Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

MACHINE LEARNING AND SIMULATION-BASED FRAMEWORK FOR DISASTER PREPAREDNESS PREDICTION

Zhenlong Jiang
Ran Ji

Department of Systems Engineering
and Operations Research
George Mason University
4400 University Dr.,
Fairfax, VA 22030, USA

Yudi Chen
Wenying Ji

Department of Civil, Environmental,
and Infrastructure Engineering
George Mason University
4400 University Dr.,
Fairfax, VA 22030, USA

ABSTRACT

Sufficient preparedness is essential to community resilience following natural disasters. Understanding disaster preparedness of residents in the affected area improves the efficiency and equity of relief operations. This research aims to develop a machine learning and simulation-based approach to predict disaster preparedness using various demographic features from multisource data. The proposed approach comprises four steps: (1) collecting and integrating various data sources, including the FEMA National Household Survey data, US census data, and county-level disaster declaration data; (2) training multiple classification models with the prepared data set and selecting the model with best prediction performance; (3) simulating resident demographic features for at the county level; (4) predicting disaster preparedness status with simulated data for a selected county. A case study is presented to demonstrate the reliability and applicability of the proposed framework.

1 INTRODUCTION

Natural disasters (e.g., hurricanes, storms, and earthquakes) often dramatically disrupt societal functions and destroy critical infrastructure, causing severe economic losses and human sufferings. In September 2018, Hurricane Florence attacked the east coast of America and caused over $17.9 billion in economic damage and 55 death toll (Stewart and Berg 2019). Recovery from devastated communities requires significant money, time, and efforts (Hillier and Nightingale 2013). According to the International Federation of Red Cross and Red Crescent Societies (2015), disaster preparedness broadly refers to any measures taken to prepare for and reduce the effects of disasters, i.e., “to predict and, where possible, prevent disasters, mitigate their impact on vulnerable populations, and respond to and effectively cope with their consequences.” Preparedness plays a vital role in reducing vulnerability and enhancing resilience to natural disasters at individual, household, and community levels. In 2003, the Federal Emergency Management Agency (FEMA) established a campaign called “Ready”, which aims to improve public attitudes toward disaster preparedness (Kohn et al. 2012). Understanding the disaster preparedness of affected communities helps to estimate the community vulnerability and their capacity to cope with an upcoming disaster. Having more knowledge about disaster impacts on affected communities, agencies can improve the efficiency and equity of relief operations through allocating resources to people who are in urgent need of relief aid.

Survey research is the most commonly used approach to investigate the household-level disaster preparedness in a specific region. Kim and Zakour (2017) conducted a survey of 719 elder adults and investigate the relationship between demographic characteristics, social support, community participation, and disaster preparedness through logistic regression. They showed that the income has more significant

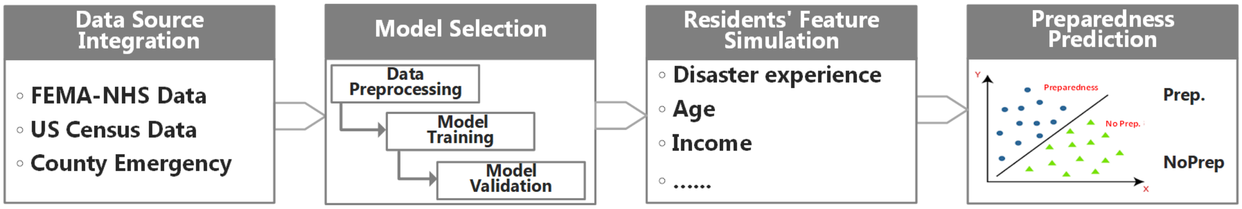
978-1-6654-3311-2/21/$31.00 ©2021 IEEE

impact on disaster preparedness than other factors. Donner and Lavariega-Montforti (2018) surveyed a sample of residents in the Rio Grande Valley, Texas to study the effects of demographic and socioeconomic factors on disaster preparedness. Their results demonstrated that factors, such as age, income, and disaster experience, are significantly associated with the level of disaster preparedness. Maduz et al. (2019) carried out a survey to analyze the effort and motivation of information seeking and preparedness behavior in the German and French speaking parts of Switzerland. Jahan Nipa et al. (2020) investigated the impacts of family income-related demographic information on student perception of disaster preparedness and disaster risk reduction education program. Despite survey research is capable of deriving statistical insights, it is time-consuming and costly to conduct unbiased studies. Moreover, the survey process takes random samples from residents across limited regions, which may not adequately cover affected areas due to the high unpredictability of disasters. Therefore, there exists a necessity to effectively utilize multi-source data (i.e., demographic characteristics and disaster preparedness survey) to measure community vulnerability and predict disaster preparedness status in face of upcoming disasters.

