Research on insurance decision-making based on the TOPSIS evaluation model and K-means clustering algorithm

Yufei Liu [1], Qiao Lei [2], Weiliang Li [3], Yongqiang Li [1], Yixuan Ma [4], Zhu Chen [4, *]

[1] School of Finance and Taxation, Inner Mongolia University of Finance and Economics, Hohhot China, 010051

[2] School of Cyber Security, Shanxi Police College, Taiyuan, China, 030401

[3] School of Law, Inner Mongolia University of Finance and Economics, Hohhot, China, 010051

[4] School of Finance, Inner Mongolia University of Finance and Economics, Hohhot, China, 010051

* Corresponding author: chenzhusxy@126.com

Abstract. Increasing losses caused by extreme weather events led to property insurance costs soaring and becoming more challenging to obtain, leaving insurance companies and property owners in a severe crisis. This study aims to address the challenges posed by climate change to the insurance industry and explore the impact of extreme weather events on property insurance. We introduce the TOPSIS evaluation model and K-means clustering algorithm to assess the risk level in different regions and provide the optimal basis for insurance decisions. By analyzing global disaster data and emission indicators, this paper found that some regions face the risk of rejecting underwriting, and used a clustering algorithm. The results show that the increase in property insurance costs aggravates the challenges in the insurance industry, but through the comprehensive assessment model and cluster analysis, we can provide more accurate risk assessment and decision support for insurance companies. By assessing the risk level and effective classification in different regions, it provides insurers with more accurate risk assessment and decision support. The results of this research are of great significance for dealing with the property insurance problems caused by extreme weather events and help the insurance industry better adapt to the challenges of climate change.

Keywords: Property insurance, Climate change, TOPSIS model, K-means clustering algorithm.

1. Introduction

Increasing losses caused by extreme weather events led to property insurance costs soaring and becoming harder to obtain, leaving insurance companies and property owners in a serious crisis. The growth of global property insurance costs and the protection gap continue to widen, highlighting the dual challenges of insurance industry profits and the ability of property owners to pay. The study of Morshedi M A and Kashani H. proposed a new decision model designed to address the limitations of past mitigation behavior models to effectively evaluate the owner response to mitigation policies [1] by considering building characteristics, owner attitudes, and condition-dependent responses to incentives. Rising J et al. shows that unquantified, poorly understood, and uncertain risks from climate change should be included in the economic assessment and decision-making process. We outline these risks, including interdisciplinary knowledge delays, spatial and temporal variation in climate impacts, interactions between risks, knowledge uncertainty, and unidentified risks. Highlighting the need for collaboration between the natural and social science communities to address these gaps and provide an approach to integrative assessment [2]. Rezvani S M H S et al.'s study introduced the Urban Resilience Assessment System (URES), which contains 16 indicators and 75 parameters, covering five key dimensions of environment, economy, organization, society, and technology. The method of constructing asset elasticity scoring relies on a new multi-standard decision analysis (MCDA) method adapted from Hierarchical Analysis (AHP), involving the concept of automatic rationality and consistent decision making (ARCDM), and organizes the behavioral

scenarios [3] by simulating the behavioral scenarios of a virtual expert panel. This study established a comprehensive decision model of property insurance and social stability, evaluated the risk level of different regions through the TOPSIS evaluation model, and solved the countries refusing underwriting combined with the emission indicators, to provide the optimal decision basis for insurance companies in the context of climate change. The research using the K-means clustering algorithm on the six numerical features of 30 regions, through the histogram, scatter plot, and elbow rule determined the optimal cluster number of 4, and using PCA data dimension reduction to 2 d and 3 d plane, finally divided 30 regions into 4 different clusters, and show the clusters of each region.

