An interactive visualization platform for California wildfires

Stephanie Cho[1], Dhyan Gandhi[1], Wenjia Hu[1], Huanci Yang[1], Elisa Yuen Ting Ching[1]

Georgia Institute of Technology – CSE 6242 Team 6

Introduction & Motivation

What is the problem?

Wildfires continue to devastate communities in California, with rising occurrences and climate change. Yet, there is a gap in tools that effectively predict and communicate wildfire risks.

Why should we care?

Predictive modeling and intuitive visualization tools can support emergency services, policy-makers, and the public in understanding wildfire risks. This would help enable better preparation, early response, and resource allocation.

Problem Statement & Research Gap

We aim to build a wildfire probability prediction model and visualization tool focused on California, while also covering the broader U.S. landscape. Our goal is not just to predict fire events but also to **explain when, where, and how severe** these fires might be. We also hope to effectively visualize the extent and trends of these wildfires across California exhaustively.

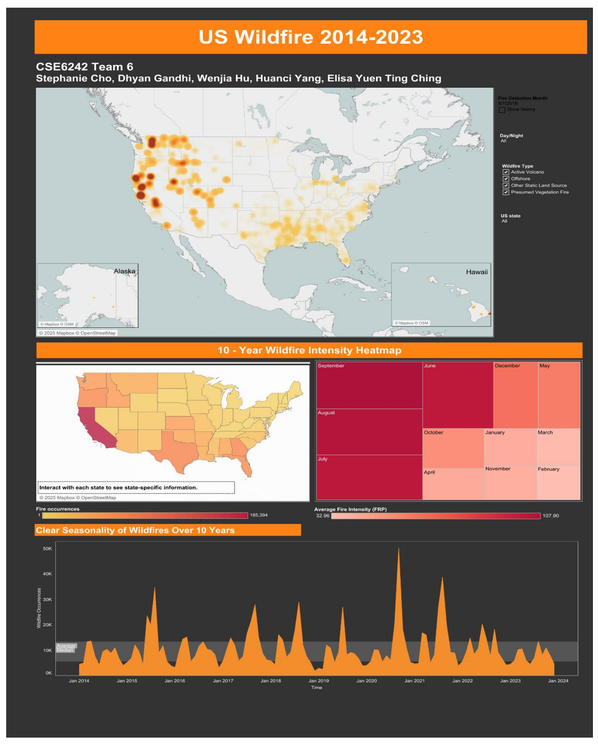
Data

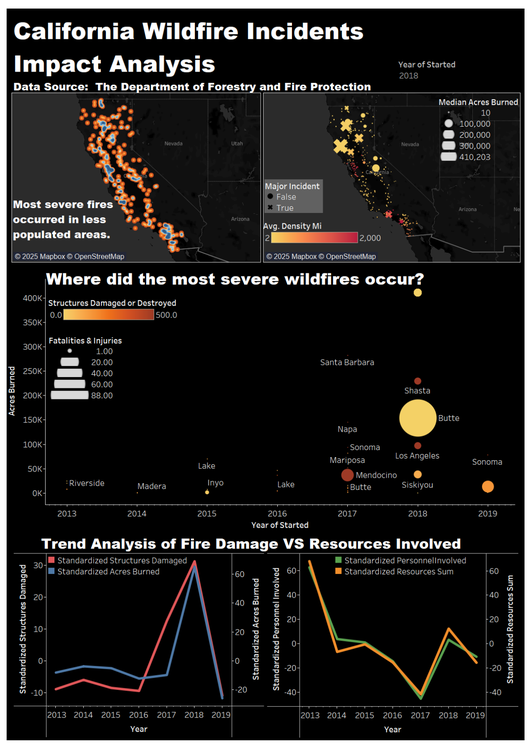
Sources downloaded from:

- CIMIS Environmental Data (OpenML): 128,126 obs.
- California Fire Incident Data (Department of Forestry and Fire Protection): 1,647 obs.
- NASA MODIS Satellite Data: 1.29 million obs.

Characteristics:

- Spanning 2013 –2023
- Weekly/daily granularity
- Environmental, temporal, and geospatial features





Our Innovation

Prior studies largely focused on binary wildfire classification or static maps. We improve upon this by:

- Using **multi-class probabilistic predictions**
- Integrating **Bayesian hyperparameter optimization**
- Developing **interactive dashboards** based on NASA and California datasets
- Dynamically tuning classification thresholds for better real-world calibration

Our Approach

Algorithms:

- XGBoost for wildfire probability prediction using environmental and geospatial data
- Bayesian optimization for hyperparameter tuning
- SMOTE to address class imbalance

Visualizations:

- Tableau dashboards to show national and California-specific wildfire trends, impacts, and model predictions

How They Work:

- CIMIS environmental data (e.g., temperature, humidity, solar radiation) is used to train a machine learning model to predict wildfire risk.
- Dashboards allow our users to explore both spatial and temporal wildfire patterns and model insights interactively.