In our previous study, we have proved that the integration of multisource data improves the effectiveness of prediction models for damage assessment (Chen and Ji 2021). Inspired by this, in this study, we propose a machine learning and simulation-based framework to predict the county-level disaster preparedness through integrating various data sources. To be more specific, we (1) compile a data set with various demographic information, including data of FEMA National Household Survey, US census, and county-level disaster declaration; (2) build four classification machine learning models (i.e., logistic regression, support vector machine (SVM), random forest, and XGBoost) to learn the relationship between a series of demographic factors and disaster preparedness; (3) utilize the US census data and governmental disaster declaration data to mimic the demographic factors of affected residents using a simulation approach; and (4) evaluate the out-of-sample performance of the prediction model using the simulated data. The remainder of this paper is organized as follows. Section 2 introduces the proposed framework covering data integration, machine learning procedure, and resident feature simulation. A case study is presented in Section 3 to depict the feasibility and applicability of the proposed approach. Conclusion, contribution, and future work are discussed in Section 4.

2 METHODOLOGY

The structure of the proposed framework is outlined in Figure 1. The proposed framework consists of four modules, including data source integration, model selection, resident feature simulation, and preparedness prediction. First, multiple data sets, including FEMA National Household Survey (NHS) data, US census data, and county-level disaster declaration data, are collected and integrated. Second, machine learning models are selected via a training-validation procedure. Meanwhile, the affected county's resident demographic feature is simulated based on US census data. Finally, with the selected classification model, the resident preparedness of target county is predicted and tested via the simulated demographic data.



Data Source Integration
- FEMA-NHS Data
- US Census Data
- County Emergency

Model Selection
- Data Preprocessing
- Model Training
- Model Validation

Residents' Feature Simulation
- Disaster experience
- Age
- Income
- ......

Preparedness Prediction

Figure 1: The proposed framework for preparedness prediction.

2.1 Data Source Integration

In our study, 2017 and 2018 FEMA National Household Survey (NHS) data are used to train the classification models and validate their performance. For out-of-sample testing purpose, the US census data and disaster declaration data are employed to simulate the features of residents of target county.

The Federal Emergency Management Agency (FEMA) was formed in 1979 to coordinate the response to a disaster occurring in the United States. The National Household Survey is conducted by the FEMA to study the progress in personal disaster preparedness by investigating American public's preparedness actions, attitudes, and motivations via a telephone interview.

In NHS of 2017 (Federal Emergency Management Agency 2019), FEMA interviewed 5042 adults (aged 18 years and older), among which 1006 were randomly selected across the country, and 4036 interviewees were from living areas with a high risk of disasters (including tornado, flood, hurricane, wildfire, earthquake, and nuclear explosion event). In NHS of 2018 (Federal Emergency Management Agency 2020), FEMA interviewed 5,003 adults with 2,000 randomly selected across the country, and 3003 from high-risk disaster attacking areas.

The NHS comprises of three major types of questions asking the interviewee regarding the attitude of hazard preparedness, hazard experience and demographic information. For hazard preparedness, the NHS focuses on seven aspects: attending training, talking with others about preparation, seeking information, developing a household plan, stocking supplies, taking a part in a drill and saving money for emergency and holding property insurance. The interviewee's attitude about hazard preparedness behaviour is categorized into 5 stages: Pre-contemplation, Contemplation, Preparation, Action, and Maintenance. In terms of hazard experience, interviewee's past experience of different types of disasters are survyed. Regarding demographic information, the NHS data covers the interviewee's age, income, gender, education level and ethnicity.

