Prediction modeling of Household's Preparedness of Natural Hazards Mitigation

Chen Xia[1], Yuqing Hu[2]

[1]Department of Architectural Engineering, The Pennsylvania State University, University Park, PA 16802, United States of America; e-mail: cpx5037@psu.edu
[2]Department of Architectural Engineering, The Pennsylvania State University, University Park, PA 16802, United States of America; e-mail: yfh5204@psu.edu

Abstract

Natural disasters are showing an increase in the magnitude, frequency, and geographic distribution. Studies have shown that individuals' self-sufficiency, which largely depends on household preparedness, is very important for hazard mitigation in at least the first 72 hours following a disaster. However, for factors that influence a household's disaster preparedness, though many studies are trying to identify from different aspects, we still lack an integrative analysis on how these factors contribute to a household's preparation. This paper aims to build a classification model to predict whether a household has prepared for a potential disaster based on their characteristics and the environment they located. We collect data from the Federal Emergency Management Agency's National Household Survey in 2018 and train four classification models - logistic regression, decision trees, support vector machines, and multi-layer perceptron classifier models- to predict the impact of personal characteristics and the environment they located on household prepare for the potential natural disaster. Results show that the multi-layer perceptron classifier model outperforms others with the highest scoring on both recall (0.8531) and f1 measure (0.7386). In addition, feature selection results also show that among other factors, a household's accessibility to disaster-related information is the most critical factor that impacts household disaster preparation. Though there is still room for further parameter optimization, the model gives a clue that we could support disaster management by gathering publicly accessible data.

Keywords: Household Disaster Preparation; Natural Hazards Mitigation; Prediction Model

Introduction

In recent years, natural disasters have shown an increase in the magnitude, frequency, and geographic distribution aspects. According to a World Bank report, nearly 3.8 million km^2 and 790 million individuals are exposed to at least two natural disasters (Dilley,2005). Millions of people in the world are exposed to the growing multi-hazard environment, which increases the importance of disaster preparation to mitigate damage, especially in a disaster-prone area. Disaster preparation needs collaboration from multiple social units:

1

households, public organizations, and local and federal disaster management departments. The preparedness could be reflected both in disaster risk perceptions and disaster preparedness practices and plays an important role in other phases (Jagnoor,2019). For individuals and households, preparedness actions before a disaster greatly reduce their risk of getting trapped into severe trouble and enable them to respond actively when disaster does happen (Das, 2018). For an organization in charge of disaster management, besides their plans for all levels of the emergency chain of action (Khorram-Manesh, 2020), information on how households prepared for disaster also provide support for their response strategies such as the allocation of resources and urgent evacuation (Khorram-Manesh, 2020). Therefore, as a basic unit to respond to disasters, households play an important role, and how to increase the households' engagement for disaster preparation is critical. However, though many factors have been put forward, we still lack an integrative analysis of how these factors contribute to the household's preparation. This paper aims to build a classification model to predict whether a household has prepared for a potential disaster based on their characteristics and the environment they are located, to provide information for governments to carry out more targeted resource distribution strategies in post disaster.

Literature Review

In recent decades, numerous studies have been focused on assessing individuals' levels of preparedness for natural hazards, and the factors that promote the adoption of preparedness measures. Bronfman surveyed individuals' preparedness for different natural hazards and revealed that participants are significantly better prepared for earthquakes than floods (Bronfman, 2019). Different theoretical frameworks have also been put forward to conceptualize the adoption of preparedness measures to face natural hazards. The most cited models are the Protective Action Decision Model and the Social-Cognitive Model. The first model reveals that people respond to natural hazards depending on environmental and cues, warnings, as well as receivers' characteristics (Lindell, 2012). The social cognitive model focuses on the role of motivational factors on the decision to adopt preparedness actions (Aton,2005). While these models may have different emphases on modeling an individual's preparation and response to hazards, general factors such as individual characters and environmental impacts have both been involved.

