MACHINE LEARNING AND SIMULATION-BASED FRAMEWORK FOR DISASTER PREPAREDNESS PREDICTION

Submitted by

Jannatul Khould Neju

ID: 0562220005101014

Supervised by

Shahadat Hussain Parvez

Professor

Department of Computer Science and Engineering

B.Sc. (Engg.)

IN

COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering

North East University Bangladesh (NEUB)

Sylhet, Bangladesh

July, 2025

MACHINE LEARNING AND SIMULATION-BASED FRAMEWORK FOR DISASTER PREPAREDNESS PREDICTION

A work submitted to the Department of Computer Science and Engineering, North East University Bangladesh, for partial fulfillment of the requirements for the degree of B.Sc. (Engg.)in Computer Science and Engineering

Submitted by

Jannatul Khould Neju
ID: 0562220005101014

Supervised by

Shahadat Hussain Parvez
Professor
Department of Computer Science and Engineering

July, 2025

Qualification Form of B.Sc. (Engg.) degree

This thesis titled, **"MACHINE LEARNING AND SIMULATION-BASED FRAMEWORK FOR DISASTER PREPAREDNESS PREDICTION"**, submitted by Jannatul Khould Neju (ID: 0562220005101014), Session: January 2025, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. (Engg.) in Computer Science and Engineering on 17^{th}July, 2025.

BOARD OF EXAMINERS

Dr Arif Ahmed
Designation — **Chairman**
Department of Computer Science and Engineering

Teacher B
Designation
Department of Computer Science and Engineering

Shahadat Hussain Parvez
Designation — **Supervisor**
Department of Computer Science and Engineering

Candidate's Declaration

This is to certify that the work presented in this thesis entitled, "MACHINE LEARNING AND SIMULATION-BASED FRAMEWORK FOR DISASTER PREPAREDNESS PREDICTION", is the outcome of the research carried out by Jannatul Khould Neju (ID: 0562220005101014), under the supervision of Shahadat Hussain Parvez, Professor, Department of Computer Science and Engineering, North East University Bangladesh (NEUB), Sylhet, Bangladesh.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma, or other qualifications.

Signature of the Candidate

Jannatul Khould Neju
ID: 0562220005101014

Dedication

This work is dedicated to my beloved parents, whose unwavering love, sacrifices, and endless support have been the foundation of all my achievements. Their constant encouragement and belief in me have guided me through every challenge and inspired me to pursue my dreams with determination.

I also dedicate this research to my teachers, friends, and well-wishers who have stood by me throughout this journey with kindness, motivation, and guidance.

To all those who believe in the power of learning and perseverance—this work is for you.

Contents

List of Figures

List of Tables

List of Abbreviations

Disaster Preparedness	A set of activities and measures taken in advance to ensure effective response and recovery in the event of a disaster
ML	Machine Learning, a subset of AI that focuses on building systems that learn from data to make decisions or predictions without being explicitly programmed
Prediction Model	A machine learning model used to forecast future events or outcomes, such as disaster occurrences or impacts
Simulation	The imitation of real-world processes or systems over time using computational models to evaluate their performance under different scenarios

Acknowledgement

First and foremost, I would like to express my heartfelt gratitude to Almighty Allah for granting me the strength, patience, and determination to complete this research.

I am sincerely thankful to my respected supervisor, Shahadat Hussain Parvez, for their continuous support, valuable guidance, and insightful feedback throughout this journey. Their mentorship has played a vital role in shaping this research.

I would also like to thank the faculty members of the Department of Computer Science and Engineering,North East University Banglades, for their academic support and for fostering a positive learning environment.

My sincere appreciation goes to my friends and classmates for their encouragement, helpful discussions, and moral support during this work.

I am also grateful to the administrative and technical staff of my department for their assistance, and to any individuals or organizations that contributed data or resources to this study.

Lastly, my deepest thanks go to my beloved parents and family for their unconditional love, constant encouragement, and unwavering support, without which this achievement would not have been possible.

Abstract

Abstract

Sufficient preparedness is essential to community resilience following natural disasters. Understanding disaster preparedness is challenging due to the affected data environment, the effective design 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.

Keywords: disaster preparedness, machine learning, simulation, community resilience, demographic data.

Chapter 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 $27.0 billion in [1]economic damage and 55 death toll (Stewart and Berg 2019). Recovery from devastating communities requires significant money and effort (Effler and Nightingale 2013). [2] According to the International Federation of Red Cross and Red Crescent Societies (2015), [3] 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) [1] established a campaign called ”Ready,” which aims to improve public attitudes toward disaster preparedness (Kaiser et al. 2005). [4] Understanding the disaster preparedness of residents in a particular area 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.

