Predicting California Wildfires

Author: Malcolm Katzenbach

Outline

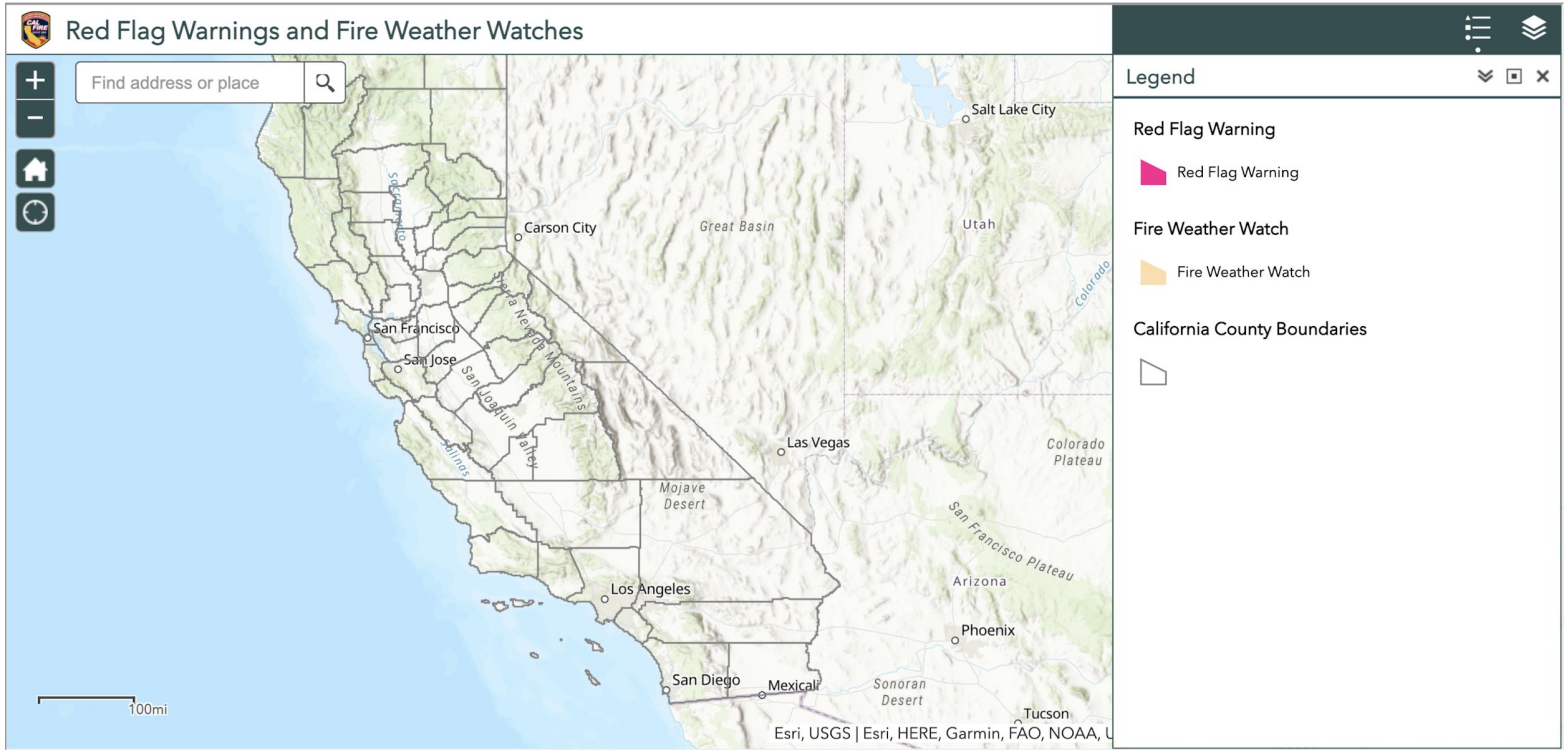
- Introduction
- Business Problem
- Method
- Data Understanding
- Exploratory Data Analysis
- Modeling
- Evaluation
- Conclusion / Future Steps

