In Harm's Way: A Critical Analysis of Machine Learning and Ethical Allocation of Transitional Shelter Assistance

Tara Hinton

2024-12-12

Introduction

In the wake of a presidential disaster declaration, the Federal Emergency Management Administration (FEMA) can aid impacted survivors by providing temporary housing assistance for those without shelter via the Transitional Shelter Assistance (TSA) program under the Individual Assistance (IA) program (FEMA, 2021). This expense reimbursement is essential to directly meeting the needs of communities with structurally vulnerable housing and circumvent the costs of hotels and short-term rentals. In the context of a warming climate, increasing disaster severity poses greater displacement risk to populations in harm's way. In the "twin crisis" of climate change and poor housing infrastructure, efficient and just allocation of FEMA funds is an ever-more pressing issue (Rao & Hornstein, 2023). Currently, TSA eligibility is determined via home damage assessments, a process that can be both costly and time-consuming. This paper examines the methodology, conclusions, and ethical concerns of the study "Machine learning-based FEMA Transitional Shelter Assistance (TSA) eligibility prediction models" by Mahdi Afkhamiaghda and Emad Elwakil, in which the authors develop a supervised machine learning model for TSA eligibility prediction (2021). Their project, which aims to expedite TSA eligibility classification, raises numerous ethical concerns about equitable allocation of disaster aid and the automation of disaster decision-making. In this paper, I critique and apply Afkhamiaghda and Elwakil's methodology to predict Rental Assistance Eligibility, part of the IA program, and provide an analysis of normative concerns.

Analysis of Methods

The authors initially formulated three models (logistic regression, decision tree, and KNN) to predict whether or not an applicant was TSA eligible. This study uses the National Emergency Management Information System database, which shows 4.8 million registries of Individual Aid applications for survivors between 1998 and 2017. This study's final dataset, cleaned of outliers and normalized using a min-max scaling technique, considers eight predictors (*Roof damage amount*, *Household composition*, *Destroyed* (binary), *Residence type*, *Repair amount*, *Flood insurance* (binary), *Special needs* (binary), and *Foundation damage amount*). Min-max scaling is critical for models like KNN, SVM, and logistic regression, because it scales all features between 0 and 1 to ensure equal contribution to the models. An accuracy, sensitivity, and precision confusion matrix indicated that supervised machine learning techniques like KNN and decision trees yielded better results than logistic regression, with KNN having the most consistent accuracy overall (70%). Accuracy indicates how often a classifier correctly predicts the outcome. Sensitivity shows the rate of correct predictions when the actual value is positive, while precision indicates how often prediction is correct when a positive value is predicted. All three models have low sensitivity, meaning that they often under-predict the number of individuals who have received transitional shelter aid. Despite this, KNN models were found to have significantly higher sensitivity (16.80%) than logistic regression (2.98%) and decision trees (1.30%). The author's logistic regression model demonstrates a precision score (70.73%) than both the KNN and decision models (43% and 45% precision, respectively), so this model may consistently under-predict

when participants receive TSA. My novel analysis applies the authors' methods of logistic regression and decision trees to predicting Rental Assistance eligibility.

Novel analysis

Data pre-processing

Data cleaning occurred in several phases, and computer processing power proved to be the primary barrier to a more comprehensive data analysis. Valid IA registrations from 2019-2020 first yielded 1.05 million observations, so additional cleaning was undertaken. Following the authors' methods, Florida, NC, and Texas were selected due to their high concentration of disaster declarations, and any rows containing all null or zero values were deleted. In methods described by Afkhamiaghda and Elwakil, a random sample of 200,000 observations from the dataset ensured maximized computational speed and model performance, while min-max normalization techniques allowed for scaling across variables. TSA was not provided to any of these 1.05 million registrants between 2019-2020, so a proxy, Rental Assistance eligibility was used. Rental Assistance provides money, rather than shelter, for disaster survivors to rent a new place to live for up to eighteen months after the disaster. While TSA is notably a more short-term program, Rental Assistance extends the authors model to different dimensions of post-disaster housing.

Logistic Regression and Decision Tree

Novel analysis for logistic regression and decision trees followed the methods of Afkhamiaghda and Elwakil, engaging first in feature importance analysis for each model. The same variables were significant under a chi-squared test at levels of 0.001, including: Roof damage amount (roof damage amount observed by FEMA), Household composition (number of individuals living in a household at the time of damage), Destroyed (Is structure permanently uninhabitable?), Residence type Damaged (dwelling residence type), Repair amount (amount of repair assistance in dollars), Flood insurance (Does the applicant have flood insurance?), Special needs (if the applicant requires special accommodations to use FEMA assistance), and Foundation damage amount.

After cross-validation using AIC, logistic regression found the combination of all of these variables to be the best. The confusion matrix (Table 1) and accuracy, sensitivity, and precision metrics (Table 2) are listed. A decision tree was executed and pruned using cross-validation (folds = 10, best = 9) to decrease the potential for overfitting, which resulted in a final model with residence damage amount as the sole important predictor. The confusion matrix (Table 1) and accuracy, sensitivity, and precision metrics (Table 2) are listed.