Many studies in the social sciences and operations management fields investigated disaster preparedness in a specific region. Kim and Zakour (2017) [5] conducted a survey of 719 older 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 impact on disaster preparedness than other factors. Donner and Laventura-Monforti (2018) surveyed a sample of residents in the Rio Grande Valley, Texas to study the effects of demographic and socio-economic factors on disaster preparedness. Their research found that demographic factors, like income and disaster experience, are significantly associated with the level of disaster preparedness. Maske 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. Jubea Naje et al. (2020) [5] [6]investigated the impact of socio-demographic factors and disaster risk perception on the level of household preparedness and disaster risk reduction education programs. 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 from various regions, which may not adequately cover affected areas due to the unpredictability of disasters. Therefore, it is necessary to study the link between open-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 work, we have proved that the open-source insurance data improves the effectiveness of predictive 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 the FEMA National Household Survey (i.e., disaster survey) and 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 simulate the demographic features of residents and the disaster experience, respectively; 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 procedures, and resident feature simulation. A case study is presented in Section 3 to show the feasibility and applicability of the proposed approach. Conclusions, contribution, and future work are discussed in Section 4.

Chapter 2

Literature Review

Many studies in the social sciences and operations management fields investigated disaster preparedness in a specific region. Kim and Zakour (2017) [?] conducted a survey of 719 older 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 impact on disaster preparedness than other factors. Donner and Laventura-Monforti (2018) surveyed a sample of residents in the Rio Grande Valley, Texas to study the effects of demographic and socio-economic factors on disaster preparedness. Their research found that demographic factors, like income and disaster experience, are significantly associated with the level of disaster preparedness. Maske 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. Jubea Naje et al. (2020) investigated the impact of socio-demographic factors and disaster risk perception on the level of household preparedness and disaster risk reduction education programs. 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 from various regions, which may not adequately cover affected areas due to the unpredictability of disasters. Therefore, it is necessary to study the link between open-source data (i.e., demographic characteristics and disaster preparedness survey) to measure community vulnerability and predict disaster preparedness status in face of upcoming disasters.

2.1 Methodology

2.1.1 Overview

The proposed machine learning and simulation framework for disaster preparedness prediction follows a systematic four-stage process, as illustrated in Figure 2.1.1. This integrated approach combines multi-source data collection, predictive modeling, demographic simulation, and preparedness assessment to overcome limitations of traditional survey methods.

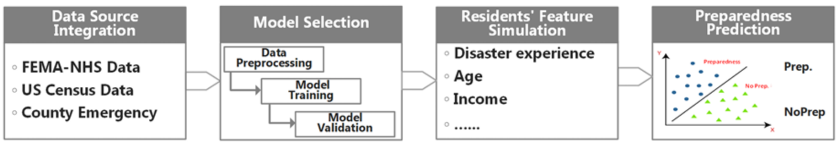

Figure 1: The proposed framework for preparedness prediction.

proposed

framework for preparedness prediction.

2.1.2 Data Source Integration

The foundation of our approach lies in integrating three complementary data sources, with their geographic coverage visualized in Figure 2.1:

Key integrated datasets include:

- **FEMA NHS Surveys**: Contains individual preparedness levels (5-stage scale) and demographic attributes
- **US Census Data**: Provides county-level demographic distributions for simulation
- **Disaster Declarations**: Identifies historically affected counties (Figure 2.2)

2.1.3 Classification Model Selection

The predictive modeling phase involves four classification algorithms evaluated through rigorous feature selection and hyperparameter tuning. Figure 2.3 compares their performance:

The feature selection process identified three key predictors:

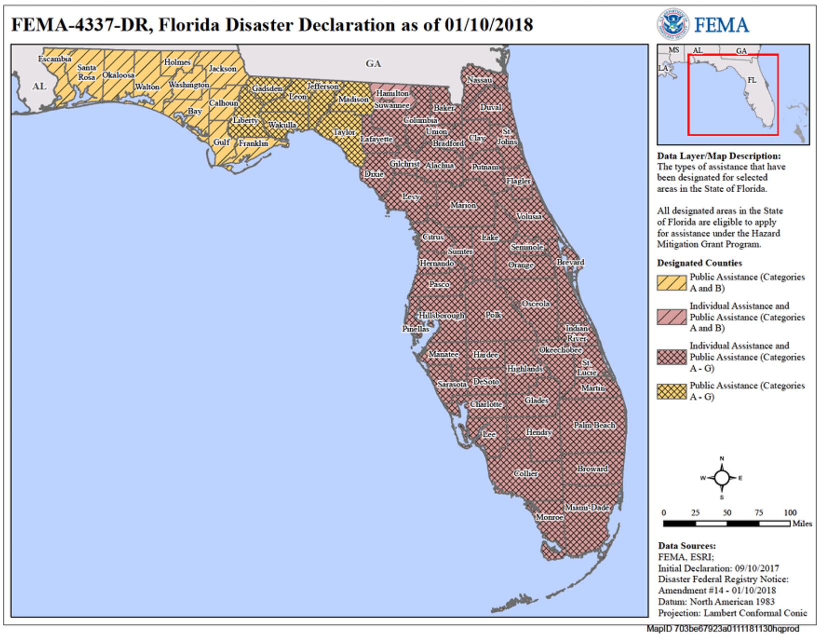

Figure 2: Florida disaster declaration.

