Article

Research on the Public's Support for Emergency Infrastructure Projects Based on K-Nearest Neighbors Machine Learning Algorithm

Caiyun Cui, Huan Cao, Qianwen Shao, Tingyu Xie and Yaming Li *

Architectural Engineering College, North China Institute of Science and Technology, Langfang 065201, China; 200800890ccy@ncist.edu.cn (C.C.); caohuan@ncist.edu.cn (H.C.); shaoqianwen@ncist.edu.cn (Q.S.); xty@ncist.edu.cn (T.X.)
* Correspondence: liym@ncist.edu.cn

Abstract: The public's support for emergency infrastructure projects, which will affect the government's credibility, social stability, and development, is very important. However, there are few systematic research findings on public support for emergency infrastructure projects. In order to explore the factors influencing the public's support and the degree of influence of each factor on the public's support, this paper employs K-Nearest Neighbors (KNN), a learning curve with m-fold cross-validation, grid search, and random forest to study the public's support for emergency infrastructure projects and its influencing factors. In this paper, a prediction model of the public's support for emergency infrastructure projects is developed based on KNN from data drawn from a questionnaire survey of 445 local residents concerning Wuhan Leishenshan Hospital, China. Two optimization algorithms, the learning curve with m-fold cross-validation and the grid search algorithm, are proposed to optimize the key parameters of the KNN predictive model. Additionally, quantitative analysis is conducted by using the random forest algorithm to assess the importance of various factors influencing public support. The results show that the prediction accuracy and model stability of the KNN prediction model based on the grid search algorithm are better than those using a learning curve with m-fold cross-validation. Furthermore, the random forest algorithm quantitative analysis shows that the most important factor influencing the public's support is government attention. The conclusions drawn from this paper provide a theoretical reference and practical guidance for decision making and the sustainable development of emergency infrastructure projects in China.

Keywords: emergency infrastructure project; public's support; K-Nearest Neighbors; random forest; machine learning

1. Introduction

The World Bank defines emergency infrastructure projects as urgent and unforeseen infrastructure initiatives aimed at addressing emergencies. For instance, in response to COVID-19, the establishment of emergency hospitals such as Huoshenshan Hospital, Leishenshan Hospital, and Xiaotangshan Hospital enabled timely treatment of the rising number of patients. These projects play a crucial role in responding to emergencies, ensuring the quality of public life, and maintaining the smooth functioning of socioeconomic activities [1]. Consequently, emergency infrastructure projects have drawn intense public attention, and the evaluation criteria for public projects have shifted from 'hard indicators' such as the construction period and resource allocation to 'soft indicators' like public satisfaction [2]. However, the short decision making time for emergency infrastructure projects makes it challenging to fully consider all public demand and willingness [3], leading the public's support to be low. This can generate significant public attention and opinion fluctuation and erode public trust in the government, which may lead to social unrest [4]. For example, residents worried about the potential pollution of drinking water because of the proximity of

Leishenshan Hospital to the water source, which led to conflict between the public and the government. This highlights the critical importance of the public's support for emergency infrastructure projects to enhance the projects' capacity for sustainable development.

So far, there have been some effective evaluations of public support in the literature. For instance, Liu et al. [5] clarified the core influencing factors and mechanisms by analyzing the public's support for the banning gasoline vehicles sales policy (BGVSP). Yao et al. [6] confirmed that the deficit model and the response model could be used to study the public's support towards environmentally friendly initiatives. The existing research on emergency infrastructure projects has focused more on achieving rapid delivery from the perspective of technology and management rather than enhancing the public's support [7], including evaluating emergency capabilities, optimal site selection, resource allocation decisions for emergency infrastructure, and infrastructure digitization. For instance, Zhu et al. [8] proposed an approach to assess emergency capabilities by constructing a scenario-based method for urban critical infrastructure disasters. Yu et al. [9] and Yuan et al. [10] developed optimal site selection and resource allocation schemes using the grey wolf optimization algorithm and the maximum preparedness coverage model, respectively. Jin et al. [11] carried out infrastructure digitization to promote the digital transformation of China's social governance.

Research on public support for emergency infrastructure projects currently remains far from sufficient in two aspects: (1) limited research of factors influencing public support for emergency infrastructure projects and (2) insufficient quantitative description of the relationships between these factors and public support.

In terms of research methodologies, the relevant literature has seen the application of various technologies to investigate the public's support. For instance, Mao and Wen [12] employed the Theory of Planned Behavior (TPB) model to assess scholars' support for academic entrepreneurship. Ren et al. [13] proposed an opinion evolution analysis model based on Gradient Boosting Regression Trees (GBRT), which accurately predicts the public's support. Wazirali [14] adopted an innovative approach that combined hyperparameter tuning with five-fold cross-validation to enhance the algorithm accuracy of the KNN intrusion detection system. Similarly, Li et al. [15] introduced a hyperparameter optimization algorithm called MARSAOP. Moreover, Kim and Park [16] employed grid search for parameter optimization in the gradient-boosting machine learning algorithm. Their study achieved highly efficient predictions of the public's support for lifelong learning. The research results show that employing machine learning techniques has a more comprehensive analysis of public support than the TPB model. The data volume and dimensions of this study are relatively small, and there is a certain relationship between the influencing factors and the public's support. Therefore, KNN is faster than GBRT and can save time. The KNN model has a single key parameter, k (the number of nearest neighbors), and using m-fold cross-validation and grid search saved computational memory resources compared to MARSAOP, making it more efficient.

To make up for the research gap, the current study aims to (1) explore the factors influencing the public's support for emergency infrastructure projects and (2) analyze the quantitative influence of each factor on the public's support. The research findings will help the government and relevant departments to gain a full understanding of public demand and willingness while identifying their own issues. This facilitates them in making better-informed decisions, which provides a reference basis for the sustainable development of emergency infrastructure projects.

The subsequent sections of this paper are organized as follows: Section 2 details the data collection process and research methodology. Section 3 presents the research results and an in-depth discussion, offering well-founded policy recommendations. Finally, conclusions and research limitations are provided.

2. Methods

2.1. Framework

In this section, the research process was first divided into various stages. Subsequently, data collection and processing were conducted using literature analysis and a questionnaire survey, and then an optimized KNN model was constructed based on the KNN algorithm, learning curve with m-fold cross-validation, and grid search. Finally, quantitative analysis was performed using the random forest.

