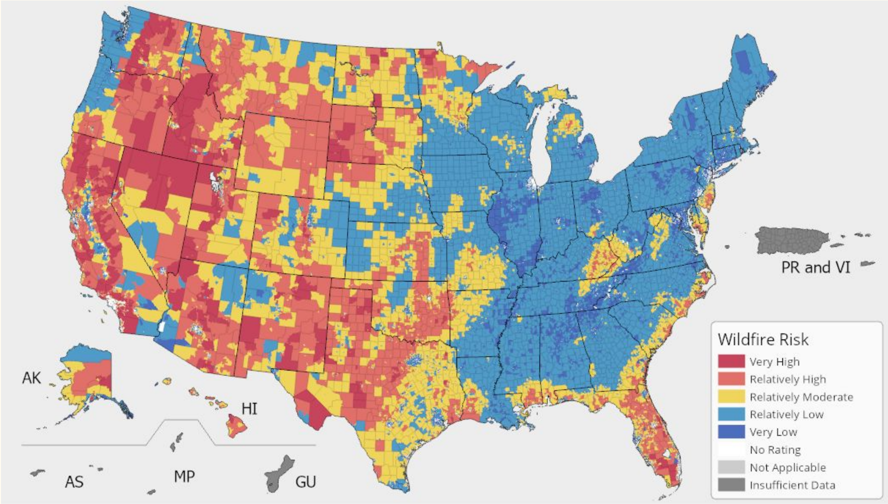
04.21.25

Wildfire Structural Damage Prediction

By Kevin Specht, Evan Ferguson, Ruichen Kong, Andres Barei, & Kiana Burton

Overview

- Wildfires are increasingly common around the country, causing massive property damage.
- Our goal is to use structural and location data of buildings to predict whether those buildings will be destroyed by wildfires.
- Our KPI (Key Performance Indicator) is accuracy of predictions for whether a building impacted by a wildfire will be destroyed or not.



Stakeholders

- City planners, building inspectors, and other government workers determining how to make buildings and cities more durable against wildfires.
- Real estate and insurance agents appraising a building's value in the context of wildfire risk.
- Homeowners and renters determining their building's level of safety against wildfires.
- Emergency workers determining how wildfires are likely to impact the zones they're operating in.

Dataset

- California Department of Forestry & Fire Protection (CAL FIRE) promotes healthy forests, improves wildfire safety practices, & surveys areas damaged by wildfires.
- CAL FIRE Damage Inspection Program (DINS) database surveys structures damaged & destroyed by wildfires in California since 2013
    - Documented by CAL FIRE & partner agencies
- Features include various characteristics of structures' location, materials, value, damage, etc.

