Predictive Model for Damage Caused by Wildfires

Katrina Suherman
University of California, San Diego
San Diego, USA
ksuherman@ucsd.edu

Rheka Narwastu
University of California, San Diego
San Diego, USA
rnarwastu@ucsd.edu

Introduction

Wildfires pose a significant threat to structures, causing widespread damage and financial losses. The severity of wildfire damage is influenced by various factors, including building materials, structural attachments, and property value. Understanding these factors can help in assessing wildfire risk and improving fire mitigation strategies.

In this project, we will build a machine learning model to predict the level of wildfire damage a structure may sustain based on its physical attributes, construction materials, and external features. We aim to develop a predictive system that can assist homeowners, policymakers, and firefighters in identifying high-risk properties and making informed decisions to enhance fire resilience.

1 Dataset

1.1 Identify Dataset

The California Wildfire Damage Dataset (’`fire-data.csv`’) is sourced from the California Open Data Portal (data.ca.gov) and maintained by CAL FIRE. It contains structured damage inspection data collected during major wildfire incidents.

The dataset consists of 130,722 rows and 46 columns, documenting structural characteristics, construction materials, and wildfire damage severity. The data is collected through CAL FIRE’s Damage Inspection (DINS) Program, which assesses structures impacted by wildfires within 100 meters of the fire perimeter. Inspections follow a systematic process, though access limitations or extreme fire damage may lead to missing values.

To evaluate fire impact, structures are classified into five damage levels:

Damage Percentage	Description
1-9%	Affected Damage
10-25%	Minor Damage
26-50%	Major Damage
51-100%	Destroyed
No Damage	No Damage

The dataset is regularly updated based on new wildfire incidents. More details on methodology and structure are available in the DINS database dictionary on the California Open Data Portal.

1.2 EDA

1.2.1 Basic Statistics The California Wildfire Damage Dataset (’fire-data.csv’) consists of 130,722 rows and 46 columns, documenting structural assessments from wildfire incidents across California. The dataset contains detailed damage inspection reports collected by CAL FIRE’s Damage Inspection (DINS) Program, which are used to evaluate the impact of wildfires on buildings and infrastructure.

The dataset includes various structural and property characteristics, such as:

- Building materials, including roof construction, exterior siding, and window type.
- Structural attachments, such as fences, and eaves.
- Property attributes, including assessed property value.
- Damage classification, which categorizes structures from no damage to completely destroyed.
- Geographical details, including county and ZIP code.

These features help identify patterns in wildfire damage and allow for the development of predictive models to assess fire risk based on structural attributes.

1.2.2 Data Cleaning

- **Column Renaming:** Several columns in the dataset contained an asterisk (‘*‘) at the beginning of their names. To improve readability and ease of access when working with the dataset programmatically, we removed these asterisks. For instance:
  - `* Damage` → `Damage`
  - `* Roof Construction` → `Roof Construction`
- **Columns Dropped:** After an initial exploration of the dataset, we decided to remove:
  - Columns related to geographical positioning (e.g., city location, exact coordinates), as the focus is on structural materials rather than location.
  - Metadata columns (e.g., inspector-specific notes, internal identifiers) that do not contribute to structural resilience analysis.
- **Row Removal:**
  - Rows where the damage classification was marked as "Inaccessible" were removed.
  - These rows accounted for only 476 out of 130,722 entries (a negligible proportion).
  - Since these cases represent structures that could not be assessed due to fire conditions, they do not provide useful information for modeling wildfire damage severity.
- **Outlier Removal:**
  - The "Assessed Improved Value (parcel)" column contained extreme values that could skew analysis.
  - We applied an outlier removal process using the 5th and 95th percentiles as boundaries.
  - This approach removed only 10% of the dataset (5% from each tail), ensuring that diverse data is retained while filtering out extreme values.