For individual characters, many studies tried to figure out how perception influences people's behavior in both pre-and post-disaster periods (Bronfman,2016; Tobin, 2011). However, perception is subjective and difficult to measure based on a unified standard. While several studies have concluded that previous experience of disaster is positively related to risk perception of natural hazards ((Plapp, 2006; Miceli, 2008). We extend the scope of experience and assume that age and education level could also play a role. Previous research also reveals that household preparedness has a positive relationship with family income since income is positively related to access to better and safer housing, low-income households are at greater risk from many hazards (Das, 2018). Therefore,

individual features of age, education level, family income level, and hazard experience have been chosen as part of the input for the prediction model.

For the environmental factors, regions with a long history of natural hazards may have more attention from the government and institutions, which will lead residents in these areas to be more aware of potential natural disasters. Take Japan, a country with frequent earthquakes, as an example, many places in Japan have established special earthquake prevention centers, which mainly to popularize knowledge related to earthquake and first aid methods to the residents, especially primary and secondary school students. Studies also suggest that those residing in chronic hazardous environments are more likely to have disaster experience than those living in an area where only one event had occurred in recent times (Tobin, 2011). On the other hand, a household's access to a community may also matter, as providing information about hazards and associated protective measures will lead to people preparing (Paton, 2009).

Though many factors have been put forward in existing research, there is still a lack of integrative and systematic analysis on how these factors contribute to a household's preparation. In this paper, utilizing data from FEMA's 2018 National Household Survey, we build and compare the performance of four classification models for predicting households' preparedness for natural disasters based on their characteristics and the environment in which they are located. The results create a quantitative relationship between factors household's preparedness and also show which features matter more

Methodology

Classification model

In this study, we trained four widely used classification models: logistic regression, decision trees, support vector machines, and multi-layer perceptron classifier models. The support vector machine uses kernel functions and edge-dependent support vectors to map low-dimensional variables to high-dimensional variable spaces and has strong theoretical foundations and numerous practical successes (Koo, et al., 2019). The decision tree repeatedly splits the data set according to a criterion that maximizes the separation of the data, resulting in a tree-like structure, which is not black-box models and can easily be expressed as rules compared with other machine learning models (Breiman, 1984). Logistic regression and multi-layer perceptron classifier differ from the other two algorithms in the sense that they all need a function form f and parameter vector x to train the model (Dreiseitl,2002). The difference between the two models is that the contribution of parameters in logistic regression (coefficients and intercept) can be interpreted, whereas this is not always the case with the parameters of a neural network (weights)

Data set

The data was collected from the Federal Emergency Management Agency (FEMA)'s National Household Survey (NHS) in 2018. FEMA was formed in 1979 to coordinate the response to a disaster that has occurred in the United States and that overwhelms the

resources of local and state authorities. They have conducted this survey through a telephone interview to assess how personal disaster preparedness and resilience have changed over time in the United States since 2007. Subjects, which includes 5003 adults in the 2018 NHS, in their survey are coming from certain areas of the country that are at higher risk of one of six hazards (Tornado, Flood, Hurricane, Wildfire, Earthquake, Urban Event).

The Survey includes not only the factors that may potentially influence a household's preparedness but also the detailed information of how they prepared for each hazard type, for prediction modeling, in this research we only extracted information of potential factors. Table 1 lists the definition and measurement of each variable. For hazard preparedness, it is reflected in many aspects, such as the financial insurance, documents copied, and suppliers preserved, which, however, are often hard to access and quantify in a uniform metric. The survey provides the stage of preparedness to measure how well the household has been prepared for the disaster regardless of concrete actions. It was a 5-degree classification where 1 means "not prepared and do not intend to prepare in the next year", and 5 means "have been prepared for more than a year and will continue preparing". Personal demographics such as age, education level, and family income are collected as the actual information provided.

For hazard experience, this variable could be explained in many measurements, such as how often the family experienced a disaster in past years or how many natural hazards they have experienced. In addition to looking at whether a disaster has been experienced, the study also analyzed whether the time that has passed since the disaster has any effect on the preparedness. So we used the year of last experience of natural hazard for the measurement of hazard experience. For regional hazard history, we only include the natural hazard, thus using a binary variable to denote whether there is a natural hazard happening in the region where the family is located. For information accessibility, the information mainly involves how to get better prepared for a disaster. A binary measurement is also used here.