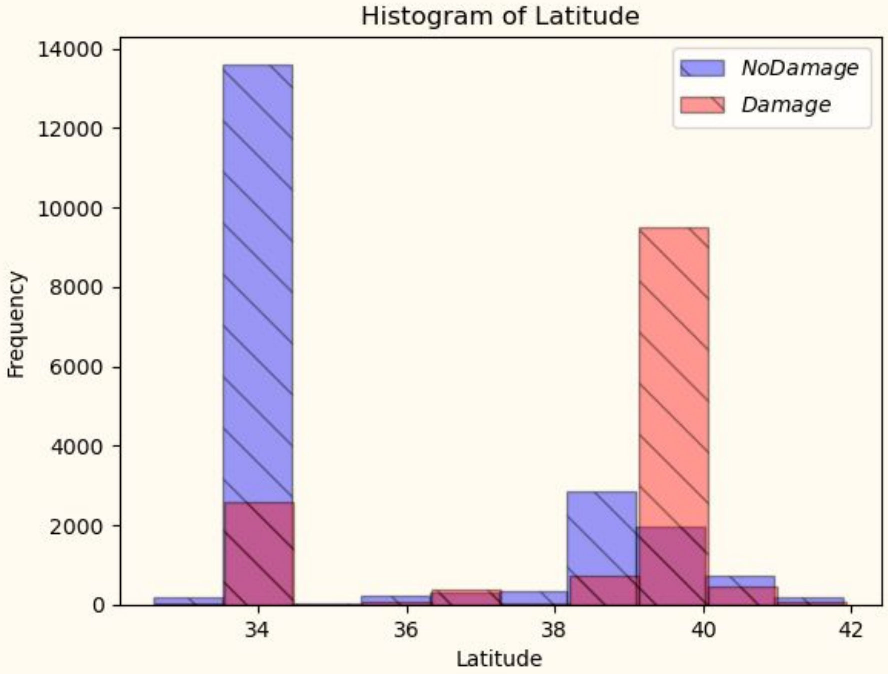
Data Cleaning

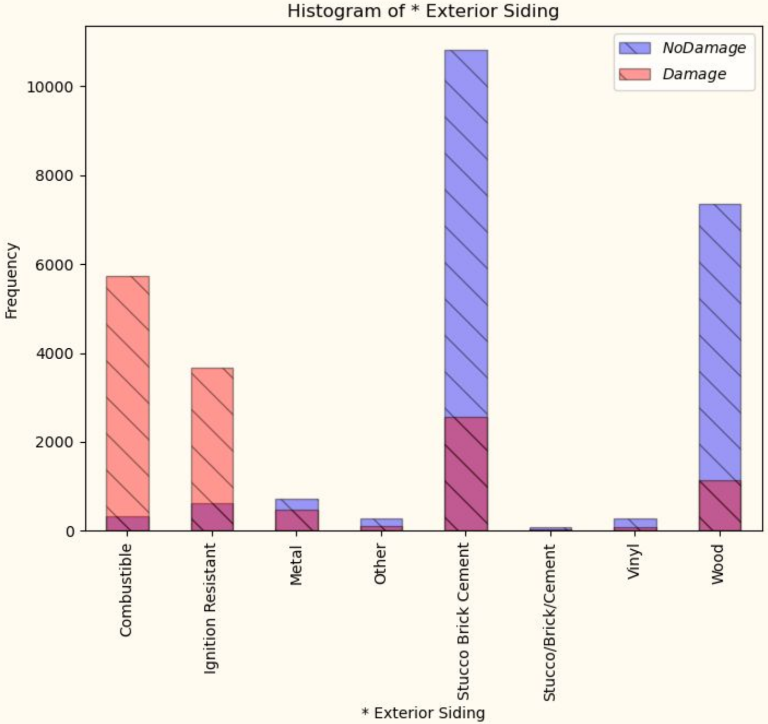
- Excluded pre-2018 data since houses without damage were not recorded then.
- Removed various features that were redundant or non-useful (ID, incident names, features with mostly missing data, etc).
- Fixed or removed observations with mislabeled or irrelevant information (ex. Damage caused by earthquakes) data.
- Transformed some features ex. *Damage* became *Destroyed or Not Destroyed*, *Year Built* became *Age*, etc.
- Removed observations with missing values.

Exploratory Data Analysis

- Our target classification became *Destroyed or Not Destroyed*.
- Frequency histograms of each feature revealed possible correlations between most available features and our target.
- No high impact from any singular features was observed.
- K-Fold Cross-validation used in some models to highlight relevant features.