2.2. Divide the Research Stage

This study employed a systematic process to review the public's support for emergency infrastructure projects and its influencing factors. Figure 1 divides the research framework into a three-stage process. Data collection and processing were conducted in Stage 1. In Stage 2, an optimized KNN prediction model was constructed to make predictions of new sample data. Stage 3 is the quantitative analysis of the influencing factors of public support. The flow of the research framework of the current study comprised the following nine steps.
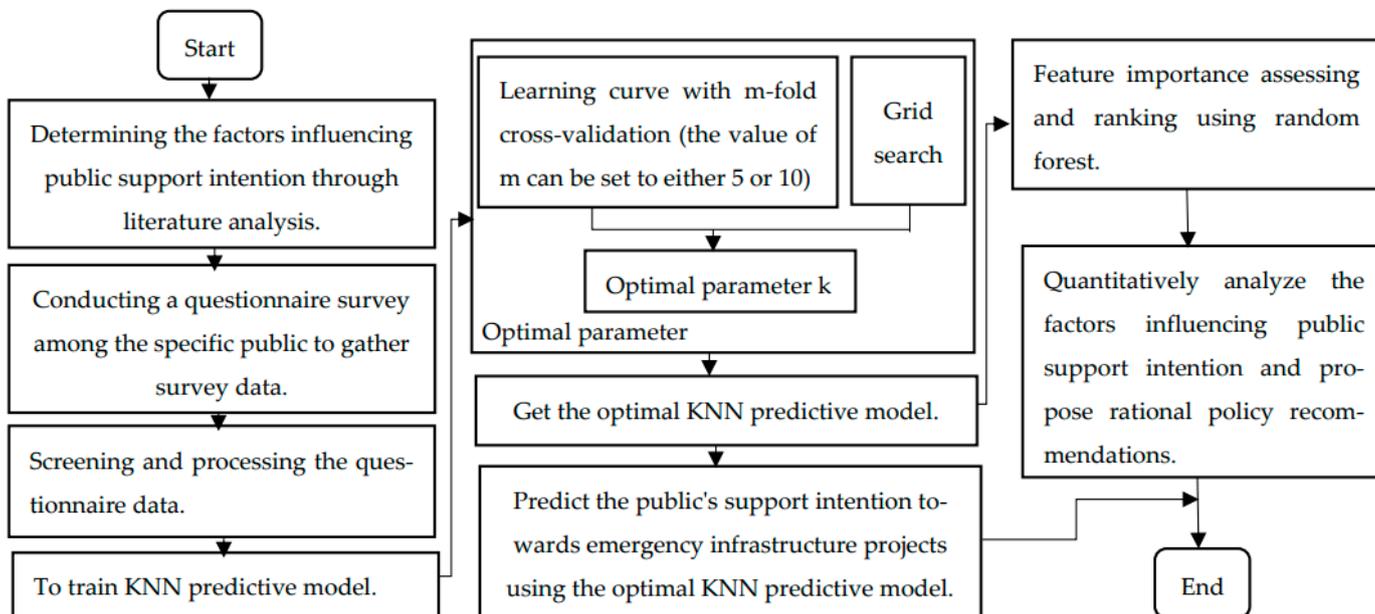

Figure 1. Flowchart of research.

2.3. Stage 1: Data Collection and Processing

Step 1 This study identifies the factors influencing public support through a compreh ensive literature analysis. This involves refining and finalizing the questionnaire items by drawing from well-established measurement items used in relevant domestic and international studies while also considering the unique characteristics of our research subject.

Step 2 This study carefully selected a specific public sample residing in the vicinity of a particular emergency infrastructure project, and we conducted a thorough questionnaire survey.

Step 3 The data collected from the survey underwent rigorous screening and processing to ensure quality and reliability. For instance, incomplete, insincere, or inconsistent questionnaires were excluded from the analysis. Meanwhile, SPSS 25.0 was used to test the questionnaire data.

2.4. Stage 2: Construct an Optimized KNN Prediction Model

Step 4 This study employed the KNN algorithm to construct a predictive model of public support for emergency infrastructure projects using the processed questionnaire data. The KNN algorithm is a classification method that selects the k-nearest neighbors to an unknown sample based on their distances. It then assigns a class label to the unknown

sample based on the majority class of its k-nearest neighbors. Let the number of samples be N. K-nearest neighbors are $k_1, k_2, \ldots, k_c$. Then, the discriminant function can be defined as follows:

$$g_i(x) = \max(k_i) \qquad\qquad i = 1, 2, \ldots c, x \in N \qquad (1)$$

Here, c indicates the class number.

Step 5 In this study, to tackle imbalanced datasets (where the distribution of samples across different categories is uneven), accuracy, recall, precision, and an F-measure were employed as performance evaluation metrics for the model. Their respective mathematical expressions are shown in the Equations (2)–(5). In order to achieve the optimal performance evaluation metrics for the model, the most important thing is the choice of parameter k. This study applied two-parameter optimization techniques: learning curve with m-fold cross-validation and grid search to determine the optimal value for the parameter k.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \qquad (2)$$

$$Recall = TP/(TP + FN) \qquad (3)$$

$$Precision = TP/(TP + FP) \qquad (4)$$

$$F - measure = \left(\alpha^2 + 1\right) Recall \times Precision / \left(\alpha^2 Recall + Precision\right) \qquad \alpha = 1 \qquad (5)$$

To further understand the above evaluation metrics, the concept of a confusion matrix [17] was introduced. The confusion matrix is shown in Table 1 differentiating between the positive class (the minority) and the negative class (the majority).

Table 1. Confusion matrix.

	Prediction Positive	Description	Prediction Negative	Description
Reference Positive	True Positive (TP)	Predicted as positive class. Correctly predicted.	False Positive (FN)	Predicted as negative class. Incorrectly predicted.
Reference Negative	False Positive (FP)	Predicted as positive class. Incorrectly predicted.	True Negative (TN)	Predicted as negative class. Correctly predicted.

Accuracy, as shown in Equation (2), refers to the ratio of instances correctly classified by the classifier to the total number of samples in the given dataset. According to Equation (3), recall measures how many positive instances were correctly classified among all true positive instances, with a greater focus on the minority class. In Equation (4), precision measures how many actually true positive instances are among all predicted positive instances, with a greater focus on the majority class. It is challenging to simultaneously achieve high recall and high precision. The F-measure combines precision and recall to strike a balance between the two and find the optimal combination.

Step 6 The above two-parameter optimization algorithms help determine the optimal value for the parameter k. Retraining the KNN model using the optimal value of k ultimately obtains an optimal KNN prediction model.

Step 7 The optimal KNN predictive model was utilized to make accurate predictions regarding the public's support for emergency infrastructure projects.

2.5. Stage 3: Quantitative Analysis

Step 8 This study leverages the random forest algorithm to assess and rank the importance of the various factors influencing public support.

Step 9 This quantitative analysis allows us to propose targeted policy recommendations based on solid empirical evidence.