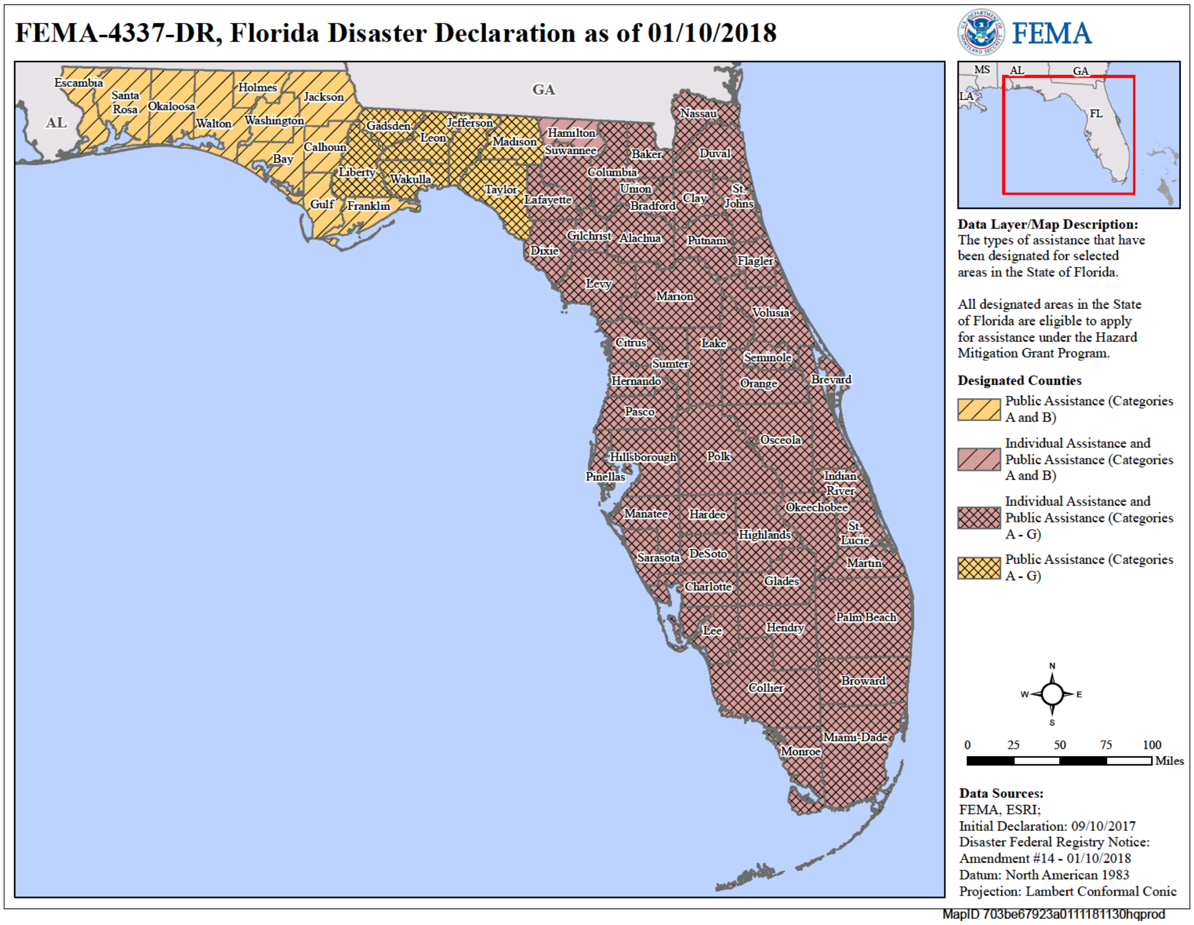
The disaster declaration data is collected from disaster declarations page of FEMA website (Federal Emergency Management Agency 2021). The US President can declare the emergency for any occasion or instance when the President determines federal assistance is necessary. When an emergency status is declared, FEMA can provide two types of assistance (i.e., public or individual) in the affected regions. Figure 2 displays the eligibility of counties in Florida to apply for public or individual assistance during Hurricane Irma.

The US census data set is collected from 2018 American Community Survey Single-Year Estimates constructed by US Census Bureau (U.S. Census Bureau 2019), which contains county-wise distributional information of residents income, age, and education level. The United States Census Bureau (USCB) is a principal agency of the U.S. Federal Statistical System, aiming to produce data about the American population and economy. The USCB conduct the American Community Survey every year to provide many important statistics including language, education, and income for every community in the nation.