Table 1: Table 1. Confusion Matrices for Logistic Regression and Decision Tree Models

Model	Actual 0, Predicted 0	Actual 0, Predicted 1	Actual 1, Predicted 0
Logistic Regression - Hinton	961132	5041	73206
Logistic Regression - Afkhamiaghda and Elwakil	36165	2225	22796
Decision Tree - Hinton	36165	2225	638
Decision Tree - Afkhamiaghda and Elwakil	22796	192	10966

	Table 2. Metrics Comparison		
Metric	Model	Hinton Model	Afkhamiaghda and Elwakil Model
Accuracy (A)	Logistic Regression	92.53778	65.97
Precision (P)	Logistic Regression	30.40350	70.73

Sensitivity (S)	Logistic Regression	60.37267	2.98
Accuracy (A)	Decision Tree	92.84250	68.83
Precision (P)	Decision Tree	30.40350	43.03
Sensitivity (S)	Decision Tree	60.37267	1.30

KNN

Though the authors implemented an additional KNN machine learning model, this analysis critiques, rather than applies, the techniques outlined in the authors' paper. KNN is a classification technique, whereby all data is classified based upon its proximity to a K-number of nearest data points (Afkhamiaghda & Elwakil, 2021) . KNN is most relevant when the sample size for data training is small. The accuracy of the KNN model increases as K increases until it reaches a point of diminishing returns with computational cost. Often, the K value is calculated by taking a square root of the number of data points, ensuring that the result is odd to avoid 'ties' in classification. It is not clear which K value the authors used in this paper as they do not directly state this value. As the number observations in our training test is high, cross-validation should be used to find the preferred K value.

Distance metrics help the model determine which neighboring points are nearest. The authors use Euclidean distance (ED), a variation of the Pythagorean Theorem, in order to find the distances between the new input and existing data points and is a common distance metric. This choice is inappropriate to the dataset, which contains many binary observations. Using ED in KNN on data with binary variables will often lead to too many 'ties' for nearest neighbors and a model failure. I would encourage the authors to use the Hamming distance, which calculates the number of positions at which data differs and is thus well suited for binary data.

Normative Considerations in ML Disaster Allocation

By developing and assessing logistic regression and decision tree models to predict TSA eligibility, this study introduces critical ethical questions about the allocation of aid during the climate crisis. In 2006, the Government Accountability Office released a report stating that an estimated 600 million to 1.4 billion dollars had been spent on improper individual assistance (IA) payments (GAO, 2007). Currently, FEMA relies upon expert knowledge and home inspection to determine eligibility for TSA funds, a process often rife with racial and socioeconomic bias (Afkhamiaghda & Elwakil, 2021; FEMA, 2021). In addition to redlining, studies suggest that FEMA home appraisals have historically undervalued homes in Black communities, impacting the quality and amount of aid that homeowners may receive (FEMA, 2021). Further, pre-existing home damage in low-income settings may decrease the extent of damage that inspectors award post-disaster (FEMA, 2021). These historical disparities provoke compelling questions about the equitability of machine learning models given historical FEMA data. applications of machine learning in moments of disaster must consider principles of distributive justice. In this section, I will analyze the deontological considerations of applying ML to IA funding and argue that significant equity considerations should be made before it is moral to use models like Afkhamiaghda and Elwakil's for decision-making.

Machine learning algorithms run the risk of treating applicants as merely data points, or means to an end. Kantian deontology upholds humans as moral agents and ends in and of themselves. When we formulate prediction models for post-disaster aid, we use the data points gathered from millions of impacted individuals. These 'data points', then, are deeply personal to each individual and should be treated with as much compassion and dignity as possible. Afkhamiaghda's and Elwakil's models aim to generate more efficient TSA allocations, using the individual's experience of disaster as a means to get there. In this iteration of their project, there is little consideration of regional or tract-level variations in TSA – the sole purpose seems to hinge on protecting the financial pot of FEMA. In this way, individuals may become party to an algorithm that may harm them, while receiving little benefit in return. A deontological perspective would

require machine learning applications that aim to honor human dimensions by considering the variations and equity differences in aid allocation.

Further, deontological values of fairness and justice saturate the use of ML models to predict and disburse aid. While TSA eligibility models might help decision-makers expedite aid disbursement post-disaster, these models may be rooted in data that disadvantaged minority and low-income communities, raising concerns about inequality exacerbation. Other machine learning models, such as COMPAS, have been heavily critiqued for reproducing social inequalities in decision outcomes. These models may feed in data like recidivism rates that are impacted by systemic inequalities. Similarly, FEMA data historically exhibits preference for white, wealthy homeowners– so it is imperative that we question the statistical and ethical performance of this study. For example, this analysis found a negative relationship between mobile home residence types and receiving assistance from FEMA's IA program.

