Predicting Fire Severity

David Coons, Jessely Enriquez, David Marmor

Problem Statement

Wildfires are a serious problem faced by communities around the country. In 2021 there were over 50,000 wildfires nationwide (U.S. Wildfire Statistics | Bankrate). Due to the effects of climate change wildfires will only get worse.

As citizens against climate change, we as a team wanted to provide a tool to identify and predict how severe a fire will be based on a variety of factors. By predicting the severity of a fire, we can then give these communities time to prepare and take preventative measures. Some of the considerations that we want to include when predicting these communities of vulnerability are their temperature, precipitation, drought, and location.

Features of Interest

- Temp
- Precipitation
- Acres burned (used to make target variable)
- Drought intensity
- Year/month
- Lat/long/state & state area/county
- Population

Dataset

Sources:

- State Size (Ref)
- Population Estimates (Ref)
- US Weather Events (Ref)
- National Interagency Fire Center (Ref)
- U.S. Drought Monitor (Ref)

Data cleaning

- Involve:
    - averaging to month per year scale
    - imputing nulls
    - combining multiple data tables

Target Variable

The target variable is Fire Classification based on acres burned. We used the classification system used by the National Wildfire Coordinating Group.

Value	Description
A	Greater than 0 but less than or equal to 0.25 Acres
B	0.26 to 9.9 Acres
C	10.0 to 99.9 Acres
D	100 to 299 Acres
E	300 to 999 Acres
F	1000 to 4999 Acres
G	5000 to 9999 Acres
H	10000 to 49999 Acres
I	50000 to 99999 Acres
J	100000 to 499999 Acres
K	500000 to 999999 Acres
L	1000000 + Acres

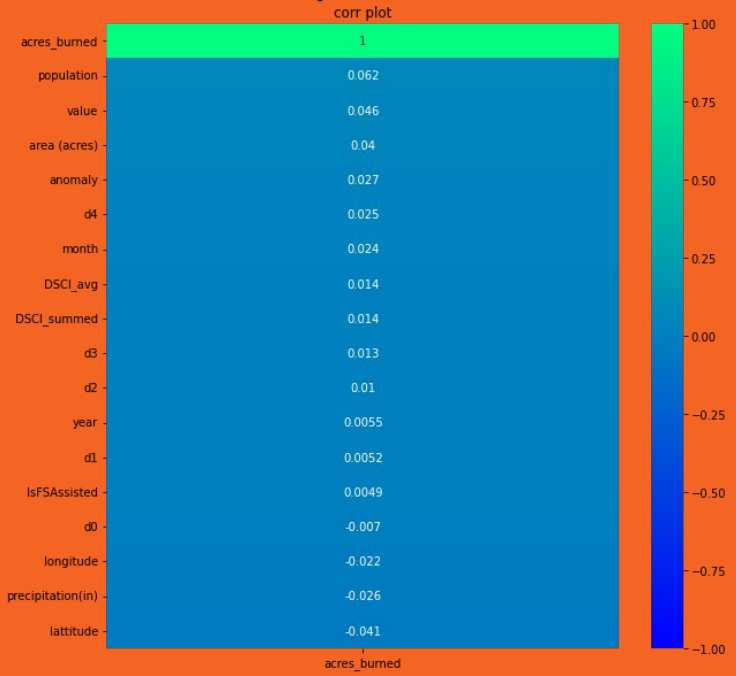
We combined classes J, K, and L since there are very few K and L fires.

```
dummified_1['fire_class']

0    68627
1    48909
2    31448
4    22336
5    22086
3    18613
7      804
6      688
8      155
9      121
Name: fire_class, dtype:
```

