International Theory and Practice in Humanities and Social Sciences
2025 Volume2,Issue2 ISSN 3078-4387

Machine learning-based multi-model fusion for investment value protection

Luyu Ma[1] **Haoyang Zhou**[1] **Hongyi Ling**[1]
[1]*Nanjing Forestry University, Nanjing, Jiangsu Province, China*

Article Info

Accepted:01 January 2025

Keywords:
Risk Prediction; TOPSIS;
Risk assessment; ARIMA;
Indicator Evaluation;

Corresponding Author:
Haoyang Zhou

http://doi.org/10.70693/itphss.v2i2.256

Abstract

Purpose: Extreme weather affects insurance risks and is highly uncertain. The protection of ancient buildings is an issue with prominent investment problems. This study aims to establish an assessment model to reduce risks in the insurance industry and raise the government's awareness of protection.
Method: First, data from the United States and Australia from February 2020 to 2024 was collected and processed to solve problems such as multi-source heterogeneity. The ARIMA model was used to predict natural disaster data, and the parameters were determined after pre-processing. Based on EWM-TOPSIS to establish a regional risk assessment model to calculate weights and give underwriting strategies; based on the underwriting model to analyze building protection strategies, establish an ESC model, select indicators from multiple dimensions, use K-means and AHP hierarchical modelling to determine cluster categories, and calculate the comprehensive value score within the scope.
Findings: Take Borobudur in Central Java, Indonesia as an example. The regional risk assessment model score is 32.6 (not recommended for insurance underwriting), and the building protection ESC model score is 6.65716 (the government needs to strengthen its protection).
Research value: Insurance risk models help insurance companies manage risks. The government should pay attention to the value of scenic spots, provide effective strategies for building protection investment, and promote the scientific development of related work.

1.Introduction

The domestic insurance industry has achieved remarkable development over the past decades and has become an important part of the financial system. To adapt to changing market demands, China's insurance industry is actively promoting market-oriented reforms, including the gradual relaxation of market access thresholds, encouraging foreign-funded insurance companies to set up branches, and promoting the diversification of insurance market players. These reform measures are not only aimed at enhancing market competitiveness but have also made significant progress in upgrading the level of insurance supervision and service quality. In the context of globalization, the expansion of the international insurance market and the emergence of new insurance products have made the insurance market more complex and diversified. To cope with this change, insurance regulators in various countries are strengthening standardized management of the insurance industry, aiming at enhancing risk management capabilities and promoting the stability and sustainability of the insurance market (Cacciotti et al., 2021).

At the same time, global climate change poses new challenges to the underwriting strategies of regional insurance companies. How to effectively reduce risks while ensuring the profitability of the company has become a pressing issue for insurance companies. The protection and insurance assessment of regionally distinctive buildings are also receiving increasing attention. In a disaster-prone situation of cultural heritage, insurance companies need to formulate appropriate underwriting strategies based on the socio-economic value and risk profile of the buildings to achieve a balance between risk minimisation and profit maximisation.

Overall, the insurance industry at home and abroad is actively responding to these market changes and challenges, promoting industry innovation, improving service quality, and safeguarding the long-term and stable development of the insurance market.

2. Literature Review

2.1 Climate change and insurance risk assessment studies

With global climate change, the impact of extreme weather on the insurance industry is becoming more and more significant, and there are also an increasing number of related studies. Cacciotti et al. (2021) pointed out that disasters caused by climate change pose risks to cultural heritage and other areas, and emphasised the importance of research on coping strategies for the insurance industry. In terms of insurance risk assessment and prediction, time series analysis methods such as the ARIMA model have been widely used. Hu, X. (2024) used the ARIMA model to predict the stability of slopes during heavy rainfall, and Wang et al. (2024) used the ARIMA-GM model to predict mine water inflows. These studies provide an effective method for predicting natural disaster data and provide a reference for the construction of a regional catastrophe prediction model in this study.

2.2 Application of the multi-indicator in insurance risk assessment

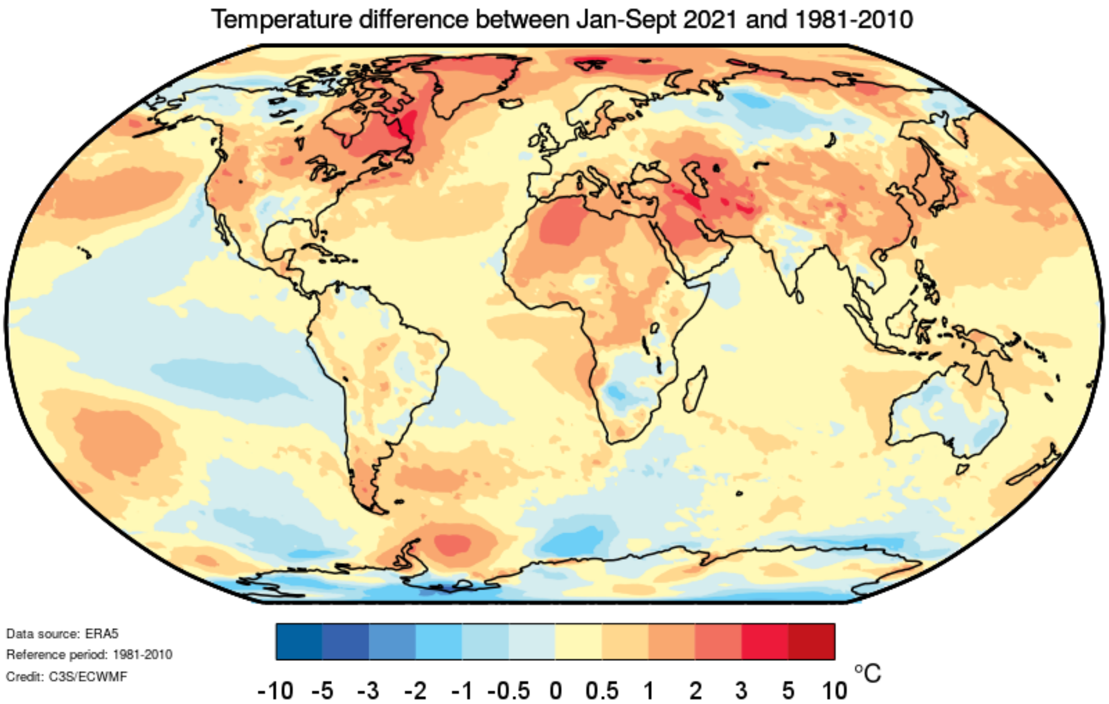
Multi-indicator decision-making methods play an important role in insurance risk assessment. The EWM-TOPSIS method, which combines the entropy weight method (EWM) and the distance method of the best and worst solutions (TOPSIS), has attracted much attention. Zhao et al. (2024) used the AHP-entropy weight method to evaluate the benefits of the Dianchi Lake wetland. Xiaofei, Y. (2022) studied the mechanism of circular economy development and the role of green finance based on the entropy weight method and big data. These studies all reflect the value of such methods in multi-indicator decision-making and provide a useful reference for objectively calculating the weight of indicators when constructing a regional risk assessment model in this paper.

