systems

MDPI

Article

Predicting Earthquake Casualties and Emergency Supplies Needs Based on PCA-BO-SVM

Fuyu Wang [1,2], Huiying Xu [1], Huifen Ye [1], Yan Li [1,*] and Yibo Wang [1]

[1] School of Management Science and Engineering, Anhui University of Technology, Ma'anshan 243032, China; ahutwfy@ahut.edu.cn (F.W.); xhytiger9166@163.com (H.X.); yehuifen1762020356@163.com (H.Y.); wangyibo1018@163.com (Y.W.)

[2] Key Laboratory of Multidisciplinary Management and Control of Complex Systems of Anhui Higher Education Institutes, Anhui University of Technology, Ma'anshan 243032, China

* Correspondence: liyan4@ahut.edu.cn

Abstract: The prediction of casualties in earthquake disasters is a prerequisite for determining the quantity of emergency supplies needed and serves as the foundational work for the timely distribution of resources. In order to address challenges such as the large computational workload, tedious training process, and multiple influencing factors associated with predicting earthquake casualties, this study proposes a Support Vector Machine (SVM) model utilizing Principal Component Analysis (PCA) and Bayesian Optimization (BO). The original data are first subjected to dimensionality reduction using PCA, with principal components contributing cumulatively to over 80% selected as input variables for the SVM model, while earthquake casualties are designated as the output variable. Subsequently, the optimal hyperparameters for the SVM model are obtained using the Bayesian Optimization algorithm. This approach results in the development of an earthquake casualty prediction model based on PCA-BO-SVM. Experimental results indicate that compared to the GA-SVM model, the BO-SVM model, and the PCA-GA-SVM model, the PCA-BO-SVM model exhibits a reduction in average error rates by 12.86%, 9.01%, and 2%, respectively, along with improvements in average accuracy and operational efficiency by 10.1%, 7.05%, and 0.325% and 25.5%, 18.4%, and 19.2%, respectively. These findings demonstrate that the proposed PCA-BO-SVM model can effectively and scientifically predict earthquake casualties, showcasing strong generalization capabilities and high predictive accuracy.

Keywords: principal component analysis; prediction of earthquake casualties; Bayesian optimization; support vector machine; emergency material demand forecasting

check for updates

Academic Editor: Vladimír Bureš

Received: 13 November 2024
Revised: 29 December 2024
Accepted: 31 December 2024
Published: 2 January 2025

Citation: Wang, F.; Xu, H.; Ye, H.; Li, Y.; Wang, Y. Predicting Earthquake Casualties and Emergency Supplies Needs Based on PCA-BO-SVM. *Systems* **2025**, *13*, 24. https://doi.org/10.3390/systems13010024

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Earthquakes, as sudden natural disasters, have a profound impact on human society, resulting in casualties, structural damage, and economic losses. The accurate forecasting of earthquake casualties and resource requirements is crucial for effective rescue operations and post-disaster reconstruction efforts. Therefore, thoroughly considering the relationship between these two aspects and conducting scientific and rational predictions and planning can enable rescue organizations to more effectively strategize their relief efforts, optimize resource allocation, and enhance rescue efficiency, thereby minimizing the humanitarian disasters caused by catastrophic earthquakes [1–4]. Establishing a robust earthquake casualty prediction system and integrating scientific methodologies and statistical models holds significant importance for improving disaster response capabilities and mitigating disaster losses [5,6].

In general, there are two primary methods for predicting emergency material demand: direct and indirect approaches. Among the direct prediction methods identified in past research are the Relief Demand Forecast Model, the Fuzzy Clustering-based Model, the Bayesian Decision Framework, and the Fuzzy Rough Set Approach over Two Universes [7–10]. In order to make the model still have better expressiveness in different scenarios and in the absence of information, some new methods have emerged in recent years. Among them, Zheng et al. [11] proposed a co-evolutionary fuzzy deep transfer learning (CoFDTL) method for disaster relief demand prediction, which was applied to two real-world disaster events in China in 2018, which can be used to solve complex multi-task transfer learning problems with insufficient samples and uncertain information. Considering that it is challenging to collect and transmit the needs to remote shelters, a post-disaster demand forecasting system based on smartphone-based DTN using principal component regression analysis and case-based reasoning is proposed [12]. Rawls et al. [13] developed a heuristic called the Lagrangian L-shaped method (LLSM) to address large-scale instances of determining in advance the location and quantity of various types of emergency supplies to be deployed without being sure if or where a disaster occurred. Based on the background of public health emergencies, Ma et al. [14] established a metabolic gray model (1,1) (GM (1,1)) and a material demand prediction model to predict the number of infections and material demand and established a dual-objective optimization model to help improve the fairness of material distribution, but the prediction model is based on the first-order differential equation, assuming that the data change is exponential, and its applicability mainly lies in describing linear or near-linear changes, and it is difficult to capture complex nonlinear relationships. In the context of flood disaster, Chen et al. [15] used the improved ant colony optimization algorithm combined with the BP neural network to model the demand for emergency supplies, and the experimental results showed that the proposed algorithm IACO-BP is the most efficient and stable among the three algorithms, which can better assist the disaster situation. Al-Ghamdi et al. [16] proposed a PSO-ANN water demand forecasting model, which adjusted the hyperparameters of the artificial neural network by using PSO, and verified that the model is an accurate model for multivariate time series forecasting. Rostami-Tabar et al. [17] studied the effectiveness of parametric and non-parametric demand forecasting methods and found that the guided method has high performance in achieving higher service levels. Lin et al. [18] proposed a big data-driven dynamic estimation model of disaster relief material demand, which fuses Baidu big data and the dynamic population distribution of the multilayer perceptron (MLP) neural network to improve its accuracy and timeliness. In the earthquake scenario, Fei et al. [19] proposed a demand forecasting model for emergency supplies that combined the Dempster–Shafer theory (DST) and the case-based reasoning (CBR) method and established a dynamic demand forecasting model by using inventory management theory. Riachy et al. [20] proposed a framework for demand forecasting in the presence of a large number of data gaps and validated the proposed method on a real dataset of UK footwear retailers. Focusing on the seismic background, Dey et al. [21] used seismic data to analyze and predict intelligent solutions. In order to improve people's ability to cope with earthquake disasters, Wei et al. [22] developed a semi-empirical model for casualty estimation. Horspool et al.'s [23] research focuses on understanding the background and causes of casualties caused by earthquakes.

Although the above methods have certain effectiveness in directly predicting the demand for emergency supplies, in the context of earthquakes, the method of indirectly predicting the demand for materials by predicting earthquake casualties has its unique advantages. Firstly, indirect forecasting can provide a more holistic view of the human impact of a disaster, allowing for a more accurate picture of changes in material needs. Direct forecasting methods tend to focus on a single demand estimate, while indirect forecasting

combines multiple factors such as the number of casualties and the extent of damage to better adapt to complex disaster scenarios. Secondly, indirect forecasting methods are more flexible in the face of uncertainty. In particular, the suddenness and unpredictability of earthquakes make direct prediction more challenging, while indirect prediction can more effectively deal with the problem of information asymmetry by analyzing casualty data and historical cases, and then providing more reliable material demand estimates. In addition, the indirect prediction method can be derived from existing casualty prediction models in the absence of sufficient data, thereby reducing the dependence on a large amount of high-quality data. In summary, indirect forecasting has higher adaptability and flexibility than direct forecasting in emergency material needs assessment, which can provide strong support for decision-making in complex and uncertain environments, and lay the foundation for subsequent emergency material distribution and emergency response.

In order to scientifically and reasonably predict the demand for emergency supplies at each disaster site, experts and scholars from various countries have proposed a variety of prediction models for earthquake casualties in recent years, so as to indirectly predict the demand for emergency supplies. Based on the gray prediction model, Ma et al. [24] proposed a prediction method for the demand for emergency supplies based on the gray Markov model in order to improve the accuracy of forecasting the number of people affected by disasters and the demand for emergency materials, and the relative error of the proposed model was verified by comparing with the gray Markov model and the gray model. Chen et al. [25] proposed a gray prediction model with fewer parameters, which combined the three factors in the earthquake with the mortality rate to form the GM (0,3) model, which is a great improvement compared with the traditional model. However, the GM (0,3) model is more stringent in data conditions, especially for data with little or unstable historical data, and its prediction effect may be seriously affected. Wang et al. [26] constructed a multi-objective optimization model for the dispatch of emergency materials in multiple periods for uncertain sudden disasters and designed and improved the NSGA-II algorithm to solve the problem according to the characteristics of the proposed model, based on which the real data of the 2008 Wenchuan earthquake were combined with the simulation data and the corresponding examples, and, compared with the other three algorithms, the proposed model and algorithm were verified to effectively solve the problem of large-scale post-disaster emergency material distribution. Instead of making a reasonable prediction of the corresponding material demand, the method of combining simulated data with real data may lead to a certain deviation between the distribution and the actual demand. Hu et al. [27] constructed a prediction model for the demand for earthquake emergency supplies in the case of unclear information in the disaster area, used the PSO-BP hybrid neural network algorithm to predict the number of earthquake casualties, and combined the results with the inventory management knowledge to indirectly complete the prediction of the demand for materials in the disaster area. However, the BP neural network is sensitive to initial parameters, prone to overfitting, and prone to fall into local optimum, while SVM has advantages in dealing with multi-dimensional, small-sample, and nonlinear problems. By constructing a Genetic Algorithm based on Principal Component Analysis to improve the Support Vector Machine model to predict the number of earthquake casualties [28] and establish a robust wavelet v-SVM prediction model for the number of earthquake casualties, the model obtained has fast learning speed, strong stability and high prediction accuracy [29]. At the same time, compared with the swarm intelligence optimization algorithm, the Bayesian Optimization algorithm has significant advantages in dealing with high-dimensional and nonlinear problems, which can find the global optimal solution faster and with higher accuracy. In 2012, Snoek J et al. [30] first introduced a method of using the Bayesian Optimization algorithm to automatically adjust the hyperparameters in machine

learning algorithms, which surpassed traditional manual tuning or brute force search methods and improved the performance of machine learning algorithms. Tang et al. [31] established a Bayesian network adaptability evaluation model and verified the applicability of the model in earthquake-prone counties in China.