EDA

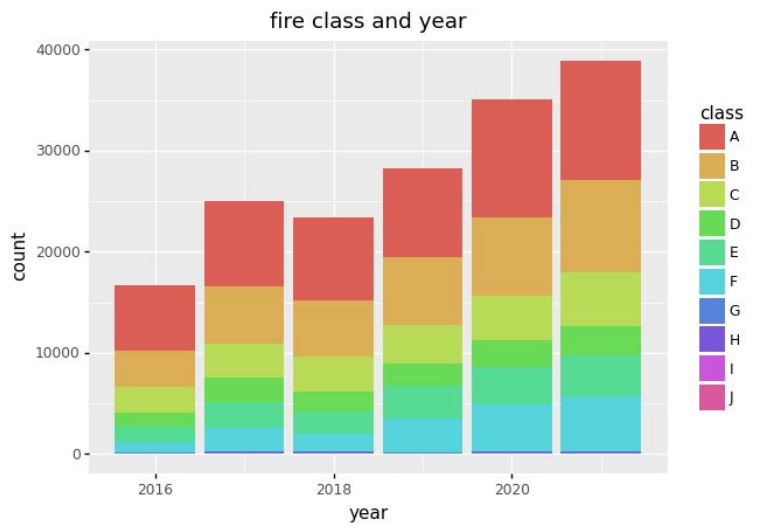
- heat map

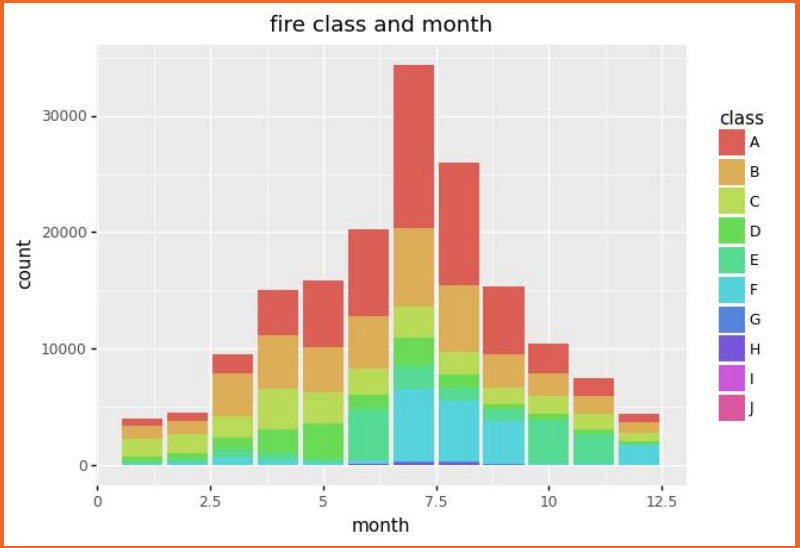


None of our variables are highly correlated with acres burned

EDA

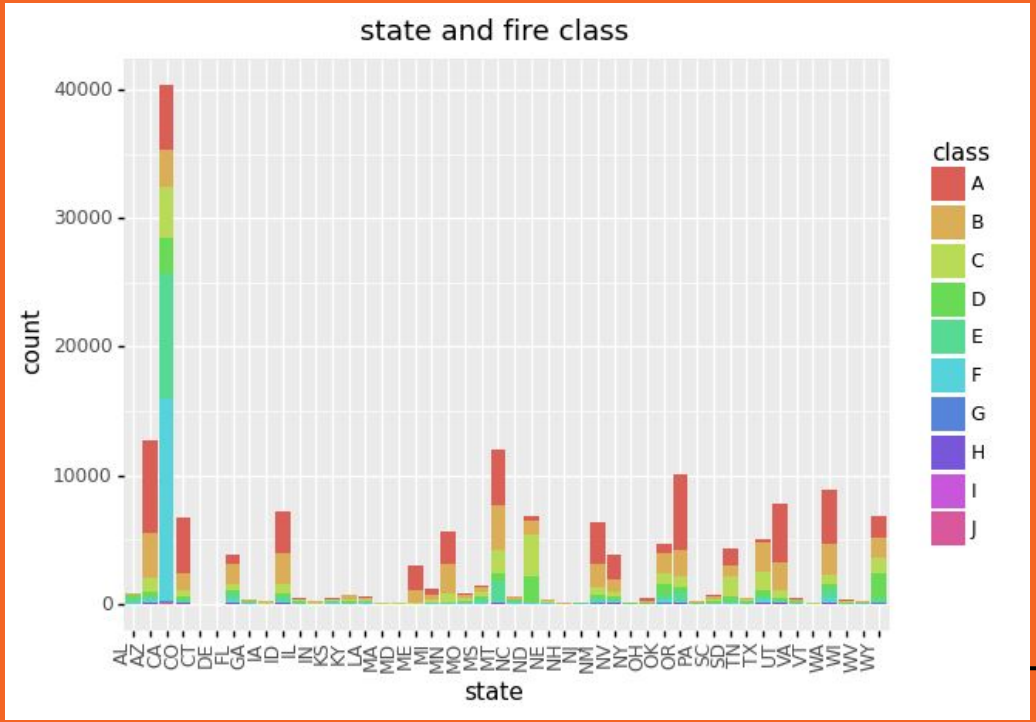
- target and time period





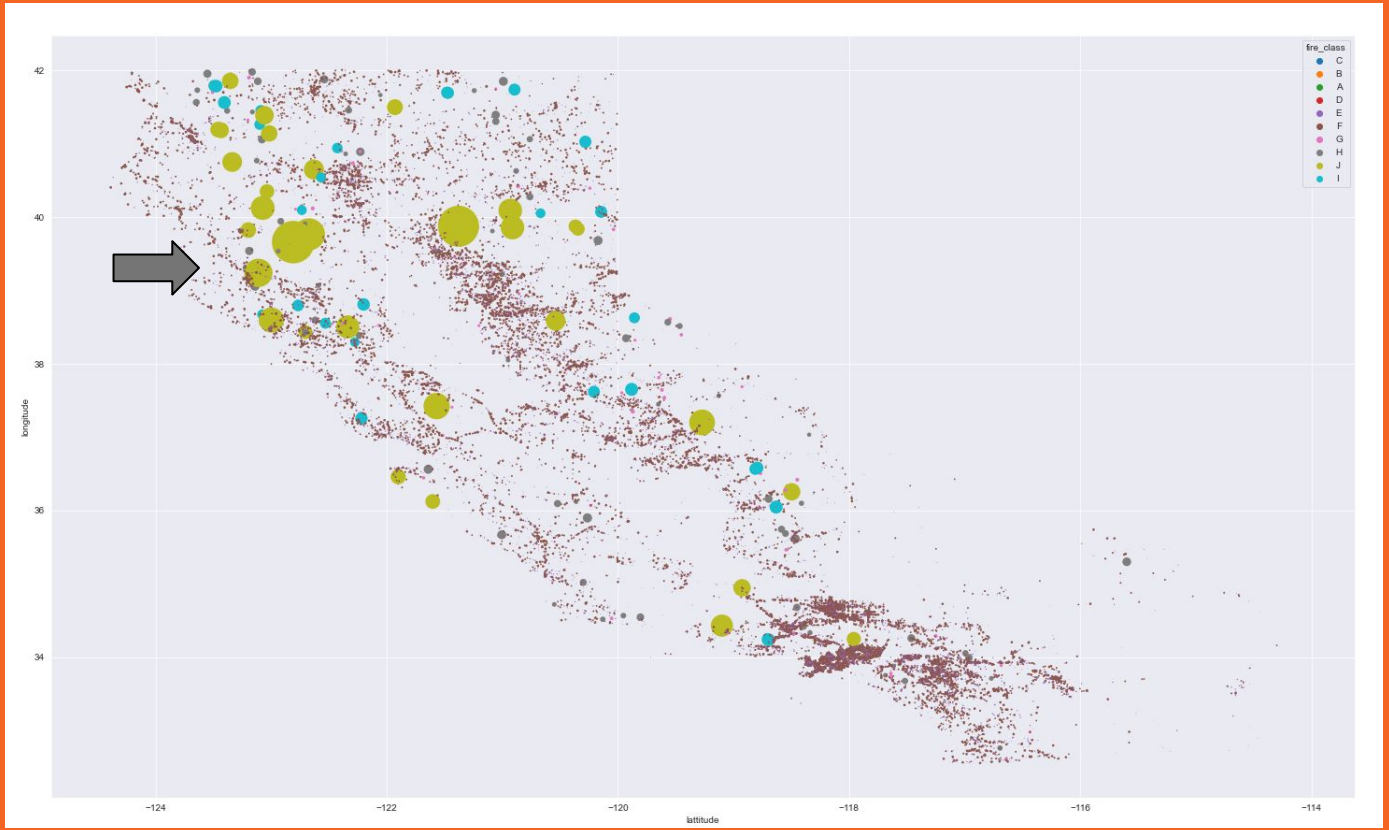
EDA

- target and region



EDA

- target and location


fire_class
C
B
A
D
E
F
G
H
J
I
42
40
38
36
34
longitude
−124
−122
−120
−118
−116
−114
lattitude

Modeling

Models used:

- random forests, gradient boosting, adaboosting & xgboost
- neural network
- knn

The best models were xgboost and random forests.