2.3 Building conservation and value assessment studies

Research in the field of building protection and value assessment is carried out from multiple dimensions. Dai et al. (2024) optimised the extraction process of soothing and calming ointment by response surface method and UPLC-MS/MS, which reflects the research and protection ideas of cultural relics. Chen et al. (2022) investigated the current situation of typical modern buildings in the West Lake Scenic Area of Hangzhou and explored protection techniques, emphasising the need for a multi-dimensional assessment of building protection. These studies provide ideas for this paper to analyse building protection strategies based on underwriting models, comprehensively evaluate building value, and formulate protection measures by the government (Karthick et al., 2023).

The research results of predecessors provide strong support for this study in many ways, which helps to explore in-depth issues related to insurance risk assessment and building protection strategies, to provide more valuable references for the Insurance industry and government decision-making.

Figure 2.1: Global annual temperature difference



3. Methodology and Procedures

3.1Data collection and processing

To improve the insurance company's underwriting strategy, it is necessary to establish a regional catastrophe prediction model to confirm the number of future catastrophes in the region and for the prediction model, through the establishment of the relevant model can be realized to confirm the future period of extreme weather in a certain region and then brought into the evaluation and analysis of the model to determine the underwriting strategy. Crawling techniques were used to retrieve disaster data from the Global Catastrophe Data Platform (GCDP) for Texas, USA and Sydney, Australia for the period February 2020-2024, and then relevant issues were explored. The data were preprocessed using interpolation to calculate approximate substitutes to compensate for missing values, deletion to directly ignore invalid values and outliers, and multiple methods were considered for supplementing relevant missing values.

3.1.1 Multi-source heterogeneous integration of data

Multi-source heterogeneous data is a unified integration of different types of data, and the selection of evaluation indicators is more diverse, there are some obvious data clutter phenomena in the data of global natural disasters. This is mainly due to the wide range of sources of disaster data, the variety of data collection methods, the lack of uniform data statistical standards, the obvious data segmentation, the lack of uniform standards, the lack of uniformity in data types and other criteria, so the integration of diversified data is the key to data analysis and prediction to make full use of the data.

3.2.2 Data preprocessing

There are more missing values for the number of natural disaster occurrences in a region, natural disasters are often sudden and it is difficult to obtain information on all disaster events promptly, and some of the smaller impacts of disaster events may be overlooked or not posted on public platforms, which can lead to missing data being found. In addition, there is a false component of social groups, where some populations may be more willing to report more disaster events than they are, which can lead to more disaster events being included in the dataset. Disaster data may come from different data sources, and the quality and accuracy of these data sources may vary due to imbalances in regional development, some underdeveloped regions with limited disaster monitoring capacity may not be able to obtain full information on disasters promptly. Therefore, the nearest-neighbour value method, linear interpolation or machine learning method are used for interpolation and supplementation.

For outliers resulting from errors or mistakes in the sample collection process that cause the data to deviate from the normal range so that certain abnormal weather or events may lead to abnormal fluctuations in the data in the area, errors or omissions in the data counting process that lead to outliers, accidental errors or human intervention cannot be ruled out, and in some cases the collection and recording of the data may be affected by accidental errors or human intervention. In this study, the data were cleaned to remove or use interpolation to replace the data points that were anomalous.

3.2 Natural Disaster Data Forecasting

To improve the insurance company's underwriting strategy, this study establishes a regional catastrophe prediction model to identify the number of future catastrophes in the region and to provide a prediction model that identifies the extreme weather in a region for a certain period in the future and then brings it into the evaluation and analysis model to determine the underwriting strategy. The ARIMA (p,d,q) model of time series analysis is used to meet the requirement of data smoothness, and the difference method can be used to reduce the fluctuation of data. The values of p,d, and q are determined by the trailing and truncation of ACF and PACF images. The related data are fitted with standardisation, and the final result of optimising the predicted data is obtained (Hu, 2024).

3.2.1 ARIMA modelling

To improve the insurance company's underwriting strategy, it is necessary to build a regional catastrophe prediction model to identify the number of future catastrophes in the region and to build a prediction model that identifies the extreme weather in a region in the coming period and then bring it into the evaluation analysis model to determine the underwriting strategy.

According to the conditions of this question, it is appropriate to use the ARIMA (p,d,q) time series analysis model. Regarding the model, first of all, to satisfy its requirement of data smoothness, the difference method can be used to reduce data fluctuation. The trailing and truncation of ACF and PACF images determine the values of p,d, and q. The related data are standardized and fitted for the optimized prediction data conclusion.

In actual complex disaster environment conditions, due to the influence of many force majeure factors, predicting the number of disasters becomes a critical problem to be solved. Therefore, combining the known information on the number of relevant extreme weather occurrences and the primary data, establishing the relevant prediction model, as the differential autoregressive moving average model, is particularly important for the subsequent scheduling work.