1.2.3 Findings in EDA We conducted an exploratory data analysis (EDA) to examine the relationships between structural attributes and wildfire damage levels. Our findings reveal several key patterns regarding the vulnerability of different materials and construction styles.

Fence Attached to Structure and Exterior Siding: Structures without a fence often have a wood exterior siding, which could increase the risk of fire.

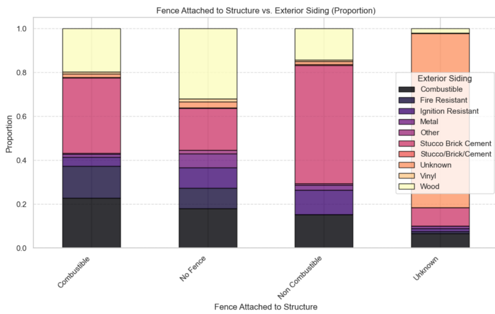

Figure 1: Relationship between Fence Attached and Exterior

Structure Category and Damage: In the dataset, the structure category consists of 7 unique values (Single Residence, Other Minor Structure, Multiple Residence, Nonresidential Commercial, Mixed Commercial/Residential, Infrastructure, or Agriculture). Based on the graph below, Single Residences and Other Minor Structures are the most commonly affected types, whereas other structure types have a higher "No Damage" cases.

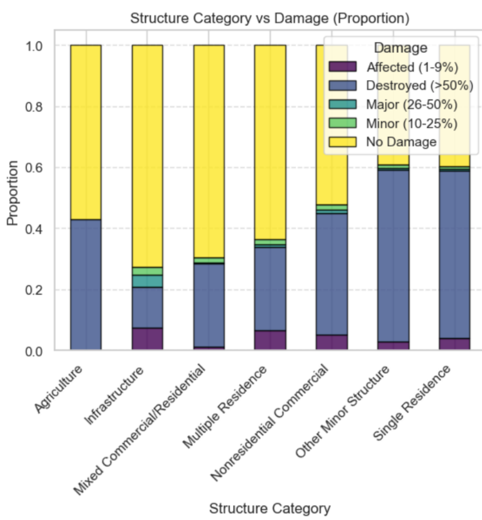

Figure 2: Relationship between Structure Category and Damage

Roof Construction and Damage: is the type and material used for the roof. Interestingly, non-combustible roofs are dominantly destroyed, followed by fire-resistant roofs that appear to have a significant number of destroyed cases as well.

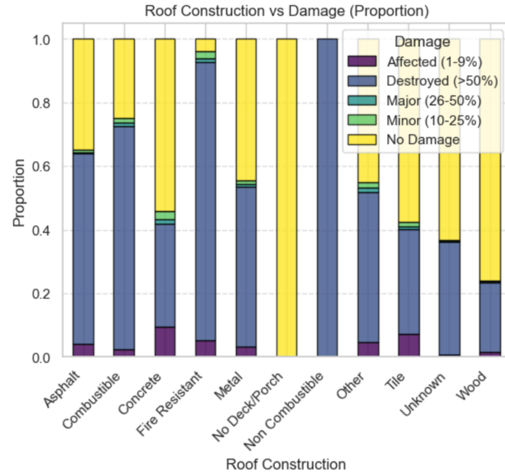

Figure 3: Relationship between Roof Construction and Damage

Eaves and Damage: are the edges of a roof to protect the structure from rain and sun. Unknown and Non-applicable eaves appear to have the highest number of destroyed cases, which is significantly larger than its other damage levels. Meanwhile, other eave categories show a higher number of "No Damage" cases compared to other damage levels.