Despite these critiques, ML offers great benefits to emergency management. In alignment with deontology's categorical imperative, models like Afkhamiaghda and Elwaki's apply a formulaic rule to fund dispersal. The application of ML models could standardize and ensure consistency in FEMA, while ensuring that IA funds are applied in consistent manners. Similarly, prediction might help reduce political bias in the decision-making process, providing data-backed suggestions for where aid should be provided. These positive attributes may help justify the use of ML; however, FEMA must make considerable strides toward incorporating equity in these models.

Over time, due to population growth and climate change, the need for post-disaster temporary shelter is projected to grow (Afkhamiaghda & Elwakil, 2021). The low sensitivities of Afkhamiaghda's and Elwakil's models combined with climate change may point to the insufficiency in utilizing solely FEMA data to allocate aid for future needs in an increasingly disaster-prone world. As we elevate machine learning to solve bias and inefficiencies in aid allocation, the broader automation of disaster decision-making challenges our best and worst moral instincts. Should a need-based system prioritize allocations of aid based upon biased historical data? What does equitability look like in a disaster zone? On one hand, implementing machine learning in TSA allocation might free homeowners from the differential burden of poorly-trained and stressed damage inspectors. Conversely, on-site home visits allow damage inspectors to get a more complete picture of homeowner losses. While the authors of this paper tend toward a utilitarian framework of maximizing economic efficiency, applications of machine learning in moments of disaster must consider principles of distributive justice. Ultimately, a deontological answer dictates that the predictive models minimize harm to individuals who need aid most.

Conclusion

Impact of Paper

Though preliminary, Afkhamiaghda and Elwakil's paper demonstrates the theoretical utility of machine learning to Transitional Shelter Assistance allocation. Their novel application of machine learning in an emergency management context establishes a framework for triaging and predicting disaster needs quickly. The critical impact of this paper extends to forms of aid beyond TSA, including Rent Assistance Eligibility, under FEMA's IA (Individual Assistance) program. With refinement, their machine learning approaches may bridge the mismatch between allocated, wasted, and critical FEMA spendings, providing checks to unnecessary spending during disaster.

Future Work

Afkhamiaghda and Elwakil's study prompts interesting questions for the future and role of machine learning in disaster relief. First, FEMA's current data format and unavailability limits the amount of work that researchers can do in this field. Large numbers of missing fields in the FEMA IA dataset may skew any statistical analyses, while the massive quantity of data conversely limit the utility of some machine learning

approaches like KNN, which require high computation power to execute at bigger scales. Future machine learning work should apply alternative methods to analyses and predictive modeling, such as random forest, and use machine learning for database management. Conversely, as historical data used to train predictive models do not account for climate change in predictions, machine learning may significantly underpredict aid allocated in the future. Researchers should incorporate the impact of climate change on the frequency and severity of disaster conditions in the U.S. to ground predictions in the real disaster landscape.

Second, it is critical that future research considers the utility of models to the needs and concerns of everyday people living in a changing climate. While questions of whether or not someone will receive government aid after disaster are important, how much aid is given is often an equally pressing question for society (Hersher & Kellman). Machine learning techniques can help answer these 'how much' questions in the face of future disasters by further predicting amounts of rental, small business, and housing assistance.

Afkhamiaghda and Elwakil's paper makes the first step toward this future, but requires significant statistical refinement. Deontological principles of justice and human agency caution us away from relying on these early types of ML for disaster aid allocation, but future work may improve upon the statistical and moral merit of these early models.

References

Afkhamiaghda, Mahdi & Elwakil, Emad. "Machine learning-based FEMA Transitional Shelter Assistance (TSA) eligibility prediction models," *Journal of Emergency Management*, (2021): 10.5055/jem.2020.0000.

FEMA. "FEMA: Transitional sheltering assistance." Available at https://www.fema.gov/transitional-shelter-assistance. Accessed October 25, 2024

Hersher, R., & Kellman, R. (2021, June 29). Why FEMA Aid Is Unavailable To Many Who Need It The Most. NPR.org. https://www.npr.org/2021/06/29/1004347023/why-fema-aid-is-unavailable-to-many-who-need-it-the-most

Kutz GD, Ryan Hurricanes Katrina and Rita disaster relief: Prevention is the key to minimizing fraud, waste, and abuse in recovery efforts, statement of Gregory Kutz, Managing Director Forensic Audits and Special Investigations, Testimony before the Committee on Homeland Security and Governmental Affairs, US Senate. United States Government Accountability Office, no. GAO-07-418T. United States Government Accountability Office, 2007.

Rao, Arya, & Shira Hornstein. "How The Twin Crises of Climate Change and Poor Public Housing Are Harming People's Health." STAT, August 29, 2023. https://www.statnews.com/2023/08/29/climate-change-public-housing-health-consequences/#:~:text=Millions%20of%20people%20are%20being,individual%20health%20of%20hom