follows:

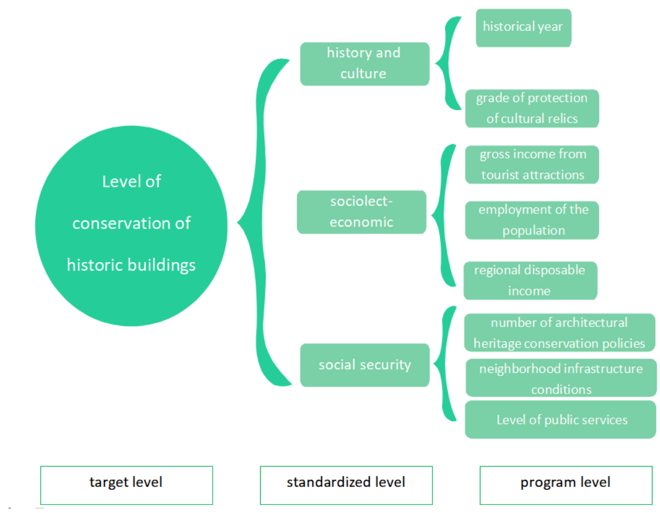

Figure 6. AHP indicator layer

The process of AHP data deduction and analysis is shown below, which is used for quantitative analysis of the data

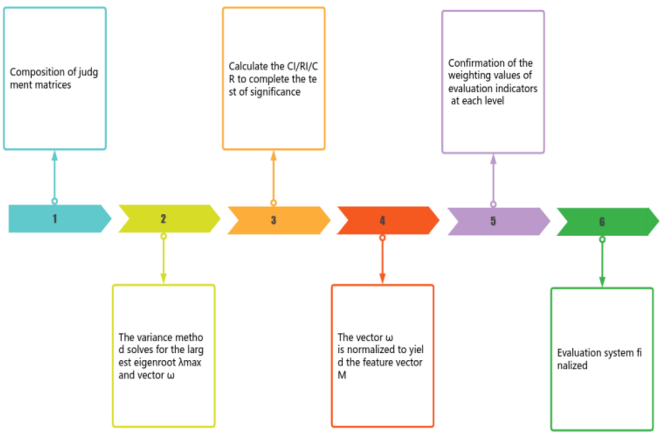

Figure 7. AHP Analysis Flow Chart

The partial score results are shown in the table below：

Table 18. AHP Evaluation Score

Serial number	Score	Serial number	score
1	73.85	41	63.08
2	55.38	42	70.78
3	70.77	43	49.23
4	66.15	44	67.69
…	…	…	…

XG boost regression prediction of building protection level

Establishment of XGboost Regression Model: Based on the evaluation scores of each historical building obtained, this article uses this score as the dependent variable and the evaluation index system of historical buildings as the independent variable to establish an XGboost regression model [5]. The following text will be used to predict the level of protection of other historical landmarks. XG Boost is a machine learning method based on ensemble learning principle Boosting, which uses a CART decision tree as the base classifier. By continuously adding CART trees to the model to split the features, a new function formed by the newly added trees is used to fit the residuals of previous predictions. Then, the predicted results of all trees are added together as the final prediction result.

Solution of XGboost Regression Model: This article selects 80% of the data as the training set, tests and cross validates the data, and we obtain the feature importance map as follows：

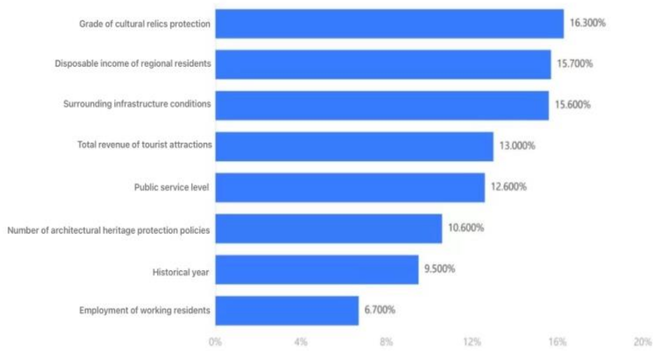

Figure 8. Feature Importance

We found that the importance of most indicators is greater than 10%, among which the level of cultural relic protection is the most important because it symbolizes the degree of importance that the national government attaches to the protection of this historical landmark. Secondly, the disposable income of local residents and the surrounding infrastructure conditions are 15.7% and 15.6%, respectively. Higher disposable income of residents and more complete infrastructure conditions indicate that the local economic level of the landmark is high and the protection measures are complete, so these two indicators are also more important. The impact on the employment situation of residents is relatively small, only 6.7%, indicating that the protection of historical landmarks depends more on the conditions possessed by the national government, rather than the economic benefits it brings to the country.

Based on the original data as the training set, the fitted image obtained is as follows:

Figure 9. Fit Image

According to the above figure, it can be found that the fitting effect is good, indicating that the XGboost regression model established is reasonable. In the following text, historical landmark data under extreme weather will be predicted.