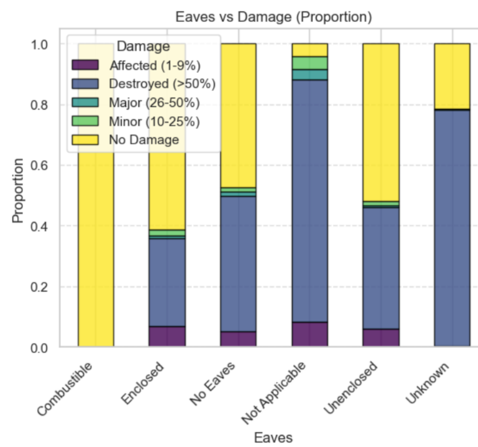

Figure 4: Relationship between Eaves and Damage

Vent Screen and Damage: Vents are essential for ventilation and also serve as a fire defense mechanism by preventing direct flame and ember intrusion, especially when properly screened. However, based on the observed relationship, structures with vents tend to have more cases of severe damage compared to those without vents.

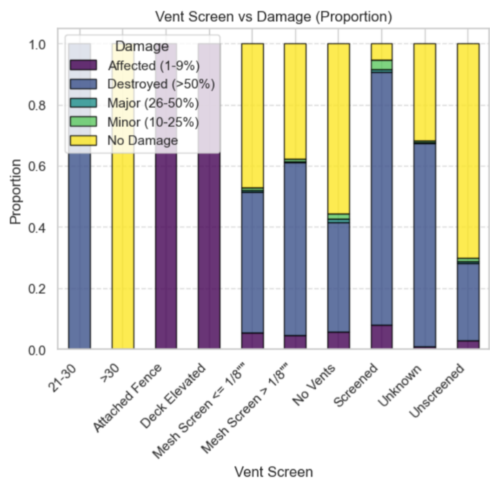

Figure 5: Relationship between Vent and Damage

Exterior Siding and Damage: Fire-resistant and ignition-resistant sidings still experience some destruction, with a significant difference of Destroyed cases than other damage levels. This indicates that fire-resistant materials alone might not be sufficient to prevent destruction in severe wildfire conditions.

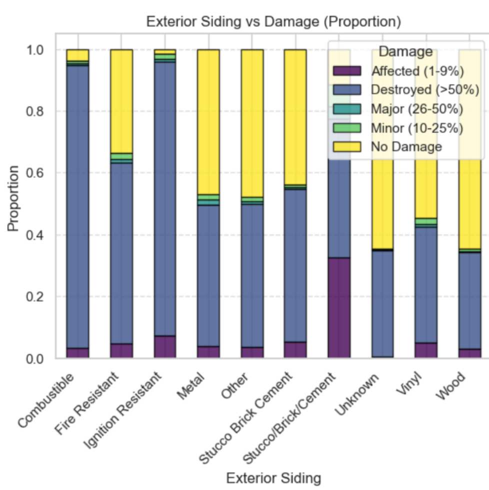

Figure 6: Relationship between Exterior Siding and Damage

Fence Attached and Damage: Combustible and non-combustible fences perform similarly in wildfire conditions, meaning material alone may not be a strong predictor of fire resilience.

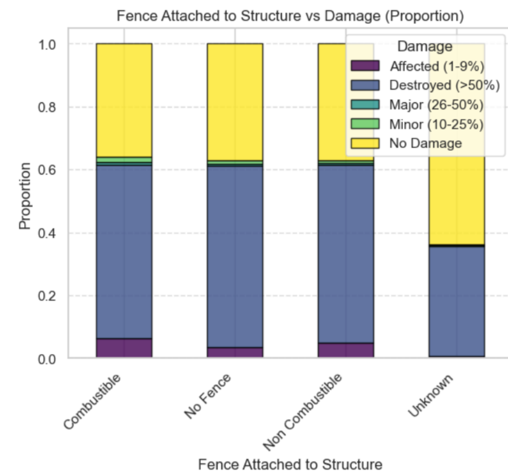

Figure 7: Relationship between Fence Attached and Damage

Roof Construction and Exterior Siding: The figure below shows that asphalt roofs tend to be the most common combination with brick / stucco sidings, followed by asphalt roofs and wood sidings.