Table 1. Definition and measurement of variables

Variable	Definition	Measurement	Data type
Hazard preparedness	How well the household prepared for potential natural hazards.	Stages of preparedness (5 categories)	Category
Age		Code actual age	Numerical
Education level	Household basic information	Highest completed level of education	Category
Family income		Total monthly household income, before taxes,	Category
Hazard experience	Whether the family ever experienced the impacts of a disaster	The year of last experience of natural hazard	Numerical

Regional hazard history	Hazards counted are a tornado, flood, hurricane, wildfire, and earthquake.	Whether there is at least one of these hazards has ever happened.	Category
Information accessibility	Information about how to get better prepared for a disaster.	In the past six months, whether the interviewee has read, seen, or heard any information.	Binary

Exploratory data analysis

For the 5003 samples provided by the survey, we first filter and clear those focusing on an urban event as man-made disaster is not included in the research scope. Initial data processing left 4503 subjects. Also, we delete those points with missing values on variables, which are mainly caused by the answer "do not know" and "refused to answer" responses. In this step, another 2148 subjects have been left out. Finally, 2355 samples are included in our research. To better visualize how these data points are distributed in the whole country, this paper uses postcodes to locate each interviewee on the map. Figure 1 shows that these data points distribute quite evenly over a large part of the county in America which reduces the bias on those places with high hazard frequency.
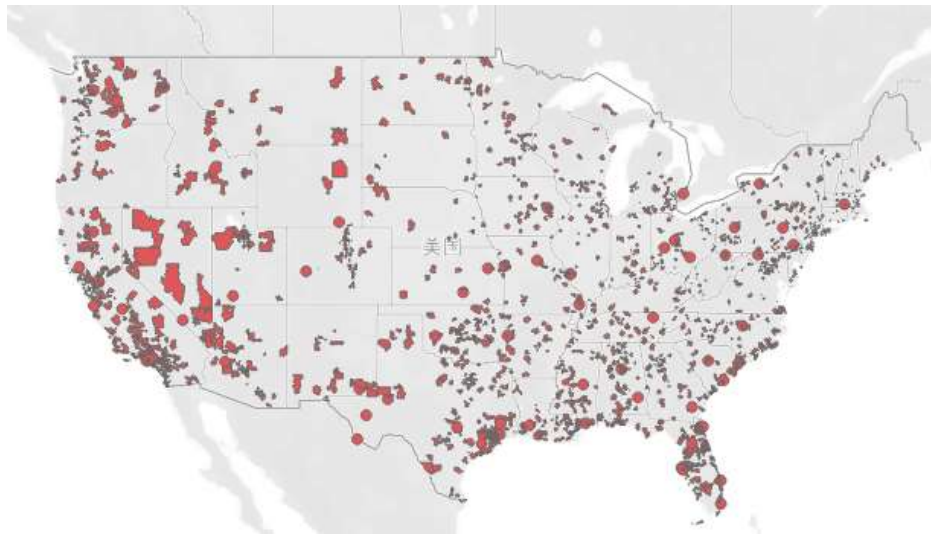

Figure 1 data point distribution in America

For category variables, Table 2 gives the information of their value category and proportion. For hazard preparedness, due to the biased distribution of samples, to build the prediction model, we simplify it into two main categories as prepared and not prepared. Generally, answers with multiple choices distribute quite randomly and no special pattern exists. Figure 2 shows how the numerical data distributes. The individuals who complete the whole interview are all above 18 and kind of following a normal distribution. For the convenience of analysis, samples in hazard experience marked as no are given with a value of 0.1 to it. Though Figure 2(b) shows that most experience seems to have happened in recent years since the interview question is the most recent hazard experience year.

