

Predicting Wildfire Size Category

Madeline Vossbrinck

AGENDA

- Background
- Data and Tools
- Methodology and Key Metrics
- Results
- Conclusions and Next Steps

BACKGROUND

- 2015 Fire Damage
 - 3,280 deaths
 - 15,700 injuries
 - \$15.1 billion in losses (2018 dollars)
 - Source: <https://www.usfa.fema.gov/data/statistics/#tab-4>
- Is it possible to predict which fires will become large so that more resources can be diverted to extinguishing them?

DATA SET

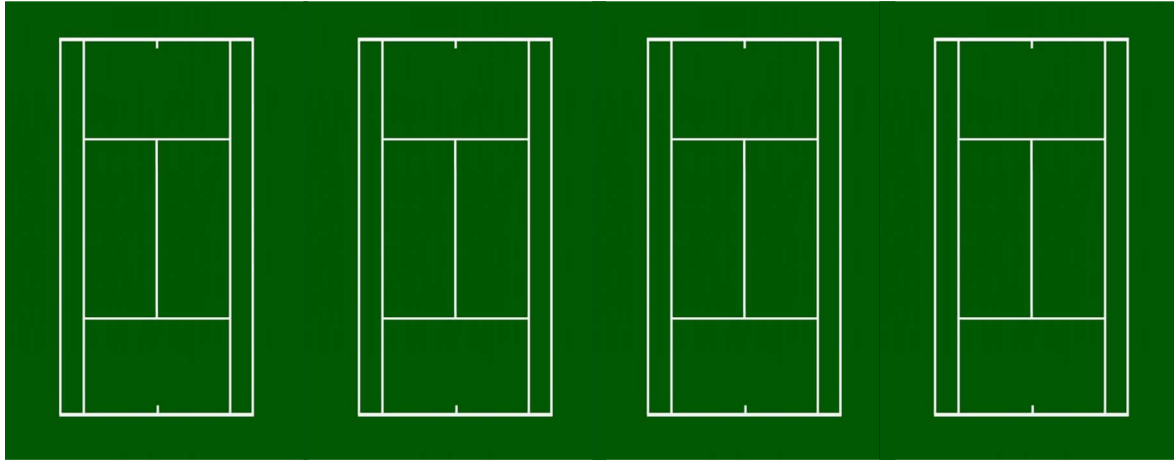
- US wildfires between 2000-2015
- At least 5 hours to contain
- At least 0.25 acres large
- Target:
 - Small fires (≤ 10 acres)
 - Large fires (> 10 acres)
- Features
 - Location
 - Fire Cause
 - Discovery Date

TOOLS

- Pandas
- Numpy
- Matplotlib
- Seaborn
- Scipy
- Statsmodels
- Sklearn
- XGBoost
- Tableau

