Transactions on Economics, Business and Management Research
ISSN: 2960-1789, eISSN: 2960-2254 | Volume 9
GBMS 2024

A study of the impact of extreme weather events on property insurance based on ARIMA time series model and SVM classifier

Qiyu Lei*

International Business College (IBC), South China Normal University, Foshan, China, 528225

* Corresponding Author Email: qiyu.lei@m.scnu.edu.cn

Abstract. The thesis aims to tackle the challenges posed by the frequency of extreme weather events globally to the property insurance industry. The essay effectively predicts the frequency of extreme weather events and their economic impact on the insurance industry by developing an integrated analytical framework that combines insurance and coverage models. Firstly, an ARIMA time series model was used to forecast future extreme weather events and combined with the Spearman correlation coefficient (SCC) to quantify the relationship between insurance company revenues and socio-economic factors in the insured region, which were used as inputs to a support vector machine (SVM) classification model to assess risk and determine the amount of insurance coverage. In addition, the variables affecting decision-making were adjusted by the entropy weight method to further enhance the accuracy and usefulness of the model. The results of the study show that the constructed model has a high accuracy of 95% in predicting extreme weather events and their impact on the insurance industry, providing a powerful risk management and resource allocation tool for insurance companies, and helping to improve the resilience of property and the overall resilience of the insurance industry.

Keywords: ARIMA; SVM; Extreme weather; Insurance industry.

1. Introduction

In recent years, the frequency and intensity of extreme weather events such as hurricanes, floods, droughts, heat waves, and cold snaps have increased due to climate change [1]. This poses a significant threat to homeowners' property, including their homes, vehicles, and commercial buildings [2]. Insurers face challenges as a result of increased extreme weather events, including increased claim frequency and payouts, which can impact their financial position and profitability. To tackle these challenges, insurers may raise premium rates, restrict coverage in certain areas, or create new insurance products to cover the risk of extreme weather events [3]. Additionally, insurers will need to reevaluate their risk models to ensure accurate assessment and management of the risk of extreme weather events [4]. In summary, climate change has resulted in an uptick in extreme weather events, which present risks and challenges for insurers and properties [5].

The essay reviews and analyses the relevant bibliographies to improve understanding of the context of extreme weather in the property insurance industry. The analysis in the essay focuses on three of these bibliographies. The first bibliography, entitled "Insurance and Climate Change" [6], provides an in-depth discussion of the significant challenges posed by climate change to the insurance industry, in particular the specific impacts of acute and chronic weather risks on different segments of the insurance sector. The article reviews the mitigation, adaptation, and innovation measures taken by the insurance industry to address these challenges and highlights the need for future research and advancements to support the resilience of the global insurance industry. Furthermore, the article considers the evolving role of the insurance industry in climate-related risk management, and how it can enhance its capacity to manage climate risks by improving risk assessment models, adopting advanced analytical techniques, and developing innovative insurance products and strategies. The second bibliography, entitled "Climate Risk Assessment in Insurance: A USA and Africa Review" [7], identifies the challenges and opportunities faced by the two regions in addressing the risks posed by climate change. It compares the practice of climate risk assessment in the mature insurance market in the United States with that of the emerging insurance market in Africa. The study indicates that the US insurance industry is adapting to more frequent and severe weather events such as hurricanes, wildfires, and floods through the use of technology, data analytics, and collaboration with climate scientists. In contrast, Africa requires collecting localized data, implementing community resilience initiatives, and developing innovative insurance products to cope with extreme weather events such as droughts and storms. The article emphasizes the importance of international cooperation in sharing best practices and facilitating the transfer of knowledge and technology and highlights the dual role of insurers in promoting climate resilience and sustainability. The findings offer valuable insights for policymakers, companies, and other stakeholders in the global insurance industry, encouraging them to take proactive measures to enhance climate resilience. The third bibliography, entitled "Impact of Climate Change on the Insurance Industry: Coping Strategies of Insurance Companies and Regulators" [8], employs a twofold approach. Firstly, it analyses IPCC reports and reinsurance company statistics to examine the impacts of global warming on the insurance industry. Secondly, it considers the increase in insured losses due to the high frequency of natural disasters and the dual impacts of climate change on insurance liabilities and assets. The article also proposes strategies that insurers and regulators should adopt to adapt to and mitigate the effects of climate change. These include disaster and loss prevention, risk transfer, and support for a low-carbon economy.