Table 2. Information of category variable value distribution

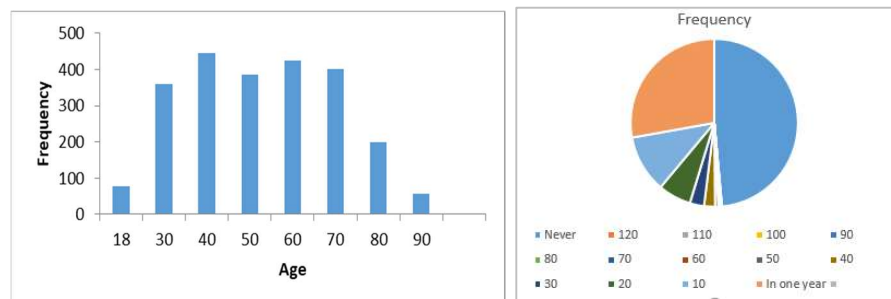| Variable	Category	Number	Percentage
Hazard preparedness	Not prepared	1078	45.77%
	Prepared	1277	54.22%
Interviewee demographics			
Education level	Less than a high school diploma	157	6.67%
	High school degree or diploma	484	20.55%
	Technical/Vocational school	120	5.10%
	Some college	529	22.46%
	College graduate	652	27.69%
	Postgraduate work or degree	413	17.54%
Family income	Under $60	32	1.36%
	$60 to $499	55	2.34%
	$500 to $999	169	7.18%
	$1,000 to $1,999	275	11.68%
	$2,000 to $2,999	279	11.85%
	$3,000 to $3,999	204	8.66%
	$4,000 to $4,999	219	9.30%
	$5,000 to $7,499	401	17.03%
	$7,500 to $9,999	203	8.62%
	$10,000 to $14,999	219	9.30%
	$15,000 to $19,999	69	2.93%
	$20,000 and over	230	9.77%
Environment demographics			
Regional hazard history	No	1045	44.37%
	Tornado	261	11.08%
	Flood	255	10.83%
	Hurricane	252	10.70%
	Wildfire	277	11.76%
	Earthquake	265	11.25%
Information accessibility	No	1124	47.73%
	Yes	1231	52.27%

Figure 2 (a) Distribution of interviewee age; (b) Hazard experience year distribution (Note: the number means experiencing in specific years. For example, '10' means experiencing in last 10 years)

6

Model building

Figure 3 shows the process of model building. First, we perform normalization for numerical data and one-hot representation for category data. Then, noticing that for some variables, the distribution of data is severely skewed, revealing that some imbalance issues have to be solved. With these processed data, we further perform the feature selection to reduce redundant or irrelevant variables. Finally, we input all these data and run the models. Results will be compared through accuracy and F1 scores.
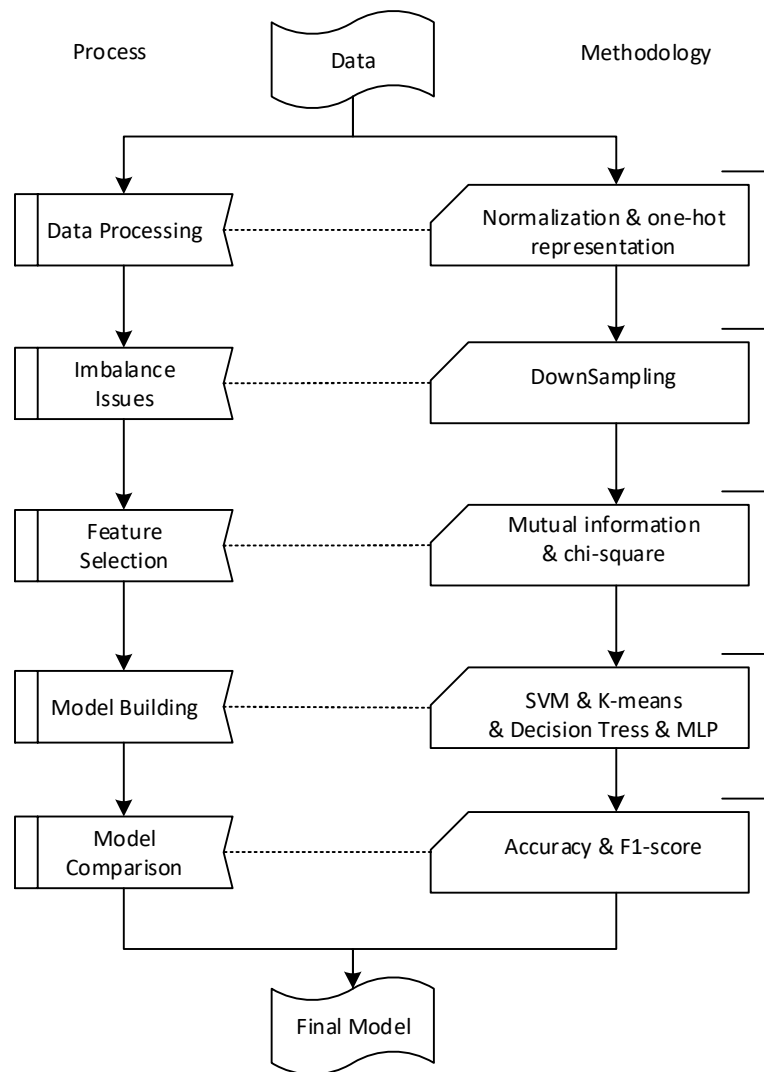