SIZE COMPARISONS

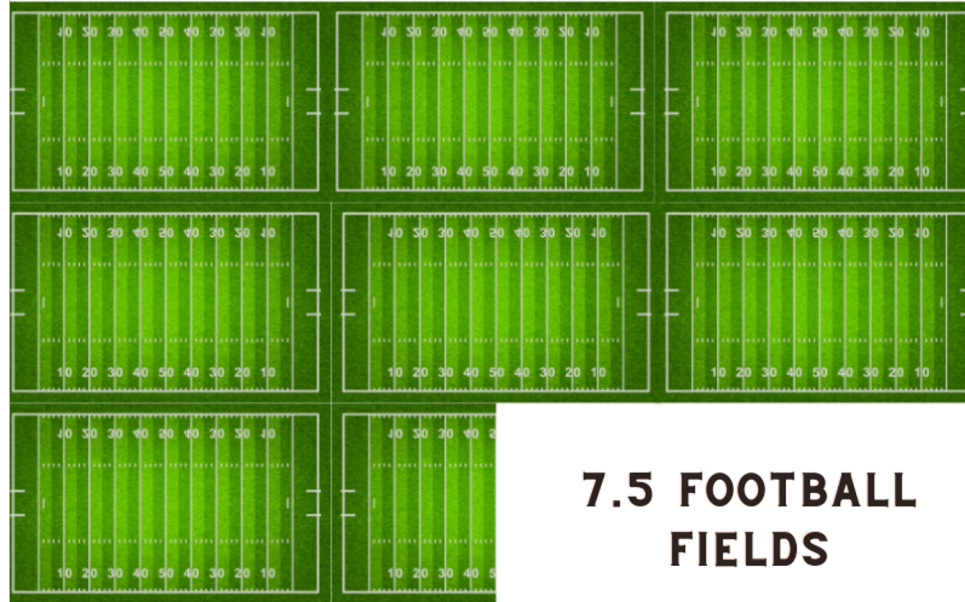
0.25 Acres



4 TENNIS COURTS

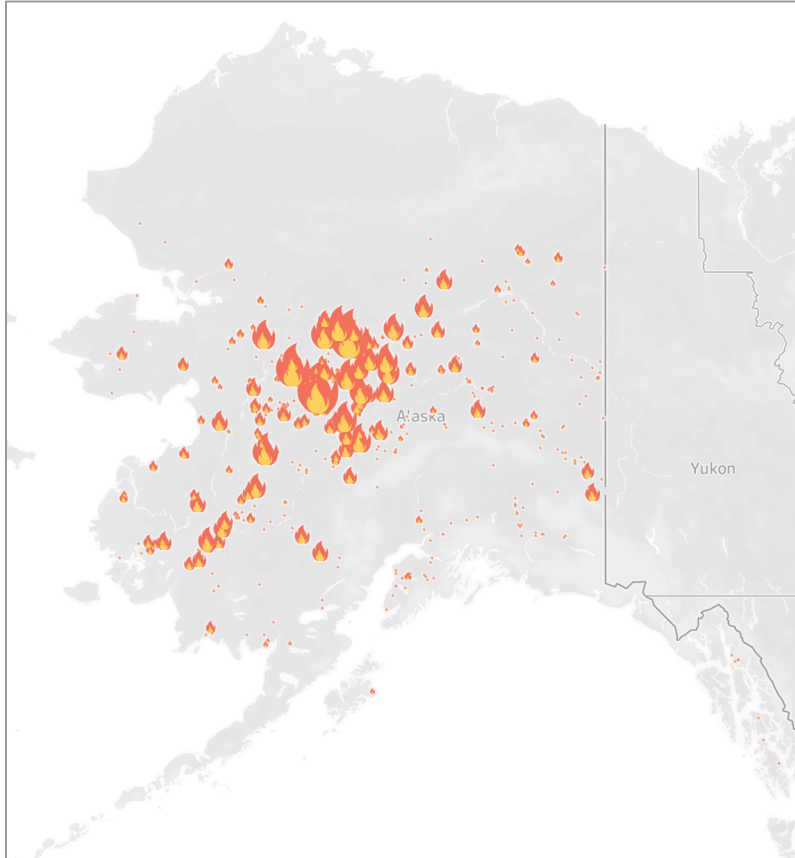
SIZE COMPARISONS

10 Acres



**7.5 FOOTBALL
FIELDS**

Alaska Fires of 2015



Acres Burned by State 2000-2015

Alaska	25,722,736
Idaho	9,976,198
California	7,768,999
Oregon	6,793,892
Nevada	5,929,174
Texas	4,801,147
Montana	4,685,209
Arizona	3,938,123
Washington	3,532,895
New Mexico	3,224,057

The largest fire was in 2012 in Oregon. It burned 558,198 acres which is larger than New York City and Los Angeles combined.

KEY METRICS

RECALL

What percent of the actual large fires are predicted correctly?

PRECISION

What percent of fires predicted to be large fires actually are large fires?

F2 SCORE

Balances recall with precision, weighs recall as twice as important

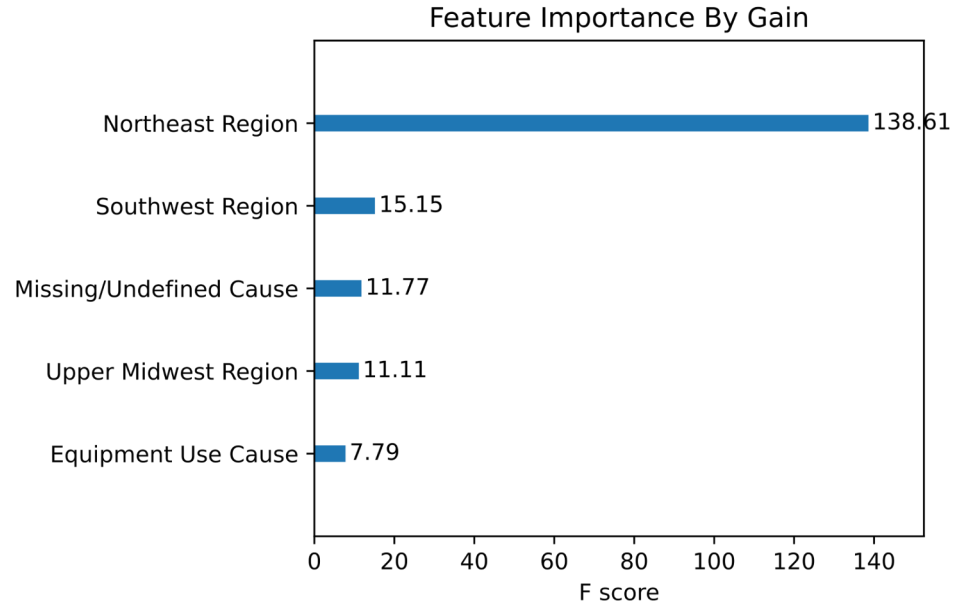
MODEL RESULTS

Model Type	Features	Recall	Precision	F2 Score
Logistic	Fire Cause, Region, and Discovery DOY	0.486	0.635	0.519
KNN (n=9)	Fire Cause, Region, and Discovery DOY	0.457	0.515	0.466
Random Forest	Fire Cause, Region, Discovery DOY, Discovery Time, Discovery Date, Latitude, and Longitude	0.610	0.695	0.625
XGBoost	Fire Cause, Region, Discovery DOY, Discovery Time, Discovery Date, Latitude, and Longitude	0.620	0.701	0.635

XGBOOST MODEL

Recall	Precision	F2 Score
0.620	0.701	0.635

n_estimators	30,000
max_depth	10
learning_rate	0.1
subsample	0.8
min_child_weight	2
colsample_bytree	0.8



PREDICTIONS

Fire Size Class	Size Range	% Correct Predictions
C	10 - 99.9 Acres	57%
D	100 – 299.9 Acres	68%
E	300 – 999.9 Acres	70%
F	1000 – 4999.9 Acres	62%
G	5000+ Acres	67%

CONCLUSIONS AND NEXT STEPS

Conclusions:

- Current model could be a tool that is useful in making a decision about resources
- Should be used in conjunction with other tools and information not available in data set

Next Steps:

- Acquire updated data
- Work on improving predictions on largest fires
- Try predicting fire size category based on other thresholds
- Use multi-class target

Questions?