There are several methodological shortcomings in the prior relevant bibliographies in studies analyzing the impact of extreme weather on the insurance industry. Firstly, studies rely heavily on historical data, which may limit the ability to predict future climate change uncertainty, as historical trends do not fully reflect future climate patterns. Secondly, although the bibliographies recognize the potential link between climate change and extreme weather events, there are difficulties in quantifying such impacts, failing to adequately account for the complexity of climate change and multi-scale interactions. In addition, studies may fail to adequately assess non-climatic factors, such as land use changes and disaster management measures, which have a significant impact on risk assessment and insured losses. What’s more, the bibliographies may have overlooked the broader economic impacts of indirect losses and secondary disasters triggered by extreme weather events. Finally, the literature lacks specific methodologies for assessing and managing long-term risks in the face of uncertainty about future climate change. In contrast, this essay directly gives strong evidence of the increasing frequency of extreme weather events through temporal regression predictions and incorporates the Spearman Correlation Coefficient (SCC) to quantify the relationship between insurance company revenues and socioeconomic factors in the insured region, which are used as inputs to a Support Vector Machine (SVM) classification model for assessing risk and determining the amount of insurance coverage. In addition, the variables affecting the decision are adjusted by the entropy weighting method to further enhance the accuracy and usefulness of the model.

2. Modelling and solution

2.1. The introduction of the ARIMA model

ARIMA model is known as the Autoregressive Integrated Moving Average Model (ARIMA). ARIMA model is a statistical model used for analyzing and forecasting time series data. ARIMA model can capture various patterns in the data such as trend, seasonality, cyclicality, etc [9].

Among them:

- AR This part of the model uses the dependence between observations and their own lagged values (i.e., observations from the previous period) to make predictions.
- This part of the model uses the difference in the observations to remove non-stationarity from the time series.
- MA This part of the model uses the dependence between the error term of an observation and its lagged error term to make predictions.

For the ARIMA model, a smooth model should be needed. The ADF test (Augmented Dickey-Fuller Test) is a commonly used unit root test to check whether the time series data is stable. Its basic principle is to test whether the root of an autoregressive model of a time series is a unit root.

The original hypothesis of the ADF test is that there is a unit root, i.e. the time series is non-stationary. If the p-value is less than the level of significance (e.g. 0.05, taken here), then the original hypothesis is rejected and the time series is considered to be stable.

The mathematical form of the ADF test is as follows.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \varepsilon_t \quad (1)$$

where y_t is the time series, Δ is the difference operation, α、 β、 γ and δ_i are the parameters to be estimated, and ε_t is the error term.

The modeling idea of the ARIMA model is illustrated in Figure 1.