Figure 3 Modeling Process

Data processing

In this step, we perform normalization and one-hot representation for the data. The normalization scales data so that it falls into a small specific range. In some comparison

and evaluation index processing, it is often used to remove the unit limit of data and convert it into dimensionless pure value, so that indexes of different units can be compared and weighed. In this research, the range of numerical data varies a lot due to different measuring units. To fix this problem, normalization is applied to numerical features to make the range of values into [0, 1]. For categorical data, because many machine learning algorithms require all input and output variables to be numeric and cannot operate on label data directly, we need to convert the categorical data to a numerical form. In this study, one-hot representation was used. It encodes N states using n-bit state registers. Each state has its independent register bit, and at any time, it has only one valid point. Samples of data form after the initial process are shown in Table 3.

Table 3. The sample of data after processing

	Number	Preparedness	Age	Education	Family income	Hazard0experience	Hazard history	Information access
0	100001	2	0.679012	3	6	0.998514	0	1
1	100002	1	0.234568	3	4	0.999009	0	1
2	100003	1	0.012346	3	11	0.996533	0	0
3	100004	1	0.703704	3	6	0.000000	0	0
4	100006	0	0.691358	5	10	0.000000	0	0

Imbalance Issues

Imbalanced data typically refers to a problem where the classes are not represented equally due to the skewed nature of data. In such problems, classes have different ratios of specimens in which a large number of specimens belong to one class and the other class has fewer specimens that are usually an essential class but unfortunately misclassified by many classifiers (Ali, 2019). If not dealt with appropriately, it will give us an illusion that the model is good with high accuracy. Note that data imbalance exists in this dataset, typically in the distribution of regional hazard history, where the number of data with a value of no is far more than others. To solve this problem, data can be resampled either using oversampling or downsampling method to construct more balanced data. (add the difference between the two methods, or clarify why downsampling methods are more suitable in our case). In this study, we used the method of downsampling and finally got 1810 samples.

Feature Selection

Variables in this study are put forward through literature review, with values coming from the National Household Survey, which may not all be suitable for the model designing. Since there may be redundant and irrelevant features, we need feature selection to remove those features, to improve the prediction performance of the predictors, and provide faster and more cost-effective predictors. In this study, we perform feature selection based on the results of correlation analysis, where chi-square is used for categorical data and mutual information is used for numerical data. Partial datasets are generated by only using top k

8

(k =1, 3, 5) most correlated categorical features and numerical features for model training (i.e., k categorical features + k numerical features). Then, the model performance of partial datasets is compared to choose the best subsets. The chi-square returns scores and p-values for each variable. The scores are better if greater, while the p-values are better if smaller. According to this metric and the results shown in Table 4, Education and Information access are chosen for further modeling. As for mutual information, it measures the dependency between two variables, and the higher values mean higher dependency. they are better if smaller. Since there is no big difference in the coefficient among the two numerical variables, both variables are kept. After this process, Age, Hazard Experience, Education, and Information access are finally included.