Introduction

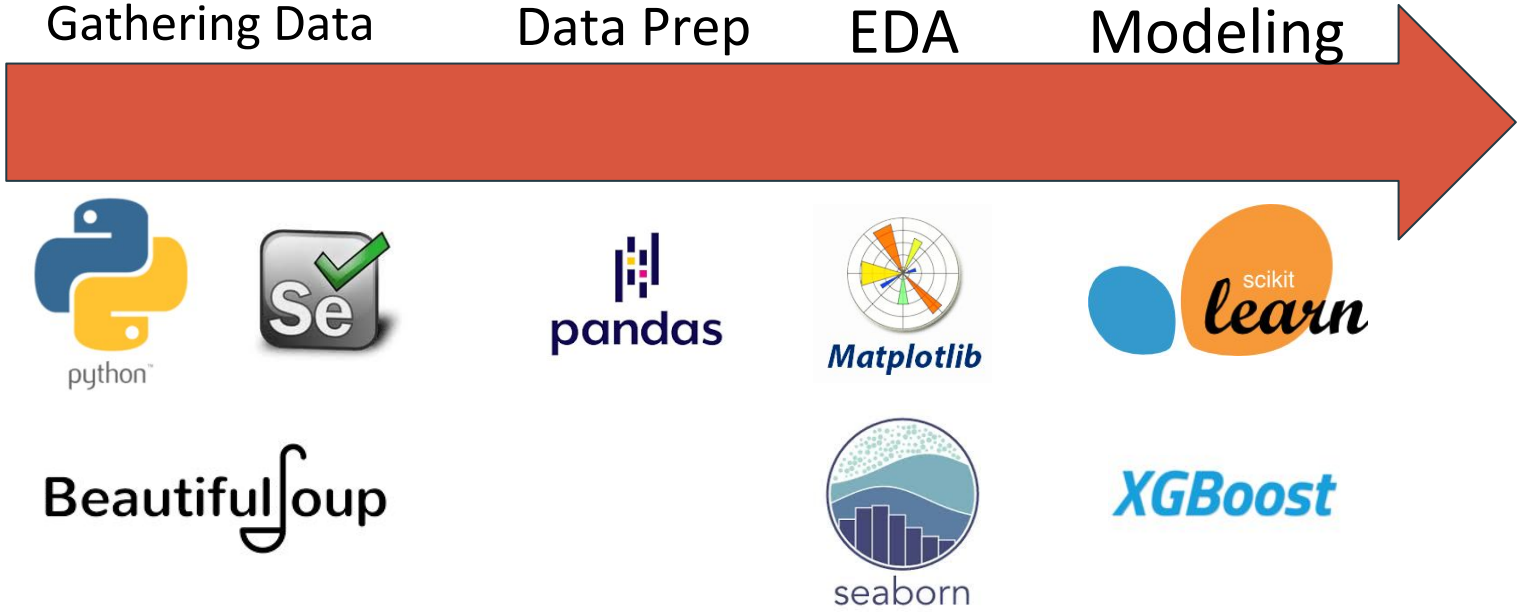
- 2020: Largest Wildfire Season in modern history
- 9.6 thousand incidents of wildfires
- 4.1 million Acres Burned
- 33 Fatalities
- 10.5 thousand Structures

Business Problem

- Weather forecast 24-72 hours

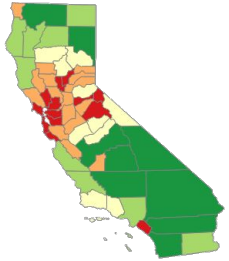


Method


Gathering Data
Data Prep
EDA
Modeling
python
Se
pandas
Matplotlib
scikit
learn
BeautifulSoup
seaborn
XGBoost