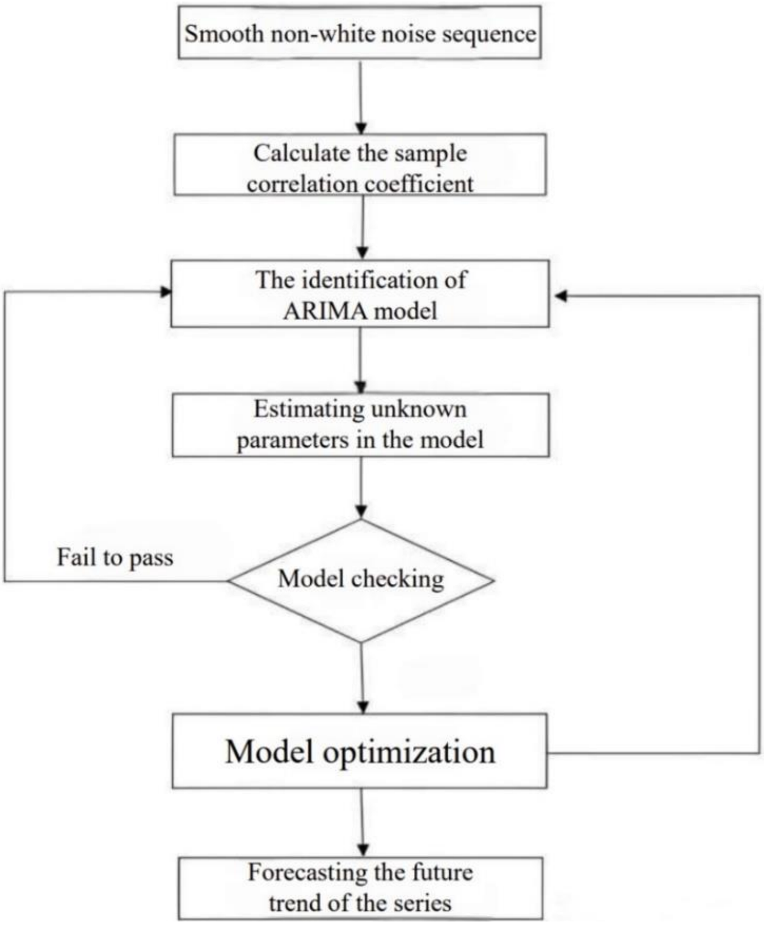

Figure 1. The architecture of ideas for ARIMA modeling

2.2. The introduction of SVM model

The full name of the SVM model is Support Vector Machines (SVM), a binary supervised learning model mainly used for classification and regression problems. It relies on finding optimal decision boundaries to distinguish between different classes or predict continuous values.

The SVM model performs as follows: to map the data into a high-dimensional space, choose the appropriate kernel function first. Afterwards, create and solve a convex optimization problem that maximizes the sample interval. This yields the separating hyperplane and the classification decision function, where the only variables that affect the model are the support vectors. Ultimately, the optimization process yields the predictor function (regression problem), which is based on the support vectors—that is, the data points that have the biggest impact on the model's decisions. The SVM model is used in regression analysis to predict continuous values using support vectors, which aids in trend analysis and numerical forecasting.

SVM models are classified as linear and non-linear. The following describes the process of deriving the classification decision function for linear SVM model.

First, the linearly separable training dataset is partitioned, where each sample x_i has an associated class label y_i. y_i can have a value of $+1$ or -1.

In the second step, find a hyperplane such that $w \cdot x + b = 0$, where w is the normal vector of the hyperplane and b is the bias term.

In the third step, for a point, its geometric interval to the hyperplane is $\frac{|w \cdot x_i + b|}{||w||}$

In the fourth step, the goal of the SVM is to find a hyperplane such that the geometric interval of this hyperplane is maximized. This can be expressed as the following optimization problem:

$$\max_{w,b} \frac{1}{\| w \|} \tag{2}$$

In order to make the problem easily solvable, the following equivalent form is usually used:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \tag{3}$$

When solving the problem, it is necessary to ensure that the hyperplane correctly classifies all training samples. For all i, there is $y_i(w \cdot x_i + b) \geq 1$

The next step is to solve the optimization problem with constraints. This is achieved by introducing Lagrange multipliers ($\alpha_i \geq 0$) to multiply each constraint, thereby obtaining the Lagrange function.

$$L(w, b, \alpha) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{n} \alpha_i [y_i(w \cdot x_i + b) - 1] \tag{4}$$

where n is the number of training samples.

Immediately after that, the expressions of w and b with respect to α are obtained by solving the minimum of the Lagrangian function by setting the partial derivatives of w and b to zero, and then the pairwise problem is obtained by substituting these expressions into the Lagrangian function:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j x_i \cdot x_j \tag{5}$$

where the constraint is $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$

In the seventh step, the optimal α value is obtained by solving the dyadic problem, and the corresponding sample point x_i is a support vector only if $\alpha_i > 0$.

The final classification decision function can be expressed as

$$f(x) = \sum_{i} \alpha_i y_i \cdot x_i \cdot x + b \tag{6}$$

where x is the new sample point to be classified and b can be computed from any support vector:

$$b = y_k - \sum_{i} \alpha_i y_i \cdot x_i \cdot x_k \tag{7}$$

where k is the index of any support vector.

2.3. The Introduction of Entropy Weight Method

To lessen subjectivity and increase the objectivity and accuracy of decision-making, the entropy weight approach, which is based on information theory, quantifies the impact of each indicator on the total evaluation by computing its entropy value [10].