3. Research Designs

3.1. Questionnaire Design

This study aims to gather questionnaire data to establish a predictive model for the public's support of emergency infrastructure projects based on the KNN algorithm. Additionally, this study seeks to analyze the factors influencing public support. Based on the literature review method, a three-part self-administered questionnaire was designed. (1) Introduction: In this part, the purpose of the questionnaire was clearly explained to the participants. They were assured that their participation was strictly for academic research and that their privacy would be protected. The aim was to alleviate any concerns participants had and ensure the authenticity and validity of the questionnaire. (2) Background Information: This part comprised seven categories of items, such as gender, age, educational level, distance from Leishenshan Hospital, etc. The specific details of background information are shown in Table 2. (3) Measurement Items: This part was developed based on a thorough literature analysis. It incorporated well-established measurement items from relevant studies while considering the unique characteristics of the research subject. By doing so, it ensured the rationality and scientific nature of the questionnaire. The section covered eight major categories, namely government attention, public concern, social comparison, emotional response, prior experience, interaction level, psychological distance, and public support. It includes 10 specific measurement items. The detailed descriptions and sources of the measurement items are shown in Table 3.

Table 2. Background information description.

Features	Items	Option	Coding	Features	Items	Option	Coding
Gen	Gender	Male	1	Occ	Occupation Type	Agricultural laborer	1
		Female	2			Self-employed worker	2
Age	Age	<30	1			Company employee	3
		30–44	2			Student	4
		45–59	3			Government employee	5
		>60	4			Other occupation	6
Edu	Educational Level	≤Junior high school	1	Dis	Distance from Leishenshan Hospital	<1 km	1
		Senior high school	2			1–3 km	2
		Junior college	3			3–6 km	3
		Undergraduate	4			6–12 km	4
		≥Graduate	5			>12 km	5
Tre	Someone you know was admitted to Leishenshan Hospital for treatment	Yes	1	Dia	Someone you know has confirmed COVID-19	Yes	1
		No	2			No	2

Table 3. Measurement item descriptions.

Categories	Features	Items	Option	Coding	Numbers	References
Government attention	G-attention	Government concern about public concerns.	Insufficient attention	0	156	[18–20]
			Extremely concerned	1	289	
Public concern	P-concern-t	Concern about the COVID-19 situation.	Insufficient attention	0	202	[21]
			Extremely concerned	1	243	
	P-concern-e	Concern about Leishenshan Hospital.	Insufficient attention	0	245	
			Extremely concerned	1	200	
Social comparison	S-comparison	Concern about comparisons with foreign countries.	Insufficient attention	0	292	[22,23]
			Extremely concerned	1	153	

Table 3. *Cont.*

Categories	Features	Items	Option	Coding	Numbers	References
Emotional response	E-response	Emotional responses lead to support for all decisions.	Insufficient attention	0	164	[24–26]
			Extremely concerned	1	281	
Prior experience	P-experience	Experienced other similar emergencies.	Heard or never experienced	0	224	[27,28]
			Personal experience	1	221	
Interaction level	I-level	Frequent participation in topical discussions and interactions.	Low participation	0	184	[25,29]
			Frequently participate	1	261	
Psychological distance	P-environment	Will not pollute the surrounding environment.	Some pollution to varying degrees	0	142	[30–32]
			Will not pollute	1	175	
			Potential pollution hazards	2	128	
	N-impact	Has not had negative impacts on life.	Some impact to varying degrees	0	95	
			No impact	1	206	
			Negligible impact	2	144	
Public's support	support	Public support for emergency infrastructure projects.	Dissatisfied	0	173	[33]
			Strongly supportive	1	272	

3.2. Sample and Data Collection

This study employed a stratified random sampling method for sample selection. Firstly, considering its significant role in combating the COVID-19 pandemic, Wuhan's Leishenshan Hospital was chosen as the subject for the emergency infrastructure project research. Secondly, Jiangxia District, where Leishenshan Hospital is located, was selected as the survey area. Finally, the survey area was divided into residential communities or villages based on their distance from Leishenshan Hospital, and residents within a range of 0 to 12 km from the hospital were randomly selected as respondents.

The questionnaire survey was conducted face-to-face with respondents in an anonymous manner from 15 April 2021 to 5 August 2022. On average, it took approximately 25 min for each respondent to complete the questionnaire. The research issued 750 questionnaires and recovered 631, with a recovery rate of 84.13%. There were two exclusion criteria: (1) a questionnaire with incomplete answers and (2) answers with obvious inconsistencies or insincerity caused by the respondents' incomprehension, even after the explanation in the face-to-face survey. As a result, 445 valid questionnaires were selected; the efficiency was 70.52%. To build the predictive model, 285 questionnaires were randomly chosen as the training set for model training, 71 questionnaires were allocated to the validation set for model calibration, and 89 questionnaires were used as the testing set to evaluate the predictive performance of the model.

4. Results and Discussion

4.1. Initial Validation of Data

Table 4 provides details of the respondents' demographic characteristics, indicating that the sample distribution is considered to be generally relevant and representative. The largest proportion of respondents by education level were high school or above (82.5%), which shows that the sample can generally understand the content of the questionnaire very well. Most respondents chose "other occupation" (47.9%), with the remaining 52.1% comprising agricultural laborers, self-employed people, company employees, students,

and government employees. Respondents from all occupations participated, which is a good indication of the diversity of the subjects. We noted that 15.5% of respondents lived within three kilometers of Leishenshan Hospital, and the remaining 85.5% lived further away, which is in line with the distribution of the local population. Respondents chose either "Yes" or "No" in answer to the question of whether they knew someone who had been diagnosed with COVID-19 and whether they knew someone who had been treated at Leishenshan Hospital.

Table 4. Respondents' demographic information.

Features	Option	Number	Percentage
Gen	Male	205	46.1%
	Female	240	53.9%
Age	<30	168	37.8%
	30–44	117	26.3%
	45–59	86	19.3%
	>60	74	16.6%
Edu	≤Junior high school	78	17.5%
	Senior high school	146	32.8%
	Junior college	110	24.7%
	Undergraduate	104	23.4%
	≥Graduate	7	1.6%
Occ	Agricultural laborer	37	8.3%
	Self-employed worker	37	8.3%
	Company employee	64	14.4%
	Student	62	13.9%
	Government employee	32	7.2%
	Other occupation	213	47.9%
Dis	<1000 m	10	2.2%
	1000–3000 m	59	13.3%
	3000–6000 m	60	13.4%
	6000–12,000 m	253	56.9%
	>12,000 m	63	14.2%
Dia	True	69	15.5%
	False	376	84.5%
Tre	True	32	7.2%
	False	413	92.8%

The statistical results for all measurement items are shown in Table 5. The average values for government attention, emotional response, and the public's support were between 0.61 and 0.65, all greater than 0.6 and close to 1, indicating that the public gave relatively positive responses. Conversely, the average values for public concern, social comparison, prior experience, and interaction level were all between 0.34 and 0.59, implying that the public perceived that there was room for improvement in these areas. Additionally, the average value for psychological distance was between 0.97 and 1.11, less than 1.2 and far from 2, indicating relatively low satisfaction. In addition, the kurtosis coefficient and skewness coefficient of all measurement items met the data, presenting a normal distribution.