Exploratory Data Analysis

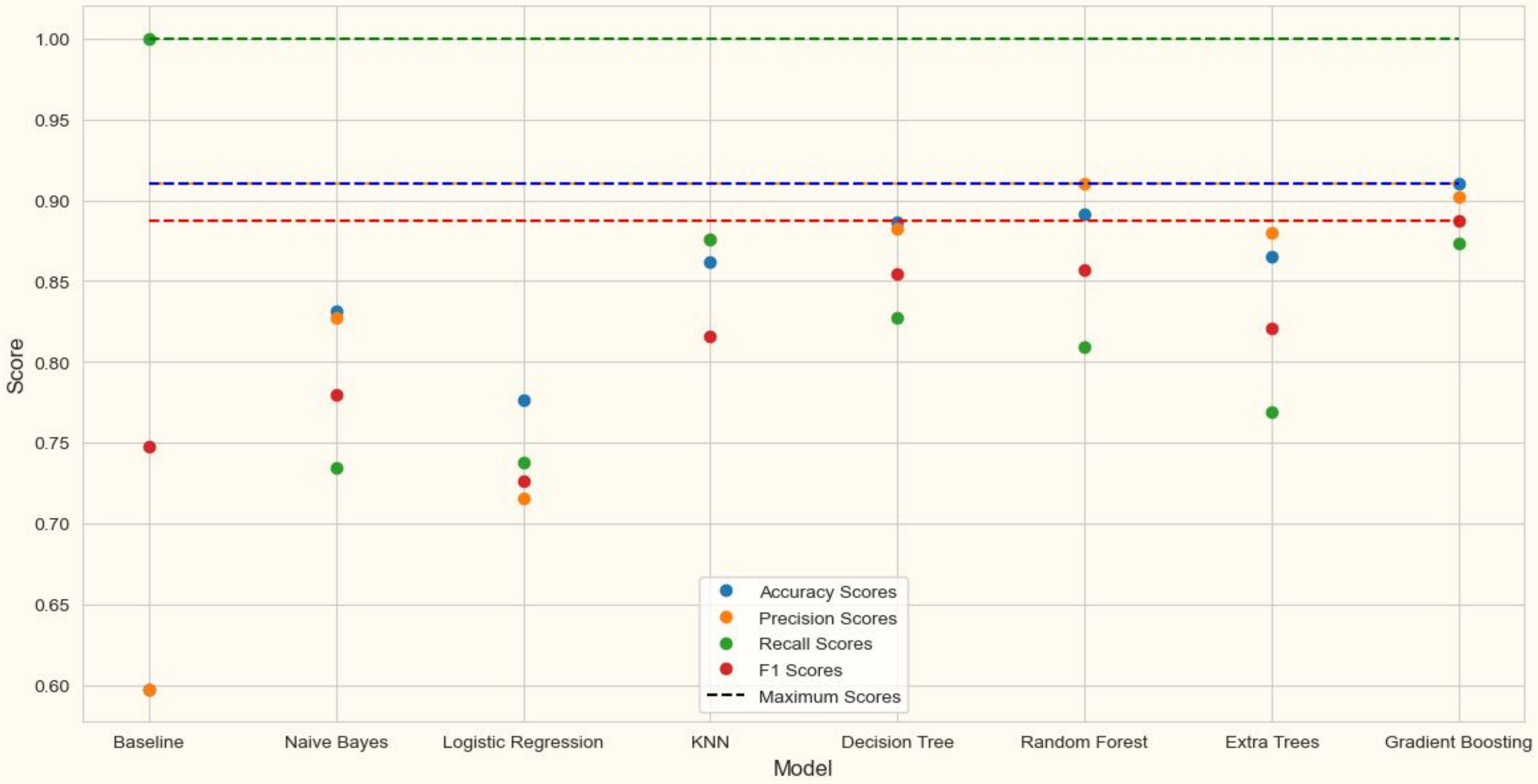




Models

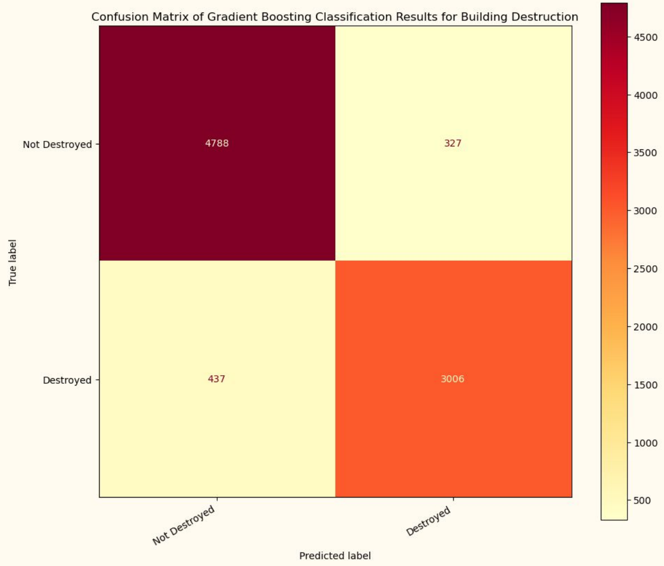
- Multiple different classifying models were trained and tested.
    - Baseline
    - Categorical Naive Bayes
    - Logistic Regression
    - k-Nearest Neighbors
    - Decision Tree
    - Random Forest
    - Extra Trees
    - Gradient Boosting
- Accuracy was the main score used to indicate performance, also used precision/recall/F1 score.
- All models except baseline obtained at least an accuracy of 77%.
- Highest accuracy was attained by Gradient Boosting Model with 91%.

Prediction Scores Comparison



Models

- All models showed some improvement over baseline in classifying destruction.
- Ensemble models (gradient boosting, random forests, etc.) outperformed other models in predicting destruction.
- Most important features seem to be latitude, longitude, exterior siding.
- Models may be impacted by small imbalance in classes (40% destroyed/60% not destroyed), bias from removal of observations with missing values, margin of error in observations due to fire damage/poor access/other factors, specific incident bias (LA fires, etc.)



Future Work

- Collect more data in order to decrease bias caused by the missing values we had to remove.
- Implement and compare additional models, imputation techniques, and data sources (geographical features of structure locations, additional sources of building or wildfire data, etc.)
- Tune hyperparameters and variable selection more to target different KPIs.
- Implement our original idea, which was to to implement ordinal multiclass classification of various damage levels for buildings affected by wildfires.
    - Would need more balanced dataset with more observations corresponding to each damage level.

Acknowledgements

We would like to thank **Hatice Mutlu** for her mentorship and feedback over the course of the project. We would also like to thank **Steven Gubkin**, **Alec Clott**, **Roman Holowinsky**, and the rest of the Erdős Institute staff for their instruction and support over the course of this bootcamp.