If the entropy value of an indicator is higher, it means that the information of that indicator is more confusing, and therefore its weight should be reduced when making decisions.

In the first step, the data were standardized to eliminate the differences in the scales of the indicators and to compress the data in the interval [0,1].

Standardized formula for positive indicators:

$$X_{ij}^{1} = \frac{X_{ij} - \min\left(X_{1j}, \dots, X_{mj}\right)}{\max\left(X_{1j}, \dots, X_{mj}\right) - \min\left(X_{1j}, \dots, X_{mj}\right)} \tag{8}$$

Standardized formula for negative indicators:

$$X_{ij}^{1} = \frac{\max\left(X_{1j}, \dots, X_{mj}\right) - X_{ij}}{\max\left(X_{1j}, \dots, X_{mj}\right) - \min\left(X_{1j}, \dots, X_{mj}\right)} \tag{9}$$

In the second step, the numerical weight P_{ij} of each indicator in each sample is calculated.

$$P_{ij} = \frac{x_{ij}^{1}}{\sum_{i=1}^{m} x_{ij}^{1}} \tag{10}$$

The third step calculates the entropy of the j th indicator e_j.

$$k = \frac{1}{\ln(m)} \tag{11}$$

$$e_j = -k \sum_{i=1}^{m} P_{ij} \ln\left(P_{ij}\right) \tag{12}$$

In the fourth step, the variance index d_j is calculated for the j th indicator.

$$d_j = 1 - e_j \tag{13}$$

In the fifth step, the entropy weight i.e. the weight of the j th indicator w_j is calculated.

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \tag{14}$$

Finally, the composite score Z_i is calculated for the i th evaluator.

$$Z_i = \sum_{j=1}^{n} w_j \cdot P_{ij} \tag{15}$$

The entropy weight method, a multi-indicator decision analysis tool, calculates each indicator's entropy value to quantify its impact on the overall evaluation. This provides an objective quantitative basis for decision-making and improves the decision-making process's fairness, transparency, and scientificity. It plays a crucial role in enhancing the effectiveness and caliber of decision-making in challenges that follow by streamlining the decision-making process, optimizing resource allocation, supporting dynamic adjustment, and being relevant to a range of decision-making scenarios.

3. Results

3.1. The result of projections of the frequency of future extreme weather events

In this session, the ARIMA model is employed for predictive modeling purposes, with the anticipated frequency of extreme weather in the future being obtained as a result. This outcome thus enables the subsequent dynamic and rational allocation of insurance premiums for various regions. The resulting data visualization is depicted in Figure 2.