2. The fundamentals of the TOPSIS and K-means models

2.1. The fundamentals of the TOPSIS model

The TOPSIS model is a multi-criterion decision analysis method to determine the best selection scheme. It is based on evaluating the alternatives by multiple indicators and determines the best choice by comparing the similarity between each alternative and the ideal solution. TOPSIS Models are widely used in decision problems, especially in scenarios where multiple factors need to be considered. Weight is the degree to which a given factor should be considered (default weight for all factors = 1). For example, you want one factor to be more weighted than the other factor, so the factor can be 2, while the other can be 1. Impact means a given factor has either a positive or a negative influence. Just as you want factor A to be as high as possible but factor B to be as low as possible, you will assign “+” weights to factor A and”-” weights to factor B., this method can be applied to rank machine learning models by various factors, including correlation, R^2, accuracy, and root mean square error. A given data value for a given factor will be considered standard units. Always label-code [4-6] for any non-digital data type.

The first step is calculating the normalization matrix and the weighted normalization matrix. This article normalizes each value in the following way: where the number of rows in the dataset is the number of columns. I varies along the row and the column.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{kj}}} \text{ , =1,2,…,}m\text{, }j\text{=1,2,…,}n \tag{1}$$

This article then multiplies each value in the column by the corresponding weight. Calculate the ideal best, ideal worst, and ideal worst values for each row, as well as the Euclidean distance. First of all, this article needs to find out the ideal optimal value and the ideal worst value: now we need to look at the impact, i.e. whether it is a "+" or "-" effect. If the "+" affects a column, the ideal optimal value is the largest, the ideal worst value is the minimum value in the column, and vice versa for the "-" effect.

Now this article needs to calculate the Euclidean distance of all the elements in the row from the ideal best and the ideal worst, where is the worst distance calculated in line i, where is the value of the element, and is the ideal distance in the column worst. Similarly, this article can find d_{ib}, which is the optimal distance calculated online.

$$d_{iw} = \sqrt{\sum_{j=1}^{n} (t_{ij} - t_{wj})^2}, \ i = 1, 2, ..., m, \tag{2}$$

Finally, the Topsis score and ranking are calculated. Now that this article has positive and negative distances, let this article calculate the Topsis score for each row based on them. The TOPSIS score = $d_{iw} \ / \ (d_{ib} \ + \ d_{iw})$ for each row is now ranked by the TOPSIS score, i.e., the higher the score, the better the ranking.

2.2. The basic fundamental of the K-means clustering algorithm

The K-means clustering algorithm is a commonly used unsupervised learning method for dividing samples in a dataset into different classes. Its basic structure includes initialization, assigning samples, updating cluster centers, and iteration. First, it is necessary to select K initial clustering centers, which can be chosen randomly or based on data characteristics. Each sample is then assigned to the category to which it belongs to the nearest cluster center. Next, calculate the new clustering center for each category, i.e., the mean of all samples in that category. Then, repeat the steps for sample assignment and cluster center update until a stop condition is reached, such as the maximum number of iterations or the cluster center no longer changes. The resulting clustering center is the result of clustering, and each sample is assigned to the category to which the clustering center belongs [7-9]. The core idea of the K-means clustering algorithm is to cluster the samples in the dataset by continuously iteratively assigning the samples to the nearest clustering center and updating the clustering center. The algorithm is widely used in various fields, such as data mining, pattern recognition, and image segmentation. It has a wide range of applications in data mining, data analysis, anomaly detection, pattern recognition, financial risk control, data science, intelligent marketing, and data operation [10].

K-means often determine the K-value based on SSE and profile coefficients. To calculate SSE, given a data set containing data objects, the set of categories produced by clustering by the K-means algorithm is defined as a set$C = \{C_1, C_2, \cdots, C_n\}$. The formula for SSE is as follows:

$$\mathrm{SSE}(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} |x_i - m_k|^2 \tag{3}$$

Where m_k is the center point of the clusterc_k, which is calculated as follows:

$$m_k = \frac{\sum\limits_{x_i \in C_k} x_i}{|C_k|} \tag{4}$$

Each sample has a corresponding profile factor, which consists of two scores:

(1) The average distance of the sample from other sample points in the same cluster class (i.e., quantifying coherence);

(2) The average distance of the sample from all sample points in the closest cluster class (i.e., quantitative resolution). The profile factor for each sample is defined as:

$$S = \frac{b - a}{\max(a, b)} \tag{5}$$

Where$S = [-1, 1]$. The smaller the SSE, the better the clustering effect; The larger the contour coefficient, the better the clustering effect.

