

Team No. 5
Professor: Dr. Amir Jafari
Topic: Natural Disaster Damage Prediction
Date: 1st November 2021

1. What problem did you select and why did you select it?

A new research field emerged in climate science in the early 2000s that wanted to explore the increasing prevalence of extreme weather events like floods, storms, cyclones, etc. The field is known as "extreme event attribution" and has gained momentum in recent years in media in addition to the scientific world. There is mounting evidence that human activity is to blame for the increased risk of these extreme weather-type events. Researchers have also given importance to analyzing the economic costs linked to the human contribution to weather events. A study in 2020 approximated that nearly \$67bn of damages caused by Hurricane Harvey in 2017 could attribute to human influences on climate. There are numerous methods to carry out attribution analysis. One way is to record instances of an extreme weather event and see their frequencies change with changes in environmental factors. We aim to build a model that accurately predicts the estimated damage to property while considering various event-related factors, in addition to external factors that might be influencing the extent of the damage.

2. What database/dataset will you use? Does it need to be cleaned?

- **Source:** [NOAA \(National Oceanic and Atmospheric Administration\)](#)
- **Data Availability:** January 1950 to August 2021
- **Number of Observations:** 1,710,146
- **Number of features (before EDA):** 51
- **Number of features (post EDA):** 189
- **Target Variable:** DAMAGE_PROPERTY

We'll do a cleaning, transformation, and feature engineering.

3. What data mining algorithm will you use? Will it be a standard form, or will you have to customize it?

- a. Linear Regression
- b. Random Forest Regressor
- c. XGBoost Regressor
- d. Ensemble Model – Voting Regressor

4. What packages will you use to implement the model? Why?

- a. Pandas: Importing data and manipulation
- b. Sklearn: For ML model building, training, and prediction
- c. Matplotlib and Seaborn: For plotting and graphs

5. What reference materials will you use to obtain sufficient background on applying the chosen model to the specific problem that you selected?

1. <https://www.carbonbrief.org/mapped-how-climate-change-affects-extreme-weather-around-the-world>

2. <https://towardsdatascience.com/a-quick-and-dirty-guide-to-random-forest-regression-52ca0af157f8>
3. <https://www.oreilly.com/library/view/tensorflow-machine-learning/9781789132212/d3d388ea-3e0b-4095-b01e-a0fe8cb3e575.xhtml>
4. <http://en.wikipedia.org/wiki/Storm>
5. [sklearn.linear_model.LinearRegression — scikit-learn 1.0.1 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)
6. <https://towardsdatascience.com/why-1-5-in-iqr-method-of-outlier-detection-5d07fdc82097>
7. <https://machinelearningmastery.com/principal-component-analysis-for-visualization/>
8. <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html>
9. https://scikit-learn.org/stable/auto_examples/ensemble/plot_voting_regressor.html

6. How will you judge the performance of your results? What metrics will you use?
The result will be measured based on regression error metrics such as:

- a. R-squared
- b. Mean Squared Error

7. Provide a rough schedule for completing the project.

Topic Proposal: 11/1/2021

EDA Analysis: 11/7/2021

Data Preprocessing: 11/20/2021

Model Selection: 11/25/2021

Dashboard creation: 11/30/2021

Final Project Report: 12/5/2021