<div align="center">Table 4 Results of feature selection</div>

Feature	Type	Method	Coefficient	P-value
Age	Numerical	mutual information	0.03750	/
Hazard Experience	Numerical	mutual information	0.02130	/
Education	Category	chi-square	25.7431	3.9000
Family Income	Category	chi-square	16.2378	5.5867
Hazard History	Category	chi-square	0.1237	7.2503
Information access	Category	chi-square	76.4180	2.2954

Model Training

In this process, we use these data to train logistic regression, decision trees, support vector machines, and multi-layer perceptron classifier models. To build the model, we use K-Fold cross-validation for experiments. The k-fold cross-validation method labeled data D (of size N) into k equal-sized partitions (or folds). During the ith run, one of the partitions of D is chosen as D.test(i) for testing, while the rest of the partitions are used as D.train(i) for training. A model m(i) is learned using D.train(i) and applied on D.test(i) to obtain the sum of test errors (Tan, 2016). The right choice of k in k-fold cross-validation depends on several characteristics of the problem. A small value of k will result in a smaller training set at every run, which will result in a larger estimate of generalization error rate than what is expected of a model trained over the entire labeled set. On the other hand, a high value of k results in a larger training set at every run, which reduces the bias in the estimate of generalization error rate. In this study, k=5 is applied to the model.

For each model, parameter fine-tuning is further performed. Classification performance is not only affected by the models used but also by their parameter settings. As Table 5 shows, we perform hyperparameters tuned for each model. For logistic regression models, parameters of penalty and C are fine-tuned. Choices of penalty are options of regularization

terms applicable to the classifier, which may improve numerical stability. It also helps to prevent overfitting. C is to control the magnitude of the "actual cost", relative to the regularization term. By applying small values of C, the regularization strength is increased which will create simple models. By applying big values of C, the power of regularization is decreased, increasing model complexity (and potentially overfitting the data). After training, the group ('C': 0.1, 'penalty': 'l2'}is chosen for this model.

Table 5 Parameters fine-tuning of each model

Model	Parameters	Selected group
Logistic Regression	penalty: ['l1','l2'] C: [0.001, 0.01, 0.1, 1, 10, 100, 1000]	{'C': 0.1, 'penalty': 'l2'}
Decision Trees	max_depth: [3, 5, None] min_samples_leaf: [1,2,3,4,5] criterion: ["gini", "entropy"]	{'criterion': 'gini', 'max_depth': 3, 'min_samples_leaf': 1}
Support Vector Machines	gamma: [0.01,0.1,1,10] C: [0.01,0.1,1,10]	{'gamma': 0.01, 'C': 10}
Multi-layer Perceptron	hidden_layer_sizes: [(50,50,50), (50,100,50), (100,)] activation: ['tanh', 'relu'] solver: ['sgd', 'adam'] alpha: [0.0001, 0.05] learning_rate:['constant','adaptive']	{'activation': 'relu', 'alpha': 0.0001, 'hidden_layer_sizes': (100,), 'learning_rate': 'constant', 'solver': 'adam'}

For the decision tree, parameters of criterion, max_depth, and min_samples_split are fine-tuned. Criteria are the function to determine a split. Max_depth and min_samples_split determine the maximum depth of the decision tree and the minimum number of samples required to be at a leaf node respectively. After training, a group of {'criterion': 'gin, 'max_depth': 3, 'min_samples_leaf': 1}is applied in the model. The SVM model has two very important parameters C and Gamma, where C is the penalty coefficient. Generally, the higher c is, the model is less tolerable for error and tends to be overfitting and vice-versa. Gamma implicitly determines the distribution of the data mapped to the new eigenspace. The larger the gamma, the less the support vector, whose number affects the speed of training and prediction. After training, a group of {'gamma': 0.1, 'C': 10} has been chosen for the final SVM model. For multi-layer perceptron classifier, parameters of hidden_layer_sizes, activation, solver, alpha, learning_rate have been fine-tuned, results shows that groups of {'activation': 'relax, 'alpha': 0.0001, 'hidden_layer_sizes': (100,), 'learning_rate': 'constant', 'solver': 'adam'} should be applied in the model.