The review of the aforementioned literature reveals that significant research outcomes have been achieved in casualty and emergency material demand forecasting across various contexts, which serve as references for this study. This work is based on the background of earthquake disasters. In past earthquake scenario studies, most scholars have utilized swarm intelligence optimization algorithms in combination with BP neural networks or Support Vector Machines (SVM) for casualty predictions. This study focuses on earthquake disasters and examines the prediction of casualties in past earthquake scenarios. Most researchers have employed swarm intelligence optimization algorithms in conjunction with BP neural networks or Support Vector Machines for casualty prediction. Existing methods include GA-SVM (Genetic Algorithm Support Vector Machine) and BO-SVM (Bayesian Optimization Support Vector Machine). Among these, GA-SVM has several disadvantages, including the risk of overfitting, high computational costs, and slow convergence rates. This model utilizes Genetic Algorithms for parameter optimization, which can lead to excessive complexity, particularly when training data are limited, thereby increasing the risk of overfitting. Additionally, the iterative process of the Genetic Algorithm requires substantial computation, resulting in prolonged training times, especially when dealing with large datasets. Due to the stochastic nature of Genetic Algorithms, the convergence of the model may necessitate a significant number of generations, which reduces efficiency.

Hence, based on previous studies, this paper proposes an earthquake casualty prediction model based on PCA-BO-SVM, and indirectly predicts the demand for emergency supplies based on the predicted number of casualties, which has the following advantages compared with the existing models:

(1) The Principal Component Analysis method introduced by the model has the ability to reduce dimensionality, and PCA can reduce the dimension of input features, reduce the complexity of the model, reduce the risk of overfitting, and improve the computational efficiency, which makes the model more efficient when processing high-dimensional data. Through Principal Component Analysis, the model retains the most important information, avoids feature redundancy, and helps improve prediction accuracy.

(2) Combined with Bayesian Optimization, the constructed PCA-BO-SVM model can be hyperparameter optimized in a lower dimension, reducing the probability of local optimum, and improving the global robustness of the model. The model can flexibly adapt to different feature combinations to improve the accuracy of earthquake casualty prediction.

(3) The proposed model enhances the performance of the Support Vector Machine in earthquake casualty prediction by combining PCA dimensionality reduction technology and Bayesian Optimization hyperparameter selection. PCA plays an important role in reducing dimensions, improving computational efficiency, and improving model interpretability, while Bayesian Optimization ensures the optimization of hyperparameter selection, thereby improving the accuracy and credibility of the prediction results, effectively overcoming the computational complexity and local optimal problems existing in GA-SVM and BO-SVM, and providing a more reliable and efficient solution for earthquake casualty prediction.

The organization of this study is structured as follows: Section 2 presents the key steps and processes involved in the overall research. Section 3 describes the principles used to construct the PCA-BO-SVM model. Section 4 is dedicated to the experimental part,

where the developed model is utilized for casualty prediction and result analysis. Section 5 focuses on estimating the demand for emergency materials. Finally, Section 6 provides the conclusions and future outlook of this study.

2. Research Design

Figure 1 shows the overall flow chart of the post-earthquake casualty and emergency material demand forecast based on PCA-BO-SVM. The key steps are as follows:

Step 1: Identify the types of emergency resources to be predicted in the aftermath of an earthquake. To alleviate the survival pressures faced by affected populations and to ensure their basic living needs are met, this study references the "Catalogue of Key Emergency Support Materials (2020)" and selects resources categorized under life rescue and livelihood assistance as the focus of the research.

Step 2: Develop the earthquake casualty prediction model based on PCA-BO-SVM. By utilizing historical data from 171 destructive earthquakes in China between 1966 and 2023, Principal Component Analysis is employed to reduce the dimensionality of the original data. Subsequently, the Bayesian Optimization algorithm is used to identify the optimal hyperparameters for the Support Vector Machine, thereby establishing the PCA-BO-SVM-based earthquake casualty prediction model.

Step 3: Determine the typical emergency resources required and their respective demand quantities. Considering the most urgent needs of disaster victims, the resources are categorized into consumable and non-consumable types. Consumable resources include drinking water, compressed biscuits, and medical supply kits, while non-consumable resources consist of tents and clothing. Finally, based on the developed model, this study predicts the number of casualties resulting from an earthquake and estimates the demand for typical emergency resources.

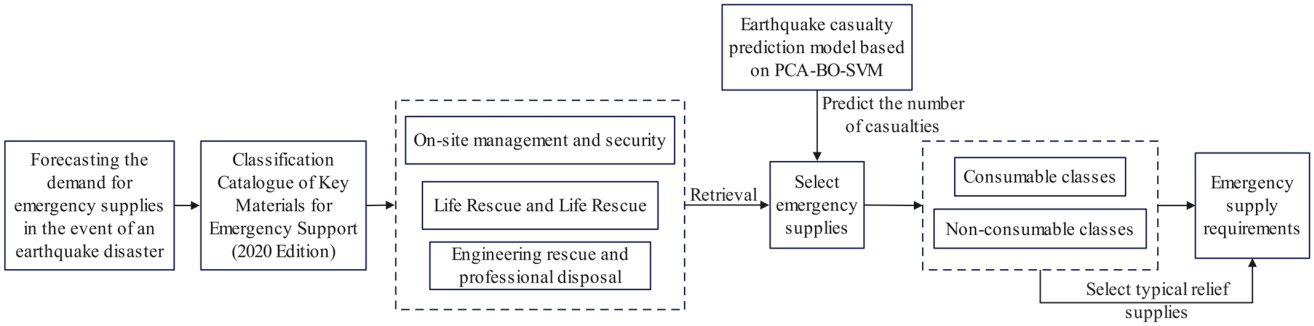


Figure 1. Process of forecasting the demand for emergency supplies during earthquake disasters.

3. Fundamental Principles

3.1. Principal Component Analysis

Principal Component Analysis (PCA) is a widely used multivariate statistical analysis method and a linear algebra technique based on eigenvalue decomposition. Its primary objective is to transform the original variables into a set of independent principal components through linear transformations, thereby reducing the dimensionality of the data while retaining critical information. This is achieved by preserving the first few principal components, which capture the essence of the data. PCA aids in identifying patterns and structures within the data, enhancing interpretability and explainability, making it a common technique in data analysis and data mining. The core concept of PCA is to identify the direction of maximum variance among the data variables and to project the original data in these directions, allowing the new principal components to explain the variability of the original data to the greatest extent. This process facilitates a better understanding

and interpretation of the patterns and relationships present in the data. PCA determines the weight coefficients of the principal components by calculating eigenvalues and eigenvectors, thus achieving the goals of dimensionality reduction and feature extraction. The specific methodology is as follows:

(1) Assume that X consists of m samples $(X_1, X_2, X_3, X_4, \ldots, X_m)$, with each sample being n-dimensional, which corresponds to the following matrix:

$$X_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{mn} \end{bmatrix} \quad (1)$$

In Equation (1), X_{mn} denotes the n feature for the m sample.

(2) Feature Parameter Standardization: To eliminate the dimensional and variance differences among different feature parameters and ensure a relatively fair influence of each feature parameter on the results of Principal Component Analysis, standardization is employed. This process enables the data distribution of each feature parameter to have similar scales and variances, thus preventing any feature parameter with a large numerical range or high variance from disproportionately affecting the computation of principal components. Consequently, this approach guarantees the accuracy and reliability of the principal components. The standardized data are more conducive to performing Principal Component Analysis, resulting in outcomes that are more interpretable and stable.

$$var(x_j) = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2,\ (i = 1, 2, \ldots,\ n;\ \ j = 1, 2, \ldots,\ n) \quad (2)$$

$$\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad (3)$$

$$x_{ij}^* = \frac{x_{ij} - \overline{x}_{ij}}{\sqrt{var(x_j)}} \quad (4)$$

In Equation (4), x_{ij}^* represents the standardized feature indicator; $\overline{x}_j$ denotes the mean of feature indicator j; $var(x_j)$ indicates the variance of feature indicator j.

