# Wildfire Risk Prediction


Jeffrey Sun
jeffsun1@stanford.edu

Yi Hou
yihou@stanford.edu

Neeraj Mathur
mathurn@stanford.edu

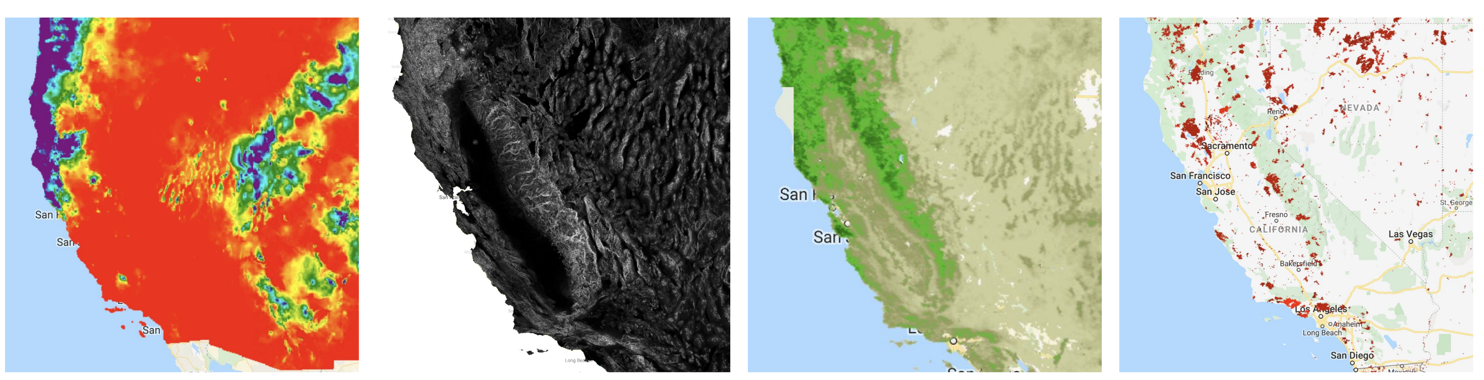

## Introduction

In the past couple of decades, the climate has influenced the wildfires significantly and has contributed to the increase in **magnitude, frequency and duration** of the wildfires. The primary **objective** of this project is to predict **wildfire risk** in the **State of California,** using various weather and geography-based metrics, like temperature, forest cover, etc.

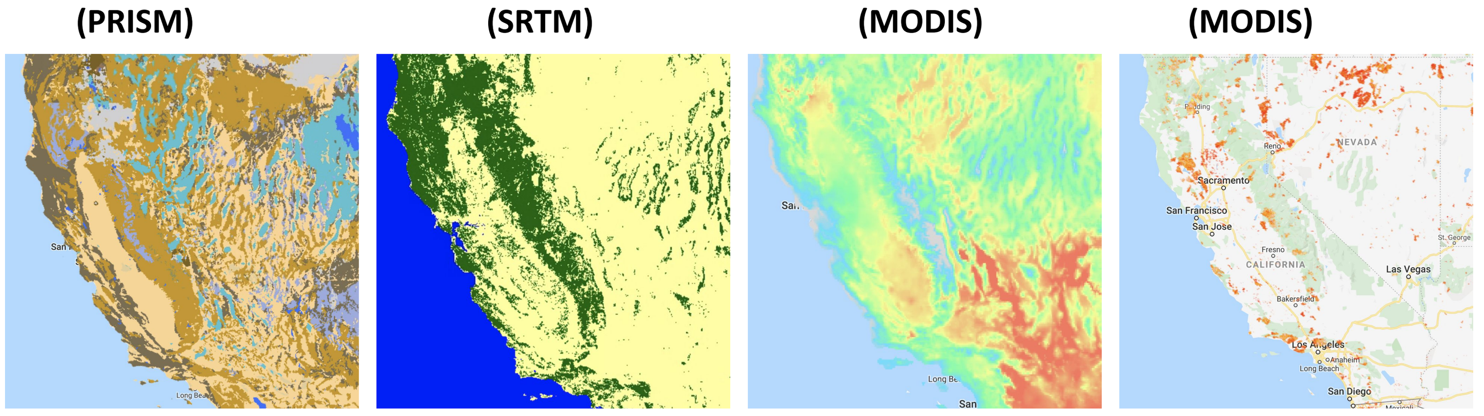
**Social Impact:** The **danger** and **high economic & social cost** of the wildfires motivated us to pursue the development of wildfire risk prediction models that could potentially be used to aid **fire prevention policy** especially in the State of California.