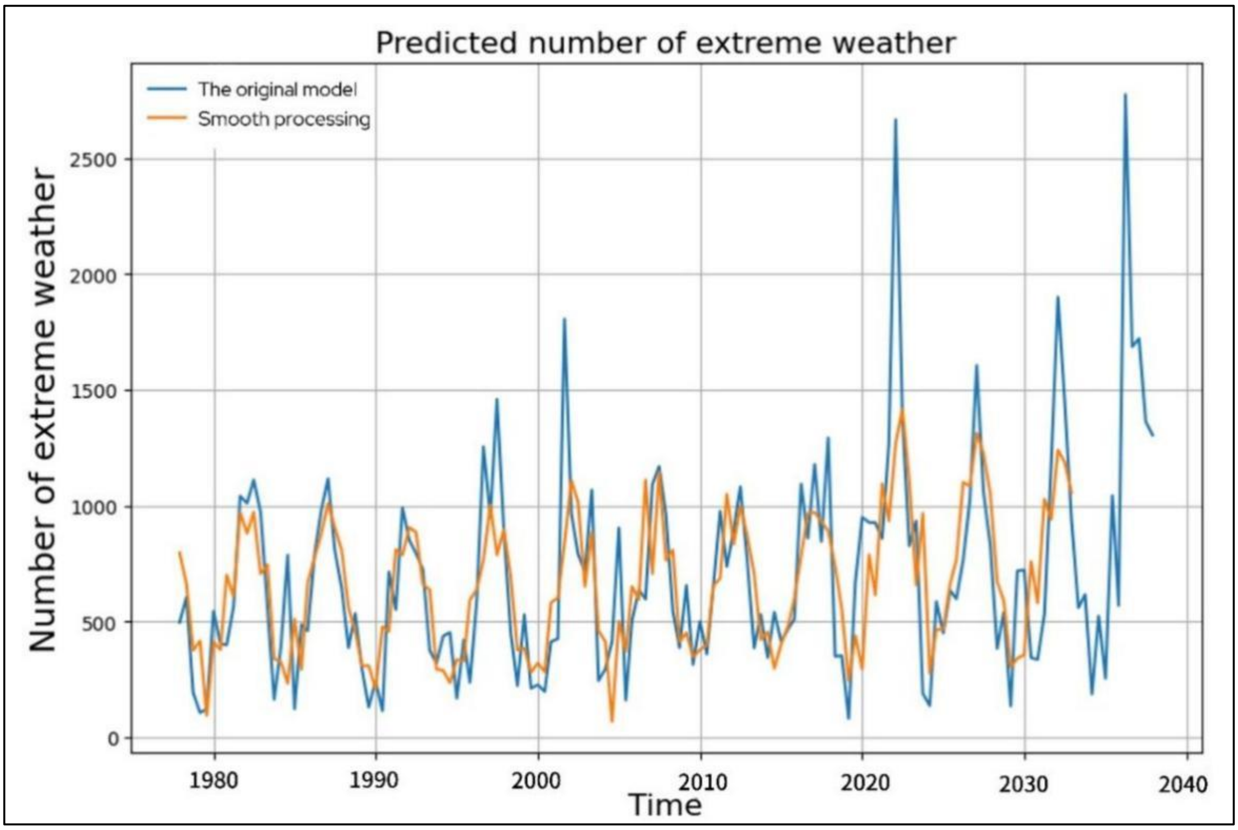

Figure 2. Map of predicted future frequency of extreme weather events

A statistical analysis of data on extreme weather occurrences globally over the past 40 years has been conducted, and the ARIMA model has been employed to make a reasonable prediction of the frequency of extreme weather occurrences over the next 20 years.

The result indicates that the frequency of extreme weather events has increased significantly over the past four decades, largely due to global warming and climate change. The ARIMA model predictions suggest that this trend will continue at a rapid and alarming rate over the next 20 years unless climate change is effectively addressed. The significant increase in the frequency of extreme weather events has the potential to significantly impact the revenue situation of the insurance industry. The mechanism can be understood as follows: a significant increase in the frequency of extreme weather events will increase the number of insurance claims made by individuals for their personal property, in cases where they have property insurance. This will lead to a significant increase in the insurance company's expenditure on compensation. However, their income, i.e. the number of insurance services ordered, also increases, but the profitable nature of insurance companies is such that the business insured is not honored as often. The frequency of extreme weather will result in the insured business being honored more frequently, thereby reducing the insurance company's income. This is not conducive to the growth of the insurance industry.

3.2. The Result of modeling dynamic adjustment of insurance amount based on regional risk

Subsequently, the obtained extreme weather prediction data is combined with the Spearman Correlation Coefficient (SCC) to quantify the relationship between insurance company revenues and socioeconomic factors in the insured areas [11]. These inputs are then used as inputs to a Support Vector Machine (SVM) classification model for assessing risk and determining the amount of insurance coverage, as well as to build a model of the amount of insurance coverage for the dynamic areas.

From a theoretical standpoint, the use of a Gaussian kernel function is optimal for addressing regional property risk assessment problems that exhibit nonlinearity, a limited amount of data, and complex features. This approach yields superior classification outcomes. Consequently, the RBF kernel function is the most suitable for this task [12]. In the process of training the SVM vector machine classification model, the following parameters were set in Table 1.

Table 1. Model Summary Table

Name	parameter name	parameter value
Model Parameter Setting	Data preprocessing	None
	Proportion of training sets	0.8
	Error term penalty coefficient	1.0
	Kernel	Rbf
	factorization value of the kernel function	0.01
	Multi-categorical decision function	Over
	Model convergence parameters	0.001
	Maximum number of iterations	2000
	Accuracy	74.51%
Model Evaluation Effectiveness	Precision rate (combined)	78.81%
	Recall rate (combined)	74.51%
	Fl-score	0.73

As illustrated in the accompanying table, the final model achieved an accuracy of 74.51% on the test set, a precision (combined) of 78.81%, a recall (combined) of 74.51%, and an f1-score (combined) of 0.73. The model is effective. The final results, which are presented in this paper, are used to partially classify whether policies are covered in different regions of the United States.