(3) Calculate the correlation coefficient matrix $R = (r_{ij})_{n\times n}$:

$$r_{ij} = \frac{\sum_{k=1}^{n} x_{ki}^* \cdot x_{kj}^*}{n-1}\ (i, j = 1, 2, 3, \ldots, n) \quad (5)$$

According to Equation (5), the correlation coefficient matrix can be calculated as follows:

$$R_{n\times n} = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ r_{21} & r_{22} & \ldots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \ldots & r_{nn} \end{bmatrix} \quad (6)$$

In Equation (6), r_{ij} $(i, j = 1,\ 2,\ \ldots,\ n)$ represents the correlation coefficients of the standardized feature indicators, $r_{ii} = 1$ and $r_{ij} = r_{ji}$.

(4) Calculate the eigenvalues and contribution rates to determine the principal components.

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3, \ldots, \lambda_n \geq 0$ of the correlation coefficient matrix are computed, with their corresponding eigenvectors denoted as $\mu_1, \mu_2, \ldots, \mu_n$, where $\mu_j = \left(\mu_{1j}, \mu_{2j}, \ldots, \mu_{mj}\right)^T$, resulting in n new indicator variables formed from the eigenvectors:

$$\begin{cases} y_1 = \mu_{11}\widetilde{x}_1 + \mu_{21}\widetilde{x}_2 + \ldots + \mu_{m1}\widetilde{x}_m \\ y_2 = \mu_{12}\widetilde{x}_1 + \mu_{22}\widetilde{x}_2 + \ldots + \mu_{m2}\widetilde{x}_m \\ \cdots\cdots\cdots\cdots \\ y_n = \mu_{1n}\widetilde{x}_1 + \mu_{2n}\widetilde{x}_2 + \ldots + \mu_{mn}\widetilde{x}_m \end{cases} \tag{7}$$

In Equation (7), y_n $(n = 1, 2, \ldots, n)$ refers to the n principal component, and the contribution rate and cumulative contribution rate of the eigenvalue λ_k $(k = 1, 2, \ldots, n)$ can be calculated as described in Equations (8) and (9):

$$\frac{\lambda_i}{\sum_{k=1}^{n} \lambda_k} \ (i = 1,\ 2, \ldots, n) \tag{8}$$

$$\frac{\sum_{k=1}^{q} \lambda_k}{\sum_{k=1}^{n} \lambda_k} \tag{9}$$

In summary, Principal Component Analysis (PCA) is a statistical technique employed for dimensionality reduction. It achieves this by projecting the original data into a new coordinate system through linear transformations, where the first principal component accounts for the maximum variance, followed by subsequent components in decreasing order of variance.

The integration of PCA into Support Vector Machines (SVM) enhances performance for several compelling reasons:

(1) Dimensionality Reduction and Information Retention: PCA effectively compresses redundant features within high-dimensional datasets, thereby reducing dimensionality while preserving the most critical information. This process not only alleviates computational burdens but also lowers model complexity, consequently reducing the risk of overfitting.

(2) Improved Computational Efficiency: The training process of SVM can become exceedingly time-consuming and complex in high-dimensional spaces. By employing PCA for dimensionality reduction, the number of features input into the SVM is significantly decreased, resulting in reduced training times and enhanced computational efficiency.

(3) Enhanced Model Interpretability: Following dimensionality reduction, the principal components provided by PCA often offer clearer insights into the underlying data structure. This clarity aids researchers in better understanding the key factors influencing earthquake casualties, thereby supporting informed decision-making.

Overall, the application of PCA not only optimizes the performance of SVM but also contributes to a more efficient and interpretable modeling process in the context of earthquake casualty prediction.

3.2. Bayesian Optimization Algorithm

Bayesian Optimization (BO) is a powerful method for optimizing black-box functions, based on Bayes' theorem and Gaussian processes. It demonstrates significant advantages when addressing high-dimensional and nonlinear problems. Compared to other algorithms, Bayesian Optimization seeks the global optimum by constructing a probabilistic model that represents the objective function, with the goal of minimizing or maximizing an unknown objective function. The core of Bayesian Optimization lies in establishing a probabilistic

model that represents the objective function and iteratively updating this model to identify the global optimum. In each iteration, the algorithm selects a sample point for evaluation and, based on the results, updates the probabilistic model. This process is repeated until a predefined termination condition is met.

The Bayesian Optimization algorithm exhibits high sampling efficiency and robustness, making it widely applicable in various fields, including hyperparameter optimization, automated parameter tuning, and the training of machine learning models. For hyperparameter optimization using Bayesian methods, the optimizer iteratively conducts a limited number of experiments within the hyperparameter search space to optimize f and obtain the optimal combination of hyperparameters:

$$x^* = \underset{x \epsilon \chi}{arg}\ min f(x) \tag{10}$$

Bayesian Optimization employs the principles of Bayesian statistics to guide the process of hyperparameter selection by constructing a probabilistic model of the objective function, typically represented as a Gaussian process. This process encompasses several key steps:

(1) Building the Surrogate Model: In the initial phase, Bayesian Optimization randomly selects several hyperparameter combinations and evaluates the corresponding model performance metrics (e.g., prediction accuracy). The results of these evaluations are then utilized to construct a surrogate model that approximates the objective function within the hyperparameter space.
(2) Utilization of Acquisition Functions: Based on the existing data, Bayesian Optimization employs acquisition functions (such as expected improvement or upper confidence bound) to determine the next hyperparameter combination to evaluate. These acquisition functions facilitate a balance between exploration (testing new hyperparameter combinations) and exploitation (optimizing known high-performing hyperparameter combinations).
(3) Iterative Optimization: Through the continuous iteration of these steps, Bayesian Optimization progressively narrows the hyperparameter search space and ultimately identifies the optimal hyperparameter combination.

The choice of hyperparameters has profound implications for model performance, primarily reflected in several aspects:

(1) Model Complexity: Hyperparameters, such as the regularization parameter C in SVM, play a crucial role in adjusting model complexity. The selection of C influences the model's susceptibility to overfitting or underfitting. A high C value may lead to exceptional fit on training data but result in poor generalization to testing data, thereby increasing the risk of overfitting. Conversely, a low C value may prevent the model from capturing complex patterns in the data, manifesting as underfitting. Thus, careful selection of C is essential, and Bayesian Optimization provides a precise mechanism for tuning this parameter.
(2) Kernel Function Selection and Parameters: In SVM, the choice of the kernel function and its parameters (e.g., σ in Gaussian kernels) significantly affect model performance. Different kernel functions and their associated parameters determine the flexibility of the decision boundary. Selecting inappropriate kernel parameters may result in a loss of accuracy when modeling complex data structures. Bayesian Optimization can effectively explore kernel parameters, facilitating the model's adaptation to underlying data patterns.
(3) Model Generalization Capability: The selection of hyperparameters directly influences the model's ability to generalize, which refers to its performance on unseen data. Fine-

tuning hyperparameters through Bayesian Optimization can enhance the model's adaptability to new data, thereby reducing its susceptibility to overfitting on a specific training set and improving predictive accuracy.

In the context of earthquake casualty prediction, Bayesian Optimization serves as an efficient method for hyperparameter selection, significantly enhancing the performance of models such as Support Vector Machines (SVM). The selection of hyperparameters, which are predetermined prior to model training, is critical for the overall effectiveness of the model.

3.3. Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm based on statistical learning theory, which demonstrates significant advantages in handling nonlinear, small-sample, and high-dimensional data [8]. The core idea is to transform the input vectors of the training sample set through a nonlinear mapping into feature vectors in a high-dimensional space, thereby converting the nonlinear classification problem into a linear one in that space. By applying the nonlinear mapping $\phi(X_i)$, the input data X_i is projected into a high-dimensional space, which can be expressed mathematically as follows: $y_i = W \cdot \phi(X_i) + b$. By optimizing to solve for the weight vector and threshold, the Support Vector Machine obtains the optimal decision hyperplane, allowing for effective classification of the samples. The weight vector w and the threshold b can be derived from the following function to obtain the optimal solution:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \tag{11}$$

$$s.t. \begin{cases} y_i - w^T \varphi(X_i) + b \le \varepsilon + \xi_i^* \\ w^T \varphi(X_i) + b - y_i \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0 \qquad (i = 1, 2, \ldots, m) \end{cases} \tag{12}$$

In Equation (11), C represents the penalty parameter, which plays a crucial role as an important tuning parameter in the Support Vector Machine model. It controls the complexity of the model, prevents overfitting, regulates the margin size, and addresses class imbalance. By constructing the Lagrangian function, the complex quadratic programming problem described above can be transformed into its dual problem. The solutions are obtained by solving for the original variables and the Lagrange multipliers, utilizing the Karush–Kuhn–Tucker (KKT) conditions to achieve the optimal solution. This transformation simplifies the solving process and enhances computational efficiency. The dual problem can be expressed as follows:

$$\begin{aligned} &L(w, b, \xi, \xi^*, a, a^*, r, r^*) \\ &= \tfrac{1}{2} w^T w + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) - \sum_{i=1}^{m} a_i^* (y_i - w^T \cdot \varphi(X_i) - b - \varepsilon - \xi_i^*) \\ &- \sum_{i=1}^{m} a_i (W^T \cdot \varphi(X_i) + b - y_i - \varepsilon - \xi_i) - \sum_{i=1}^{m} r_i - \sum_{i=1}^{m} r_i^* \end{aligned} \tag{13}$$

In Equation (13), the kernel function of the Support Vector Machine is denoted as $K(X_i \cdot X_j)$, while a_i^* and a_i represent the Lagrange multipliers. The optimal solution can be obtained as follows:

$$b = y_i - \sum_{j=1}^{l} \left(\hat{a}_j^* - \hat{a}_j \right) K(X_i \cdot X_j) + \varepsilon \text{ or } b = y_i - \sum_{j=1}^{l} \left(\hat{a}_j^* - \hat{a}_j \right) K(X_i \cdot X_j) - \varepsilon \tag{14}$$

The radial basis function (RBF) kernel of the SVM can be expressed as follows:

$$K_g = (|X - X_i|) = exp\left(-g|X - X_i|^2\right) \tag{15}$$

After calculations, the final regression function of the SVM is given by the following:

$$f(X) = \sum_{i=1}^{m} (a_i^* - a_i)\varphi(X_i)^T \varphi(X_i) + b \tag{16}$$

In Equation (16), $\varphi(X_i)^T$ and $\varphi(X_i)$ represent the kernel function; a_i^* and a_i denote the Laplace operators; and the bias term is represented by b.