## Dataset

**Data Source:** We've used the dataset available from **Google Earth Engine** whereby data is collected from various satellite imagery and geospatial data sources like MODIS, PRISM, SRTM, ERGo, PALSAR, GRIDMET, etc.



**Temperature (PRISM)** | **Elevation (SRTM)** | **Leaf Area (MODIS)** | **Wildfire locations (MODIS)**



**Soil Type (ERGo)** | **Forest Area (PALSAR)** | **Radiation (GRIDMET)** | **Fire severity map (MODIS)**

**Data Sampling:** We sampled **100,000 locations** from within the California region. For each location, we gathered total of **45 features** from 2017 related to temperature, precipitation, elevation, leaf area, soil type, human modification, forest/non-forest, and radiation data.

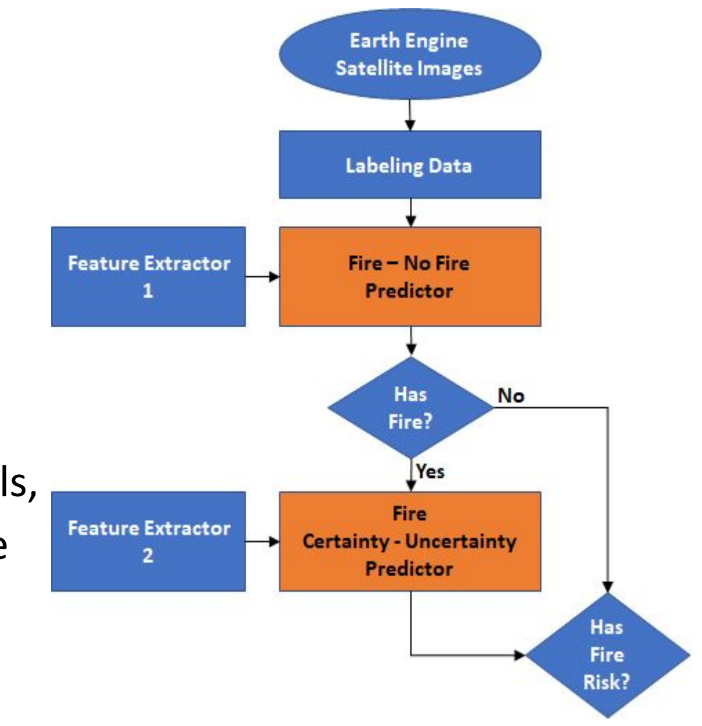
## Challenges

1. **Data gathering:** We needed to collect a breadth of relevant features from different datasets, and to match these features with each other on large number of random locations. Furthermore, these features needed to be from the same time period so that the model can be used to predict future fire occurrences.
2. **Data labeling:** to combine sampled features from different sources.
3. **Skewed Data:** The data samples are skewed such that the vast majority of samples are non-fire cases, whereas it's important for the model to predict fire cases.
4. **Model selection:** Given the nature of the problem and dataset, we had to experiment with various types of linear, non-linear and probability based models.
5. **Pipelining:** to apply results predicted by upper level onto lower level predictor.
6. **Hyperparameters tuning**: to exhaustively search optimal hyperparameters for both set of predictors independently; analyze deviance to avoid overfitting; adjust number of estimators to balance accuracy and computation time.

## Approach

Since most of our dataset contains no-fire samples, we've divided our fire risk assessment task into two levels of a pipeline:

A. **The upper level:** a **fire/no-fire** predictor, which will judge the likelihood of whether a sample will cause a fire.
B. **The lower level** will do a more detailed analysis and give a prediction for the **uncertainty** of that sample if there is a fire.
C. Based on the results of various algorithms at both levels, a **detailed features analysis** is performed to determine their impact and relationship.