Before the model is established, the relevant data are first preprocessed, and the interpolation and deletion methods are used to make up for the missing items and abnormal data in the data; the interpolation method is used to calculate the approximate replacement values to make up for the missing values, while the deletion method directly ignores the invalid values and abnormal values.

The ARIMA model combines three basic approaches:

The regression model (AR) describes the relationship between current and historical values, using historical time data of the variable to predict itself. Preprocessing of known relevant data reveals through observation that the current value of the time series has no external disturbing quantities but is determined only by the value of the series at past moments, and an autoregressive model describes this linear relationship:

$$x_t = \delta + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t \tag{1}$$

Among these denotes the constant term of the AR model, the coefficient of the regression term of order I, and the sequence residual. And the sequence value at the current moment.

The moving average (MA) model focuses on the accumulation of the error term in the autoregressive model, which is combined with the moving average model based on the AR model to form the ARMA model, that is:

$$x_t = \delta + \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \mu_i \varepsilon_{t-i} \tag{2}$$

The ARIMA(p,d,q) model can then be obtained by doing d-order differencing on the ARMA model:

$$y_t = \delta + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \mu_i \varepsilon_{t-i} \tag{3}$$

3.2.2 Forecast Result Display

Below is a graphical representation of disaster prediction data for Texas, USA and Sydney, Australia, with only the hail prediction data selected as representative to demonstrate the validity and reliability of the modelling (Tamimi, 2021).

Table 3.1: ARIMA parameter estimation results

	Value	StandardError	T-statistic	PValue
Constant	552.01	1599.1	0.3452	0.42994
AR{1}	0.048574	0.895	0.022812	0.1818
MA{1}	-0.84616	1.1171	-0.39968	0.0894
MA{2}	-0.15384	1.373	-0.064832	0.04831
Variance	1.426e+07	0.0054342	2.6242e+09	0

The small p-values for MA1, MA2 and Variance in the table indicate a strong statistical trend suggesting that the parameter has some explanatory power for the model (Wang, 2024).

The comparisons revealed that the model had a goodness of fit R^2 of 0.821, the model performed relatively well and was statistically significant, and it was considered that it was found to be more appropriate to predict the data using the ARIMA model.

Below are trailing truncated plots of the autocorrelation function and partial autocorrelation function for the correlation prediction in Hailstorm

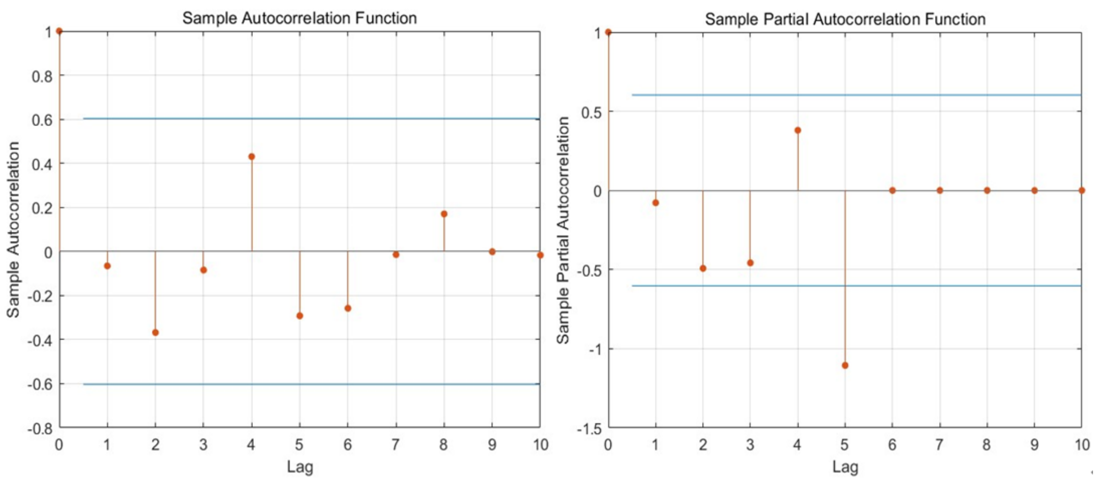


Figure 3.1: ARIMA PACF and ACF Charts

3.3Establishment of a regional risk evaluation model based on EWM-TOPSIS

3.3.1 Weight calculation based on entropy weight method

(1) Determine metric weights based on EWM methodology

In the multi-indicator decision-making problem, each indicator is represented by the importance of the indicator weight coefficients. Firstly, according to the insurance indicator system, the multi-tree clusters decision matrix is built and normalized to get the matrix X :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & x_{21} \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{4}$$

Where i = 1, 2, ..., m; j = 1, 2, ..., n is the value of the j-th indicator parameter for the i-th indicator.

(2) Standardization of decision matrices

Since the units of the parameters in the decision matrix are not consistent, the matrix needs to be dimensionless and standardized, and the resulting standardized matrix, Z, is standardized using the standardization method as follows:

(5)

$$Z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^{2}}} \tag{6}$$

(3) Calculate the probability matrix P

P_{ij} is the weight of each object for the first indicator, which is considered the probability used in the relative entropy calculation:

$$P_{\mathrm{ij}} = \frac{Z_{\mathrm{ij}}}{\sum_{\mathrm{i=1}}^{\mathrm{n}} Z_{\mathrm{ij}}} \tag{7}$$

(4) Using the associated probability matrix

For each indicator, an information entropy value is calculated to represent the expected value of the amount of information for that indicator:

$$\mathrm{e_j} = -\frac{1}{\ln \mathrm{m}} \sum_{i=1}^{\mathrm{n}} P_{ij} \ln(P_{ij}) \quad （\mathrm{j=1, 2..m}） \tag{8}$$

(5) Calculation of entropy weights based on information utility values

The larger the information utility value, the more information is available for the first indicator, so we can normalize the information utility value to determine the entropy weight of each indicator:

$$d_j = 1 - \mathrm{e_j} \tag{9}$$

$$W_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \quad （\mathrm{j} = 1, 2 \ldots m） \tag{10}$$