In this study, the Cronbach's alpha coefficient was employed as an indicator to assess the questionnaire's reliability. Generally, Cronbach's alpha coefficient above 0.7 indicates a high level of questionnaire reliability; values between 0.6 and 0.7 are considered acceptable, and values below 0.6 are not acceptable. In this study, SPSS 25.0 was used to calculate the Cronbach's alpha coefficient of the questionnaire as 0.705, which shows that the questionnaire used in this study has high reliability.

Table 5. Statistical results of the measurement items.

Categories	Features	N	Mean	Standard Deviation	Skewness	Kurtosis
Government attention	G-attention	445	0.65	0.478	−0.629	−1.612
Public concern	P-concern-t	445	0.55	0.498	−0.186	−1.974
	P-concern-e	445	0.45	0.498	0.204	−1.967
Social comparison	S-comparison	445	0.34	0.476	0.660	−1.572
Emotional response	E-response	445	0.63	0.483	−0.547	−1.709
Prior experience	P-experience	445	0.50	0.501	0.014	−2.009
Interaction level	I-level	445	0.59	0.493	−0.353	−1.884
Psychological distance	P-environment	445	0.97	0.779	0.055	−1.349
	N-impact	445	1.11	0.725	−0.171	−1.086
Public's support	support	445	0.61	0.488	−0.458	−1.798

This study employed the Pearson correlation coefficient to analyze the correlation between public background information and the measurement item "emotional response". Table 6 shows the correlations between these two factors. The correlational analyses show that the public emotional response to emergency infrastructure projects is not influenced by Gen, Age, Edu, Dis, and Tre and does not vary significantly based on their differences. However, public emotional response was found to be positively correlated with occ (0.105, $p < 0.05$) and negatively with Dia (−0.121, $p < 0.05$).

Table 6. Pearson correlation coefficients.

	Gen	Age	Edu	Occ	Dis	Dia	Tre
ER1	0.014	0.069	0.014	0.105 *	0.065	−0.121 *	−0.055

Notes: $N = 445$, * $p < 0.05$.

4.2. Predictive Model for Public's Support for Emergency Infrastructure Projects Based on KNN

In this study, 16 items were defined as features for the classification predictive algorithm. Among these items were seven background information items and nine measurement items (excluding the 'support' variable). These features served as sample inputs to establish a predictive model of the public's support for emergency infrastructure projects. The specific steps for constructing the predictive model were as follows:

(1) Firstly, the historical data from the questionnaire survey were carefully preprocessed, and incomplete, insincere, or inconsistent responses were excluded from the dataset, ensuring that the final dataset contained only reliable and valid information.

(2) Next, the relationship between the factors influencing the public's support and the corresponding public support was established as a set called W within the entire dataset. Set W contained i samples, where each sample comprised p influencing factors of public support and one public's support denoted as Q. In this study, the value of p was 16, which included the seven background information items mentioned in Table 2 and the nine measurement items listed in Table 3 (excluding 'support'). The value of Q was either 0 or 1, representing the two different categories of public support in the questionnaire. This relationship can be mathematically represented as shown in Equation (6):

$$W = \begin{pmatrix} [X_{11}, X_{12}, \ldots, X_{1p}, Q_1] \\ [X_{21}, X_{22}, \ldots, X_{2p}, Q_2] \\ \ldots \\ [X_{i1}, X_{i2}, \ldots, X_{ip}, Q_i] \end{pmatrix} \tag{6}$$

(3) Finally, the factors (X) influencing public support were defined as the target sample for prediction. In the KNN classification predictive algorithm in this study, the process begins with traversing the entire sample set W and computing the distances between the target sample and each sample in set W. These distances were then

sorted in ascending order to identify the top k-nearest neighbors. Subsequently, the corresponding public support set, $Q = [Q_1, Q_2, \ldots, Q_k]$, of these k-nearest neighbors was obtained. Ultimately, voting was performed on set Q. In this step, each public support in set Q equaled one vote. The public's support Q_k with the highest number of votes was then assigned as the public's support for the target sample. In this study, the Euclidean distance metric was used for this purpose. Euclidean distance is mathematically represented as shown in Equation (7):

$$L_2(x_i, x_j) = \left(\sum_{i=1}^{n} |x_i - x_j|^2 \right)^{\frac{1}{2}} \tag{7}$$

In this study, achieving an optimal predictive model required careful selection of the nearest neighbor parameter k. The value of k played a crucial role in the KNN algorithm's performance. If k is too small, the model may become overly sensitive to noise in the data, leading to overfitting. This means that the model will perform well on the training set, but its performance will be significantly worse for new, unseen data (test and validation sets), indicating low generalization ability. On the other hand, if k is too large, the model may oversimplify the underlying patterns in the data, leading to underfitting. In this case, the model will have increased approximation errors during the learning process and may not accurately capture the intricacies of the relationships between the influencing factors and the public's support. To overcome these challenges and find the optimal value of k, this study adopted two methods: learning curves with m-fold cross-validation and grid search. These methods help in selecting the most appropriate value of k that will maximize the model's predictive performance. The choice of k in the KNN predictive model has a significant impact on its performance, resulting in variations in various evaluation metrics.