## Features Analysis

After narrowing down to Random Forest and Gradient Boosting models for Fire/No-Fire predictor and Gradient Boosting for Fire Uncertainty predictor

*Figure 1: Feature Importance for Fire Occurrence and Fire Uncertainty models*

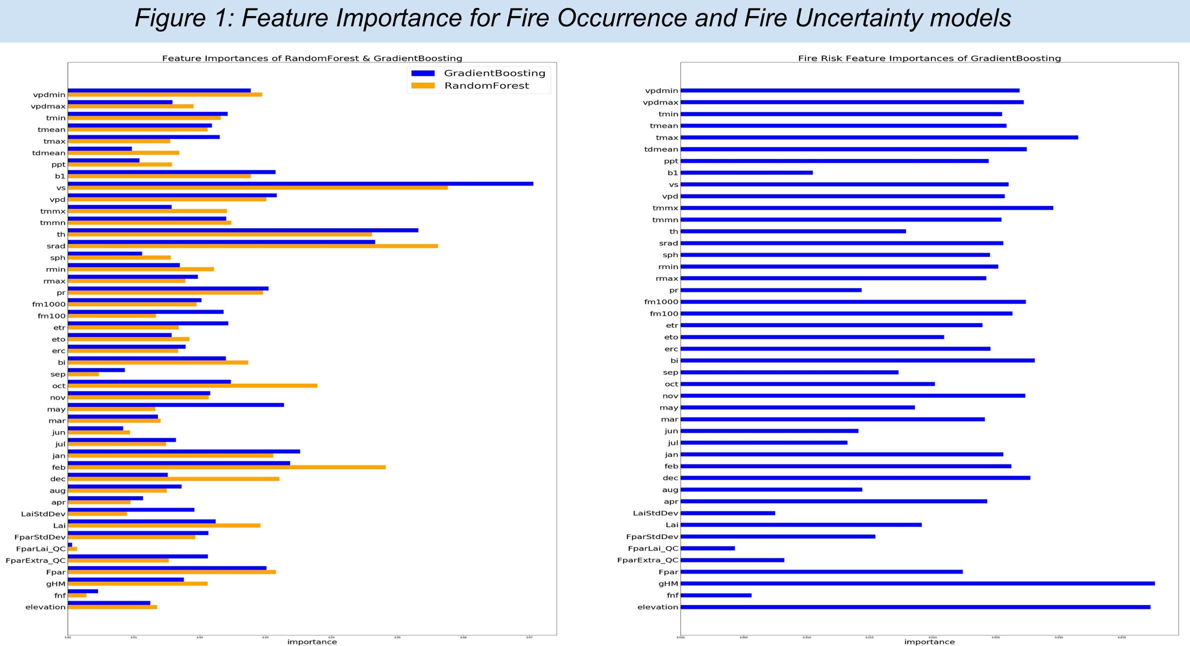


*Figure 2: Top 3 features causing wildfires*

*Highlights the impact of top 3 features on causing wildfires*
1. *Surface downward shortwave radiation*
2. *Wind direction*
3. *Wind velocity*

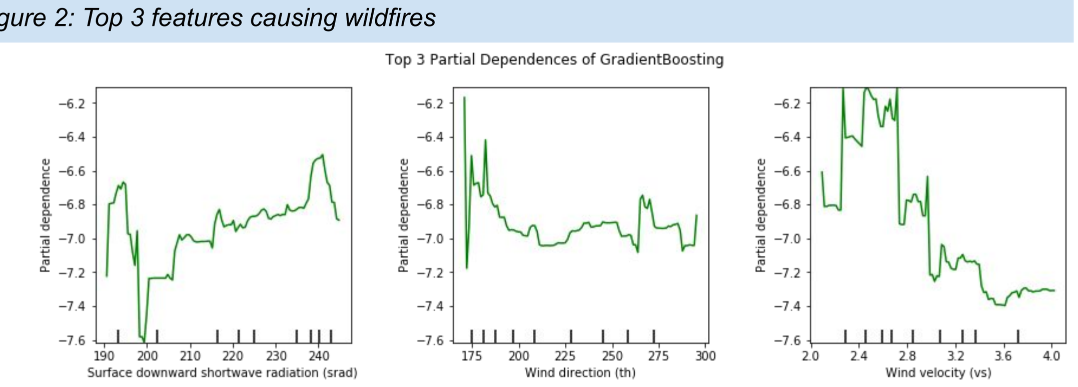
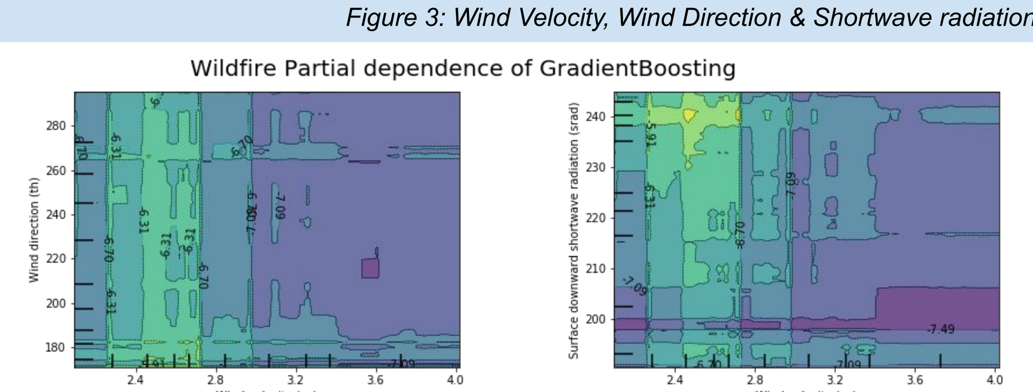