The final weights were calculated as follows:

Table 3.2: Factor weighting

Factor	Weight
Hailstorms	0.141311402
Tornadoes death number	0.100936847
Wildfires	0.107546188
Profit	0.394919529
Population	0.255286034

Data visualisation and statistical analysis of relevant conclusion scores 错误!未找到引用源。:

Based on the results of the analysis of the EWM-TOPSIS model, the weights assigned to each of the assessed factors were finally derived (see Table 3). These weights represent the relative importance of each factor in the decision-making process. Among them, population density has the highest weight (0.3949), indicating that this factor dominates the assessment of whether to insure or not, probably because areas with higher population density are exposed to greater risks, which in turn have a greater impact on the insurance decision. This is followed by the profit weight (0.2553), reflecting the importance of economic efficiency in decision-making (Xiaofei, 2022).

Comparatively speaking, the weights of hail (0.1413), mountain fire (0.1075), and tornado deaths (0.1009) are low, indicating that these natural disaster factors have a relatively weak influence on the overall assessment. By combining the weights assigned to subjective and objective factors, the EWM-TOPSIS model provides decision-makers with a more scientific and reasonable reference basis for more effectively carrying out the selection and optimisation of insurance solutions.

3.3.2 Giving Underwriting Strategies Based on Superior and Inferior Solution Distance Methods

TOPSIS is a multi-attribute evaluation model whose basic principle is calculating the distance between the best and the worst solutions to find the most desirable alternative closest to the positive ideal and furthest from the negative one. By maximizing the profit criterion and minimizing the cost criterion to achieve the best value solution for each alternative, we can also score each alternative to determine the insurance score for each region, according to the specific score, to give the insurance company's advice on insurance. The steps are as follows:

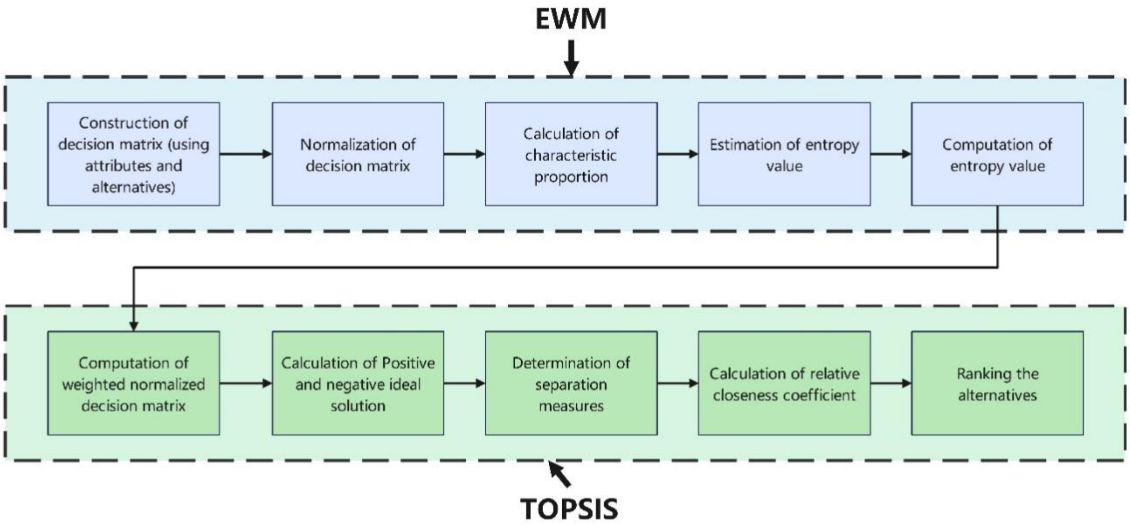


Figure 3.2: EWM-TOPSIS Flowchart

Define the distance of the $i(i=1,2,...,n)$ evaluation object from the maximum value D_i^+, then:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \omega_j (Z_j^+ - z_{ij})^2} \tag{11}$$

Similarly, the minimum distance is obtained as D_i^-. The un-normalised score for the $i(i=1,2,...,n)$ evaluation object is calculated.

The un-normalized score for the $i(i=1,2,...,n)$ evaluation subject is calculated as follows:

$$S_i = \frac{D_i^-}{D_i^- + D_i^+} \tag{12}$$

The scores were normalized and magnified one hundred times for grading purposes:

$$T_i = \frac{S_i}{\sum_{i=1}^{n} S_i} \tag{13}$$

Due to the overwhelming amount of data overall, consideration was given to displaying the insured scores for some districts:

Table 3.3: Model I Training set score

Serial Number	Score
1	58.51938084

2	51.65485698
3	57.15233102
4	58.32015208
5	51.84633048
6	49.77941234
7	22.30783494
8	37.1023103
9	42.93571277

Meanwhile, the Table gives the specific insured scores for selected districts. Areas numbered 1, 3 and 4 have scores of 58.52, 57.15 and 58.32 respectively, indicating that these areas are closer to or above average in terms of relatively low-risk or high insurance needs. In contrast, area number 7 has a significantly lower score of 22.31, suggesting that this area may be more at risk or that its insurance needs are lower.

Combining the data from the charts and tables shows that there is a wide variation in insurance scores between districts, reflecting the diversity of risk profiles and insurance needs in different districts.