4.2.1. Learning Curve with m-Fold Cross-Validation Results

In the first step, the program was executed multiple times with all possible k values ranging from 0 to 20 to construct learning curves for the established KNN predictive model. The best value of k was then determined based on the point where the model exhibited the best performance on the learning curve. However, in practical research, the learning curves vary each time the program is executed. This suggests that the established predictive model's generalization ability is not optimal. To address this issue and enhance the model's generalization ability, this study employed learning curves with m-fold cross-validation due to the limited size of the dataset [34]. This process helps to mitigate the impact of variations in the learning curves and ultimately results in an optimized KNN predictive model with improved generalization ability and better suitability for real-world applications. The principle of m-fold cross-validation is shown in Figure 2.
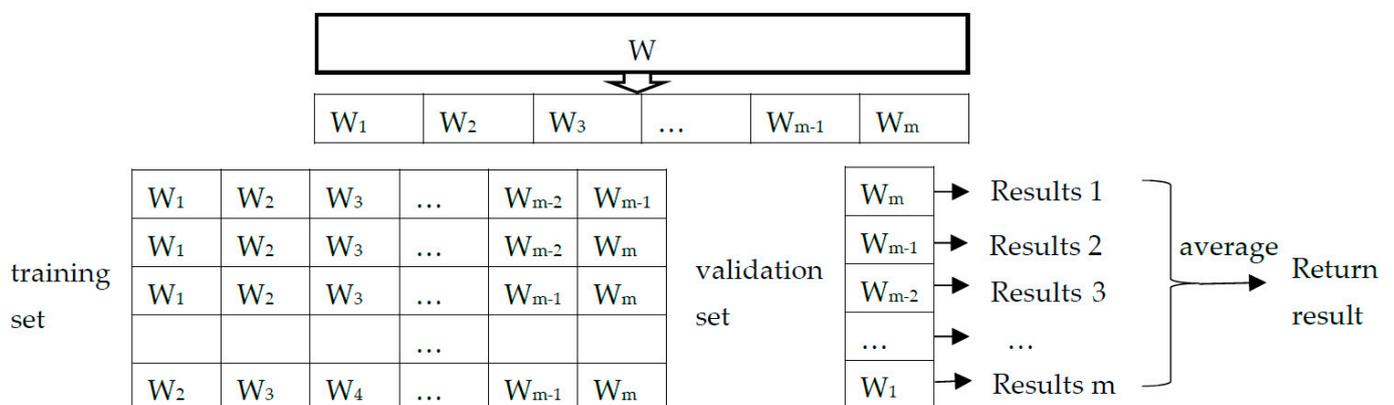

Figure 2. The principle of m-fold cross-validation.

Using the above methods, we retrained the existing KNN predictive model. Finally, the average classification accuracy of the m models was computed as the model's final classification accuracy. The value of m can be set to either 5 or 10 [35]. Different values of m result in different means and variances, which correspond to different average effects and stability of classifiers, consequently affecting various metrics of the KNN model. Such dataset partitioning allows all samples in the dataset to serve in both the training set and the validation set, which significantly enhances the model's generalization ability, resulting in an optimized KNN predictive model. Meanwhile, the best value of the nearest neighbor parameter k was determined by selecting the parameter value that corresponded to the optimized performance point on the learning curve. The results of the learning curves with m-fold cross-validation are shown in Figure 3. The horizontal axis represents the nearest neighbor parameter k with values ranging from 0 to 20, while the vertical axis represents the mean, reflecting the average effect of the KNN model. According to Figure 3a, the KNN model performs best when k is set to 12, achieving an average effect of 92.76%. According to Figure 3b, the KNN model performs best when k is set to 14, with an average effect of 93.25%.
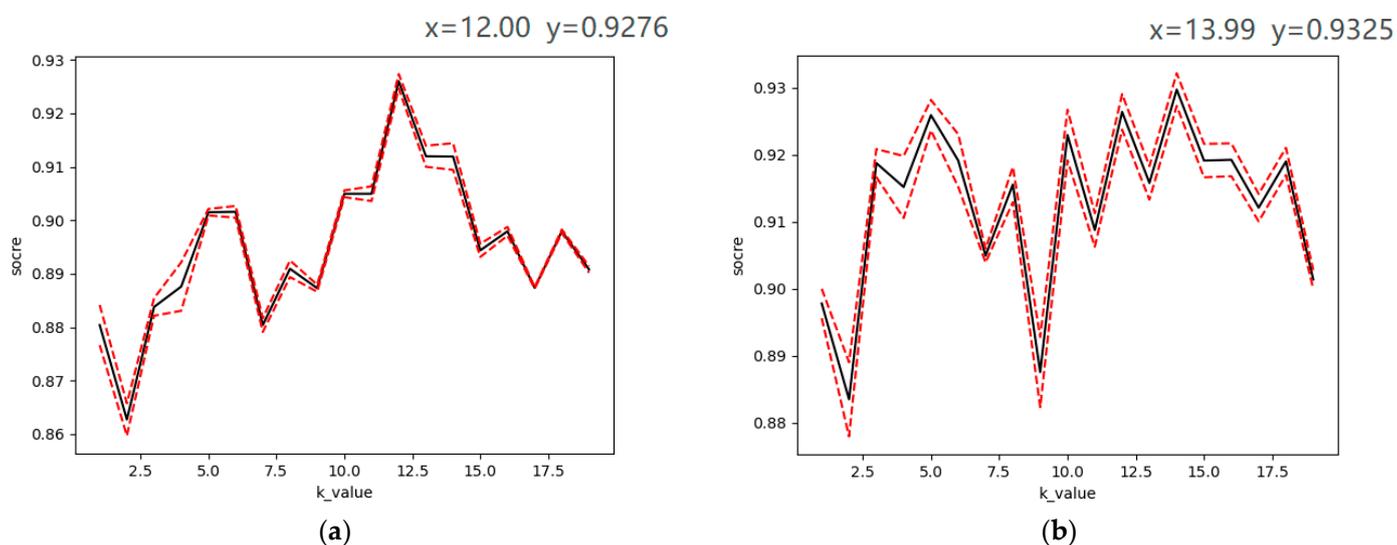

Figure 3. Learning curve with m-fold cross-validation results. (**a**) Learning curve with five-fold cross-validation results. (**b**) Learning curve with ten-fold cross-validation results.

4.2.2. Grid Search Results

This study employed the grid search algorithm to determine the optimal value of k, with the parameter search range set from 0 to 20. The grid search algorithm utilized an exhaustive search approach, where the program explored all possible values within the specified parameter range. Through iterative traversal, it attempted every possibility and selected the parameter value that exhibited the best performance. Simultaneously, this study employed the m-fold cross-validation method to calculate the algorithm's accuracy. Ultimately, the k-value demonstrating the best overall performance was chosen, leading to the selection of an optimized KNN predictive model for our specific dataset.

In the program implementation, this study defined a function named 'grid_search' that utilized the GridSearchCV method from the Sklearn machine learning library for automated parameter tuning. The parameter options are shown in Table 7. This efficiently searched for the best value of k and evaluated its overall performance using the m-fold cross-validation method. In this study, the value of m was set to 5 or 10. Finally, the program outputs the optimal value of k and the accuracy achieved by the grid search algorithm. The results of the grid search are shown in Table 8.

Table 7. The parameter options of the GridSearchCV method.

Parameter of GridSearchCV Method	Options	Parameter of GridSearchCV Method	Options
estimator	KNeighborsClassifier	n_jobs	1
param_grid	n_neighbors: range [0,20]	verbose	0
cv	5 or 10	refit	True
scoring	accuracy	iid	True

Table 8. Grid search results.

m-Fold Cross-Validation	Value of Nearest Neighbor Parameter k	Grid Search Accuracy
Five-fold cross-Validation	12	92.25%
Ten-fold cross-Validation	8	93.66%

4.2.3. KNN Model Performance with Different k Values

After employing two different methods to determine the optimal nearest neighbor parameter k for the KNN model, the model's performance metrics were as shown in Table 9 for different values of k. The selected values of k were 8, 12, and 14, and it was observed that the model's performance metrics were relatively better when k was set to 8 or 14.