4. Earthquake Casualty Prediction Model

Figure 2 illustrates the modeling process for predicting earthquake casualties based on PCA-BO-SVM. The relevant steps are as follows:

Step 1: Utilize Principal Component Analysis (PCA) to perform dimensionality reduction, correlation analysis, and normalization on the collected raw data, and to determine the principal components.

Step 2: Select the fitness function. The quality of the parameters is assessed using the Mean Absolute Percentage Error (MAPE) as the criterion for the selection of the fitness function.

Step 3: Employ the Bayesian Optimization (BO) algorithm to optimize the hyperparameters of the SVM. Set up the Bayesian optimizer to search for the optimal SVM hyperparameters, and optimize the BO-SVM model using the normalized training samples.

Step 4: Output the best hyperparameters, train the SVM model, and construct the PCA-BO-SVM prediction model for earthquake casualties. Subsequently, perform predictions on the test samples and analyze the results.

Step 5: Analyze and compare the running efficiency and prediction accuracy of different models to draw conclusions.

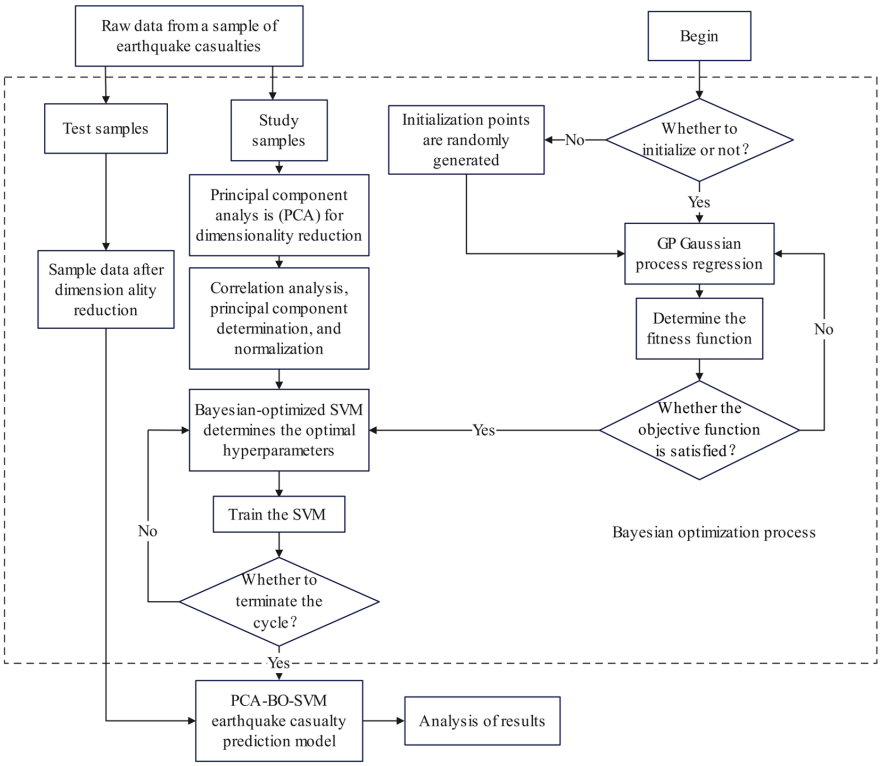


Figure 2. PCA-BO-SVM flow chart.

4.1. Data Preparation

In this paper, 171 destructive earthquakes of magnitude 5 or higher that occurred in China from 1966 to 2023 were selected as sample data; the first 163 earthquakes were used as learning samples and the last 8 earthquakes were used as test samples. These data come from the "China Earthquake Cases" published by the Seismological Press and the following websites: the National Seismological Data Center of China (https://data.earthquake.cn), the China Earthquake Network (https://news.ceic.ac.cn), the Ministry of Emergency Management of the People's Republic of China (https://www.mem.gov.cn), and the Chinese Statistical Database (https://www.hongheiku.com). Detailed information is provided in Table 1. Based on the analysis of factors influencing earthquake casualty predictions discussed in [25], this study focuses on seven influencing factors for destructive earthquake disasters: time, earthquake intensity, occurrence time, magnitude, population density in the disaster area, seismic resistance level of buildings, and earthquake forecasting capability. For the time variable, values 1 and 2 represent night and day, respectively, where night refers to earthquakes occurring between 19:00 and 6:00, during which escape is hindered, while day refers to earthquakes occurring between 6:00 and 19:00. Earthquake intensity is expressed using Roman numerals IV to XI, with larger numbers indicating greater intensity and destructiveness in the affected areas; the minimum earthquake intensity in the selected samples is IV (4). Regarding magnitude, larger numbers correspond to higher magnitudes; this study considers earthquakes with magnitudes ranging from 4.0 to 8.1 that caused destruction. The population density in the disaster area refers to the ratio of the number of people affected in the area after a natural disaster, expressed in "people/area". Population density is an important indicator that reflects the degree of population concentration and density in the disaster area. In similar circumstances, high population density often implies a greater number of people affected, which is significant for disaster management and rescue efforts. For seismic resistance levels, higher numbers indicate greater seismic resistance of buildings. In this study, building seismic resistance is categorized into nine levels, represented by the numbers 1 to 9. The forecasting capability is indicated by values 0, 1, and 2, where 0 denotes no warning, 1 indicates a warning but with low accuracy, and 2 represents a warning with high accuracy.

Table 1. Raw data.

NO.	Location	Time	Intensity	Magnitude	Population Density (People/km^2)	Seismic Fortification Level	Forecast Level	Hypocenter Depth (km)	Casualties
1	He Jian	2	VII	6.3	680.0	7	0	30	935
2	Yang Jiang	1	VIII	6.4	329.1	7	0	5	1033
3	Li Yang	2	VI	5.5	524.7	7	1	18	222
4	Yan Yuan	1	VIII	6.7	40.5	7	2	19	495
5	He lin Ge Er	1	VII	6.3	58.5	7	0	18	907
6	Song Pan	1	IX	7.2	7.8	8	2	15	797
7	Jie Xiu	2	VII	5.2	580.6	8	2	9	181
8	Wu Yuan	1	VII	6.0	87.8	7	0	30	108
9	He Ze	1	VII	5.9	705.5	7	2	12	5184
10	Xun Wu	2	VII	5.5	119.1	7	1	13	84
11	Da Tong	1	VIII	6.1	218.7	7	1	14	160
12	Xiao Jin	1	VIII	6.6	11.6	7	1	15	152
13	Chong Ming	2	VI	5.0	506.9	8	1	10	73

Table 1. *Cont.*

NO.	Location	Time	Intensity	Magnitude	Population Density (People/km^2)	Seismic Fortification Level	Forecast Level	Hypocenter Depth (km)	Casualties
14	Yao An	1	VII	5.6	88.2	7	2	33	152
15	Li Jiang	2	VII	5.7	60.6	8	0	15	142
16	Ning Lang	1	VIII	6.2	46.4	7	1	10	1980
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…
151	Yong Sheng	2	VII	6.0	68.0	8	1	10	221
152	Zhao Su	2	VIII	6.1	14.0	8	1	18	54
153	Ning Er	1	VIII	6.4	52.0	7	2	5	422
154	Pan Zhi Hua	2	VIII	6.1	163.4	7	1	10	1051
155	Yao An	1	VIII	6.0	88.2	7	2	10	373
156	Xiang Ge Li La	2	VIII	5.9	16.0	7	1	10	41
157	Pi Shan	2	VIII	6.5	8.1	7	1	10	266
158	Jiu Zhai Gou	1	VIII	7.0	12.6	8	1	20	550
159	Wei Yuan	2	VI	5.4	529.9	6	1	10	64
160	Lu Xian	1	VIII	6.0	501.2	6	2	10	160
161	Lu Shan	2	VIII	6.1	99.9	7	2	17	47
162	Shi Mian	2	VI	4.2	42.6	8	1	15	106
163	Ji Shi Shan	1	VIII	6.2	262.5	7	2	10	1134
164	Yan Jin	2	VI	5.1	151.0	7	1	9	136
165	Yan Jin	2	IV	5.1	151.0	7	1	22	69
166	Ying Jiang	1	VIII	5.9	59.0	7	2	10	173
167	Yong Sheng	1	VI	5.4	68.0	8	1	15	113
168	Shu Fu	1	VII	6.2	103.9	8	2	39	139
169	Wu Qia	1	VIII	6.8	2.6	9	1	27	214
170	Dang Xiong	2	V	6.6	3.8	9	0	8	103
171	Mei Shan	2	VII	8.0	414.7	7	0	14	325

4.2. Principal Component Analysis

This study employed Principal Component Analysis (PCA) to conduct a correlation analysis and dimensionality reduction on seven influencing factors: time, earthquake intensity, occurrence time, magnitude, population density in disaster areas, seismic resistance of buildings, earthquake prediction capability, and focal depth. The results of this processing are illustrated in Figure 3, where the correlation coefficient heatmap uses color intensity to represent the magnitude of the correlation coefficients. Darker colors indicate stronger correlations, while lighter colors indicate weaker correlations. This visualization allows for an intuitive assessment of which variables exhibit strong correlations, thereby facilitating further analysis of the relationships and influences between the variables.