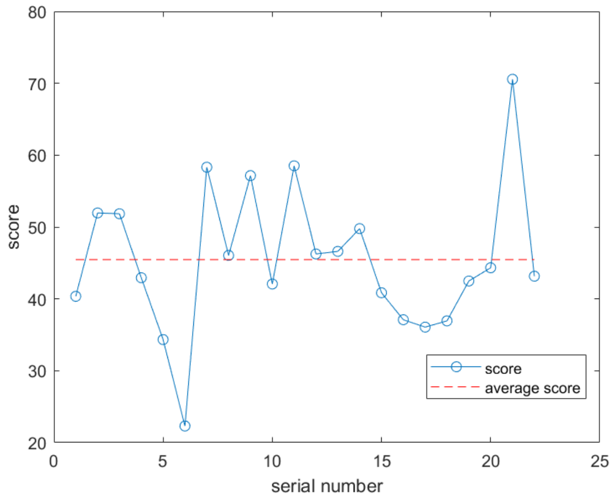


Figure 3.3: EWM-TOPSIS Flowchart

Chart 3 shows the insured scores for different districts (with consecutive numbering as the horizontal coordinate and scores as the vertical coordinate) and how these scores compare to the average score. The blue line graph labels the specific scores for each region, and the red dotted line represents the average score for all regions.

As can be seen from the graph, the scores of some districts fluctuate widely. For example, regions 7 and 9 scored below average, while region 20 scored significantly higher than other regions. The chart also shows that the scores fluctuate somewhat across multiple regions, which may be related to differences in risk factors or other environmental variables between regions.

Based on the scores, the areas were considered using fuzzy cluster analysis and divided into 3 classes based on the scores, thus dividing the assistive strategy levels into three classes. The definitions are as follows: Class I: recommended; Class II: may be considered; Class III: not recommended. The specific strategies are shown in the table below:

Table 3.4: Decision Interval Classification

Hierarchy	I	II	III
Recommended Strategies	Recommended	Considered	Not Recommend
Insure(S)	[54.6,75)	[33.4, 54.6)	[0,33.4)

Consider a case demonstration for the Sydney area versus the Texas area in the U.S. The model was used to solve for the insured score of:

Table 3.5: Area Scores and Ratings

Area	Score	Recommended Strategies
Sydney, Australia	43.1563872575671	II
Texas, USA	70.561198346518	I

From the enrollment scores, Sydney is the area to be considered for coverage, while Texas, USA, is the recommended area. This may be because Sydney has more natural disasters, while Texas has a lower overall risk of extreme weather.

3.4 Analysis of the problems of architectural conservation strategies based on underwriting models

As monument preservation is often a detailed direction of the underwriting issue, the underwriting strategy model will be used as a basis for a more targeted assessment of the value of designated attractions or monuments, and to provide advice to the local government on the extent to which preservation measures should be taken in respect of the buildings. Protective measures for attractions not only ensure that these cultural heritages are properly preserved and passed on but also attract a large number of tourists to develop the tourism industry to increase the government's fiscal revenue and reduce the government's fiscal deficit in times of economic downturn.

3.4.1 Building conservation ESC modelling based on underwriting strategies

To assess the cultural, historical, economic and community significance of the building for underwriting and preservation from different dimensions, the value of the building was classified into several dimensions, i.e., patronage, tourism revenue, using the clustering and hierarchical process (Analytical Hierarchy Process) to cluster and score several indicators hierarchically, and to determine the preservation strategy for the relevant building based on the scores, number of jobs provided, etc. To quantify the indicators of different dimensions into scores under the same criteria, a conversion factor is introduced. The highest value index in a value type is the maximum value index for that value type. They indicate the importance of different values, the higher the value, the greater the VI (Dai et al., 2024).

The evaluation of the value of classic architecture should be analysed comprehensively from three major perspectives: economic value, social value and cultural value. Economic value is usually reflected in the ability to attract a large number of tourists and drive local economic development. Social value refers to the social impact and contribution of the attraction to the local community and neighbouring residents. Attractions can provide employment opportunities for residents, drive local population growth, and enhance community cohesion and sense of belonging. Cultural value covers the history, culture, art and other values of a region or a country, and through the protection of local cultural values, it can enhance the cultural heritage of the city and contribute to the history of mankind.

Specific evaluation indicators are divided into three categories: economic value, social value and cultural value. In the economic value, passenger flow and economic income are the main indicators. The social value is determined by the population and the amount of employment provided. Cultural value is determined by the scale of cultural activities.

The quantitative calculation of the three categories of value is as follows:

Economic value is determined indirectly through the number of visitors, admission prices and related service revenues, and is collectively referred to as profitability value. The economic value of a monument refers to the wealth that can be created for people that can be measured in monetary terms, hence the modelling of the economic value: （σ: Income from tourism-related services）.

$$economic\ value\ \{ \begin{matrix} Tourism\ revenue(VI_r) \\ Numbers\ of\ visitors(VI_v) \end{matrix} \tag{14}$$

$$VI_r = \lambda_1 \times [\ visitor\ flow \times ticketprice + \sigma] \tag{15}$$

$$VI_v = \lambda_2 \times visitor\ flow \tag{16}$$

The social value can be judged by the attractiveness of the building through the resident population of the relevant place of residence, and then indirectly reflected by the social contribution of the building through the growth of the employed population, both of which are important indicators for judging the degree of social contribution, and the definition of social value is as follows:

$$social\ value\ \{ \begin{matrix} Population\ growth(VI_p) \\ Employment\ offering(VI_e) \end{matrix} \tag{17}$$

$$VI_p = \lambda_3 \times (\ Population_{i+1} - Population_i) \tag{18}$$

$$VI_e = \lambda_4 \times (Employment_{i+1} \text{ - } Employment_i) \tag{19}$$

Population: is the resident population of the area in a year。

Employment: is the number of people employed in the area in a year i.