Table 9. KNN model performance with different k values.

Evaluation Metrics		Learning Curve with m-Fold Cross-Validation		Grid Search	
		Five-Fold (k = 12)	Ten-Fold (k = 14)	Five-Fold (k = 12)	Ten-Fold (k = 8)
Accuracy		94.44%	95.83%	94.44%	95.83%
Recall	0	93.00%	96.00%	93.00%	96.00%
	1	96.00%	96.00%	96.00%	96.00%
Precision	0	93.00%	93.00%	93.00%	93.00%
	1	96.00%	98.00%	96.00%	98.00%
F1-score	0	93.00%	95.00%	93.00%	95.00%
	1	96.00%	97.00%	96.00%	97.00%

4.2.4. Validation of Model Prediction Performance

In this study, the test set comprised 89 valid questionnaire responses, as described in Section 2.2. It was utilized to evaluate the prediction performance of the KNN models using different k values. For validation purposes, a random sample of 20 valid questionnaire responses was chosen. The prediction results are shown in Table 10. Overall, the KNN models demonstrated good prediction performance, achieving an average accuracy of over 90%. Notably, the KNN model with a k-value of 8 exhibited more stable prediction results and displayed a superior ability to predict the public's support for emergency infrastructure projects.

Table 10. Prediction of public's support.

Actual Public Support Intention	1	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	1	1	0	1
Model Prediction Result (k = 8)	1	1	0	0	0	0	0	1	1	0	1	1	1	0	1	1	1	1	0	1
Model Prediction Result (k = 14)	1	1	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	1

4.3. Feature Importance Assessment and Ranking Results

This study utilized the random forest algorithm, which consists of multiple decision trees, to calculate the contribution of each of the 16 features to the w decision trees in the random forest [36]. This allows this study to conduct a feature importance assessment. The assessment was carried out using the Out-Of-Bag (OOB) error rate. In the program

implementation, this study leveraged the Sklearn machine learning library to investigate how each feature contributed to reducing the impurity of the w decision trees within the random forest, thereby quantifying the importance of each feature. The model was trained by executing the program, automatically computing the importance of each feature, and generating the feature importance ranking. Notably, the sum of the importance values for all features equaled 1. During model training, this paper employed the Bootstrap sampling technique to create training subsets and construct the random forest. This technique involves randomly selecting n samples with replacements from the sample set to form a training subset and repeating this process w times, generating w training subsets. The feature importance assessment and ranking results are shown in Figure 4.
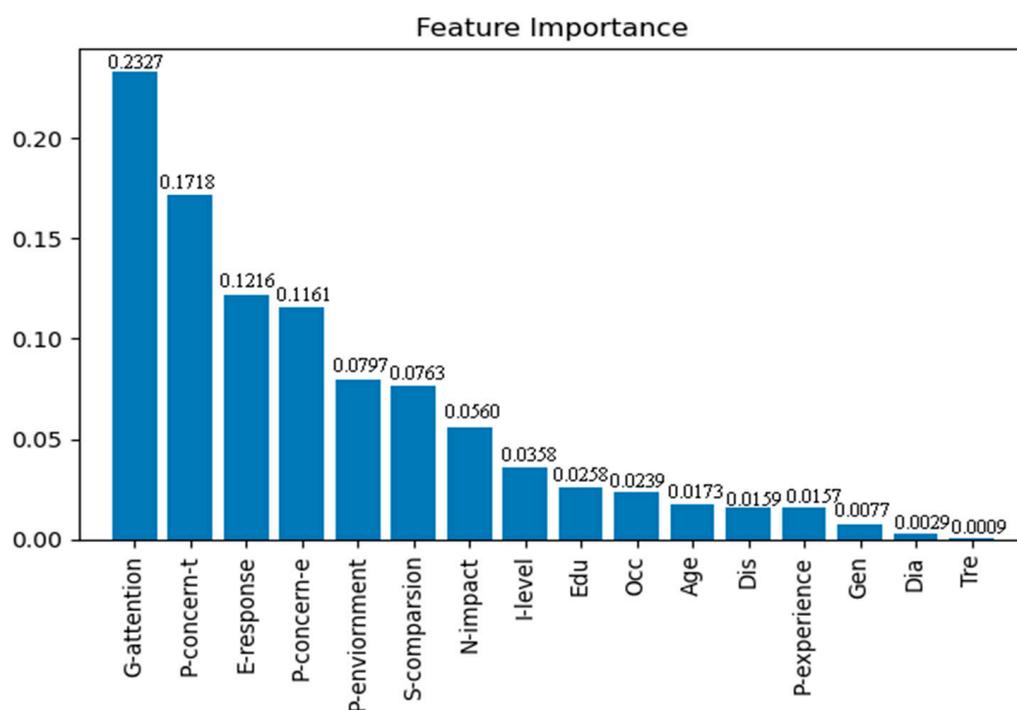
Figure 4. Feature importance assessment and ranking results.

In Figure 4, the horizontal axis represents the 16 features influencing the public's support, while the vertical axis represents the importance of each feature, arranged in descending order of importance. From Figure 4, it can be seen that government attention, public concern, and emotional response have the most substantial impact on public support, with their importance all exceeding 10%. In particular, government attention was the most significant influence, with an importance of 23.27%. Following closely behind were psychological distance and social comparison, both with importance values exceeding 5%. Finally, the impact of features like interaction level, background information, and prior experience on public support was relatively minor, with their importance all being less than 5%. Among the factors, knowing someone diagnosed with COVID-19 and knowing someone receiving treatment at Leishenshan Hospital had the least impact, with their importance being less than 0.3% (specifically 0.29% and 0.09%, respectively).

4.4. Discussion

Based on the results from the KNN prediction model and the random forest feature importance assessment, it becomes evident that government attention, public concern, and emotional response have the most significant impact on public support.

Government attention, which pertains to the government's acknowledgment of public concerns, emerges as the most critical factor influencing public support. This finding aligns with previous studies that emphasize the importance of establishing a positive government

image [37]. A positive government image fosters a strong sense of happiness among the public, thereby fostering strong public support [37].

Similarly, public concern, which reflects the level of attention the public pays to the COVID-19 pandemic and the establishment of Leishenshan Hospital, also stands out as a primary factor influencing public support. This research finding validates the discoveries of Xu et al. [38]. Public concern reflects increased awareness among the public regarding emergency infrastructure projects, leading to strong public support [38].