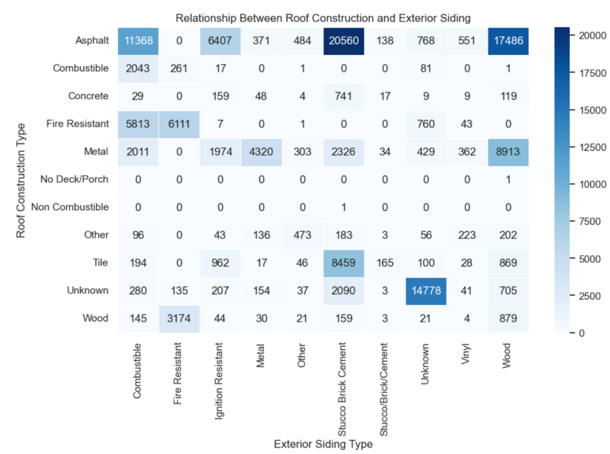

Figure 8: Relationship between Exterior Siding and Roof Construct

2 Predictive Task

2.1 Predictive task

After performing data cleaning and feature selection, our goal is to predict the severity of wildfire damage to a structure based on its physical attributes and construction materials. This is a classification problem where the target variable, `Damage`, represents different levels of structural damage. The dataset categorizes damage into five classes:

- No Damage
- Affected (1-9%)
- Minor (10-25%)

- Major (26-50%)
- Destroyed (>50%)

2.2 Evaluation

To assess the performance of our predictive model, we will use accuracy and F1-score as our primary evaluation metrics. Since our task is a multi-class classification problem, these metrics provide a comprehensive way to evaluate model performance across all wildfire damage categories.

- Accuracy: Measures the proportion of correctly classified structures out of the total predictions. It provides a straightforward assessment of overall model performance.
- F1-score: A balanced metric that considers both precision and recall. Since wildfire damage classification involves an imbalanced dataset (e.g., fewer structures in the "Major" and "Destroyed" categories), the F1-score helps ensure that the model does not disproportionately favor the majority class.

2.3 Generalization

To ensure the generalizability of the model, we implement cross-validation techniques, including GridSearchCV, to optimize hyperparameters and prevent overfitting. Instead of relying on a single train-test split, we use k-fold cross-validation, where the dataset is divided into multiple subsets to evaluate the model's performance across different training and testing scenarios. This helps ensure that the model does not learn patterns specific to a particular wildfire incident but instead captures general trends in wildfire damage.

GridSearchCV is applied to systematically explore different hyperparameter combinations and select the best-performing configuration. By tuning parameters such as the depth of decision trees, the number of estimators in ensemble models, and regularization factors, we improve the model's robustness and adaptability to unseen wildfire events.

2.4 Baseline

Baseline dataset: California Wildfire Data
Baseline features:

- *Structure Category* - Classification of the structure type.
- *Roof Construction* - Type of roof construction, which can influence fire resistance.
- *Eaves* - Design of the eaves, as open eaves may increase fire vulnerability.
- *Vent Screen* - Presence and type of vent screening, which affects ember intrusion.
- *Exterior Siding* - Exterior wall materials that impact fire spread.
- *Window Pane* - Window type and material, which influence heat resistance.
- *Fence Attached to Structure* - Presence of an attached fence, which may act as a fire conduit.
- *Assessed Improved Value (parcel)* - Assessed value of improvements on the property.

Before training, categorical features were transformed using one-hot encoding to convert them into a numerical format suitable for machine learning models. This process allows models to interpret categorical attributes effectively without imposing an arbitrary ordinal relationship. Additionally, all numerical features were standardized to ensure uniform scaling across different attributes.

Four models were used as baselines: Logistic Regression, Random Forest Classifier, XGBoost Classifier, and LightGBM Classifier.