Cultural values can be determined by the number of corresponding cultural activities as well as their duration, and the cultural value of a building usually reflects the values and unique cultural heritage of the local community, as well as specific reflections of historical periods and cultural perceptions.

$$\text{cultural value-Scales of cultural activities}(\mathrm{VI}_C) \tag{20}$$

n： Number of cultural events held at the site in a year.

$$Vc = \lambda_5 \times \sum_{i=1}^{n} (\text{Numbers of partipations} \times \text{Duration of activities}) \tag{21}$$

3.4.2 K-means and AHP-based hierarchical modelling

For the above multidimensional indicators, select the representative 120 countries as the research object, the above five indicators based on the systematic clustering algorithm autumn of the best classification categories, and then specify the number of categories as the K-means clustering category, and ultimately for the division of the five indicators of the number of categories to formulate VI (subscripts, i = 1,2,3,4,5) evaluation of the value of the range, based on the above algorithm can be Based on the above algorithm, the VI values of each of the five indicators for 120 countries are obtained, and finally, the hierarchical analysis method is used to assign weights to the five indicators and carry out weighted average to obtain the final VI value of the comprehensive evaluation index (ZHAO et al., 2024)。

The optimal number of data clusters can be obtained by using the systematic clustering algorithm and the Elbow Method (Elbow Method) as follows:

(1) Suppose n samples are divided into classes. Is the k -th class, define its degree of distortion D_K.

$$D_K = \sum_{i \in C_k} |x_i - \mu_k|^2 \tag{22}$$

(2) Define the total degree of distortion of all classes as the aggregation factor J:

$$J = \sum_{k=1}^{k} \sum_{i \in C_k} |x_i - \mu_k|^2 \tag{23}$$

The functional relationship between K and J is embodied in the correlation contour plot, and the value of k corresponding to its turn is the number of optimal sub-clustering categories, as shown in the following figure:

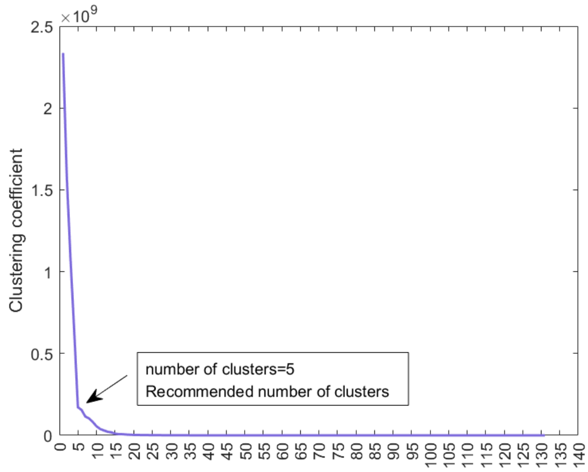


Figure 3.4: Number of clusters

As we know from the figure, a clear inflection point occurs when the value is 5, so the number of clusters is determined to be 5.

The selection of representative countries as the research object has strong universality and specificity. By using the K-means algorithm to cluster the dataset into five categories, i.e., the five metrics are used as the vertical coordinates of the five scatter plots, and the horizontal coordinates are processed as random numbers so that the clustering ranges corresponding to the subsequent VI values are based on the vertical coordinates of the scatter plots.

The output of the scatterplot of the five different indicators and the metric value of the clustering centre corresponding to each VI value are shown below.

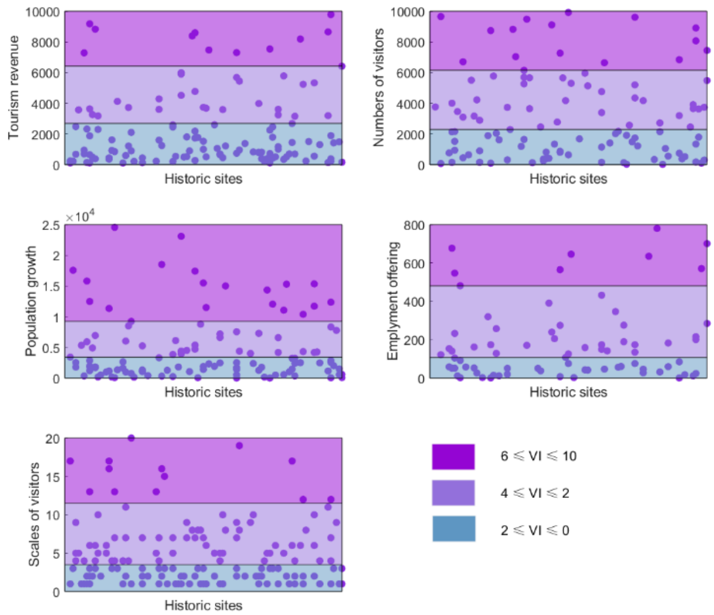


Figure 3.5: clustering diagram

For the 120 countries analysed to obtain a K-means clustering map, and then for each indicator VI value range division, divided into a total of five levels, the higher the VI value, the higher the overall score of the indicator (Abdulnassar & Nair, 2023). The specific division is shown in the following table:

Table 3.6: VI information based on the K-means algorithm

Em / VI	Tourism Revenue	Numbers of Visitors	Population Growth	Employment Offering	Scales of Visitors
2	[0, 2689)	[0, 2293)	[0, 3429)	[0, 108)	[0, 3)
4	[2689, 4773)	[2293, 4351)	[3429, 9246)	[108, 285)	[3, 7)
6	[4773, 6417)	[4351, 6150)	[9246, 12489)	[285, 481)	[7, 11)
8	[6417, 7539)	[6150, 8070)	[12489, 18508)	[481, 701)	[11,13)
10	[7539, 10000)	[8070, 10000)	[18508, 25000)	[701, 800)	[13, 20)

After obtaining the division interval of the VI value, it is necessary to assign the five VI values to get the comprehensive value score of a building. Hierarchical analysis was first proposed by Professor Satai of the University of Pittsburgh in the United States in the 1970s, which can achieve the purpose of comprehensive evaluation by scoring a variety of complex indicators. In investment, hierarchical analysis can be used to evaluate the level of investment risk in different regions, with a high degree of flexibility, which is widely recognised by the public. Based on this, this study adopts the hierarchical analysis method for empowerment.

For each of the five indicators, a consistency test was performed and the consistency indicator was calculated:

$$CI = \frac{\lambda \max - n}{n - 1} \tag{24}$$

Is the square matrix order the maximum eigenvalue of the matrix?