Subsequently, a simple sensitivity analysis model was constructed. The objective of this analysis was to assess the sensitivity of the insurance amount allocation strategy to changes in different variables, particularly with respect to the expected profit and the amount of compensation required per person. Consequently, the resulting dynamic insurance amount allocation model was optimized. The optimized model is presented in Figure 3.

STATE	Yes(1) or No(0)
ALABAMA	1
ALASKA	1
AMERICAN SAMOA	1
ARIZONA	0
ARKANSAS	1
ATLANTIC NORTH	1
ATLANTIC SOUTH	1
CALIFORNIA	0
COLORADO	1
CONNECTICUT	0
DELAWARE	0
DISTRICT OF COLUMBIA	1
E PACIFIC	1
FLORIDA	1
GEORGIA	0
GUAM	1
GULF OF ALASKA	1
GULF OF MEXICO	0
HAWAII	1
IDAHO	1
ILLINOIS	0
INDIANA	0
IOWA	1
KANSAS	0
KENTUCKY	1
LOUISIANA	1
MAINE	0

Figure 3. Classification results

3.3. Results after optimizing the model by entropy weighting method

The developed model is currently undergoing improvement with the objective of providing property risk assessment recommendations that are more specific and responsive to the needs of communities and property developers. The entropy weighting method is employed to assign weights to different indicators, thereby enabling the model to reflect more objectively the degree of influence of each factor in property risk. This method assesses the importance of the indicators based on the magnitude of the information entropy of the evaluation indicators, thus avoiding subjective bias and ensuring the fairness and accuracy of the weight allocation. After applying the weights determined by the entropy weighting method to the input data of the classification model, the model was found to be more accurate in assessing property risk, especially in identifying high-risk areas. The high-risk areas are now more accurately assessed as having a property risk that is too high, and thus property development and construction are not recommended there. This improvement provides the community with a more reliable tool for assessing property risk and also helps property developers avoid potentially high-risk areas in their investment decisions, increasing the success and sustainability of their projects. The result is shown in Figure 4.

In-depth comparison with the classification results of Problem 1, it is easy to find that the model's output produced a significant change after the entropy weighting method was used to assign weights to different indicators. This change is mainly reflected in the fact that more districts or regions are assessed as having excessive real estate risk, which tends to face a higher frequency of extreme weather, higher regional crime rates, or poor performance on key indicators such as per capita insurance expenditure.

The implications of this change are far-reaching, as it directly impacts the decision-making process of property developers. In the previous model, some higher-risk areas were incorrectly assessed as low or medium-risk areas due to the weighting of certain key risk factors being ignored, leading to a higher risk of investment loss and potential social problems. In the optimized model, by increasing the weights of these key indicators, the model can more accurately reflect the actual risk profile of the region, providing the insurance industry with a more reliable decision-making basis for assessing regional insurance returns.

The resulting graph is analyzed and recommendations are provided. Both insurers and property developers face a number of challenges and opportunities when considering the impact of extreme weather on the insurance industry. Insurers must assess the potential risks of extreme weather events on real estate projects and provide appropriate insurance products to protect investors and owners from financial loss. It is imperative that developers consider these risks at the site selection and design stage. This necessitates the choice of locations with stable environments, a low risk of natural disasters, and the utilization of building materials and construction techniques that can withstand extreme weather. Concurrently, insurers can attract customers by offering innovative risk management solutions, such as premium incentives for green building and sustainable design, to encourage developers to adopt more environmentally friendly and weather-resistant construction methods. In addition, insurers can collaborate with developers to offer a range of insurance products, including risk insurance during construction, natural disaster insurance, and property warranty insurance, to cover the entire property lifecycle from construction to maintenance. When providing their services, developers should consider how to enhance homebuyers' confidence through insurance products, such as offering all-risk insurance programs that include extreme weather events such as earthquakes, floods, and typhoons. By implementing these measures, developers will not only be able to reduce their customers' concerns about potential losses resulting from extreme weather events, but also enhance the market competitiveness of their projects. Ultimately, this cross-border collaboration benefits the entire real estate industry and the insurance industry, enabling them to better adapt to the challenges posed by extreme weather and achieve sustainable development.

