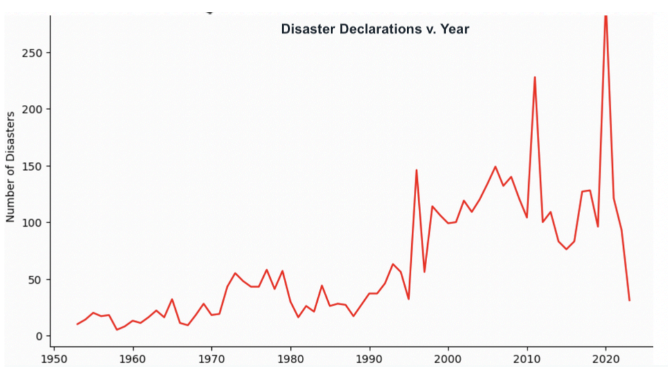
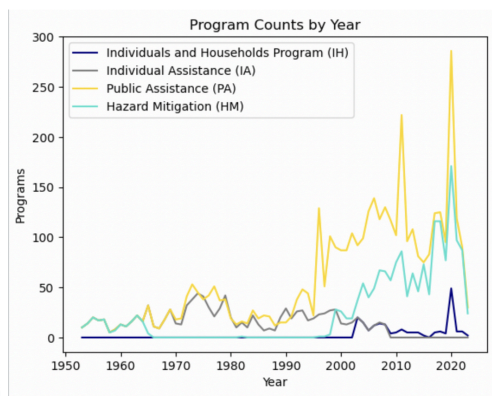
Predicting FEMA Disaster Declarations

By Austin Cohen, Caden EspindaBanick, and Kobe Rose

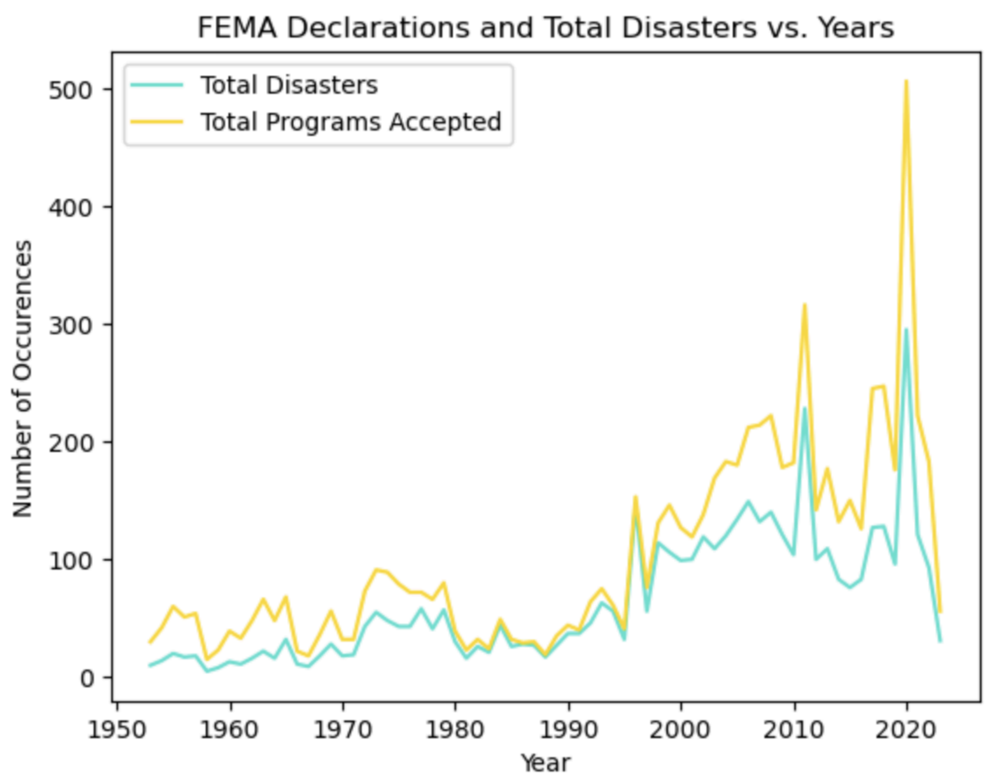
Natural disasters are occurring more frequently and with more intensity than ever before, resulting in significant damage to American infrastructure. When disasters strike, the Federal Emergency Management Agency (FEMA) assumes responsibility for delivering relief to impacted communities and individuals. FEMA provides disaster relief through its four main programs: Individual and Household which provides financial and/or direct assistance to eligible applicants; Individual Assistance which offers assistance such as grants for temporary housing, home repairs, and other disaster-related expenses to individuals and households affected by disasters; Public Assistance which provides grants to governments and certain nonprofit organizations to assist with response and recovery efforts following a disaster; and Hazard Mitigation which aims to reduce the risk and impact of future disasters by implementing measures to mitigate or prevent loss of life, property damage, and economic disruption.