The average stochastic consistency index is found based on the square matrix order n. It is used to calculate the consistency ratio, which can reflect the degree of inconsistency of the matrix:

$$CR = \frac{CI}{RI} \tag{25}$$

The consistency passes as CR = 0.00063 < 0.10. The weights are found by the eigenvalue method and the maximum eigenvalue of the matrix $\lambda \max$ G and its corresponding eigenvector: and the weight of each indicator can be obtained by normalizing β. The rest of the two judgement matrices' specific judgement processes as above consistency test are passed.

Using AHP, the weights of each factor in the relative index and the weights of small categories in the overall index can be obtained, the final VI value can be calculated and the corresponding strategy rating can be carried out (Liu, Zhao& Miao, 2024). The specific performance is as follows:

Table 3.7: Weighted scale of values

Wight / Type of value	Total weight	Subindex	Weighting of Subscripts
Economy	0.5096	Tourism revenue	0.619
		Numbers of visitors	0.381

Society	0.3116	Population growth	0.6
		Employment offering	0.4
Culture	0.1787	Scales of cultural activities	1.0

Final scores for each index were calculated:
Economic Value Index:

$$VIeconomy = 0.619VI\mathrm{r} + 0.381VI\mathrm{v} \tag{26}$$

Social Value Index.

$$VI\mathrm{society} = 0.6VIp + 0.4VIe \tag{27}$$

Cultural Value Index:

$$VIculture = Vc \tag{28}$$

Final total value index:

$$VI = 0.510Veconmy + 0.3Vsociety + 0.179VIculture \tag{29}$$

The model innovatively combines the k-means clustering machine learning method to measure the comprehensive value of a scenic spot from three perspectives: economic, social, and cultural and makes constructive suggestions for the local government to decide how much protection measures should be taken according to the size of the specified scenic spot's VI value. A smaller scenic spot's comprehensive value is lower, and therefore, the government should vigorously protect the scenic spot (Chen et al., 2024). Define the correspondence between the VI value and the positive degree of protection measures as follows:

When the VI value is in [2,3.2) need to take active protection measures. Objective protective measures are required when the VI value is [3.2,4.6). Negative protective measures are required when the VI value is [4.6, 6].

4. Results and Discussion

4.1 Specific case studies and applications

In the centre of Java, Indonesia, there exists a ‘Borobudur’, which derives its name from the Sanskrit word ‘Vihara Buddha Ur’, meaning ‘hilltop stupa’, also known as a temple on a hill. It is also known as a temple on a hill. Borobudur is the world's largest Buddhist monumental stupa, also known as a temple on a hill. The reason for choosing this target architectural area is that the economic and cultural data of this area is complex and relatively low dimensional, and the impact of extreme weather is high so that the model can be fully used to propose appropriate strategies for complex environments, and due to the Borobudur Temple Compounds has a high tourism value and valuable cultural value, it will be very meaningful to study architectural preservation strategies in this area.

The regional risk evaluation model and architectural preservation ESC model proposed in this paper calculates the regional scores for the relevant building complexes, and after simulation using the model, the value and preservation strategy for this region will be proposed, as follows:

When considering the evaluation of historic landmarks, it is necessary to make model predictions in multiple dimensions, i.e. to analyse building underwriting and preservation strategies from multiple perspectives through different scores.

Based on the bank underwriting problem model above the number of extreme weather occurrences around the building is trained and analysed accordingly, with a final score of 32.6, which is ranked as Class III, i.e. conservation measures are not recommended. Considering the low latitude characteristics of Indonesia, which results in more extreme weather events such as storms, and the high incidence of fire and earthquakes in Indonesia from a geographic perspective, this score is delineated with a high degree of accuracy and credibility. The score is only 0.5 points different from consideration, indicating that the gap with consideration of coverage is not large, i.e., the government can improve the recommendation by strengthening measures such as prevention of disasters (Yan & Zhao, 2021). Using the architectural monument conservation model, calculations were made and scores were assigned according to the results, resulting in the following results:

Table 4.1: Scores for each factor

Norm	Tourism revenue	Numbers of visitors	Population growth	Employment offering	Cultural value
Score	10	6	6	4	4

From the analysis of the results, the building has a high income from tourists and a medium to high level of tourist arrivals, which represents a strong attraction value and a high level of income, with a high value added to the products sold and a strong economic benefit. The site has a medium level of population growth but a low to medium level of employment provision, indicating a certain level of contribution to employment and population growth, which is of relative value to social stability. The final VI value is obtained from the scores of each index:

Table 4.2: ESC indicators and scores

confluence of factors	Economy value	Social value	Cultural value	Final Score	NPR
Score	8.476	5.2	4	6.65716	32.9（III）

According to the scores of each index, this building is considered to have high economic value and medium social value and has certain cultural value. In the ESC model, the final score is 6.65716, which is considered to provide protection, and the underwriting model rating of III means that natural disasters in this area are more frequent and intense, so to protect the safety and integrity of the building, it is more necessary to provide a series of protection. From the results of the strategy, Borobudur Temple Compounds has a high economic value, and the underwriting model of III is more related to the local climate problems. The Indonesian side in recent years has been strengthening the protection of Borobudur Temple Compounds from extreme weather while preserving the monuments of the building by restoring the interior, reinforcing the exterior, and de-etching.

4.2 Advantages and extension of the model

4.2.1 Advantages of the model

1. The ARIMA model is very effective in dealing with time series data and can effectively use historical data to predict future trends and patterns.
2. The established mathematical model reduces the influence of subjective judgement on weight allocation through the calculation of information entropy, and the overall evaluation model has sufficient objectivity.
3. The model has a wide scope of application and can be used for decision-making in a variety of regions and under complex environmental conditions.
4. The model is flexible and adaptable and can be adapted to changing application scenarios by adding or reducing the influence of corresponding factors through the adjustment of relevant parameters.