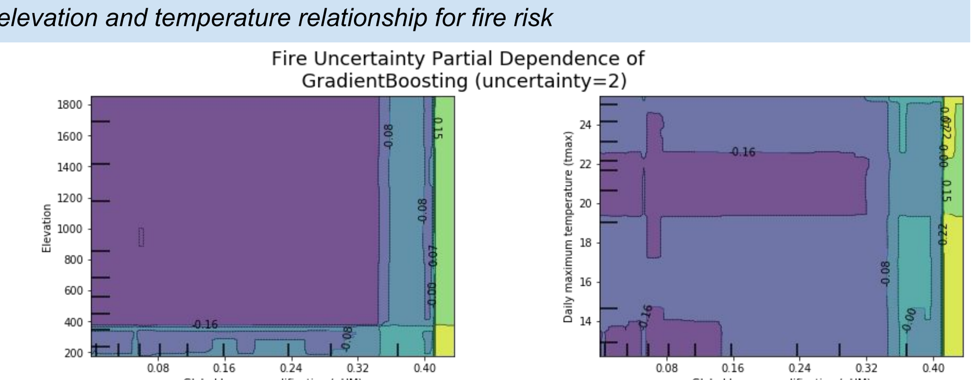

*Figure 3: Wind Velocity, Wind Direction & Shortwave radiation relationship for fire occurrence*




*Highlights the impact of **low wind speed** and **high radiation** in causing wildfires*

*Figure 4: Human modification, elevation and temperature relationship for fire risk*

*Highlights:*
1. *humans like to live at **low elevations***
2. ***human modifications** like settlement, electrical infrastructure, agriculture, transportation, etc. causes **higher temperature** leading to higher wildfire risk*




## Results

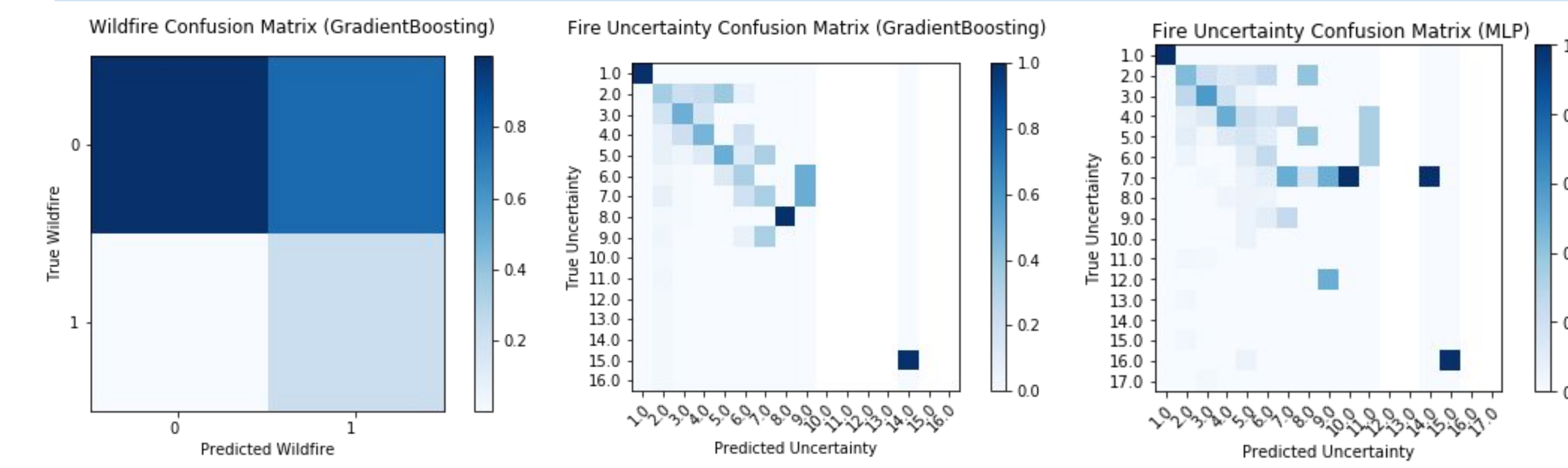
***Table 1: Classifiers to predict Fire and No-Fire cases***