- **Logistic Regression** assumes a linear relationship between features and wildfire damage levels. It is less likely to overfit and serves as a good benchmark for evaluating more complex models.
- **Random Forest Classifier** utilizes ensemble learning to improve classification performance. It captures non-linear relationships between features and damage severity. Additionally, it effectively handles both categorical and continuous variables and automatically mitigates missing values.
- **XGBoost Classifier (Extreme Gradient Boosting)** is an optimized gradient boosting framework that combines multiple weak learners to achieve strong predictive performance. It is particularly efficient for medium-sized datasets and has built-in mechanisms to handle missing values and feature importance ranking.
- **LightGBM Classifier** is another gradient boosting model designed for efficiency and scalability. It performs exceptionally well on large datasets, reduces training time, and provides high accuracy while being computationally less expensive than XGBoost.

Baseline models are trained on the same dataset using default hyperparameters to establish a fair comparison. The categorical features were one-hot encoded, and numerical features were standardized before training. The performance of each model was evaluated using accuracy and F1-score on the same test dataset.

2.4.1 Baseline Results The performance of each model is summarized in Table 1.

Model	Accuracy	F1-score
Logistic Regression	0.839083	0.818776
Random Forest Classifier	0.806386	0.803298
LightGBM Classifier	0.849651	0.830065
XGBoost Classifier	0.850731	0.831744

Table 1: Baseline Model Performance

From the results, logistic Regression performs well with an accuracy of 0.839083 and an F1-score of 0.818776, indicating strong baseline performance despite being a linear model. However, boosting-based models like LightGBM and XGBoost capture more complex patterns, achieving the highest accuracy. Given their effectiveness and efficiency, we will proceed with optimizing Logistic Regression, LightGBM, and XGBoost to enhance performance.

3 Model

Initial Setting

- Dataset: California Wildfire Data

- Features used: Same as baseline features
- Models used: Based on the baseline results, we decided to further explore Logistic Regression, LightGBM, and XGBoost due to their high accuracy.
- Evaluation metrics: Model optimization is based on improvements in accuracy and F1-score.

3.1 Basic Hyper parameter Tuning

3.1.1 *Logistic Regression*

- We employed GridSearchCV with a StratifiedKFold(n_splits=5) to systematically explore different configurations of the following hyperparameters for `LogisticRegression`:
  - `C` in {`0.01`, `0.1`, `1`, `10`}
  - `penalty` in {`l1`, `l2`}
  - `solver` = {`saga`}
- **Best Found Parameters:** After running the grid search, the optimal parameter set was:

```
'classifier__C': 10,
'classifier__penalty': 'l1',
'classifier__solver': 'saga'
```

3.1.2 *XGBoost*

- We employed `GridSearchCV` with a `StratifiedKFold()` with number of 5 splits with the following hyperparameters for XGBoost:
  - `colsample_bytree` ∈ {0.6, 0.8}
  - `gamma ∈ {0.1, 0.2}`
  - `learning_rate = {0.2}`
  - `max_depth = {7}`
  - `n_estimators ∈ {100, 200}`
  - `subsample ∈ {0.6, 0.8}`
- **Best Found Parameters**: After running the grid search, the optimal parameter set was:

```
'classifier__colsample_bytree': 0.8,
'classifier__gamma': 0.2,
'classifier__learning_rate': 0.2,
'classifier__max_depth': 7,
'classifier__n_estimators': 200,
'classifier__subsample': 0.8
```

3.1.3 *LightGBM*

- We employed `GridSearchCV` with a StratifiedKFold(n_splits=5) to systematically explore different configurations of the following hyperparameters for LightGBM:
  - `n_estimators ∈ {100, 200}`
  - `max_depth ∈ {3, 6, 9}`
  - `learning_rate ∈ {0.3, 0.5}`
  - `reg_alpha = {0.1}`
  - `reg_lambda = {1.0}`parameter growth.
- **Best Found Parameters**: After running the grid search, the optimal parameter set was:

```
'classifier__n_estimators': 100,
'classifier__max_depth': 9,
'classifier__learning_rate': 0.3,
'classifier__reg_alpha': 0.1,
'classifier__reg_lambda': 1.0
```

3.2 Feature Engineering

3.2.1 *Quantitative Data* For quantitative data, we use only one numerical feature, `Assessed Improved Value (parcel)`. This feature represents the financial valuation of structural improvements and serves as an important predictor in our model.