Figure 3 shows the correlation between the features, and it can be seen that there is a significant positive correlation between seismic intensity and magnitude. There is a significant negative correlation between the seismic level of buildings and population density, the predicted level, and the earthquake magnitude. In the context of earthquake casualty prediction, the correlation of various features reveals the impact on casualty prediction and provides a data basis for follow-up research. Figure 3 can quickly identify which features have high correlation and which features have weak relationships with each other, and strong correlation features may bring better prediction performance in model training, which is crucial for subsequent model training and prediction results.

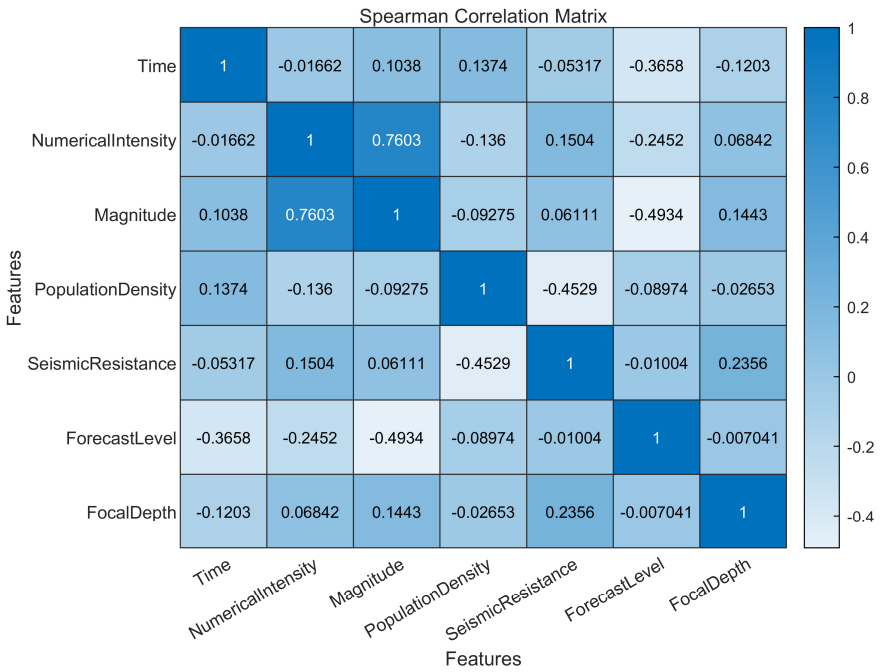


Figure 3. Correlation coefficient heatmap.

The contribution rates of each principal component, normalized eigenvalues, and cumulative contribution rates are presented in Table 2, with the contribution rates and cumulative contribution rates depicted in Figures 4 and 5, respectively.

Table 2. Normalized eigenvalues, contribution rates, and cumulative contribution rates.

Component	Normalize Eigenvalues	Contribution Rate/%	Cumulative Contribution Rate/%
Y_1	2.3798	34.00	34.00
Y_2	1.5739	22.48	56.48
Y_3	0.9625	13.75	70.23
Y_4	0.8798	12.57	82.80
Y_5	0.5584	7.98	90.78
Y_6	0.4616	6.59	97.37
Y_7	0.1841	2.63	100.00

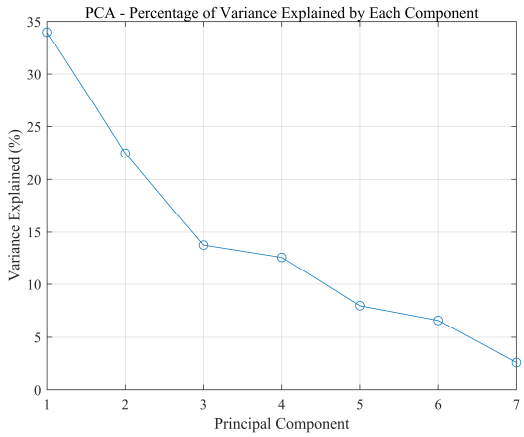


Figure 4. Contribution rate of each component.

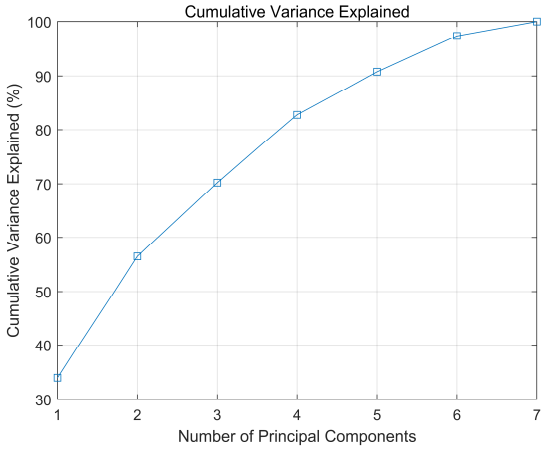


Figure 5. Cumulative contribution rate.

To capture the majority of the information contained in the original data, it is common practice to select principal components with a cumulative contribution rate of over 80%. Therefore, based on Table 2, the first four principal components were chosen, resulting in a dimensionality reduction from seven-dimensional data to four-dimensional data. Using X_1, X_2, X_3, X_4, X_5, X_6, X_7 to represent the seven original variables, the expressions for the four principal components are provided in Equation (17). The four principal components were then normalized using Equation (18), which served as the input variables for constructing the SVM model. This approach not only reduces computational complexity and enhances the model's generalization ability to avoid overfitting but also improves operational efficiency. The normalization results are presented in Table 3.

Table 3. Principal component normalization data.

No.	Y_1	Y_2	Y_3	Y_4	No.	Y_1	Y_2	Y_3	Y_4
1	0.174	−0.481	3.225	0.941	153	−0.263	−0.464	−1.603	−0.213
2	0.337	−1.197	−0.387	−1.067	154	−0.099	−0.460	−0.507	−0.325
3	0.159	−1.253	0.413	0.246	155	−0.444	−0.249	−0.981	0.164
4	−0.017	0.518	−0.116	0.773	156	−3.493	0.235	−5.786	8.287
5	−0.222	0.240	0.674	−0.601	157	−0.001	−0.035	−0.816	−0.495
6	0.741	1.316	−0.831	0.475	158	0.428	1.519	0.254	0.093
7	−1.005	−0.795	0.103	0.326	159	−1.284	−2.428	0.653	−0.094
8	−0.354	0.960	2.050	0.255	160	−0.312	−2.122	−0.137	0.768
9	−0.312	−1.806	0.611	0.921	161	−0.361	0.202	−0.196	0.670
10	−0.925	−0.264	−0.054	−0.341	162	−2.037	0.740	0.216	−0.621
11	−0.099	−0.312	0.050	0.006	163	−0.196	−0.674	−0.644	0.378
12	0.078	0.299	−0.268	−0.136	164	−1.580	−0.725	−0.223	−0.743
13	−1.374	−0.668	0.595	−0.424	165	−2.364	−0.028	1.575	−0.111
14	−0.978	1.172	1.716	1.626	166	−0.527	−0.181	−1.035	0.125
15	−0.546	0.871	0.364	−0.989	167	−1.349	0.768	0.246	−0.547
16	−0.050	−0.183	−0.756	−0.438	168	−0.527	2.446	2.377	1.944
…	…	…	…	…	169	0.404	2.868	0.991	0.418
151	−0.607	0.563	−0.502	−0.734	170	−1.455	1.234	−0.128	−1.978
152	−0.117	1.324	0.063	−0.087	171	0.864	−0.759	0.935	−0.429

In Equation (17), X_1 represents time, X_2 indicates earthquake intensity, X_3 denotes magnitude, X_4 signifies population density, X_5 represents seismic resistance, X_6 indicates prediction capability, and X_7 denotes focal depth.

$$\begin{aligned} Y_1 &= 0.5707X_1 + 0.5580X_2 + 0.1223X_3 + 0.0564X_4 - 0.2905X_5 + 0.0363X_6 + 0.5090X_7 \\ Y_2 &= 0.1138X_1 + 0.0746X_2 - 0.5700X_3 + 0.6681X_4 + 0.0306X_5 + 0.4287X_6 - 0.1595X_7 \\ Y_3 &= -0.2591X_1 - 0.0177X_2 + 0.4690X_3 - 0.0075X_4 - 0.4616X_5 + 0.6973X_6 - 0.1152X_7 \\ Y_4 &= 0.1895X_1 + 0.0372X_2 + 0.2609X_3 - 0.1116X_4 + 0.8125X_5 + 0.4536X_6 + 0.1277X_7 \end{aligned} \tag{17}$$

In Equation (18), the denominator X_{max} represents the maximum value of each principal component, while X_{min} denotes the minimum value of each principal component. The numerator X_i corresponds to the value of the i sample for each principal component, and the normalized value for each principal component is represented by X_i^*.

$$X_i^* = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{18}$$

4.3. Model Building

To obtain the optimal hyperparameters for the Support Vector Machine (SVM), this study implemented a program based on the Bayesian Optimization Support Vector Machine using MATLAB R2020b. The number of earthquake casualties was designated as the output variable, while the four principal components listed in Table 3 served as the input variables. After extensive training, the optimal fitness curve was achieved, as illustrated in Figure 6. The minimum objective value and the number of function evaluations are shown in Figure 7. Additionally, the Bayesian Optimization process took 17.42 s to identify the best hyperparameters, demonstrating a high level of operational efficiency.