4.2.2 Modelling applications

The model can help insurance companies assess the insurance risk and provide decision-making suggestions, thus increasing the income of insurance companies. According to the ARIMA prediction model, the future natural disasters in the designated area are predicted and then combined with other representative indicators to objectively calculate the investment risk index of the area for the insurance company's reference using the EWM-TOPSIS method to optimise the insurance company's decision-making, ensure the profitability of the insurance company, and reduce the risk of insuring. On this basis, it also provides the government with the attraction protection scheme based on K-means machine learning, which can quantify the decision-making scheme of how far to take protection measures and protect the value of the attraction to a certain extent. However, the use of multiple algorithms in the model, the overall model-building complexity, high computational costs, and the need for greater arithmetic support in the later stage.

5. Conclusion and Suggestion

By analysing the conclusions of this paper, a large number of regions around the world have natural disasters and other data for insured risk assessment, for multi-indicator research. The insurance risk model is an important tool for the internal risk control of insurance companies, which underwrite all kinds of risks, involving a wide range of business, large potential losses, and facing complex and diverse risks. In the first half of this study, an insurance risk model is constructed for a specific region, which is useful for improving the efficiency of risk management and helping insurance companies to better respond to the market competition and the needs of different regions, and to make corresponding dynamic changes in underwriting strategies. Insurance companies can use the model to identify and assess the impact of various types of risks on insured losses, to better reduce the risk of company losses. The second half of this study refines the assessment object, evaluates and analyses in a smaller scope, evaluates the value of scenic spots for the government, and carries out in-depth value assessment and analysis of designated scenic spots. Scenic spots with higher value can drive local economic development, increase employment, increase the cultural heritage of the host city, expand social influence, and have a very positive effect on the local government. The scenic spots with lower comprehensive value scores do not bring more benefits to the local area in the past development, so the government needs to take a series of protection measures to promote the sustainable development of economic, social and cultural values in these areas. The government can formulate relevant policies to supervise the environment and ecology of scenic spots, increase publicity, and also pay attention to its role in promoting society to ensure the sustainable use and protection of scenic spots, and at the same time promote the development of the local economy and cultural heritage.

In addition, this study also applies the above two types of models to the pagoda complex in Java, Indonesia, to predict its natural disasters, to apply the insurance assessment model to find its underwriting risk, and to bring the evaluation indexes mentioned in this paper into the ESC model of architectural protection to get a comprehensive score of its value, which provides an effective strategy for the government's protection and insurance investment.

Reference

Abdulnassar, A. A., & Nair, L. R. (2023). Performance analysis of Kmeans with modified in itial centroid selection algorithms and developed Kmeans9+ model. Measurement: Sensors, 2 5. https://doi.org/10.1016/j.measen.2023.100666

Cacciotti, R., Kaiser, A., Sardella, A., De Nuntiis, P., Drdácký, M., Hanus, C., & Bonazza, A. (2021). Climate change-induced disasters and cultural heritage: Optimizing management str ategies in Central Europe. Climate Risk Management, 32. https://doi.org/10.1016/j.crm.2021.10 0301

Chen, Y., Jiang, L., Liu, J., Hao, Z., Liu, J., Wang, Y., & Shen, L. (2022). Investigation on the status quo of typical modern buildings in West Lake Scenic Area of Hangzhou and exp loration on protection techniques. Zhejiang Architecture, 02, 61–64+68. https://doi.org/10.1587 4/j.cnki.cn33-1102/tu.2022.02.001

Dai, T., Xu, F., Wang, J., Zhao, J., Yang, Z., You, X., & Wang, Y. (2024). Optimization o f the extraction process of Shu Xin An Shen Cream based on response surface methodology and UPLC-MS/MS. Journal of Shanxi Medical University, 03, 354–361. https://doi.org/10.137 53/j.issn.1007-6611.2024.03.012

Hu, X. (2024). Predictive analysis of strong rainfall stability of slopes based on time series ARIMA model. Pearl River Water Transportation, 04, 74–76. https://doi.org/10.14125/j.cnki.zjs y.2024.04.013

Karthick, S., Kermanshachi, S., Pamidimukkala, A., & Namian, M. (2023). A review of cons truction workforce health challenges and strategies in extreme weather conditions. Internationa l Journal of Occupational Safety and Ergonomics, 29(2), 773–784. https://doi.org/10.1080/1080 3548.2022.2082138

Liu, K., Zhao, S., & Miao, Y. (2024). Green vegetable impurity image segmentation based o n SSA-Kmeans clustering algorithm. Chinese Journal of Agricultural Mechanization, 02, 151–156. https://doi.org/10.13733/j.jcam.issn.2095-5553.2024.02.022

Tamimi, O. Y. (2021). The role of internal audit in risk management from the perspective o f risk managers in the banking sector. Australasian Accounting, Business and Finance Journa l, 15(2), 114–129. https://doi.org/10.14453/aabfj.v15i2.8
Wang, J., Guo, Q., Wang, F., Aviso, K. B., Tan, R. R., & Jia, X. (2021). System dynamics simulation for park-wide environmental pollution liability insurance. Resources, Conservation and Recycling, 170. https://doi.org/10.1016/j.resconrec.2021.105578
Wang, Y., Wang, Z., Liu, C., Li, J., Bai, J., Yang, W., & Zhao, Y. (2024). Prediction of m ine water influx based on ARIMA-GM model. Coal Technology, 09, 154–157. https://doi.org/10.13301/j.cnki.ct.2024.09.031
Xiaofei, Y. (2022). Research on the action mechanism of circular economy development and green finance based on the entropy method and big data. Journal of Enterprise Information Management, 35(4–5), 988–1010. https://doi.org/10.1108/JEIM-01-2021-0024
Yan, Y., & Zhao, Z. (2021). Research on the integral conservation and development of archi tectural heritage and traditional villages - A case study of Foguang Temple and Yan Jiazhai Village in Wutai Mountain. Urban Architecture, 04, 105–108+126. https://doi.org/10.19892/j.cnki.csjz.2021.04.25
Zhao, S., Zhao, C., Jiang, N., & Zeng, S. (2024). A study on the benefit assessment of Dia nchi lagoon wetland based on the AHP-entropy weight method. Journal of Environmental Sci ence, 04, 411–420. https://doi.org/10.13671/j.hjkxxb.2023.0303