Additionally, emotional response, denoting that the public's emotional reaction prompts them to support decisions, is identified as another key factor influencing the public's support. This conclusion aligns with the findings of Oliver [39]. Emotional responses reflect the public's concerns in unfamiliar situations [40], influencing their behavioral judgments. Positive emotional responses empower the public to proactively adapt and have confidence in the government's measures in response to emergencies, leading to strong public support [41].

4.5. Practical Implications

Based on the research conclusions above, the following policy recommendations are proposed to promote the public's support:

(1) For the government, it is crucial to value and respect the expression of public opinions. This will help government departments identify issues and make corrections, thus enhancing public satisfaction with the government. Additionally, the government should pay close attention to public concerns. This can contribute to establishing a positive government image and foster trust and support from the public. Furthermore, regular education and guidance should be provided to enhance the public's psychological coping ability and response capabilities during emergencies. This can help eliminate negative emotional responses.

(2) Online media should prioritize timely and accurate reporting of social hot topics through official channels. Avoiding the dissemination of false information that could lead to social panic is crucial. Providing reliable and factual information fosters a positive social atmosphere and satisfaction with the government.

(3) It is essential for the public to approach emergencies with a scientific and proactive mindset. Analyzing and resolving problems in a rational manner helps avoid excessive panic and suspicion. This strengthens individual feelings of security and contributes to preventing negative emotional responses.

5. Conclusions

Given the low public support for emergency infrastructure projects, this study constructed an optimized KNN predictive model of public support for emergency infrastructure projects based on KNN, a learning curve with m-fold cross-validation, and a grid search. Additionally, the factors influencing public support were comprehensively evaluated and quantitatively analyzed using random forest. The main results of this study are as follows: (1) Background information, government attention, public concern, social comparison, emotional response, prior experience, interactive level, and psychological distance all influence, to varying degrees, the public's support for emergency infrastructure projects. Notably, government attention, public concern, and emotional response have the greatest impact, all exceeding 10%. Psychological distance and social comparison have a secondary influence, both exceeding 5%. The interactive level, background information, and prior experience have the least impact, all less than 5%. (2) The proposed KNN prediction model effectively predicts the public's support for emergency infrastructure projects during public health crises, achieving an average accuracy rate exceeding 90% and demonstrating good stability. (3) Using grid search with ten-fold cross-validation improved the predictive ability and generalization more than the learning curve with m-fold cross-validation. (4) Model predictions and random forest feature importance evaluation show that among

the various influencing factors, government attention has the greatest impact on public support, exceeding 20%.

The findings provide several theoretical insights and practical implications for the management of emergency infrastructure projects. This study examines emergency infrastructure projects from the novel perspective of the public. It expands the scope of traditional project management performance evaluation and broadens the research perspective on public support and infrastructure. This study employed machine learning techniques to study the public's support for emergency infrastructure projects and its influencing factors. It showed the novelty of research technology.

However, this study also had a few limitations. Firstly, the dataset used to train the predictive model was relatively small. In future research, it is essential to combine face-to-face and online questionnaire surveys to gather more data, thereby enhancing the generalizability of the predictive model. Secondly, the study's primary emphasis on emergency hospitals neglected consideration of emergency infrastructure projects of other types, leading to limitations in the results' applicability. It is hoped that, in future studies, the research scope can be broadened to include diverse projects, such as emergency shelter projects, to further corroborate the study's findings.

Author Contributions: Conceptualization, C.C.; data curation, H.C., Q.S. and T.X.; funding acquisition, C.C.; investigation, Q.S. and T.X.; methodology, H.C.; project administration, C.C. and Y.L.; software, Y.L.; supervision, Y.L.; writing—original draft, H.C.; writing—review and editing, C.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Wearne, S.H. Management of urgent emergency engineering projects. *Proc. Inst. Civ. Eng. Munic. Eng.* **2002**, *151*, 255–263. [CrossRef]
2. Bouckaert, G. Comparing measures of citizen trust and user satisfaction as indicators of 'good governance': Difficulties in linking trust and satisfaction indicators. *Int. Rev. Adm. Sci.* **2003**, *69*, 329–343.
3. Bearth, A.; Siegrist, M. Are risk or benefit perceptions more important for public acceptance of innovative food technologies: A meta-analysis. *Trends Food Sci. Technol.* **2016**, *49*, 14–23. [CrossRef]
4. Roe, E.; Schulman, P.R. Comparing Emergency Response Infrastructure to Other Critical Infrastructures in the California Bay-Delta of the United States: A Research Note on Inter-Infrastructural Differences in Reliability Management. *J. Contingencies Crisis Manag.* **2015**, *23*, 193–200. [CrossRef]
5. Liu, Y.; Dong, F.; Li, G.; Pan, Y.; Qin, C.; Yang, S.; Li, J. Exploring the factors influencing public support willingness for banning gasoline vehicle sales policy: A grounded theory approach. *Energy* **2023**, *283*, 128448. [CrossRef]
6. Yao, Q.; Chang, C.; Joshi, P.; McDonald, C. Climate change versus the water-energy-food nexus: The oldness or newness of the scientific issues as a factor in the deficit model and the hierarchy of response model. *Environ. Dev. Sustain.* **2022**, 1–18. [CrossRef]
7. Zidane, Y.J.-T.; Klakegg, O.J.; Andersen, B.; Hussein, B. "Superfast!" managing the urgent: Case study of telecommunications infrastructure project in Algeria. *Int. J. Manag. Proj. Bus.* **2018**, *11*, 507–526. [CrossRef]
8. Zhu, W.; Wang, J.; Yang, L. A Method Research on Scenario Construction of Critical Infrastructure Incidents and Emergency Capacity Evaluation. *Manag. Rev.* **2016**, *28*, 59–65. (In Chinese)
9. Yu, D.; Gao, L.; Zhao, S. Emergecny facility location-allocation problem with convex barriers. *Syst. Eng. Theory Pract.* **2019**, *39*, 1178–1188. (In Chinese)
10. Yuan, Y.; Liu, Y.; Zhu, S.; Wang, J. Maximal preparedness coverage model and its algorithm for emergency shelter location. *J. Nat. Disasters* **2015**, *24*, 8–14. (In Chinese)
11. Jin, J.; Yu, J.; Sun, Q.; Gao, Y. Modular co-evolution of digital infrastructure innovation: A case study of China's public health emergency governance. *Stud. Sci. Sci.* **2021**, *39*, 713–724. (In Chinese)

12. Mao, L.; Wen, L. The Influencing Factors of Academic Entrepreneurial Intention Research Based on TPB Model. *Oper. Res. Manag. Sci.* **2022**, *31*, 164–169. (In Chinese)