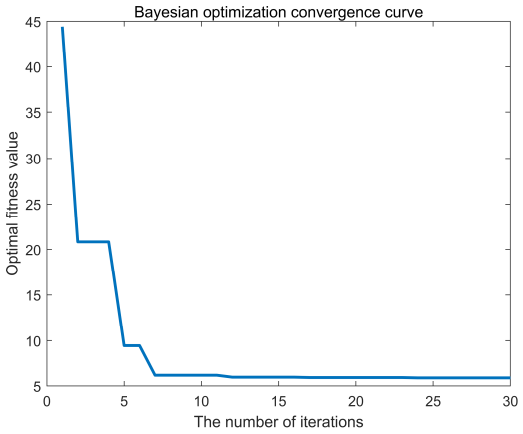


Figure 6. Bayesian Optimization.

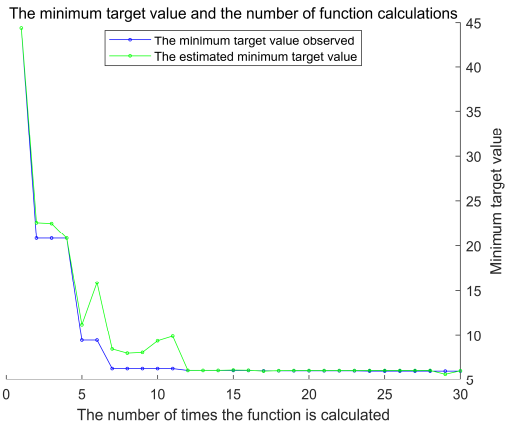


Figure 7. The minimum target value and the number of function calculations.

4.4. Prediction Results and Analysis

Using the constructed PCA-BO-SVM model, we performed predictions and validations on the test samples, and the results are presented in Figure 8. As shown in the results depicted in Figure 8, by comparing with historical data, it can be seen that there is not much difference between the predicted casualty and the actual casualty number, indicating that the PCA-BO-SVM model can effectively capture the key factors that affect the casualty number when processing the data. The average accuracy of the prediction results is 94.25%, which shows the accuracy of the prediction of the proposed model, and it can be seen that the prediction model based on Principal Component Analysis and Bayesian Optimization Support Vector Machine (PCA-BO-SVM) has a high consistency with the actual value.

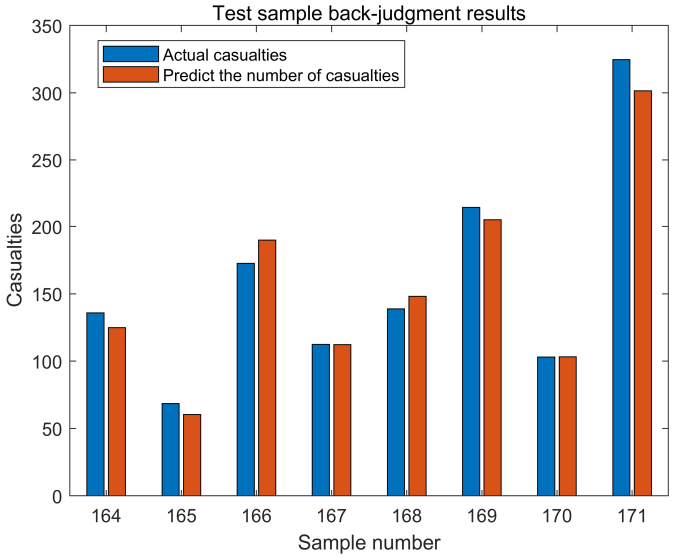


Figure 8. Results of test sample reassessment.

To further evaluate the predictive performance of the models, we employed the Genetic Algorithm Optimized Support Vector Machine (GA-SVM) model, the Bayesian Optimized Support Vector Machine (BO-SVM) model, and the Principal Component Analysis and Genetic Algorithm Optimized Support Vector Machine (PCA-GA-SVM) model to make predictions using the test samples provided in this study. This analysis allows for a comparison of the predictive performance across the different models. The prediction results for the four models are presented in Table 4, which includes the runtime for each model, predicted values, accuracy, relative error, maximum error, minimum error, average error, and average accuracy.

As shown in Table 4, the PCA-BO-SVM model constructed in this study has a runtime of 17.42 s, making it the fastest among the models evaluated, while also achieving the highest accuracy and the lowest average error. Compared to the GA-SVM, BO-SVM, and PCA-GA-SVM models, the PCA-BO-SVM model exhibits a reduction in average error of 12.86%, 9.01%, and 2%, respectively, while improving average accuracy and operational efficiency by 10.1%, 7.05%, and 0.325% and 25.5%, 18.4%, and 19.2%, respectively. These results indicate that the PCA-BO-SVM model possesses high operational efficiency, average accuracy, and strong generalization capability, providing a scientifically sound basis for predicting casualties from earthquake disasters and establishing a foundation for demand forecasting.

Table 4. Casualty prediction results from different models.

NO.	Casualties	GA-SVM Running Time T_1 = 23.37 s			BO-SVM Running Time T_2 = 21.36 s			PCA-GA-SVM Running Time T_3 = 21.55 s			PCA-B0-SVM Running Time T_4 = 17.42 s		
		Predicted Value	Accuracy/%	Relative Error/%	Predicted Value	Accuracy/%	Relative Error/%	Predicted Value	Accuracy/%	Relative Error/%	Predicted Value	Accuracy/%	Relative Error/%
164	136	146	93.2	7.2	133	97.8	2.4	126	92.6	7.1	125	91.9	7.9
165	69	134	51.5	93.9	126	54.8	82.5	69	100	0.6	61	88.4	11.7
166	173	178	97.2	2.8	201	86.1	16.3	186	93.0	7.8	190	91.1	9.9
167	113	82	72.6	27.8	115	98.3	1.7	118	95.8	4.0	113	100	0.2
168	139	135	97.1	3.0	125	89.9	10.4	152	91.4	9.5	148	93.9	6.6
169	214	221	96.8	3.3	213	99.5	0.6	205	95.8	4.2	205	95.8	4.2
170	103	151	68.2	46.4	141	73.0	37.1	109	94.5	5.9	103	100	0.2
171	325	314	96.6	3.4	319	98.2	1.8	287	88.3	11.6	302	92.9	7.2
Maximum Error/%			64.76%			56.91%			37.75%			23.48%	
Minimal Error/%			4.18%			1.30%			0.38%			0.19%	
Average Error/%			22.62%			18.77%			11.76%			9.76%	
Average Accuracy			84.15%			87.20%			93.925%			94.25%	

The relative errors of the prediction results from the four models are illustrated in Figure 9. According to the relative errors of the predictions from various models presented in Figure 9, the PCA-BO-SVM model demonstrates the best performance. Without conducting Principal Component Analysis, both the GA-SVM and BO-SVM models exhibit relatively large errors, while the results of the BO-SVM model outperform those of the GA-SVM model. This indicates that the Bayesian Optimization Support Vector Machine (BO-SVM) is more effective than the Genetic Algorithm (GA) in this context. It is evident that the BO-SVM has superior convergence and robustness when handling high-dimensional data and complex optimization problems, allowing it to find the global optimum more rapidly. Moreover, it performs exceptionally well in managing noisy data and non-convex optimization issues, making it well suited for practical applications. Figure 9 illustrates the significant improvement in performance resulting from dimensionality reduction through principal component analysis, leading to better data analysis.