In our proposed framework, the machine learning classification model is developed based on 2017 and 2018 NHS data. The 2017 NHS data is used as the training set to build models, while the 2018 NHS data is used as the validation set to tune and select the models. Despite the time-consuming and expensive nature of survey research, it may not adequately cover affected areas due to the unpredictability of the disaster; also, it may not collect enough samples for the affected areas. To mitigate the bias of scarce data, there is a necessity to enrich the NHS survey data by combining other sources to mimic the demographic feature and disaster experience of residents in affected areas. We thus simulate the local resident demographic feature for target counties based on the US census data from American Community Survey and disaster experience using the county-level disaster declaration data.

2.2 Classification Model Selection

It has been shown in existing literature (Kim and Zakour 2017; Donner and Lavariega-Montforti 2018) that demographic information such as income and disaster experience is highly associated with the resident's disaster preparedness. The NHS data contains the interviewee's disaster experience and five demographic



Data Layer/Map Description:
The types of assistance that have been designated for selected areas in the State of Florida.

All designated areas in the State of Florida are eligible to apply for assistance under the Hazard Mitigation Grant Program.

Data Sources:
FEMA, ESRI;
Initial Declaration: 09/10/2017
Disaster Federal Registry Notice:
Amendment #14 - 01/10/2018
Datum: North American 1983
Projection: Lambert Conformal Conic

MapID 703be67923a0111181130hqprod

Figure 2: Florida disaster declaration.

features including age, income, education level, gender, and ethnicity. We will feed these five demographic predictors and 1 disaster experience predictor to build machine learning models to classify the preparedness status (i.e., prepared or not prepared). We employ four widely used classification models, including logistic regression, supporter vector machine (SVM) (Meyer et al. 2019), random forest (Liaw 2018), and XGBoost (Chen et al. 2019) to and select the best model via training-validation procedure.

2.3 Resident Feature Simulation

Notably, the FEMA NHS data is survey data with residents from selected regions, which may not adequately cover affected areas with sufficient data samples. To overcome this, we employ Monte Carlo approach to simulate resident demographic feature based on county-level US census data. The purpose is to mimic the resident demographic information with other available data source. For example, when predicting the preparedness of a target county, the NHS data may lack the necessary information of resident age. We can use the distributional information provided by US census data to generate samples accordingly. The other demographic features can be generated using a similar simulation approach. Each feature is simulated independently. Finally, we create a demographic profile of a resident by randomly combining the individual simulated features. By repeating this process multiple times, we are able to construct a data set for the target county with simulated demographic information. The simulation procedure is conducted as follows.

1. Collect distributional information of demographic features in the target county from USCB website;
2. Simulate resident demographic profiles using US census data;
3. Predict disaster preparedness via simulated data.

3 CASE STUDY

3.1 Data Preparation and Exploratory Analysis

The NHS dataset is only available in 2017 and 2018, and they are highly imbalanced across different counties or states under different types and risk levels of natural hazards. Figure 3 displays two county-level maps with data points in areas affected by Atlantic hurricane season. These two figures show that the NHS only covers most major counties (such as Miami-Dade FL and Harris County TX), most smaller counties are not interviewed.

In our study, we narrow down our attention to six states (i.e., Alabama, Florida, Georgia, Louisiana, South Carolina, and Texas), which are affected by hurricanes from 2016 and 2018, to construct classification models. Filtering the data down to the six states, we have 1263 observations for 2017 NHS and 1240 observations for 2018 NHS. To clean the data, we also remove those observations containing missing values, mainly caused by responses with “do not know” and “refused to answer”. As a result, 1001 samples from 2017 NHS and 724 samples from 2018 NHS are kept to construct classification models. The description of variables is listed in Table 1.