Not every disaster necessitates the utilization of FEMA's relief programs. FEMA disaster declarations rely on Governors declaring a state of emergency for their state, which can be politically daunting. Our project aims to aid Governors in the political process by predicting what FEMA programs will be utilized for a declaration.

Our original data set, managed by [FEMA] (posted on Kaggle), contained 4738 entries and 31 columns of county-level data of all federally declared disasters in the US from 1953 through 2023. Each row contained a unique disaster with its corresponding state. Some columns we included were; state demographic data (log-population, income, area, etc.); FEMA data (date and type of disaster); and one-hot encoded independent variable columns indicating if FEMA provided funding for “Individuals”, “Public”, or “Both”. Finally, we split our data set into 80% training data, 10% validation, and 10% test. Because individual funding only occurred 6% of the time, we oversampled our training data to make each variable occur 33% of the time.



We wanted models that were either interpretable or as accurate as possible. Beginning with interpretability, we created a decision tree with a max depth of three. Every end node in our decision tree with a declaration date after 2003 predicted “public” declarations while pre-2003 declarations had all three categories in its end nodes. Secondly, according to our model, hurricanes and fires increase the probability of “public” declarations. The validation accuracy of our decision tree was 77%, 5-6% lower than our accuracy-focused models. Interestingly, the decision tree’s validation accuracy was better than its training accuracy. This is likely due to the oversampling of the “individual” and “both” categories in our training data and a relative strength predicting “public” declarations which made up a higher percentage of the validation data.

Turning to our accuracy-based models we began with four models, random forest, bagging, KNN, and gradient boosting, to select the model with the highest validation accuracy. We manually integrated through hyperparameters, attempting to maximize validation accuracy while minimizing the overfitting (determined by the difference between training and validation accuracy). Our KNN model had the best training accuracy (89.44%) but regardless of the hyperparameters severely overfit the model. This may be because, unlike the decision tree, the KNN model was best at predicting “individual” and “both” declarations which were underrepresented in the validation data. Our gradient boosting model had the best validation accuracy (83.92%) and the second-best train accuracy (88.68%). We decided to move forward with this model on the test data, resulting in a test accuracy of 79.89%. Finally, we created a confusion matrix for our final model on the test data, showing that “both” was predicted with 76% accuracy, “individual” with 62.5% accuracy, and “public” with 82% accuracy. Given the significant underrepresentation of some of our categories, we were satisfied with the models' robust prediction capabilities.

We had two major findings. First and unsurprisingly, disaster declarations are increasing as are the number of declarations per disaster. Secondly, year, log-population, and disaster types (namely fire and hurricane) were all variables of importance in predicting different declaration schemes. Recent disasters were more likely to receive public-only funding grants, while fires and hurricanes tended to decrease the likelihood of Individual Assistance declarations. Turning to future improvements, using our model to predict the future may be unwise because of the recent relative uptick in “public” declaration schemes. We attempted to do our training, validation, and test split based on year but there were not enough samples of individual-only declarations. A future improvement could be splitting based on year and using more drastic oversampling methods. This may work better in the future where recent years are not dominated by COVID-related declarations.

Name	Proposal	Coding	Presentation	Report
Austin Cohen	1	1	1	1
Caden EspindaBanick	1	1	1	1
Kobe Rose	1	1	1	1