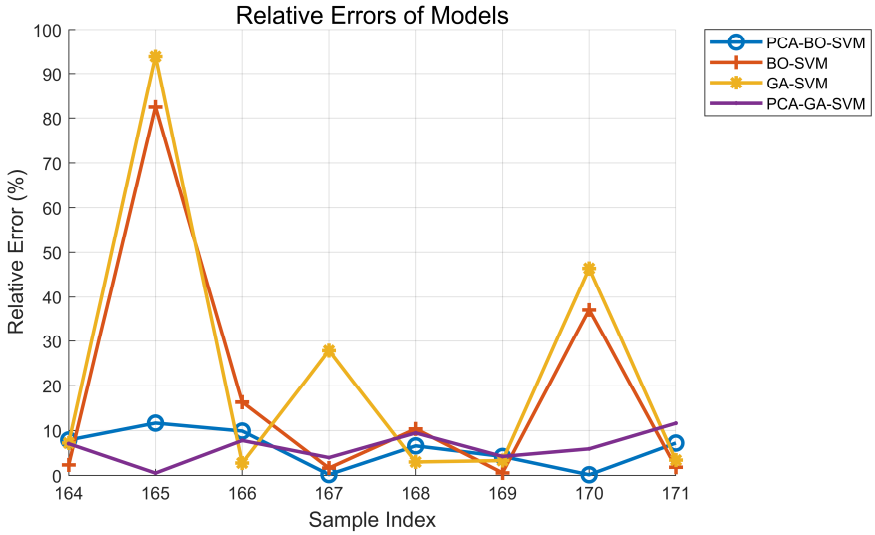


Figure 9. Relative errors of prediction results from four models.

4.5. Discussion

In general, the PCA-BO-SVM model shows higher prediction accuracy and operational efficiency in earthquake casualty prediction, and the implementation of the model can effectively improve the foresight of disaster management, marking the importance of data-driven decision-making in the field of public safety and disaster management, and also providing a certain scientific basis for government decision-makers in emergency

response, material allocation, and post-disaster recovery. However, disaster preparedness depends not only on understanding natural phenomena and data collection, but also on a comprehensive assessment of social risks. The applicability of the model has yet to be verified, and its accuracy and effectiveness are limited by data quality and model complexity. At the same time, the prediction of emergency supplies should also consider the influence of other factors in future research, so as to enrich the types of emergency supplies and items to meet the needs of disaster victims more scientifically and reasonably.

In this paper, the framework and methodology of the proposed model may also have the potential to adapt to other types of natural disasters, such as floods, typhoons, or fires. However, the specific verification also needs to be carried out through the corresponding PCA processing and Bayesian Optimization of the characteristic data of different disaster types, and the model is redesigned to predict the number of casualties or other impact indicators under different scenarios. The accuracy of prediction accuracy of the model is not only achieved through PCA dimensionality reduction and Bayesian Optimization but also depends on the quality of the data, which brings challenges to different research backgrounds and adaptation ranges. In order to solve this problem, in the future, we can explore the hybrid model method, consider more input features into the model, and use improved algorithms to integrate the model to enhance the robustness and generalization ability of the model. We can overcome bottlenecks through method innovation, data integration, and interdisciplinary collaboration, and further improve the accuracy and adaptability of the model so that it can more comprehensively handle various nonlinear relationships and adapt to different disaster scenarios.

5. Analysis of Emergency Material Demand for Earthquakes

In the initial phases of an earthquake disaster, the most pressing needs of affected populations primarily include consumable supplies such as drinking water, compressed biscuits, and bread. Non-consumable supplies are predominantly focused on items for warmth, such as tents and clothing. Additionally, medical supply kits are essential. This study selects drinking water, compressed biscuits, medical supply kits, tents, and clothing as representative emergency supplies from the aforementioned categories.

When calculating the demand for these typical emergency supplies, this study incorporates regional coefficients and seasonal coefficients [32], drawing on insights from the existing literature and research. The calculation formula is as follows:

$$D^k = \begin{cases} v^k \times R \times W^k \times Q^k \times m, & k \in Consumable\ supplies \\ v^k \times R \times W^k \times Q^k, & k \in\ Non-Consumable\ supplies \end{cases} \quad (19)$$

In Equation (19), D^k represents the total demand for supplies,v^k indicates the per capita demand for the k type of supply, R denotes the predicted number of casualties in the earthquake-affected area, W^k signifies the seasonal variation coefficient, Q^k represents the regional variation coefficient, and m indicates the number of days for which emergency supplies are needed. The following table (Table 5) presents typical standards for per capita demand for emergency supplies, including a medical supply pack that contains anti-inflammatory medications, iodine tincture, glucose, gauze, and other items.

Table 5. Typical per capita demand for emergency supplies.

Type of Material	Drinking Water (L)	Compressed Biscuits (kg)	Medical Kits (Sets)	Tents (Sets)	Clothes (Pieces)
Demand per capita	2	0.5	1	0.5	3

Estimation of Emergency Supply Demand

Using the PCA-BO-SVM model developed in this study to predict the number of casualties due to earthquakes, we conducted a demand forecast for emergency supplies, taking the prediction result for 3.4 (as shown in Table 6) as an example. In the demand forecasting formula (Equation (19)), we set the regional variation coefficient Q^k to 0.9, the seasonal variation coefficient for consumable materials W^k to 1.2, and the seasonal variation coefficient for non-consumable materials W^k to 1.5, with the required supply of emergency materials m set to 15 days. The results for the required quantities of emergency supplies are presented in Table 7.

Table 6. Comparison of actual and predicted casualties in the disaster-affected area.

No.	Disaster Site	Actual Casualties	Predict the Number of Casualties
164	Yan Jin	136	125
165	Yan Jin	69	61
166	Ying Jiang	173	190
167	Yong Sheng	113	113
168	Shu Fu	139	148
169	Wu Qia	214	205
170	Dang Xiong	103	103
171	Mei Shan	325	302

Table 7. Predicted demand for supplies at various disaster sites.

	Drinking Water (L)	Compressed Biscuits (kg)	Medical Kits (Sets)	Tents (Sets)	Clothes (Pieces)
Yan Jin	4050	1013	2025	85	507
Yan Jin	1977	495	989	42	248
Ying Jiang	6156	1539	3078	129	770
Yong Sheng	3662	916	1831	77	458
Shu Fu	4796	1199	2398	100	600
Wu Qia	6642	1661	3321	139	831
Dang Xiong	3338	835	1669	70	418
Mei Shan	9785	2447	4893	204	1224

6. Conclusions

In the aftermath of an earthquake, it is crucial to scientifically and rationally predict the number of casualties to effectively implement emergency rescue measures and estimate the demand for emergency supplies. This predictive capability facilitates a more efficient response to earthquake disasters and is vital for guiding the allocation of relief resources. Moreover, it plays a key role in minimizing both human casualties and property losses. Therefore, this study investigates the prediction of casualties resulting from seismic events and the demand for emergency resources, leading to the following conclusions:

(1) Principal Component Analysis (PCA) was employed to reduce the dimensionality of the raw data from 171 destructive earthquakes in China, which included seven influential factors: time, earthquake intensity, occurrence timing, magnitude, population density in disaster areas, building seismic resistance, and earthquake forecasting capability. This dimensionality reduction from seven to four dimensions significantly enhanced operational efficiency and reduced model complexity.

(2) The Bayesian Optimization (BO) algorithm exhibits superior convergence and robustness when dealing with high-dimensional data and complex optimization problems. It can more rapidly locate the global optimum and demonstrates exceptional performance in handling noisy data and non-convex optimization issues, making it more adaptable to real-world applications. Through Bayesian Optimization, this study con-

figured the optimizer to search for optimal SVM hyperparameters, thereby improving model performance.

(3) A PCA-BO-SVM earthquake casualty prediction model was constructed. Additionally, a comparison of the predictions from three different models revealed that the model developed in this study had a runtime of 17.42 s, an average accuracy of 94.25%, the shortest processing time, the highest accuracy, and the lowest average error. Compared to the GA-SVM, BO-SVM, and PCA-GA-SVM models, the PCA-BO-SVM model demonstrated improvements in average accuracy and operational efficiency of 10.1%, 7.05%, and 0.325%, as well as 25.5%, 18.4%, and 19.2%, respectively. Simultaneously, the average error was reduced by 12.86%, 9.01%, and 2%. These results indicate that the model effectively captures the nonlinear relationships between various indicators and earthquake casualty numbers, thus meeting practical demands and laying the foundation for subsequent resource demand predictions.

(4) Following an earthquake, the most urgent needs of disaster victims are considered in the demand forecasting for emergency supplies. These supplies can be categorized into consumable and non-consumable types. This study selects drinking water, compressed biscuits, medical supply kits, tents, and clothing from both categories as representative emergency supplies to predict their demand.

Finally, concerning resource demand forecasting, this study focused solely on typical emergency items such as drinking water, compressed biscuits, medical supply kits, tents, and clothing. It lacks consideration of other significant emergency resources, which could be addressed in future research to expand the range of resource types considered.

6.1. Policy Implications

In conclusion, this study constructs and verifies the PCA-BO-SVM model to clarify its effectiveness in predicting casualties under seismic scenarios. At the same time, it is worth noting that in addition to improving the accuracy and operational efficiency of the prediction results, the impact of the proposed model on the social and policy levels cannot be ignored.

In the context of frequent earthquake disasters, it is important to have a more accurate casualty prediction model for formulating relevant policies. The PCA-BO-SVM model constructed in this paper improves the accuracy and efficiency of prediction, enables relevant departments to better formulate emergency response strategies, improves post-disaster rescue operations, and provides a scientific basis for emergency management and rescue operations.

Policymakers can use the number of casualties predicted by the model to indirectly predict the demand for emergency supplies, optimize resource allocation, ensure that urgently needed materials and personnel arrive at the most needed places at the best time, minimize casualties, lay a foundation for disaster relief and social stability after disasters, and improve the government's response to emergencies.