3. Results

3.1. The analysis of the TOPSIS model

In this paper, a TOPSIS evaluation model is established to determine the risk level in the different regions. First, the data on disaster frequency, total deaths, total impact, and total consumption were counted in 58 countries, which were used to calculate the disaster risk level and economic implications. Then taking the emission index as an environmental factor, this paper selects Tajikistan and Turkey, two countries that refuse to underwrite, and brings the above TOPSIS model to solve the problem. The calculation results are shown in Table 1, the underwriting measures score is shown in Table 2, and the underwriting decision results are shown in Figure 1.

Table 1. Results of TOPSIS treatment

Country	Risk Level	Economic impact	Relative proximity	Underwriting strategy	bear fruit
Bhutan	8.98E-09	0	0.841069	0.546256	Standard underwriting
Timor-Leste	9.90E-06	0	0.584117	0.649631	Refuse to cover
Ukraine	4.83E-04	0.005523	0.616098	0.316149	Standard underwriting
Uzbekistan	1.12E-06	0	0.3331	0.783008	incentive measures
Yemen	2.43E-05	0.001937	0.311675	0.710312	incentive measures
Armenia	0.00E+00	0	0.852697	0.542419	Standard underwriting
Israel	1.50E-05	0.001307	0.33289	0.736377	incentive measures
Iraq	5.35E-05	0.000006	0.581387	0.650749	Standard underwriting

Table 2. Scores of underwriting measures

tactics	score
Standard underwriting	25
incentive refuse to cover measures	14
Refuse to cover	5