To preprocess this feature, we apply a structured numerical transformation pipeline. First, any missing values in `Assessed Improved Value (parcel)` are imputed using the median value of the column. This ensures that the distribution of the data remains stable while preventing extreme values from skewing the dataset.

Next, we standardize the feature using `StandardScaler()` to normalize the values.

By applying these preprocessing steps, we ensure that the numerical feature is properly prepared for predictive modeling while maintaining data integrity and consistency.

3.2.2 *Qualitative Data* Similar to the baseline model, we applied one-hot encoding to all categorical features in the dataset. To achieve this, we used the `OneHotEncoder` function with the *handle_unknown* as 'ignore'. This ensures that each categorical feature is expanded into multiple binary columns, each representing a unique category. If the model encounters an unseen category in the test set that was not present during training, it will be safely ignored instead of causing an error.

3.3 Model Results

The performance of each model is summarized in Table 2

Model	Accuracy	F1-score
Logistic Regression Tuned	0.839038	0.818731
LightGBM Classifier Tuned	0.850641	0.832199
XGBoost Classifier Tuned	0.851136	0.833018

Table 2: Model Performance

3.3.1 *Model Challenges and Limitations*

- **Marginal Gains or Diminished Returns:** While hyperparameter tuning often leads to improvements, our experiments revealed that tuning Logistic Regression did not yield substantial benefits. In some cases, the tuned version even showed a slightly lower Accuracy compared to the baseline, although the difference was minimal. This suggests that:
  - *Baseline Strength:* The baseline Logistic Regression model may already have been near its optimal settings, leaving limited room for improvement.
  - *Model Capacity:* Logistic Regression, being a linear model, can struggle to significantly boost performance without

introducing additional features or more complex transformations.

- **Trade-offs in Complexity:** While ensemble methods such as LightGBM and XGBoost benefit from more extensive tuning Logistic Regression's simpler structure may not allow for the same degree of performance increase once basic hyperparameters have been explored.

4 Literature

DamageMap: A Post-Wildfire Damaged Buildings Classifier by Marios Galanis et al.

The article *DamageMap: A Post-Wildfire Damaged Buildings Classifier* introduces a machine learning model for assessing wildfire damage using aerial imagery. The authors developed *DamageMap*, a classifier that analyzes post-wildfire images to categorize building damage levels. Using a subset of the xBD dataset, which includes pre- and post-event satellite images, their model leverages convolutional neural networks (CNNs) trained solely on post-wildfire images. Unlike other methods, this approach eliminates reliance on pre-disaster data. The model's performance is evaluated across four key metrics: accuracy, precision, recall, and F1 score.

While DamageMap focuses on post-disaster damage assessment, our approach develops a predictive system to assist homeowners, policymakers, and firefighters in identifying high-risk properties before a wildfire occurs. This enables proactive mitigation strategies, such as improved building materials, enhanced fire prevention efforts, and more effective resource allocation.

5 Result

5.1 Comparison

- The ensemble models demonstrate slight improvements after tuning. For instance, LightGBM's Accuracy increased by 0.00099 and its weighted F1-score by 0.0021339, while XGBoost showed an increase of 0.000404 in Accuracy and 0.0012739 in weighted F1-score.
- In contrast, the tuned Logistic Regression model experienced a slight performance drop of approximately -0.000045 in both Accuracy and weighted F1-score, suggesting that its baseline configuration was nearly optimal and its linear nature may limit further gains.
- The observed improvements in the ensemble methods likely stem from their capacity to capture complex, non-linear relationships, coupled with the benefits of refined preprocessing and systematic hyperparameter optimization. Although the gains are subtle, they are consistent across evaluation metrics.
- Overall, while hyperparameter tuning yields incremental benefits, the small absolute differences imply that the baseline models were robust, leaving limited scope for dramatic enhancements through tuning alone.