Intuition:

- Higher temperatures and lower humidity are known to elevate wildfire risks—ML models learn from these correlations.
- Interactivity helps emergency planners zoom into risky zones and months.

Evaluation & Results

Model Evaluation:

- Metrics: Precision, F1 Score, ROC-AUC
- Hold-out test set with stratified sampling

Model Performance:

- Precision: 0.6930
- F1 Score: 0.6365
- ROC-AUC: 0.9771 (excellent class separation)

Comparison & Observations:

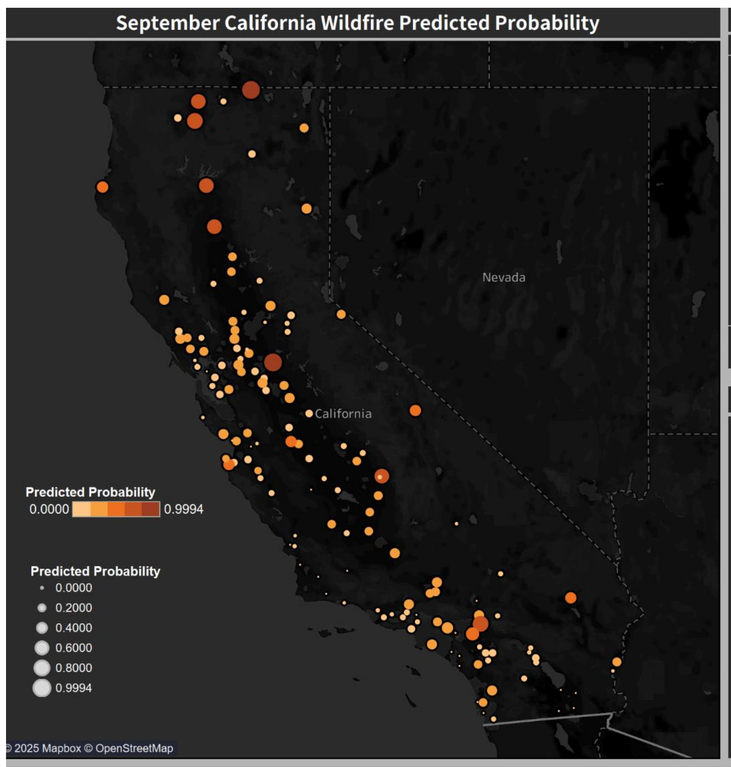
- XGBoost **outperformed traditional methods** (Random Forest, Logistic Regression)
- Dynamic thresholding enhanced minority class recall
- Top predictors: Soil Temperature, Solar Radiation, Air Temperature, Location, Wind

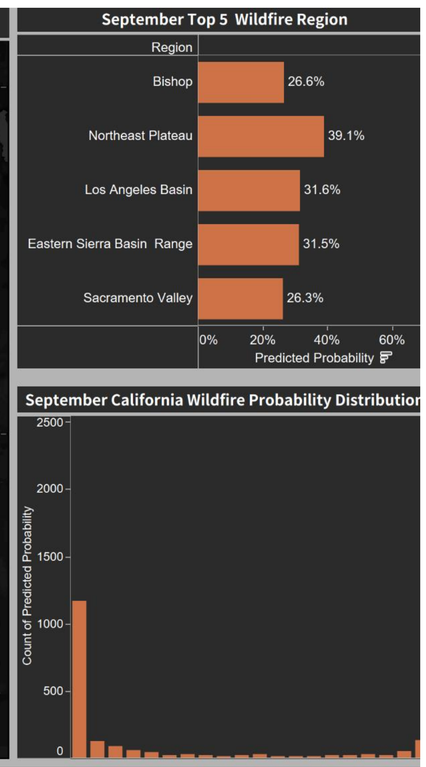
Visualization Evaluation - User Study

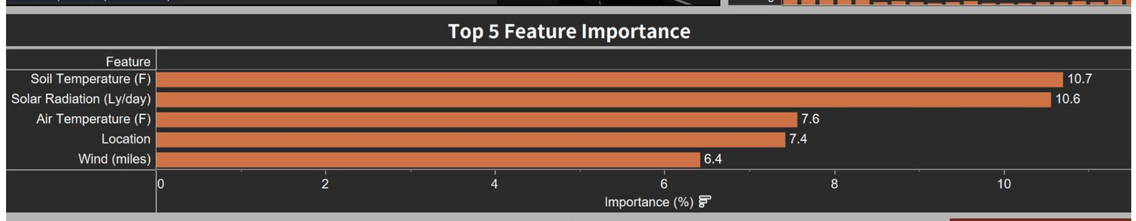
- Avg rating: 4.88/5 from 17 survey participants
- Strong agreement on clarity, design, usefulness
- Some suggestions for improving interactivity and layout