Protective measures：

Based on the established XGboost model, the protection level of historical landmarks can be determined based on the collected data. The following suggestions can be provided:

1. For historical landmarks with high regression values, it indicates that the landmark has high value and good protection conditions. There are now comprehensive landmark protection measures in place, and the probability of damage is also small. It is only necessary to improve the

community's awareness of protection and service to avoid arbitrary damage to the landmark; We can also promote deeper communication policies with insurance companies, improve risk protection systems, and ensure the integrity of historical landmarks.

2. For historical landmarks with low regression values, it indicates that the protection conditions in the area where the historical landmark is located are poor, the protection measures are not appropriate, and the infrastructure is backward, resulting in a low level of comprehensive community quality. Based on this situation, give three suggestions：

- If it has high economic benefits, it can be relocated to other areas to avoid the abandonment of historical landmarks in underdeveloped areas.
- It can improve the economic development of the surrounding areas, increase infrastructure construction, actively develop the tourism industry, and promote foreign tourists to come and play, in order to drive the economic level of the surrounding areas. And improve the social security system, actively cooperate with insurance companies, and jointly bear risks and benefits.
- This landmark does not have strong economic benefits but has high historical and cultural value. It can be relocated to developed areas in the surrounding area to drive the development of tourism industry.

TESTING OF THE MODEL

Sensitivity analysis

(1) Sensitivity analysis of logistic-based insurance probability models

Table 19. Sensitivity analysis

city	year	Before adjustment	After adjustment	Change radio
Hangzhou	2020	0.994989	0.995114	0.013%
Hangzhou	2021	0.996468	0.996723	0.026%
Hangzhou	2022	0.998466	0.998672	0.022%

In this part, this paper respectively slightly upward adjusted the values of the variables in the insurance probability model, upward adjusted QOP as an example, the results are shown in the table above, it can be found that after the upward adjustment of QOP, the impact on the final probability of the results of the smaller, indicating that the final results of the variable QOP is not sensitive to the other variables are also the same, so the model sensitivity analysis through.

(2) Sensitivity analysis of conservation models based on XGboost regression

Table 20. Sensitivity analysis

Serial number	Before adjustment	After adjustment	Change radio
1	77.6678	76.3218	1.73%
2	68.7707	67.2674	2.18%
3	65.4855	63.9983	2.27%

In this part, this paper respectively, the protection model of the variables in the value of a slight decline in the downward adjustment of peripheral infrastructure conditions, for example, the results are shown in the table above, can be found in the downward adjustment of peripheral infrastructure conditions, the final regression prediction results of the impac of the smaller, indicating that the final results of the variable is not sensitive to this variable, the other variables are also the same, so the model sensitivity analysis through.

Cross-validation

(1) Cross-validation of logistic-based insurance probability models

Table 21. Cross-validation

	Accuracy rate	Recall rate	Precision rate	F1
Training set	0.946	0.946	0.946	0.946
Cross verification set	0.908	0.908	0.856	0.874
Test set	0.848	0.848	0.648	0.741

The table of evaluation metrics through the cross validation set is shown in the table above, it can be found that the accuracy, recall and precision of the prediction evaluation metrics are better than 80% and pass the cross validation.

(2) Cross-validation of protection models based on XGboost regression

Table 22. Cross-validation

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.717	0.847	0.245	0.367	0.993
Cross verification	0.958	0.978	0.487	0.624	0.905
Test set	1.568	1.25	0.764	0.916	0.854

The table of evaluation metrics through the cross-validation set is shown in the table above, and it can be found that the scattered MSE, MAPE and MAE take small values and R^2 is close to 1, which indicates that the model is good and passes the cross-validation.

Acknowledgements: This study was funded by Innovation and Entrepreneurship Training Program for College Students of Zhejiang University of Science and Technology (2023cxcy107).

REFERENCES

1. Han Yiqing, Chen Xiaoqian. An empirical analysis of the influencing factors of premium income in China[J]. *Business Culture* (Second Half of the Month)2012, (06):166-167. Chinese
2. Zheng Zhibin, Deng Yanjun, Huang Yongping. Impacts of extreme weather and climate events on wetland ecology in Jianghan Lake Group[J]. *People's Yangtze River*, 2021, 52(Suppl 2):45-51. Chinese
3. Ma Yuxing. Auto insurance claim probability prediction based on Multiple-Lasso-Logistic regression model[J]. *Advances in Applied Mathematics*, 2023, 12(2): 563-573. Chinese
4. Chen Lu, Zhang Xueqing. Research on the quantitative evaluation model of protection and reuse of red buildings in Shanghai[J]. *Furniture & Interiors*, 2022,29(07):114-119. Chinese
5. WANG Kunzhang, JIANG Shubo, ZHANG Hao, et al. Regression-classification-regression life prediction model based on XGBoost[J]. *China Measurement & Test,* 2023, 49(8):104-109. Chinese
6. Ekaterina Bulinskaya, Julia Gusak. Optimal Control and Sensitivity Analysis for Two Risk Models [J]. Communications in Statistics - Simulation and Computation. Volume 45, Issue 5. 2016. pp 1451-1466.

****