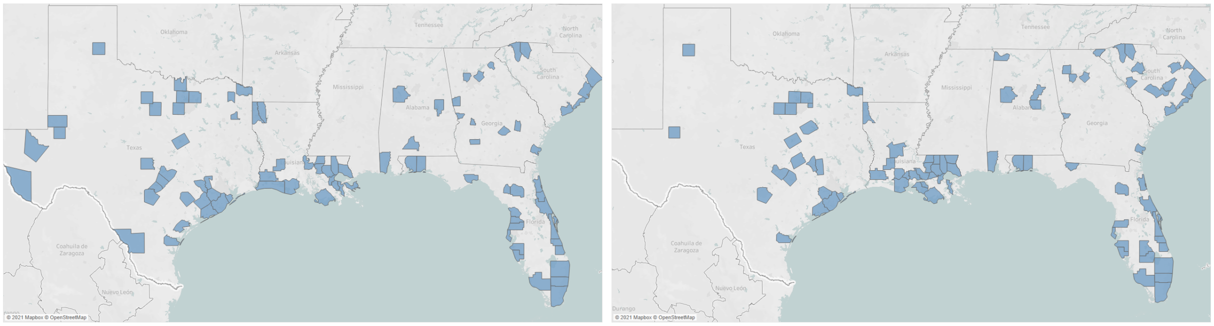


Figure 3: Left: County-level data points location for 2017 NHS. Right: County-level data points location for 2018 NHS.

Table 1: Definition and data type for selected variables

Variable	Variable’s Definition	Data type
Preparedness	How well the interviewee prepared for potential natural disaster.	Binary
Experience	Whether the interviewee or his family member experienced the impacts of a disaster	Binary
Age	Demographic feature: interviewee’s age	Numerical
Gender	Demographic feature: interviewee’s gender	Binary
Education level	Demographic feature: interviewee’s education level	Categorical
Income	Demographic feature: interviewee’s income	Categorical
Ethnicity	Demographic feature: interviewee’s ethnicity	Categorical

3.2 Feature Selection and Model Training

Variables in this study are initially selected based on literature (Kim and Zakour 2017; Donner and Lavariega-Montforti 2018), which has shown that demographic features and disaster experience are correlated with the individual disaster preparedness. By using an independent dataset (NHS data) from the literature, the employed variables may not all be suitable for the model designing. We thus conduct the feature selection to improve model performance in the study. The cleaned 2017 NHS data set is employed as the training set of feature selection. We use best subset selection method to select suitable features. For example, we train logistic regression model for all the subsets of features. Then, we predict the interviewees' preparedness attitudes on 2018 NHS data set by trained logistic regression model and compute the prediction accuracy. For logistic regression model, we find that the model with features age, income, and disaster experience has the highest prediction accuracy. We also find that the older people with high income and disaster experience are more likely to be prepared for a disaster. We follow the same procedure to conduct future selection for SVM, random forest and XGBoost models. All four models have selected age, income, and disaster experience as the significant predictors for disaster preparedness.

After feature selection, we conduct parameter fine-tuning for each model. We use the cleaned 2017 NHS data set as the training set and cleaned 2018 NHS data set as the validation set. During the parameter fine-tuning, we also utilize the prediction accuracy as the evaluation metric. During the parameter fine-tuning process, we train models with different parameter setting and test the performance of each model. The parameter setting with the highest prediction accuracy in testing data set is recorded as the final parameter setting. For example, for SVM model, we consider two different types of kernel functions: linear and polynomial. Then, we design the following groups of parameter setting for SVM: $\{\text{kernel} = linear, \text{cost} = 1\}$, $\{\text{kernel} = linear, \text{cost} = 2\}$, $\{\text{kernel} = polynomial, \text{gamma} = 0.2, \text{cost} = 1\}$, $\{\text{kernel} = polynomial, \text{gamma} = 0.3, \text{cost} = 1\}$, $\{\text{kernel} = polynomial, \text{gamma} = 0.5, \text{cost} = 1\}$, $\{\text{kernel} = polynomial, \text{gamma} = 0.2, \text{cost} = 2\}$, $\{\text{kernel} = polynomial,\ \text{gamma} = 0.3, \text{cost} = 2\}$, and $\{\text{kernel} = polynomial, \text{gamma} = 0.5, \text{cost} = 2\}$. After comparison, the parameter setting $\{\text{kernel} = linear, \text{cost} = 1\}$ that has the highest prediction accuracy is selected. Table 2 shows the candidate values of each parameter for each model. The selected value for each parameter is listed in the last column of Table 2.

Table 2: Potential parameter setting and fine-tuning results.