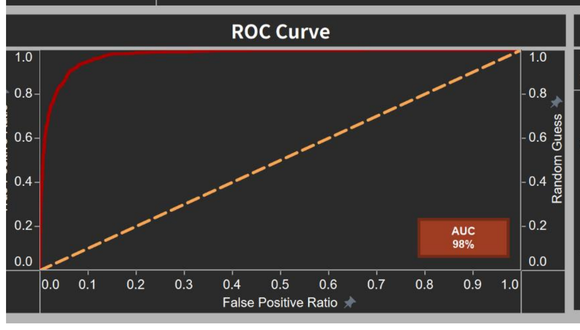
Findings:

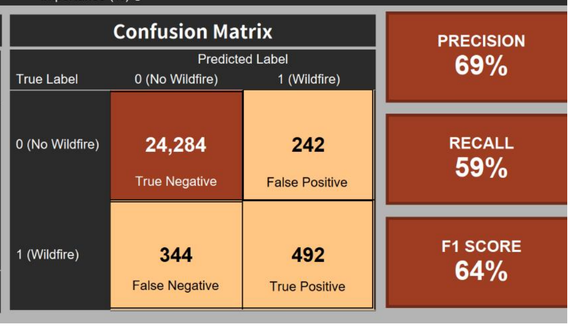
- Two seasonal peaks in U.S. wildfires (Spring and Summer)
- California's highest predicted risk occurs in August and September
- Most major wildfire incidents occurred in less populated regions in California

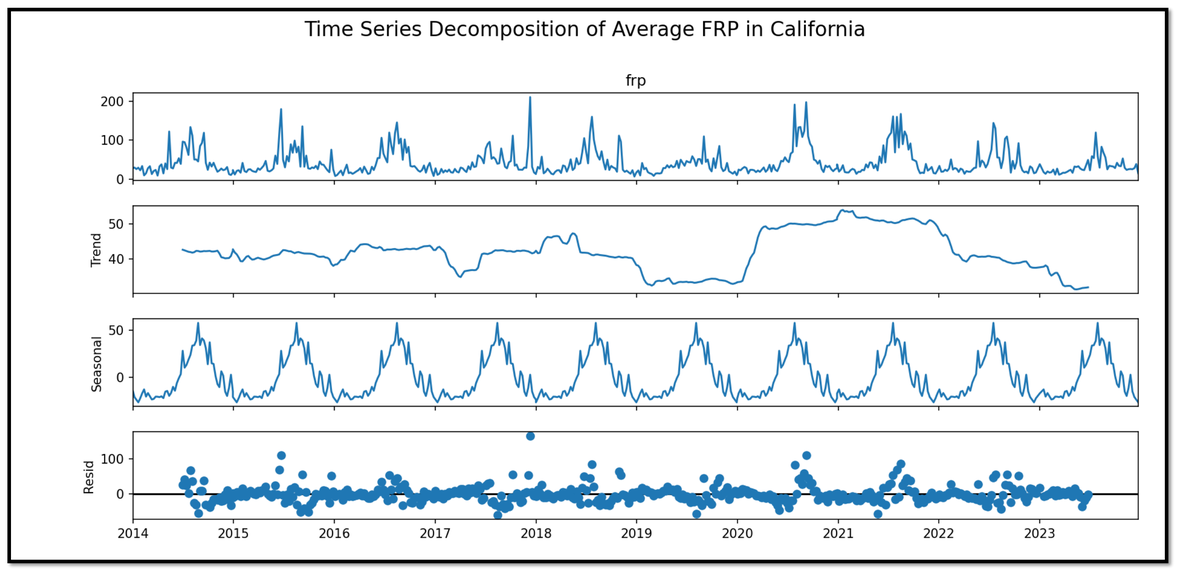












Conclusion

- Integrated ML and dashboards improve wildfire risk analysis and communication
- Highlighted misalignment in resource deployment and fire severity
- Insights applicable to emergency planning and public awareness