STATE	Yes(1) or No(0)
ALABAMA	1
ALASKA	1
AMERICAN SAMOA	1
ARIZONA	1
ARKANSAS	1
ATLANTIC NORTH	0
ATLANTIC SOUTH	1
CALIFORNIA	0
COLORADO	1
CONNECTICUT	0
DELAWARE	0
DISTRICT OF COLUMBIA	1
E PACIFIC	1
FLORIDA	0
GEORGIA	0
GUAM	1
GULF OF ALASKA	1
GULF OF MEXICO	0
HAWAII	0
IDAHO	1
ILLINOIS	0
INDIANA	0
IOWA	1
KANSAS	0
KENTUCKY	1
LOUISIANA	0
MAINE	0

Figure 4. Decision-making results after re-empowerment in selected districts

4. Conclusions

This paper employs an ARIMA time series model to forecast the frequency of future extreme weather events and a Spearman phase relationship to quantify the relationship between insurance company revenues and socio-economic factors in the insured region. The aforementioned metrics are then fed into a support vector machine classification model, which is used to assess risk and determine the amount of insurance. The entropy weighting method has been employed to optimize the model, thereby enabling a more accurate assessment of property risk and identification of high-risk areas. Although the model demonstrated high accuracy in predicting extreme weather events and their impacts on the insurance industry and provided a risk assessment and resource allocation tool for insurance companies, there were still problems with data quality and quantity, model complexity, multi-scale interactions, and insufficient consideration of non-climatic factors. To enhance the model, it is necessary to conduct further validation and tuning, undertake integrated multi-model analyses, foster interdisciplinary collaboration, and conduct in-depth studies of long-term risk management and uncertainty.

References

[1] Zhou Botao, Qian Jin. Interpretation of the IPCC AR6 report: Changes in extreme weather events[J]. Progress in Climate Change Research, 2021, 17(6): 713.

[2] Pastor-Paz J, Noy I, Sin I, et al. Projecting the effect of climate change on residential property damages caused by extreme weather events[J]. Journal of Environmental Management, 2020, 276: 111012.

[3] Kraehnert K, Osberghaus D, Hott C, et al. Insurance against extreme weather events: An overview[J]. Review of Economics, 2021, 72(2): 71-95.

[4] Wang, Xiangnan. Climate change and the insurance industry: impacts, adaptation and mitigation[J]. Financial Regulation Research, 2020, 11: 46-61.

[5] Odunaiya O G, Okoye C C, Nwankwo E E, et al. Climate risk assessment in insurance: A USA and Africa Review[J]. International Journal of Science and Research Archive, 2024, 11(1): 2072-2081.

[6] Gupta A, Venkataraman S. Insurance and climate change[J]. Current Opinion in Environmental Sustainability, 2024, 67: 101412.

[7] Odunaiya O G, Okoye C C, Nwankwo E E, et al. Climate risk assessment in insurance: A USA and Africa Review[J]. International Journal of Science and Research Archive, 2024, 11(1): 2072-2081.

[8] Chen Xiaofeng, Huang Lu. Impact of Climate Change on the Insurance Industry: Coping Strategies of Insurance Companies and Regulators[J]. Southern Finance, 2010 (5): 55-58.

[9] Wang X, Kang Y, Hyndman R J, et al. Distributed ARIMA models for ultra-long time series[J]. International Journal of Forecasting, 2023, 39(3): 1163-1184.

[10] Zhu Y, Tian D, Yan F. Effectiveness of entropy weight method in decision-making[J]. Mathematical Problems in Engineering, 2020, 2020: 1-5.

[11] Kumar K P, Reddi V. Significance of Spearman's Rank Correlation Coefficient[J]. International Journal For Multidisciplinary Research, 2023, 5(4).

[12] Bong, K.-H. Comparison of SVM classification kernel functions and parameter selection[J]. Computer Engineering and Applications, 2011, 47(3): 123-124.