Model	Parameter	Potential parameter setting	Value
Logistic	Link	logit; cloglog; probit	logit
SVM	SVM-Kernel	linear; polynomial	linear
	cost	1; 2	1
	gamma	0.2; 0.3; 0.5	-
Random Forest	ntree	1500;2000;2500;3000	2500
	mtry	2;3;4;5	3
XGBoost	max depth	2;4;6;8;10	2
	min child weight	1;2;3	1
	eta	0.2;0.3;0.4;0.5	0.3
	objective	binary:logistic; binary:hinge	binary:logistic

3.3 Evaluation Metrics and Model Comparison

We employ Accuracy, Specificity and F1 score to measure the performance of classification models.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{F1 Score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{Specificity} = \frac{TN}{TN+FP}$$

where the TP, FN, FP, TN stands for True Positive, False Negative, False Positive and True Negative, respectively. TP represent the number of observations that correctly classified to the positive class; FN calculates the number of observations which incorrectly identified in the negative category; FP represents the number of observations that incorrectly identified as the positive class; and, TN computes the number of correctly classified observations as the negative class. Precision is the ratio of correctly identified positive observations to the total predicted positive observations. Recall is the ratio of correctly identified positive observations to the actual class's total positive observations. F1 Score is the harmonic mean of Precision and Recall, which is regarded as a better measure of the incorrectly identified cases when facing the unevenly distributed data.

Table 3 shows the value of selected evaluation metrics for the classification models. From this table, the SVM outperforms other three models in terms of all evaluation metrics. Moreover, its specificity value is much higher than the other three models. In addition, the four classification models have similar ROC curve and AUC values, as shown in Figure 4. Through comparisons with all evaluation metrics, SVM model clearly displays superior performance than other three competitors and thus is selected as the best model to predict the disaster preparedness.

Table 3: The comparison of classification models' performance

Model	Accuracy	AUC	Specificity	F1 score
SVM	0.656	70.9%	0.757	0.734
Logistic regression	0.652	70.7%	0.689	0.712
Random forest	0.645	69.6%	0.709	0.714
XGBoost	0.656	69.4%	0.709	0.719

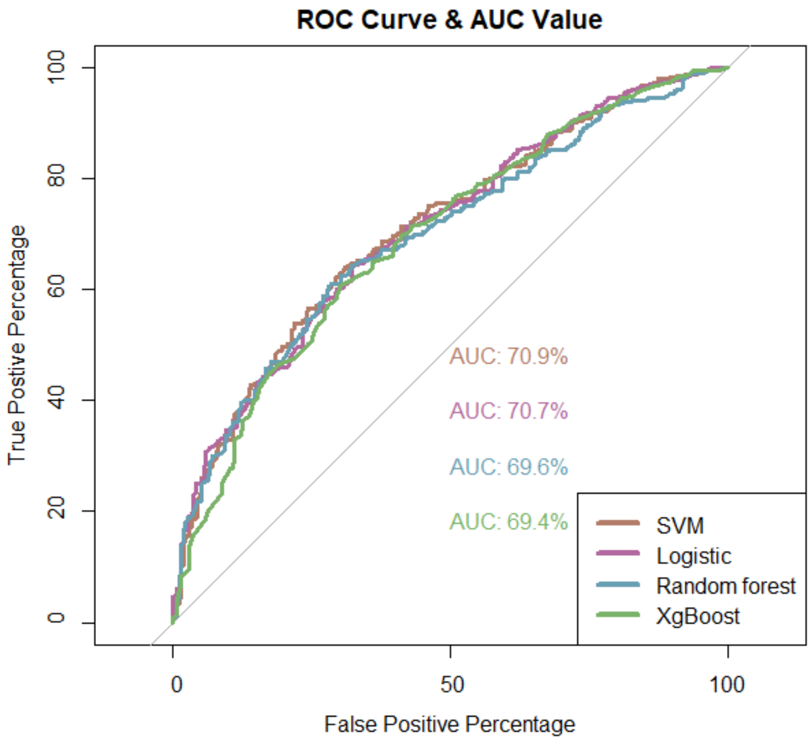


Figure 4: ROC Curve and AUC value.