5.2 Effectiveness

While the core features remained the same, the refined approach to handling these features led to better model performance.

5.2.1 Key Differences in Feature Design

- *Enhanced Feature Preprocessing*
  The baseline model applied one-hot encoding directly to categorical variables and standardized numerical features. The final model improved this process by incorporating a structured column transformer, which separately handled numerical scaling and categorical encoding, ensuring more consistent feature representation.
- *Removal of Redundant Features*
  The baseline model applied one-hot encoding with *drop_first* as 'True', which removed one category per categorical variable. In contrast, the final model retained all categories to allow for complete information capture, reducing potential information loss.

5.2.2 Effectiveness of Features in Performance To gauge the relative importance of each feature, we extracted the feature importances from the trained model

- *Exterior Siding* (Unknown, Combustible, Ignition Resistant) remains the most influential factor in predicting wildfire damage. The type of exterior siding plays a crucial role in fire resistance, which aligns with expectations in fire risk assessment.
- *Roof Construction* (Unknown, Fire Resistant, Wood) also played a significant role. The presence of fire-resistant materials or wood-based roofing appears to impact damage likelihood.
- *Window Pane* (Unknown) has lower importance but still contributes to the prediction, suggesting that window materials or their fire resistance properties may have a secondary effect on overall structural integrity.

5.3 Hyperparameter

5.3.1 Parameter Sensitivity The ensemble methods (LightGBM and XGBoost) exhibited some sensitivity to these hyperparameters; even minor adjustments in values such as `learning_rate` and `max_depth` led to measurable improvements, albeit modest. In contrast, Logistic Regression, with its simpler linear structure, showed minimal change, indicating that its baseline configuration was already near optimal. These findings emphasize that while careful tuning is critical for optimizing model performance, the extent of improvement is constrained by the robustness of the baseline and the inherent complexity of the data.

5.4 Major Takeaways

1. **XGBoost Emerges as the Best Model.** Although the improvements over the baseline are slight, the tuned XGBoost model achieves the highest Accuracy and weighted F1-score, making it the preferred choice among the models tested.
2. **Baseline Models Were Already Robust.** The baseline performances for LightGBM and XGBoost (Accuracy of 0.849651 and 0.850731, respectively) indicate that these ensemble methods were already well-suited to the dataset, leaving limited scope for large improvements via tuning.

(3) **Feature Importance and Utility.** The selected features proved highly predictive of `Damage`, reflecting both structural and financial characteristics. The combination of these structural and financial attributes contributes to the high predictive power observed across all models.

(4) **Tuning Yields Incremental Improvements.**
- *Logistic Regression:* After optimizing, the accuracy and F1-score remain virtually unchanged suggesting the baseline was near-optimal for this linear model.
- *LightGBM:* After optimizing, it shows a slight boost indicating that careful hyperparameter adjustment can still refine an already strong baseline.
- *XGBoost:* After optimizing, the Accuracy and F1-score rose slightly. Although the gain is small, it remains consistent across multiple runs.

(5) **Limited Headroom for Improvement.** The data's intrinsic complexity and the strong baseline setups mean that hyperparameter tuning can only squeeze out slight performance gains.

6 Demo

A simple user interface has been developed to allow users to interact with our predictive model. Users can input various parameters and instantly view the model's prediction for wildfire damage. Access the demo via the following link:

Demo

Since our experiments showed that the tuned XGBoost model achieved the highest Accuracy and weighted F1-score, the demo utilizes the XGBoost model to provide predictions.