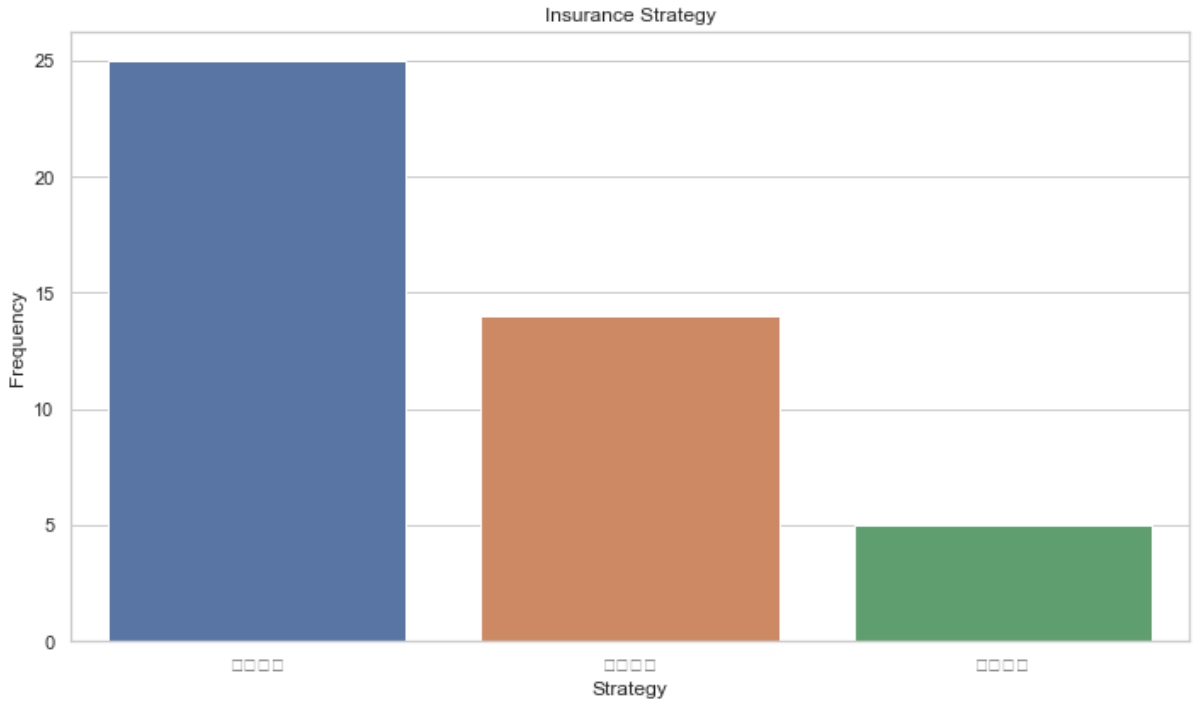


Figure 1. Distribution of underwriting decision results

Based on the distribution of underwriting decisions, we can see that 57% of countries or territories are covered by standard coverage, 31% are within the scope of incentives, and only 11% are within the scope of denial of coverage or increased premiums because this paper chose the index of emission as an environmental factor, this can be changed by man. In this paper, we can take out the above-mentioned countries that refuse to cover separately, analyze them, and bring them into the TOPSIS model to solve the problem by selecting two countries, Tajikistan and Turkey. By controlling emissions in the five countries, for example, after reducing to the original 40%, 60%, and 80%, observe the change of underwriting strategy. In the case of Tajikistan, the results are in Table 3.

Table 3. Relationship between Tajikistan's carbon emission reduction and underwriting strategy

Country	Carbon emissions reduction	Relative proximity	Underwriting strategy
Tajikistan	100%(fixedly)	0.13	Refuse to cover
Tajikistan	80%	0.29	Refuse to cover
Tajikistan	60%	0.49	Standard underwriting
Tajikistan	40%	0.77	incentive measures

The above data shows that when carbon emissions are reduced by 80%, there is a specific effect, but the underwriting strategy remains the same: when carbon emissions are reduced to 60%, you can enjoy standard coverage, and when carbon emissions are reduced to 40% of the original value, you can enjoy incentives. The same is true for Turkey's data. The results are in Table 4.

Table 4. Relationship between Turkey's carbon emission reductions and underwriting strategies

Country	Carbon emissions reduction	Relative proximity	Underwriting strategy
Turkey	100%(fixedly)	0.28	Refuse to cover
Turkey	80%	0.36	Standard underwriting
Turkey	60%	0.56	Standard underwriting
Turkey	40%	0.77	incentive measures

The above data shows that when carbon emissions are reduced by 80%, the underwriting strategy changes from denial to standard underwriting. When carbon emissions are reduced to 60%, the relative proximity is further increased, but it still does not meet the criteria for incentives. Incentives are available when carbon emissions are reduced to 40% of the original value.

But this is very difficult to do in real life. It can be concluded that when the carbon emission reduction of a country or region is reduced to a certain extent, it will be more acceptable to insurance companies. Therefore, it is recommended that property owners take some measures to reduce carbon emissions in their daily lives as much as possible, such as reducing the use of plastic bags, reducing the use of private cars switching to public transport, etc.

3.2. The analysis of K-means clustering algorithm

In this paper, the annual average precipitation (mm), annual average yearly temperature (°C), frequency of extreme weather events, altitude (m), geographical location, historical disaster data, and vegetation coverage (%) were selected as the characteristics of 30 regions, and the K-means clustering algorithm was used for clustering. Histograms and scatter-plots of the 6 numeric features are plotted, as shown in Figure 2. The final clustering results for each region are shown in Table 5.