Data Understanding

CalFire Wildfire Incidents

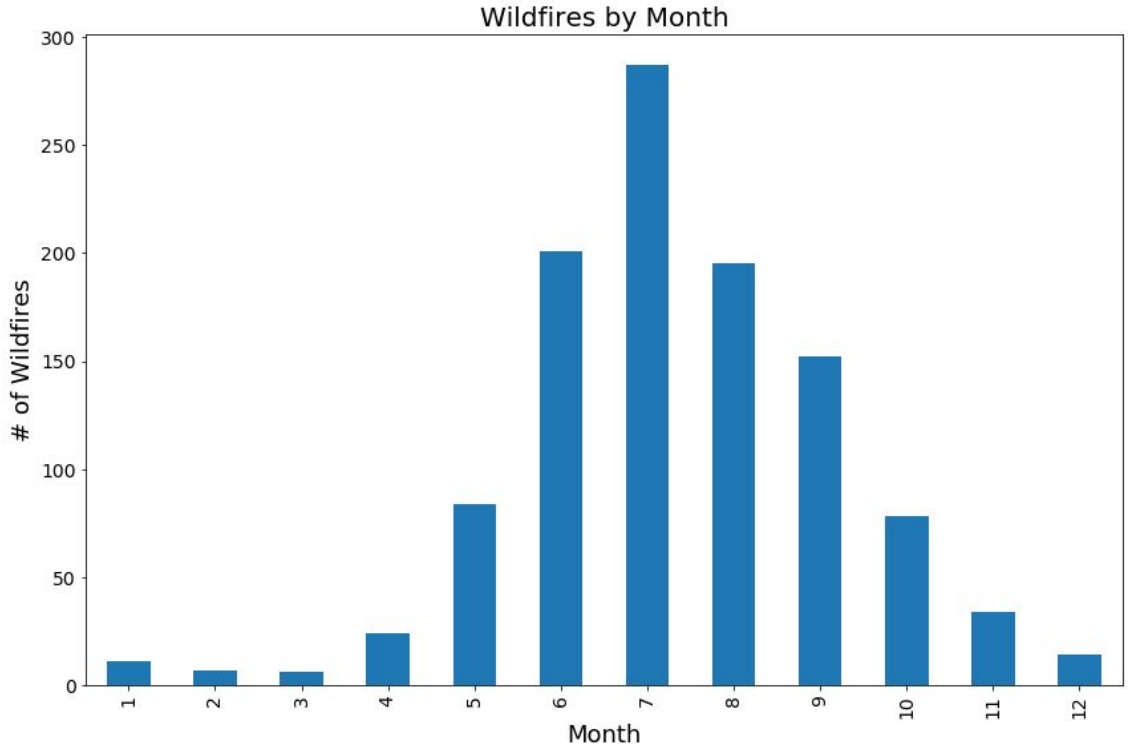


CalLands

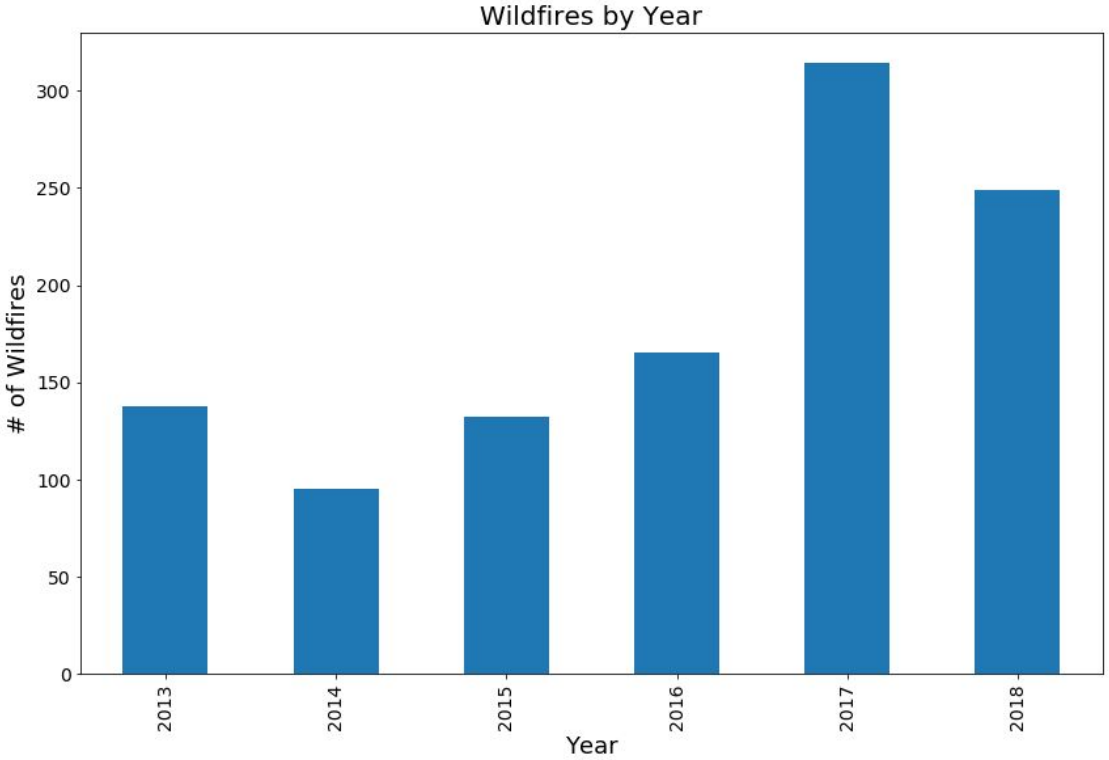
Anyplace America

CIMIS