13. Ren, Z.; Zhang, P.; Liu, J.; Lan, Y. Research on netizens' emotion evolution in emergency based on machine learning. *J. Phys. Conf. Ser.* **2019**, *1419*, 012004. [CrossRef]

14. Wazirali, R. An Improved Intrusion Detection System Based on KNN Hyperparameter Tuning and Cross-Validation. *Arab. J. Sci. Eng.* **2020**, *45*, 10859–10873. [CrossRef]

15. Li, Y.; Liu, G.; Lu, G.; Jiao, L.; Marturi, N.; Shang, R. Hyper-Parameter Optimization Using MARS Surrogate for Machine-Learning Algorithms. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *4*, 287–297. [CrossRef]

16. Kim, C.; Park, T. Predicting Determinants of Lifelong Learning Intention Using Gradient Boosting Machine (GBM) with Grid Search. *Sustainability* **2022**, *14*, 5256. [CrossRef]

17. Deng, X.; Liu, Q.; Deng, Y.; Mahadevan, S. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Inf. Sci.* **2016**, *340–341*, 250–261. [CrossRef]

18. Song, E.; Yoo, H.J. Impact of social support and social trust on public viral risk response: A COVID-19 survey study. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6589. [CrossRef]

19. Miller, A.H.; Listhaug, O. Political Parties and Confidence in Government: A Comparison of Norway, Sweden and The United States. *Br. J. Political Sci.* **1990**, *20*, 357–373. [CrossRef]

20. Soonhee, K. Public Trust in Government in Japan and South Korea: Does the Rise of Critical Citizen Matter? *Public Adm. Rev.* **2010**, *70*, 801–810.

21. Whiting, A.; Williams, D.L. Why people use social media: A uses and gratifications approach. *Qual. Mark. Res.* **2013**, *16*, 362–369. [CrossRef]

22. Zhu, D.; Wang, G. Influencing Factors and Mechanism of Netizens' Social Emotions in Emergencies—Qualitative Comparative Analysis of Multiple Cases Based on Ternary Interactive Determinism (QCA). *J. Intell.* **2020**, *39*, 95–104. (In Chinese)

23. Bandura, A.; Pastorelli, C.; Barbaranelli, C.; Caprara, G.U. Self-efficacy path ways in depression. *J. Personal. Soc. Psychol.* **1999**, *76*, 258–269. [CrossRef] [PubMed]

24. Finucane, M.L.; Alhakami, A.; Slovic, P.; Johnson, S.M. The Affect Heuristic in Judgments of Risks and Benefits. *J. Behav. Decis. Mak.* **2000**, *13*, 1–17. [CrossRef]

25. Connelly, S.; Gooty, J. Leading with emotion: An overview of the special issue on leadership and emotions. *Leadersh. Q.* **2015**, *26*, 485–488. [CrossRef]

26. Swerdlow, B.; Johnson, S. *How Will You Regulate My Emotions? A Multistudy Investigation of Dimensions and Outcomes of Interpersonal Emotion Regulation Interactions*; University of California: Berkeley, CA, USA, 2019. [CrossRef]

27. Rasmus, T.-K.; Karl, W.; Phillip, H.K. Practice makes perfect: Entrepreneurial-experience curves and venture performance. *J. Bus. Ventur.* **2014**, *29*, 453–470.

28. Alexander, A.; Richard, C.; Sourav, R. A theory of entrepreneurial opportunity identification and development. *J. Bus. Ventur.* **2003**, *18*, 105–123.

29. Preece, J. Sociability and usability in online communities: Determining and measuring success. *Behav. Inf. Technol.* **2001**, *20*, 347–356. [CrossRef]

30. Liu, J.; Geng, L.; Xia, B.; Bridge, A. Never Let a Good Crisis Go to Waste: Exploring the Effects of Psychological Distance of Project Failure on Learning Intention. *J. Manag. Eng.* **2017**, *33*, 04017006. [CrossRef]

31. Chu, H.; Yang, J.Z. Risk or Efficacy? How Psychological Distance Influences Climate Change Engagement. *Risk Anal. Off. Publ. Soc. Risk Anal.* **2020**, *40*, 758–770. [CrossRef]

32. Spence, A.; Poortinga, W.; Pidgeon, N. The psychological distance of climate change. *Risk Anal. Off. Publ. Soc. Risk Anal.* **2012**, *32*, 957–972. [CrossRef] [PubMed]

33. Li, Q.; Luo, R.; Zhang, X.; Meng, G.; Dai, B.; Liu, X. Intolerance of COVID-19-related uncertainty and negative emotions among chinese adolescents: A moderated mediation model of risk perception, social exclusion and perceived efficacy. *Int. J. Environ. Res. Public Health* **2021**, *18*, 2864. [CrossRef] [PubMed]

34. Wang, Q.; Liu, J.; Zhang, L. Study on the Classification of K-Nearest Algorithm. *J. Xi'an Technol. Univ.* **2015**, *35*, 119–124+141. (In Chinese)

35. Lyu, Z.; Yu, Y.; Samali, B.; Rashidi, M.; Mohammadi, M.; Nguyen, T.N.; Nguyen, A. Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam. *Materials* **2022**, *15*, 1477. [CrossRef] [PubMed]

36. Jia, H.; Lin, J.; Liu, J. An Earthquake Fatalities Assessment Method Based on Feature Importance with Deep Learning and Random Forest Models. *Sustainability* **2019**, *11*, 2727. [CrossRef]

37. Chen, G. Research on the practical dilemma and countermeasures of network public opinion governance of grassroots government. *Netw. Secur. Technol. Appl.* **2022**, *03*, 118–120. (In Chinese)

38. Xu, L.; Ma, Y.; Wang, X. Study on Environmental Policy Selection for Green Technology Innovation Based on Evolutionary Game: Government Behavior vs. Public Participation. *Chin. J. Manag. Sci.* **2022**, *30*, 30–42. (In Chinese)

39. Oliver, R.L. Cognitive, Affective, and Attribute Bases of the Satisfaction Response. *J. Consum. Res.* **1993**, *20*, 418–430. [CrossRef]
40. Gerrity, M.S.; White, K.P.; Devellis, R.F.; Dittus, R.S. Physicians'reactions to uncertainty: Refining the constructs and scales. *Motiv. Emot.* **1995**, *19*, 175–191. [CrossRef]
41. Dai, W.; Meng, G.; Zheng, Y.; Li, Q.; Dai, B.; Liu, X. The impact of intolerance of uncertainty on negative emotions in COVID-19: Mediation by pandemic-focused time and moderation by perceived efficacy. *Int. J. Environ. Res. Public Health* **2021**, *18*, 4189. [CrossRef]