Table 5. Cluster results for all regions

Region	Cluster	Region	Cluster	Region	Cluster
Region 1	1	Region 4	0	Region 7	1
Region 2	2	Region 5	2	Region 8	3
Region 3	2	Region 6	2	Region 9	1

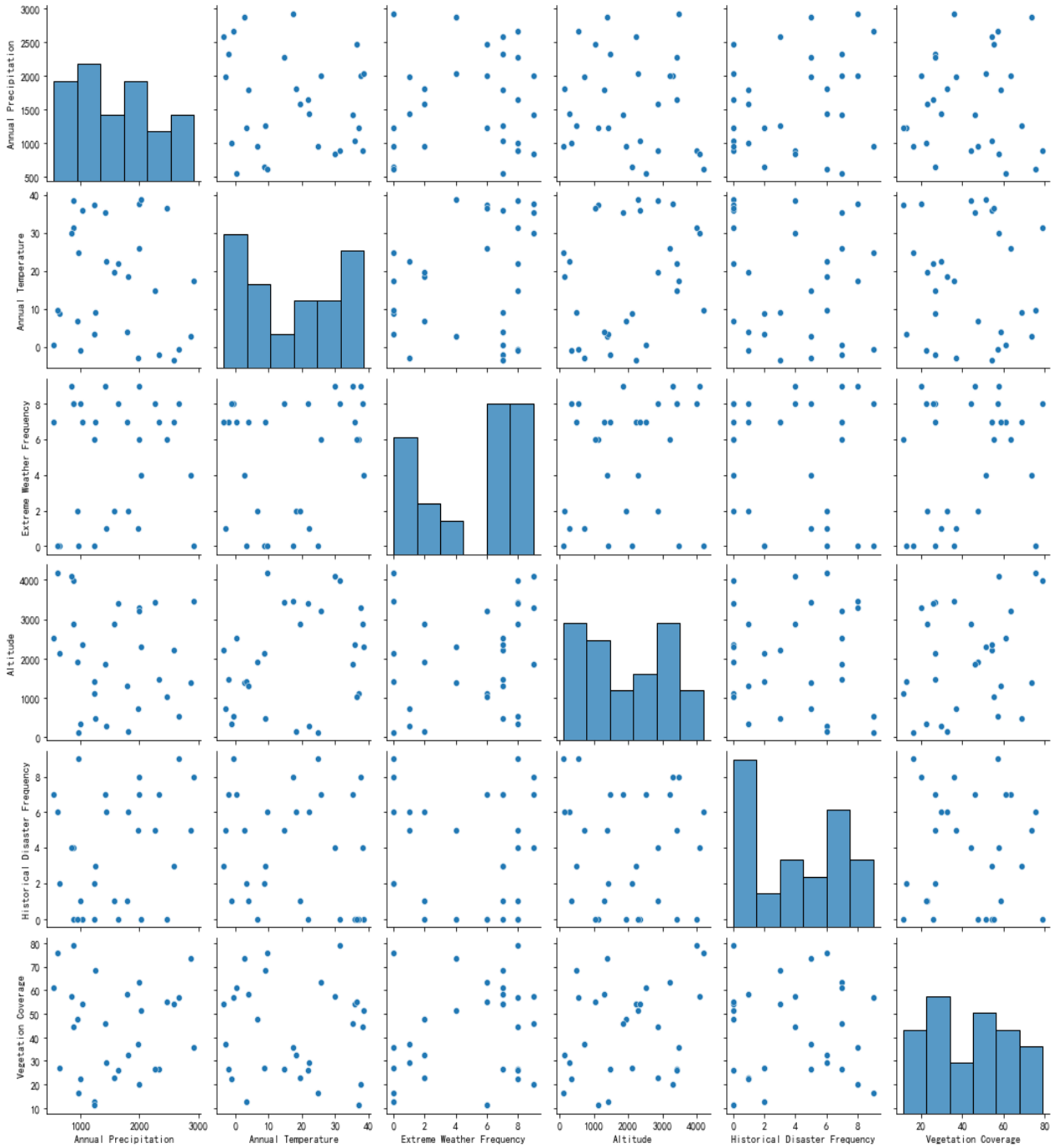


Figure 2. Relation between the numerical features

4. Conclusions and outlooks

Based on the TOPSIS evaluation model and K-means clustering algorithm, this study discusses and analyzes the decision-making problems of the insurance industry in the face of the challenges brought by climate change. By establishing a comprehensive decision-making model of property insurance and social stability, combined with the TOPSIS evaluation model to evaluate the risk level of different regions, and using the K-means clustering algorithm to analyze the six numerical characteristics of 30 areas, it provides a decision-making basis for insurance companies to optimize the allocation of insurance in the context of climate change. The results of this study show that the increasing trend of global property insurance costs has exacerbated the dual challenges of profit and property owners' ability to pay in the insurance industry. Through the analysis of the TOPSIS evaluation model and emission indicators, this study finds that some countries or regions are within

the scope of refusal to underwrite or increase premiums, and through the application of the K-means clustering algorithm, the regions can be effectively clustered, providing more accurate risk assessment and decision-making basis for insurance companies.

The selection of data selection and indicators mentioned in the paper may have certain limitations. While this paper considers factors such as the frequency of disasters, total deaths, total impacts, and total consumption, these factors may not fully cover all factors that influence climate change risk and insurance decisions. In the TOPSIS model, the choice of weights and influencing factors may have a greater impact on the results. However, this paper does not provide a specific method for determining the weights and impact factors, which may lead to the results of the model results being less robust and reliable. Therefore, more research is needed to determine the appropriate values for these parameters and to verify their impact on the model results.