6.2. Limitation and Future Walker

Finally, in future research and applications, the adaptability of the model to multiple disaster types and different socio-economic backgrounds can be considered. By improving the algorithm to integrate the model through data integration, exploring the hybrid model method, considering more input features into the model, and interdisciplinary cooperation, this model is further improved and optimized so as to provide more solid theoretical support for policy formulation and social security. At the same time, the prediction of emergency supplies should also consider the influence of other factors in future research,

so as to enrich the types of emergency supplies and items to meet the needs of disaster victims more scientifically and reasonably.

Author Contributions: Data curation, H.X. and H.Y.; methodology, F.W., H.X., and Y.L.; founding acquisition, F.W.; visualization, H.X. and Y.L.; validation, F.W. and H.X.; writing—original draft, F.W. and H.X.; writing—review and editing, F.W., Y.L., and Y.W.; supervision, F.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the National Natural Science Foundation of China (No. 72274001), the Major Project of Natural Science Research of Anhui Provincial Education Department of China (No. 2024AH040027), and the Key Project of the National Social Science Foundation of China (No. 22AJL002) and Excellent Scientific Research and Innovation Team of Anhui Provincial Education Department of China (No. 2022AH010027).

Data Availability Statement: As this work was carried out in an emergency situation, the data used were based on data on earthquake disasters that occurred in China's history. The method of obtaining information is through public data and relevant books of the Chinese government. The data are included in the article and has been explained in the data preparation phase. (The original contributions provided in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: No potential conflicts of interest were reported by the authors.

References

1. Biswas, S.; Kumar, D.; Hajiaghaei-Keshteli, M.; Bera, U.K. An AI-based framework for earthquake relief demand forecasting: Acase study in Türkiye. *Int. J. Disaster Risk Reduct.* **2024**, *102*, 104287. [CrossRef]
2. Zarghami, S.A.; Dumrak, J. A system dynamics model for social vulnerability to natural disasters: Disaster risk assessment of an australian city. *Int. J. Disaster Risk Reduct.* **2021**, *60*, 102258. [CrossRef]
3. Balaei, B.; Noy, I.; Wilkinson, S.; Potangaroa, R. Economic factors affecting water supply resilience to disasters. *Soc. Econ. Plann Sci.* **2020**, *76*, 100961. [CrossRef]
4. Altay, N.; Narayanan, A. Forecasting in humanitarian operations: Literature review and research needs. *Int. J. Forecast.* **2020**, *38*, 1234–1244. [CrossRef]
5. Huang, D.; Wang, S.; Liu, Z. A systematic review of prediction methods for emergency management. *Int. J. Disaster Risk Reduct.* **2021**, *62*, 102412. [CrossRef]
6. Abazari, S.R.; Aghsami, A.; Rabbani, M. Prepositioning and distributing relief items in humanitarian logistics with uncertain parameters. *Soc. Econ. Plann Sci.* **2021**, *74*, 100933. [CrossRef]
7. Sheu, J. An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transp. Res. E Logist. Transp. Rev.* **2007**, *43*, 687–709. [CrossRef]
8. Sheu, J. Dynamic Relief-Demand Management for Emergency Logistics Operations Under Large-Scale Disasters. *Transp. Res. E Logist. Transp. Rev.* **2010**, *46*, 1–17. [CrossRef]
9. Taskin, S.; Lodree, E.J. A Bayesian decision model with hurricane forecast updates for emergency supplies inventory management. *J. Oper. Res. Soc.* **2011**, *62*, 1098–1108. [CrossRef]
10. Sun, B.Z.; Ma, W.M.; Zhao, H.Y. A fuzzy rough set approach to emergency material demand prediction over two universes. *Appl. Math. Model.* **2013**, *37*, 7062–7070. [CrossRef]
11. Zheng, Y.J.; Yu, S.W.; Song, Q.; Huang, Y.J.; Sheng, W.G.; Chen, S.Y. Co-Evolutionary Fuzzy Deep Transfer Learning for Disaster Relief Demand Forecasting. *IEEE Trans. Emerg. Top. Comput.* **2022**, *10*, 1361–1373. [CrossRef]
12. Basu, S.; Roy, S.; DasBit, S. A Post-Disaster Demand Forecasting System Using Principal Component Regression Analysis and Case-Based Reasoning Over Smartphone-Based DTN. *IEEE Trans. Eng. Manag.* **2019**, *66*, 224–239. [CrossRef]
13. Rawls, C.G.; Turnquist, M.A. Pre-positioning of emergency supplies for disaster response. *Transp. Res. Part B Methodol.* **2010**, *44*, 521–534. [CrossRef]
14. Ma, Z.C.; Zhang, J.; Gao, S.C. Research on emergency material demand based on urgency and satisfaction under public health emergencies. *PLoS ONE* **2023**, *18*, e0282796. [CrossRef]
15. Chen, F.J.; Chen, J.Y.; Liu, J.G. Forecast of flood disaster emergency material demand based on IACO-BP algorithm. *Neural Comput. Appl.* **2022**, *34*, 3537–3549. [CrossRef]
16. Al-Ghamdi, A.; Kamel, S.; Khayyat, M. A hybrid neural network-based approach for forecasting water demand. *Comput. Mater. Contin.* **2022**, *73*, 1365–1383. [CrossRef]

17. Rostami-Tabar, B.; Hasni, M.; Babai, M.Z. On the inventory performance of demand forecasting methods of medical items in humanitarian operations. *IFAC-PapersOnLine* **2022**, *55*, 2737–2742. [CrossRef]
18. Lin, A.Q.; Wu, H.; Liang, G.H.; Cardenas-Tristan, A.; Wu, X.; Zhao, C.; Li, D. A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster. *Int. J. Disaster Risk Reduct.* **2020**, *49*, 101682. [CrossRef]
19. Fei, L.G.; Wang, Y.Q. Demand prediction of emergency materials using case-based reasoning extended by the Dempster-Shafer theory. *Soc. Econ. Plann Sci.* **2022**, *84*, 101386. [CrossRef]
20. Riachy, C.; He, M.; Joneidy, S.; Qin, S.C.; Payne, T.; Boulton, G.; Occhipinti, A.; Angione, C. Enhancing deep learning for demand forecasting to address large data gaps. *Expert Syst. Appl.* **2024**, *268*, 126200. [CrossRef]
21. Dey, B.; Dikshit, P.; Sehgal, S.; Trehan, V.; Sehgal, V.K. Intelligent solutions for earthquake data analysis and prediction for future smart cities. *Comput. Ind. Eng.* **2022**, *170*, 108368. [CrossRef]
22. Wei, H.; Shohet, I.M.; Skibniewski, M.A.; Shapira, S.; Levy, R.; Levi, T.; Salamon, A.; Zohar, M. Assessment of Casualty and Economic Losses from Earthquakes Using Semi-empirical Model. *Procedia Eng.* **2015**, *123*, 599–605. [CrossRef]
23. Horspool, N.; Elwood, K.J.; Johnston, D.; Deely, J.M.; Ardagh, M.W. Factors influencing casualty risk in the 14th November 2016 MW7.8 Kaikōura, New Zealand earthquake. *Int. J. Risk Reduct.* **2020**, *51*, 101917. [CrossRef]
24. Ma, L.; Qin, B.D.; Lu, N. Demand Forecasting Method of Emergency Materials Based on Metabolic Gray Marko. *J. Syst. Simul.* **2023**, *35*, 229–240. (In Chinese) [CrossRef]
25. Chen, X.; Liu, Z. Demand forecast of emergency supplies based on gray model. In Proceedings of the International Conference on Advances in Mechanical Engineering and Industrial, Zhengzhou, China, 11–12 April 2015. [CrossRef]
26. Wang, F.Y.; Ge, X.F.; Li, Y.; Zheng, W.C. Optimising the distribution of multi-Cycle emergency supplies after a disaster. *Sustainability* **2023**, *15*, 902–929. [CrossRef]
27. Hu, M.Q.; Liu, Q.Y.; Ai, H.Z. Research on Emergency Material Demand Prediction Model Based on Improved Case-Based Reasoning and Neural Network. In Proceedings of the IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Changsha, China, 20–22 October 2021; pp. 270–276. [CrossRef]
28. Wang, C.H.; Liu, L.S.; Ren, J.; Yuan, Y.; Wang, L.B.; Chen, K.N. Application of Support Vector Machine Model Optimized by Principal Component Analysis and Genetic Algorithm in the Prediction of Earthquake Casualties. *Earthquake* **2020**, *40*, 142–152. (In Chinese) [CrossRef]
29. Huang, X.; Zhou, Z.L.; Wang, S.Y. The prediction model of earthquake casuailty based on robust wavelet v-svm. *Nat. Hazards* **2015**, *77*, 717–732. [CrossRef]
30. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *Adv. Neural Inf. Process Syst.* **2012**, *25*, 4.
31. Tang, H.; Zhong, Q.; Chen, C.; Martek, I. The Adaptive Seismic Resilience of Infrastructure Systems: A Bayesian Networks Analysis. *Systems* **2023**, *11*, 84. [CrossRef]
32. Chen, T.; Xu, Y.R.; Long, H.Y. Demand prediction of emergency supplies for earthquake disaster. *J. Shanghai Univ. (Nat. Sci. Ed.)* **2022**, *28*, 946–956. (In Chinese) [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.