Our target metric is multiclass f1 score. Because our classes are unbalanced accuracy is a misleading metric. Multiclass f1 score gives us a better sense of how well the model can predict all classes.

Multiclass f1 score

- macro is avg f1 score (unweighted)
- micro global avg total sums of each class tp/tn/fp/fn
- weighted avg

Oversampling

```
dummified_1['fire_class']

0    68627
1    48909
2    31448
4    22336
5    22086
3    18613
7      804
6      688
8      155
9      121
Name: fire_class, dtype:
```

```
What class do we want from needs to be an INT(our case ~ 0-9)? 5
303836
0    68627
1    48909
2    31448
4    22336
3    22086
5    22086
6    22086
7    22086
8    22086
9    22086
Name: fire_class, dtype: int64
```

Model Metrics

model	Test f1 score
Random Forests	78%
KNN	77%
XGBoost	72%
Neural Network	40%
Gradient Boosting	38%
Adaboosting	20%

Random Forests model metrics

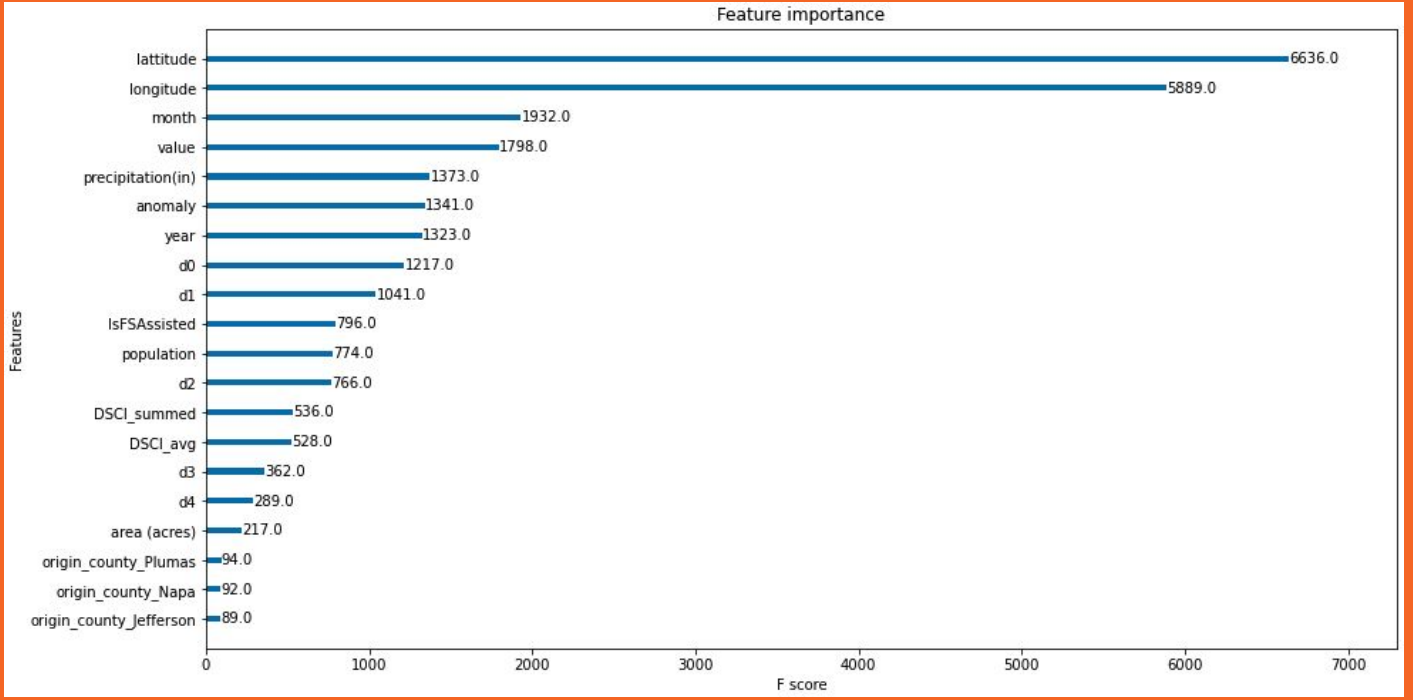
Test accuracy	.71
Test f1-score	.78

XGBoost Summary

Final Metrics	
Test accuracy	68%
Test f1-score	72 %

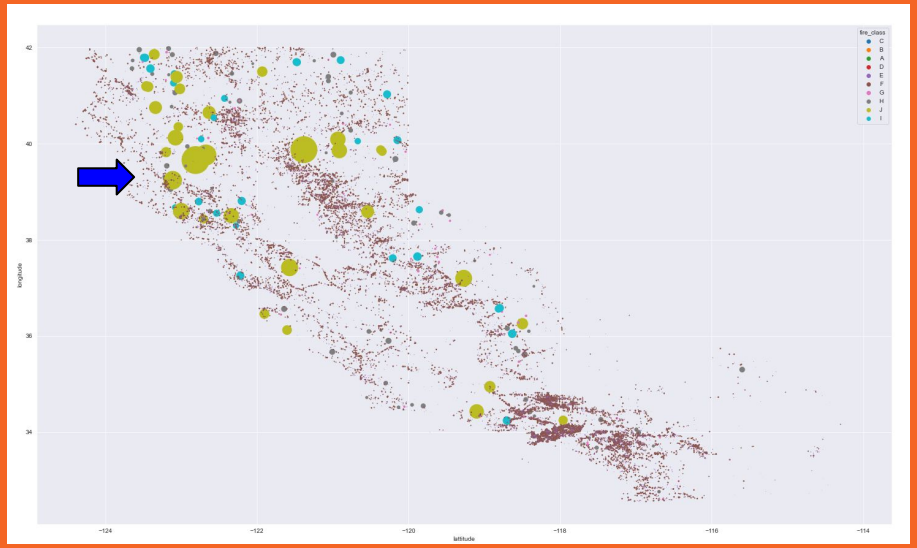
Oversampling Class Size of 3

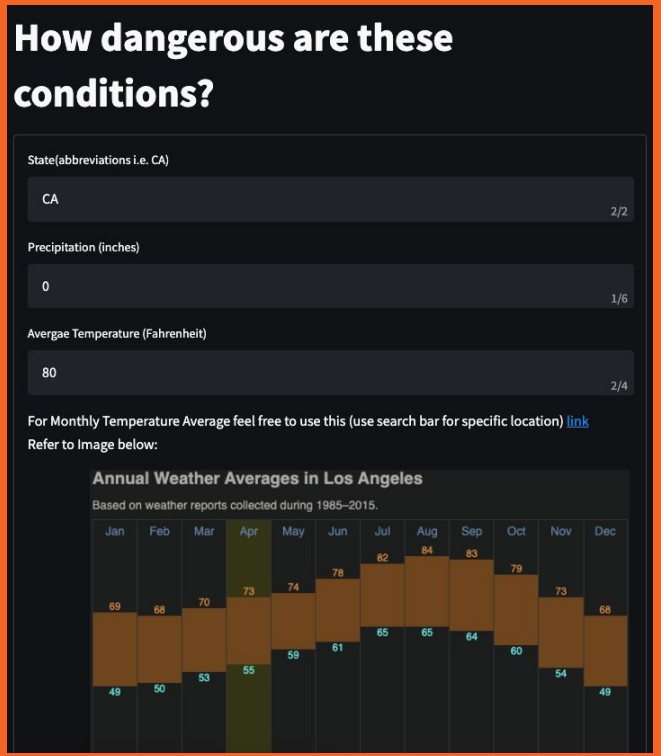
Parameters	
n_estimators	400
max depth	4
objective	'multi:softmax'



Feature Score = F score

Streamlit Demo (link)





Conclusion

The factors that seem to have the most impact on fire class are :

- Location
- Time
- Drought Conditions

Forest services, state and local governments, and concerned citizens can use our model to assess potential risk of severe fire using our app

Further Work

- Expand backend calculations based on form (highest class of 5)

- Interactive dashboard

- Times series forecasting

- Predicting probability of a fire occurring

The model should be updated yearly to maintain its usefulness.

Thank you!

Any questions?