Weather Underground

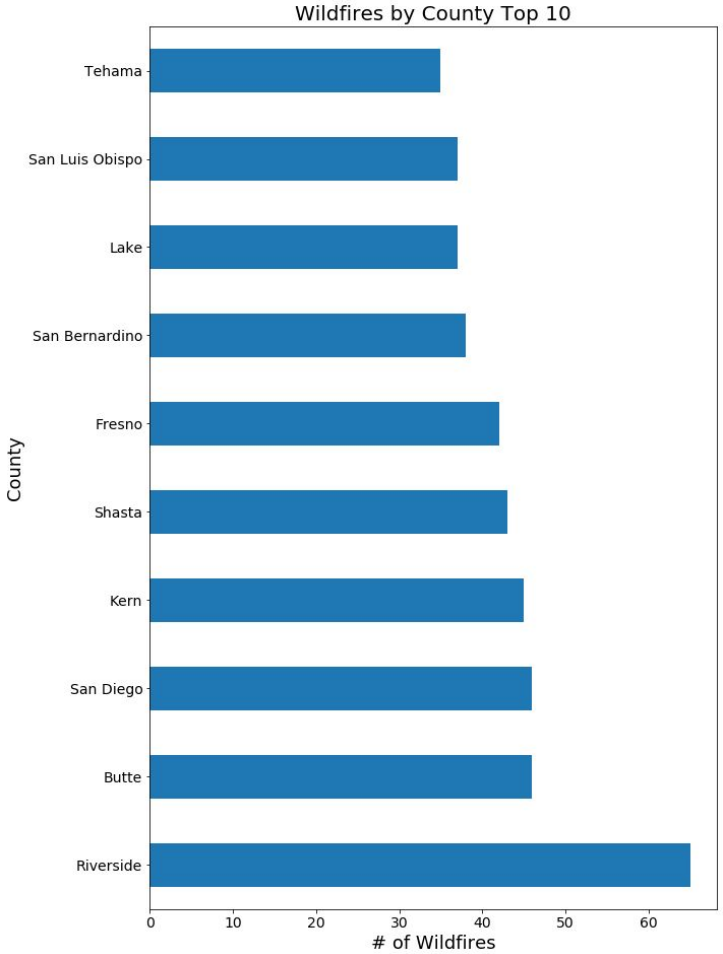
EDA


Wildfires by Month
# of Wildfires
300
250
200
150
100
50
0
1
2
3
4
5
6
7
8
9
10
11
12
Month