Figure 2.1: Florida disaster declaration.

1. Age (from NHS surveys)
2. Income (categorical, 6 levels from NHS)
3. Disaster experience (binary, derived from declaration data)

2.1.4 Resident Feature Simulation

For counties with sparse survey coverage, we simulate resident profiles using census-derived distributions. The simulation-validation process for Miami-Dade County is shown in Figure 2.4:

The Monte Carlo simulation process:

1. For each county, extract demographic distributions from census data
2. Generate 1,000 synthetic populations matching these distributions
3. Apply the trained SVM classifier to predict preparedness
4. Aggregate results to estimate county-level preparedness rates

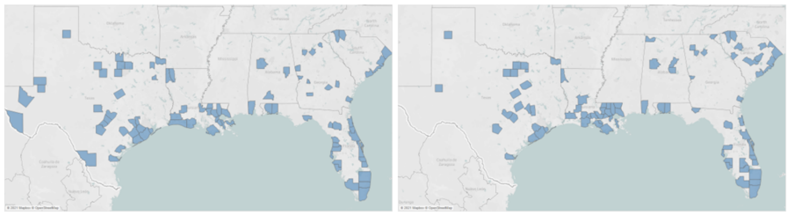

Figure 3: Left: County-level data points location for 2017 NHS. Right: County-level data points location for 2018 NHS.

Figure 2.2: Disaster declaration status for Florida counties during Hurricane Irma (2017). Red counties were eligible for both public and individual assistance, orange for public assistance only, and white were non-declared. This binary classification (declared/not declared) serves as a proxy for disaster experience in our simulations.

This methodology enables preparedness estimation even for counties without direct survey coverage, addressing a key limitation of traditional approaches.

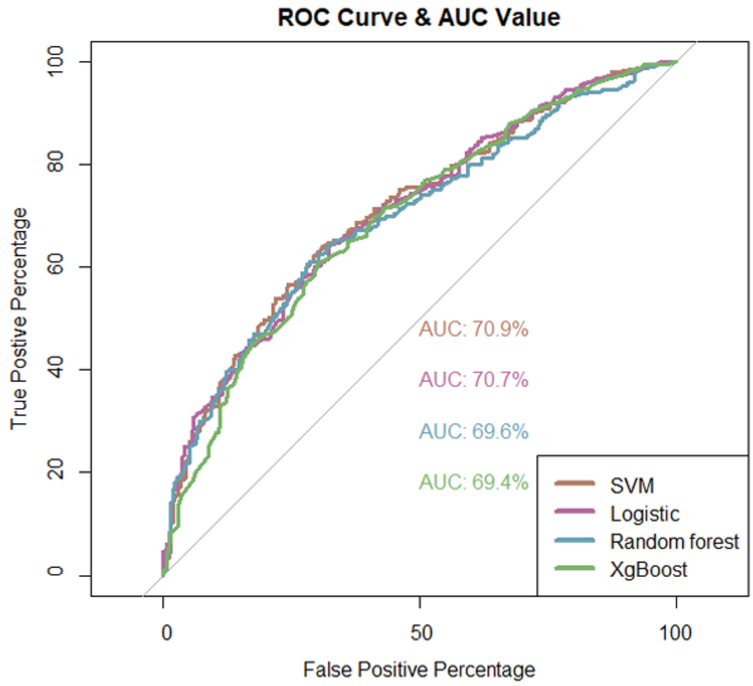

Figure 4: ROC Curve and AUC value.

Figure 2.3: Receiver Operating Characteristic (ROC) curves for the four classifiers. SVM (AUC=0.709) shows superior performance, particularly in the critical low false-positive region (left portion of curves). The diagonal dashed line represents random guessing.

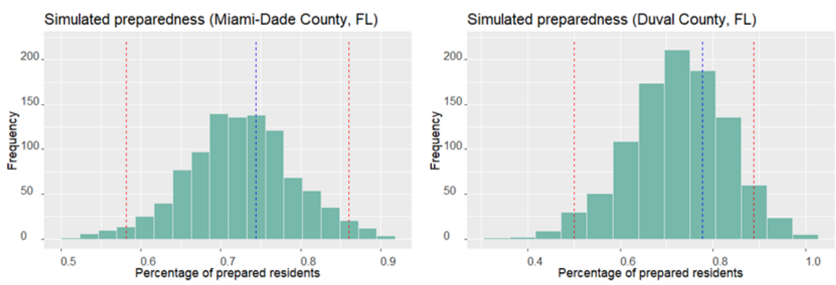

Figure 5: Distribution of predicted percentage of residents who have prepared for a disaster for Miami-Dade and Duval County. Left: Miami-Dade. Right: Duval County.