3.4 Preparedness Prediction for Miami-Dade and Duval County

We select Miami-Dade and Duval County, the two largest counties in Florida, to test the proposed model performance with simulated data. In the cleaned NHS 2018 data set, there are 43 interviewees living in Miami-Dade County, among which 32 have prepared for disasters. There are 18 interviewees in the Duval County, with 14 of them have prepared for disasters. To exam the model performance in Miami-Dade County, we simulate $43 \times 1000 = 43000$ local residents. Each resident has 3 features: income, age, and disaster experience. The income and age are simulated by the income distribution and age distribution obtained from Miami-Dade County's US census data, respectively. The feature of disaster experience for each resident is assumed to follow a Bernoulli distribution with $p = 0.75$. We assume that

$$p = \frac{m}{n}$$

where m is the number of interviewees living in an emergency declared county and report disaster experience in the 2018 NHS cleaned data set, and n represents the number of interviewees living in an emergency declared county in the 2018 NHS cleaned data set. Next, we split the 43,000 residents into 1000 data samples randomly. As a result, we have 1000 simulated data sets for Miami-Dade County. Each data set contains 43 simulated residents with the three features: income, age, and disaster experience. By using the same approach, we also generate 1000 data sets for Duval County with 18 simulated residents in each data set.

We predict the percentage of residents who have prepared for disasters using 1000 simulated data sets. Histograms are plotted based on the predicted results for Miami-Dade and Duval County. Figure 5 shows the distribution of the percentage of residents who have prepared for disasters. The blue dash line represents the percentage of residents who prepare for disasters in the cleaned NHS 2018 data set for each county. The interval between two red dash lines represents the 95% confidence interval for the mean value of the predicted percentage of preparedness. There is large prediction uncertainty at both counties, which proves the necessity of the simulations for understanding disaster preparedness reliably. Previous studies have pointed out that an increased number of samples reduces uncertainty (Ji and AbouRizk 2017; Chen et al. 2019). Consequently, a large number of surveys are recommended in order to reliably understand disaster preparedness. In Figure 5, the percentage of the surveyed residents who have prepared for a disaster in the 2018 NHS cleaned data set belongs to the 95% confidence interval. In sum, our classification model is capable of predicting county-level preparedness in an accurate and reliable manner.

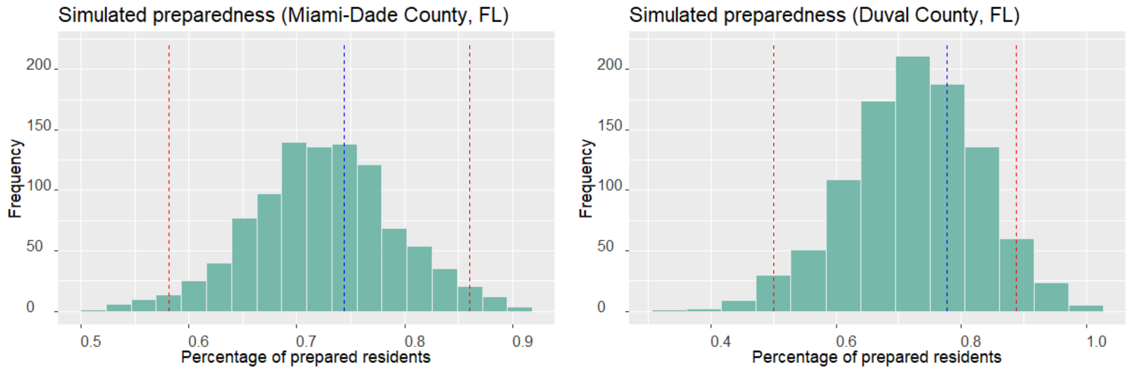


Figure 5: Distribution of predicted percentage of residents who have prepared for a disaster for Miami-Dade and Duval County. Left: Miami-Dade. Right: Duval County.

4 CONCLUSION

In this research, a machine-learning and simulation-based approach is proposed to predict disaster preparedness based on FEMA NHS data and local demographic features. For the case study, we predict resident attitudes of disaster preparedness for Miami-Dade and Duval County, FL based on county-level demographic characteristics. The case study results show that the proposed approach can predict the disaster preparedness in a accurate and reliable manner.

The proposed approach contributes to the academia by developing a machine learning and simulation-based framework for disaster preparedness prediction to overcome the limitation of survey research on disaster preparedness study. For practitioners, this research can assist decision-maker in humanitarian relief organization to estimate the community vulnerability and their capacity to cope with upcoming disasters. In the future, the authors will investigate the data enrichment approach for improved prediction accuracy.