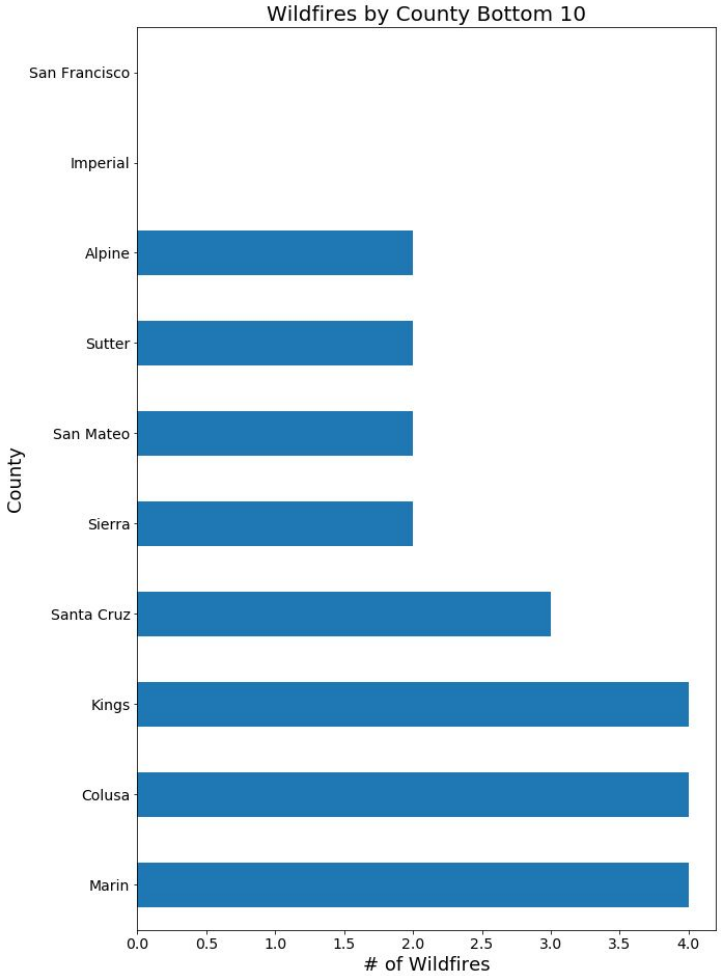
EDA


Wildfires by Year
# of Wildfires
0
50
100
150
200
250
300
2013
2014
2015
2016
2017
2018
Year

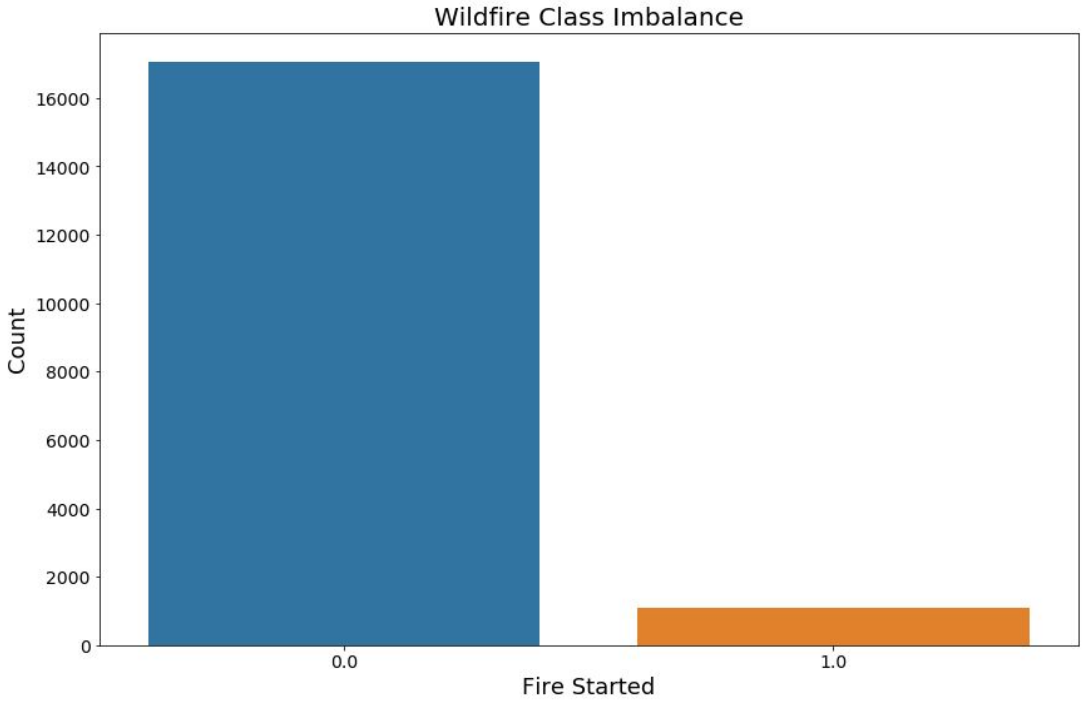
EDA


Wildfires by County Top 10
Tehama
San Luis Obispo
Lake
San Bernardino
Fresno
Shasta
Kern
San Diego
Butte
Riverside
County
0
10
20
30
40
50
60
# of Wildfires