References

[1] Morshedi M A, Kashani H. Assessment of vulnerability reduction policies: Integration of economic and cognitive models of decision-making [J]. Reliability Engineering & System Safety, 2022, 217: 108057.

[2] Rising J, Tedesco M, Piontek F, et al. The missing risks of climate change [J]. Nature, 2022, 610 (7933): 643 - 651.

[3] Rezvani S M H S, de Almeida N M, Falcao M J, et al. Enhancing urban resilience evaluation systems through automated rational and consistent decision-making simulations [J]. Sustainable Cities and Society, 2022, 78: 103612. Ma Kunlong. Short-term distributed load forecasting method based on big data [D]. Changsha: Hunan University, 2014.

[4] Xiaoli Du, Yan Wang, Fengxia Chen. Evaluation of coal-resource-based cities transformation based on CRITIC-TOPSIS model [J]. Sustainable Cities and Society, 2024, 103.

[5] Dai R, Xiao C, Liang X, et al. Spatial-temporal evolution law analysis of resource and environment carrying capacity based on game theory combination weighting and GMD-GRA-TOPSIS model. Evidence from 18 cities in Henan Province, China [J]. Journal of Cleaner Production, 2024, 439: 140820.

[6] M. R Y, S. A, S. M. Integrated AHP-TOPSIS models to evaluate the suitability and sustainable development of beach tourism: a case in Chennai City, Tamil Nadu, India [J]. Journal of Coastal Conservation, 2024, 28 (1).

[7] Rutinaldo N A, Barroca Á N, Freitas D S Y B, et al. A new hybrid optimization approach using PSO, Nelder-Mead Simplex and K-means clustering algorithms for 1D Full Waveform Inversion. [J]. PloS one, 2022, 17 (12): e0277900 - e0277900.

[8] Li Y. Empirical analysis of the impact of financial development on the income gap between urban and rural residents in the context of large data using fuzzy Kmeans clustering algorithm [J]. International Journal of Electrical Engineering Education, 2020, 002072092093683.

[9] Lin X, Xu J. Road network partitioning method based on Canopy-Kmeans clustering algorithm [J]. Archives of Transport, 2020, 54 (2): 95 - 106.

[10] Wang G, Meng D, Chen R, et al. Automatic Rice Early-Season Mapping Based on Simple Non-Iterative Clustering and Multi-Source Remote Sensing Images [J]. Remote Sensing, 2024, 16 (2):