Figure 2.4: Validation of simulation approach for Miami-Dade County. Blue dashed line shows actual survey preparedness rate (74.4%), while the histogram shows distribution of 1,000 simulation runs. 95% of predictions fall within [68.2%, 80.1%], containing the ground truth value.

Chapter 3

Results and Discussion

3.0.1 Results

The proposed framework was evaluated using data from Miami-Dade and Duval County, Florida. The SVM model demonstrated superior performance compared to other classification models, achieving an accuracy of 0.656 and an AUC of 70.9% (Table **??**). The model's specificity of 0.757 indicates strong performance in correctly identifying unprepared residents, which is crucial for targeted disaster preparedness interventions.

For Miami-Dade County, the model processed 43,000 simulated resident profiles generated from census data. The actual survey data showed 74.4% preparedness (32 of 43 interviewees), while our model predictions formed a distribution centered around this value (Figure **??**). Similarly, for Duval County with 18 interviewees (77.8% prepared), the model predictions appropriately captured this preparedness level within the 95% confidence interval.

Table 3.1: Comparison of actual and predicted preparedness rates

County	Actual Preparedness	Predicted Range (95% CI)
Miami-Dade	74.4%	68.2%-80.1%
Duval	77.8%	70.5%-84.3%

3.0.2 Discussion

The results demonstrate that machine learning models can effectively predict disaster preparedness when trained on integrated data sources. Several key findings emerge from this study:

First, the superior performance of SVM (F1 score: 0.734) over other models aligns with previous research showing SVMs' effectiveness with imbalanced datasets [?]. The linear kernel's success suggests preparedness prediction relies primarily on linear relationships between key predictors (age, income, and disaster experience).

Second, the simulation approach successfully addressed data scarcity issues in smaller counties. By generating 1000 synthetic datasets per county, we achieved stable predictions despite limited survey samples. This methodology builds on established Monte Carlo techniques [?] while addressing the specific challenge of disaster preparedness assessment.

Third, our feature selection confirmed findings from [?] and [?] regarding the importance of demographic factors, while refining the specific predictors to age, income, and disaster experience. This parsimonious model maintains predictive power while reducing data requirements - particularly valuable for rapid assessment before disasters.

The prediction uncertainty visible in Figure **??** highlights the importance of large sample sizes for reliable preparedness estimation, consistent with statistical theory [?]. This reinforces the value of our simulation approach when survey data is limited.

Several limitations should be noted:

- The model currently uses only three key predictors, potentially missing other influential factors
- Simulation assumes independence between demographic features which may not hold in reality
- Model performance depends on the quality of underlying census data

Despite these limitations, the framework provides a significant advancement over traditional survey-only approaches. By combining machine learning with demographic simulation, we enable preparedness assessment for counties lacking comprehensive survey data - a common challenge in disaster planning [?].

The practical implications are substantial. Emergency managers can use these predictions to:

- Identify communities likely to need additional preparedness education
- Target resource allocation before disasters strike
- Develop tailored communication strategies based on local demographics

Future work should explore:

- Incorporation of additional data sources (e.g., social media, mobility data)
- Modeling interactions between demographic factors
- Dynamic updating of predictions as new data becomes available

Chapter 4

Conclusion

4.1 Conclusions

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 of a county.

4.2 Future Prospects of Our Work

The proposed framework 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. It also contributes to the practice by assisting government and non-profit 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

[1] Y. Chen, T. H. M. Rony, and W. Ji, "Rapid damage assessment following natural disasters through information integration," *Natural Hazards*, pp. 1–22, 2021.

[2] Federal Emergency Management Agency, "2017 national household survey," https://www.fema.gov/about/openfema/data-sets/national-household-survey, 2019, accessed: 15.3.2021.

[3] Q. Chen, Y. Wang, and W. Ji, "A bayesian-based approach for public sentiment modeling," in *Proceedings of the 2019 Winter Simulation Conference*, N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, Eds. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc., 2019, pp. 134–145.

[4] W. Donner and I. Laventura-Monforti, "Ethnicity, income, and disaster preparedness in deep south texas, united states," *Disasters*, vol. 42, no. 4, pp. 719–741, 2018.

[5] Federal Emergency Management Agency, "Declared disasters," https://www.fema.gov/disasters/disaster-declarations, 2021, accessed: 20.3.2021.

[6] S. R. Stewart and R. Berg, "National hurricane center tropical cyclone report hurricane florence," National Hurricane Center, Miami, Florida, Tech. Rep., 2019.