Model Comparison

We selected several indicators to compare the performance of the four models from two aspects of classifying quality and complexity, including Precision (P), Recall (R), and F-measure (F). Their basic definitions are as follows:

Precision：

$$P = \frac{TP}{TP + FP}$$

Recall：

$$R = \frac{TP}{TP + FN}$$

 F measure：

$$\frac{2}{F_1} = \frac{1}{P} + \frac{1}{R}$$

Among these functions, TP refers to the number of samples that are correctly classified into the positive class. FN represents the number of samples that have been mistakenly classified into the negative category. FP refers to the number of samples that have been mistakenly classified into the positive class. TN represents the number of samples that are correctly classified into the negative class. Precision measures the test results, the recall rate focuses on the sample, and F represents the overall evaluation index.

Table 6 The performance comparison of classification models

Classification model	Precision	Recall	F measure
Logistic Regression	0.6916	0.7592	0.7244
Decision Tree	0.7071	0.7521	0.7042
Support Vector Machines	0.7013	0.6660	0.6829
Multi-layer Perceptron	0.7023	0.8531	0.7386

Table 6 shows the values of each model on indicators. The decision tree performs best on Precision with the value of 70.71%, while Multi-layer Perceptron performs best on Recall and F measures, with values of 85.31% and 73.99% on each indicator. In this study, we want to use the model to predict whether the household has prepared for a disaster to make a more responsive disaster mitigation strategy. Considering the application scenarios, when referring to safety-related issues, we would rather sacrifice some accuracy than miss those who are not prepared. Therefore, in this research, the indicator of recall is more

11

critical. In addition, F measure is a harmonic mean of model precision and recall. It, thus, can also reflect the degree of the model's precision to a certain degree. Given this preference, the Multi-layer Perceptron is selected as the final model.

Conclusion and discussion

The main objective of this study is to find a classification model for predicting a household's preparedness for natural disasters based on their characteristics and the environment in which they are located. To reach this objective, we collected data from the Federal Emergency Management Agency's National Household Survey in 2018, trained logistic regression, decision trees, support vector machines, and multi-layer perceptron classifier models, and compared their performance based on the indicators of precision, recall and F measure. Results show that the multi-layer perceptron classifier model performs best.

The prediction model could be applied in the scenarios of both disaster preparation and disaster response. Since household preparation data are publicly accessible, through the prediction results, the government could grasp information of how residents have been prepared for potential natural hazards and thus carry out a more targeted disaster response strategy. For this scenario, the intermediate result also plays a role. From the feature selection, we know 2 numerical features (hazard experience and age) and 2 category features (education and information accessibility) are finally included in the model. As the selection is based on results of correlation analysis, compared with other features, these included feature matters more on a household's disaster preparation. Though some features such as age and education level could not be changed, the information access, however, could be used well to improve a household's preparation for natural disasters, especially as it ranks the highest value of correlation results with the response variable.

To figure out whether information access could be affected by a household's characteristics, we further performed a chi-square test to evaluate whether there is a significant association between information access and other variables. From the results shown in table 7, we see that the p-value is less than the significance level of 5% for all variables. Like any other statistical test, if the p-value is less than the significance level, we can reject the null hypothesis (H_0: the variables are independent, knowing the value of one variable does not help to predict the value of the other variable) and assume that there is a correlation existing in two variables. This means both households' characteristics and regional hazard history could affect the disaster preparation information they get. while research has revealed that local government and institutions in disaster-frequent regions may attach more importance to hazard education and prepare activities than those that never experience a severe natural disaster, whether the correlation between household's characteristics and information access is significant enough and how it works could be conducted in future research to help make more customized hazard prepare information propagation strategy to improve households' preparedness for a potential natural disaster.

Table 7 Results of chi-square test between information access and other variables

Variable	Age	Education	Family income	Hazard history	Hazard experience
X-squared	105.22	47.943	30.819	56.315	142.2
P-value	0.01222	3.648e-09	0.001177	6.999e-11	7.177e-08

In addition, the model could also be applied for disaster response as existing research shows that preparedness appears to pay off. On the one hand, households with more prepared conditions appear to be cooperative with the regulations disaster mitigation-related measures. On the other hand, families unprepared for a disaster are more likely to get into trouble during a disaster due to lacking emergent knowledge and materials needed, they may need more and faster rescue than those well prepared. Therefore, with a household's disaster preparedness information, governments can make it more efficient when making response strategies towards disaster.