REFERENCES

Federal Emergency Management Agency 2019. “2017 National Household Survey”. https://www.fema.gov/about/openfema/data-sets/national-household-survey, accessed 15.3.2021.

Federal Emergency Management Agency 2020. “2018 National Household Survey”. https://www.fema.gov/about/openfema/data-sets/national-household-survey, accessed 15.3.2021.

Federal Emergency Management Agency 2021. “Declared Disasters”. https://www.fema.gov/disasters/disaster-declarations, accessed 20.3.2021.

U.S. Census Bureau 2019. “2018 American Community Survey Single-Year Estimates”. https://www.census.gov/newsroom/press-kits/2019/acs-1year.html, accessed 15.3.2021.

Chen, T., T. He, M. Benesty, and V. Khotilovich. 2019. “Package ‘xgboost”’. *R version* 90.

Chen, Y., and W. Ji. 2021. “Rapid Damage Assessment Following Natural Disasters Through Information Integration”. *Natural Hazards Review*.

Chen, Y., Q. Wang, and W. Ji. 2019. “A Bayesian-based Approach for Public Sentiment Modeling”. In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 3053–3063. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Donner, W. R., and J. Lavariega-Montforti. 2018. “Ethnicity, Income, and Disaster Preparedness in Deep South Texas, United States”. *Disasters* 42(4):719–733.

Hillier, D., and K. Nightingale. 2013. “How Disasters Disrupt Development: Recommendations for the post-2015 development framework”.

Jahan Nipa, T., S. Kermanshachi, and R. K. Patel. 2020. “Impact of Family Income on Public’s Disaster Preparedness and Adoption of DRR Courses”. In *Creative Construction e-Conference 2020*, 94–102. Budapest University of Technology and Economics.

Ji, W., and S. M. AbouRizk. 2017. “Credible Interval Estimation for Fraction Nonconforming: Analytical and Numerical solutions”. *Automation in Construction* 83:56–67.

Kim, H., and M. Zakour. 2017. “Disaster Preparedness among Older Adults: Social Support, Community Participation, and Demographic Characteristics”. *Journal of Social Service Research* 43(4):498–509.

Kohn, S., J. L. Eaton, S. Feroz, A. A. Bainbridge, J. Hoolachan, and D. J. Barnett. 2012. “Personal Disaster Preparedness: an Integrative Review of the Literature”. *Disaster medicine and public health preparedness* 6(3):217–231.

Liaw, M. A. 2018. “Package ‘randomForest”’. *University of California, Berkeley: Berkeley, CA, USA*.

Maduz, L., T. Prior, F. Roth, and M. Käser. 2019. “Individual Disaster Preparedness: Explaining Disaster-Related Information Seeking and Preparedness Behavior in Switzerland”. Technical report, ETH Zurich.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. Chang, C.-C. Lin, and M. D. Meyer. 2019. “Package ‘e1071”’. *The R Journal*.

International Federation of Red Cross and Red Crescent Societies 2015. “Preparing for Disasters”. fromhttp://www.ifrc.org/en/what-we-do/disaster-management/preparing-for-disaster/.

Stewart, S. R., and R. Berg. 2019. “National Hurricane Center Tropical Cyclone Report Hurricane Florence”. Technical report, National Hurricane Center, Miami, Florida.

AUTHOR BIOGRAPHIES

ZHENLONG JIANG is a PhD student in the Department of Systems Engineering and Operations Research, George Mason University. His research focuses on the data-driven optimization under uncertainty and data analytics, with applications in humanitarian relief and disaster management. His email address is zjiang@gmu.edu.

RAN JI is an assistant professor in the Department of Systems Engineering and Operations Research, George Mason University. His research focuses on data-driven optimization under uncertainty and data analytics, with applications in financial engineering, humanitarian relief and disaster management. His email address is rji2@gmu.edu.

YUDI CHEN is a PhD candidate in the Department of Civil, Environmental & Infrastructure Engineering, George Mason University. His research focuses on the integration of data mining and complex system simulation to enhance infrastructure and social resilience. His email address is ychen55@gmu.edu.

WENYING JI is an assistant professor in the Department of Civil, Environmental & Infrastructure Engineering, George Mason University. Dr. Ji is an interdisciplinary scholar focused on the integration of advanced data analytics and complex system modeling to enhance the overall performance of infrastructure systems. His email address is wji2@gmu.edu.