| Model | | Fire Precision | Fire Recall | Fire F1-Score | No Fire Precision | No Fire Recall | No Fire F1-Score |
|---|---|---|---|---|---|---|---|
| Logistic Regression | | 0.01 | 0.82 | 0.02 | 1.00 | 0.77 | 0.87 |
| Decision Tree | | 0.01 | 0.73 | 0.01 | 1.00 | 0.72 | 0.84 |
| Random Forest | | 0.22 | 0.76 | 0.34 | 1.00 | 0.99 | 1.00 |
| **Gradient Boosting** | **All Samples** | **0.23** | **0.83** | **0.36** | **1.00** | **0.99** | **1.00** |
| | **10,000 samples** | **0.66** | **0.83** | **0.73** | **1.00** | **0.99** | **0.99** |
| AdaBoost | | 0.13 | 0.75 | 0.21 | 1.00 | 0.98 | 0.99 |
| SVM | | 0.11 | 0.51 | 0.19 | 1.00 | 0.99 | 0.99 |
| Ridge Classifier | | 0.13 | 0.75 | 0.21 | 1.00 | 0.98 | 0.99 |

***Table 2: Models to predict wildfire certainty and uncertainty***

| Model | Accuracy (uncertainty diff < 6) | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Logistic Regression | 0.77 | 0.23 | 0.40 | 0.27 |
| Decision Tree | 0.76 | 0.32 | 0.45 | 0.37 |
| Random Forest | 0.77 | 0.37 | 0.47 | 0.41 |
| **Gradient Boosting** | **0.77** | **0.51** | **0.52** | **0.49** |
| **Multilayer Perceptron** | **0.78** | **0.52** | **0.51** | **0.51** |
| K Neighbours | 0.77 | 0.38 | 0.44 | 0.40 |
| Gaussian Process | 0.77 | 0.40 | 0.47 | 0.42 |

*Figure 5: Confusion Matrices of selected models for Wildfire occurrence and Uncertainty predictions (only uncertainty < 20 depicted)*




## Discussion & Conclusion

**Dataset:**
- The skew in the dataset (i.e. most data points are non-fire cases) can be resolved by either rebalancing the number of fire and no-fire cases or applying different weights for classes.

**Predictors:**
- Ensemble models can predict most fire cases correctly and also give low false prediction for no-fire cases.
- Random forest and gradient boosting perform better on wildfire prediction.
- Gradient boosting and multilayer perceptron can predict fire severity more desirably.
- Downsampling no-fire can improve fire prediction precision and resolve the bias of training data.

**Features:**
- Fire/no-fire prediction is dominated by a small number of features, while fire severity is based on the effect of overall features.
- Wind velocity, direction and surface temperature are important indicators for existence of fire.
- Human activity worsen fire severity.

## Future Directions

- Using longer periods of historical information for Wildfire risk prediction.
- Doing fire risk prediction at a national and global level (outside of California).
- Building a useful dashboard or map for anticipating future fire occurrences.

## References


*1) Pechony O, Shindell DT (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. PNAS,* http://www.pnas.org/content/107/45/19167

*2) Eidenshink J, Schwind B, Brewer K, Zhu Z, Quayle B, Howard S (2007). A Project for Monitoring Trends in Burn Severity. Fire Ecology,* https://doi.org/10.4996/fireecology.0301003

*3) Google Earth Engine,* https://earthengine.google.com/

*4) Heisig, J. (2018, April 12). Step by Step: Burn Severity mapping in Google Earth Engine. UN Spider,* http://www.un-spider.org/advisory-support/recommended-practices/recommended-practice-burn-severity/burn-severity-earth-engine