Limitation

In this study, data used for training for the model are collected from FEMA's National Household Survey, thus not being customized for the prediction variables. If possible, future research could be performed based on self-designed surveys to optimize the model. In addition, though with the highest score, the performance of the selected MLP model is still not very good, indicating there may be still some room for parameters optimization to make the model present the best performance, which needs to be continuously adjusted in the following research.

References:

Ali, Haseeb, et al. "Imbalance class problems in data mining: A review." *Indonesian Journal of Electrical Engineering and Computer Science* 14.10.11591 (2019).

Aton D, Smith L, Johnston D. When good intentions turn bad: promoting natural hazard preparedness. Australian Journal of Emergency Management, The. 2005; 20(1):25.

Beatty, T. K., Shimshack, J. P., & Volpe, R. J. (2019). Disaster preparedness and disaster response: Evidence from sales of emergency supplies before and after hurricanes. *Journal of the Association of Environmental and Resource Economists*, *6*(4), 633-668.

Bronfman, N. C., Cisternas, P. C., Repetto, P. B., & Castañeda, J. V. (2019). Natural disaster preparedness in a multi-hazard environment: Characterizing the sociodemographic profile of those better (worse) prepared. *PloS one*, *14*(4), e0214249.

Bronfman, N. C., Cisternas, P. C., López-Vázquez, E., & Cifuentes, L. A. (2016). Trust and risk perception of natural hazards: implications for risk preparedness in Chile. *Natural Hazards*, *81*(1), 307-327.

Das, R. (2018). Disaster preparedness for better response: Logistics perspectives. International journal of disaster risk reduction, 31, 153-159.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, *35*(5-6), 352-359.

Dilley M, ChenU, DeichmannRS, Lerner-Lam A, ArnoldM. Natural disaster hotspots: a global risk analysis. Washington DC: World Bank Publications; 2005.

Jagnoor, J., Rahman, A., Cullen, P., Chowdhury, F. K., Lukaszyk, C., ul Baset, K., & Ivers, R. (2019). Exploring the impact, response and preparedness to water-related natural disasters in the Barisal Division of Bangladesh: a mixed-methods study. BMJ Open, 9(4), e026459.

Khorram-Manesh, A. (2020). Flexible surge capacity–public health, public education, and disaster management. Health Promot, 10(3), 2.

Khorram-Manesh, A., & Burkle, F. M. (2020). Disasters and Public Health Emergencies—Current Perspectives in Preparedness and Response.

Koo, B., La, S., Cho, N.-W., & Yu, Y. (2019). Using support vector machines to classify building elements for checking the semantic integrity of building information models. *Automation in Construction, 98*, 183-194. DOI: 10.1016/j.autcon.2018.11.015

Lindell MK, Perry RW. The protective action decision model: theoretical modifications and additional evidence. Risk Analysis. 2012; 32(4):616–32.

McIvor, D., Paton, D., & Johnston, D. (2009). Modeling community preparation for natural hazards: Understanding hazard cognitions. Journal of Pacific Rim Psychology, 3(2), 39-46.

Miceli R, Sotgiu I, Settanni M (2008) Disaster preparedness and perception of flood risk: a study in an alpine valley in Italy. J Environ Psychol 28(2):164–173

Paton, D., Houghton, B. F., Gregg, C. E., McIvor, D., Johnston, D. M., Bürgelt, P., ... & Horan, J. (2009). Managing tsunami risk: Social context influences on preparedness. *Journal of Pacific Rim Psychology*, *3*(1), 27-37.

Plapp, T., & Werner, U. (2006). Understanding risk perception from natural hazards: examples from Germany. In RISK21-coping with risks due to natural hazards in the 21st century (pp. 111-118). CRC Press.

Tobin, G. A., Whiteford, L. M., Jones, E. C., Murphy, A. D., Garren, S. J., & Padros, C. V. (2011). The role of individual well-being in risk perception and evacuation for chronic vs. acute natural hazards in Mexico. *Applied Geography*, *31*(2), 700-711.

Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Belmont, CA: Wadsworth. International Group, 432, 151-166.