EDA


Wildfires by County Bottom 10
San Francisco
Imperial
Alpine
Sutter
San Mateo
Sierra
Santa Cruz
Kings
Colusa
Marin
County
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
# of Wildfires

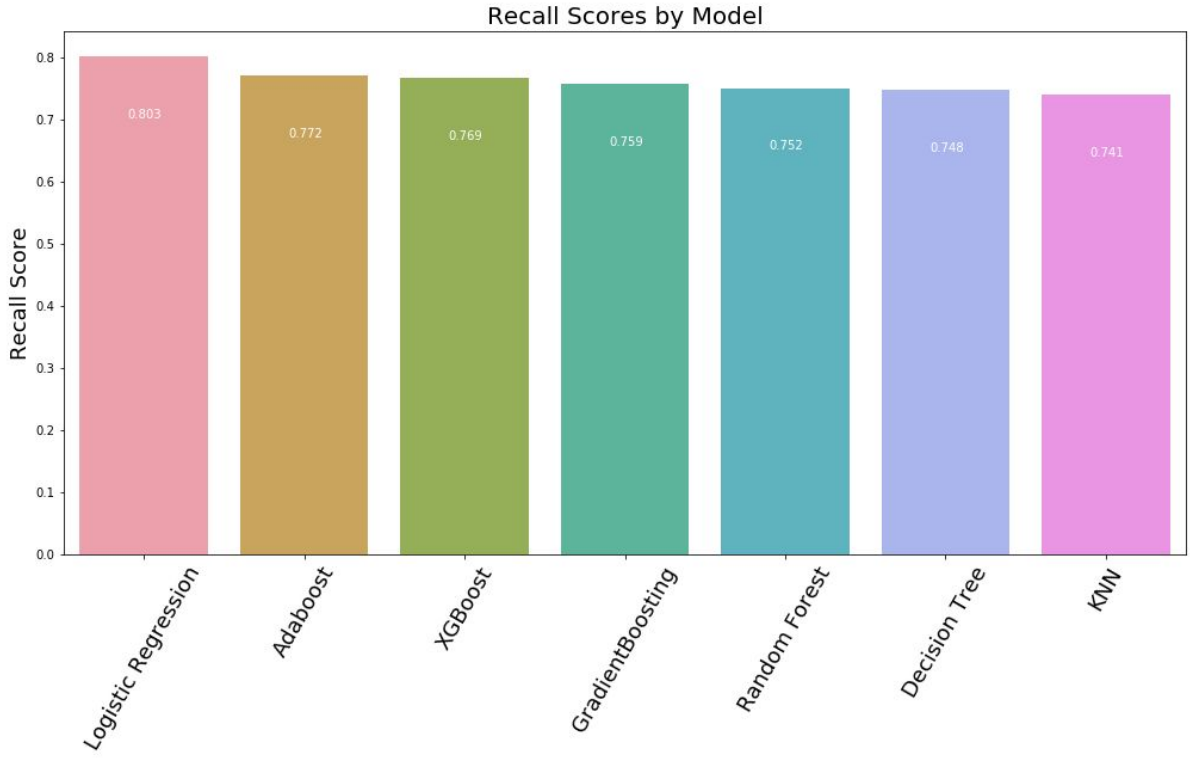
Modeling



UpSampling

Modeling

Recall - Minimizing False Negatives



Evaluation

Conclusion and Next Steps

- Model predicts 4/5 Wildfires
  - Use with CalFire Fire Warning System
- Future Steps
  - Feature Engineering
  - County and Industry Relevant Data

Thank You

Author: Malcolm Katzenbach

GitHub: Malcolm206

Email: mkatzenbach8@gmail.com