ANALYZING WILDFIRES

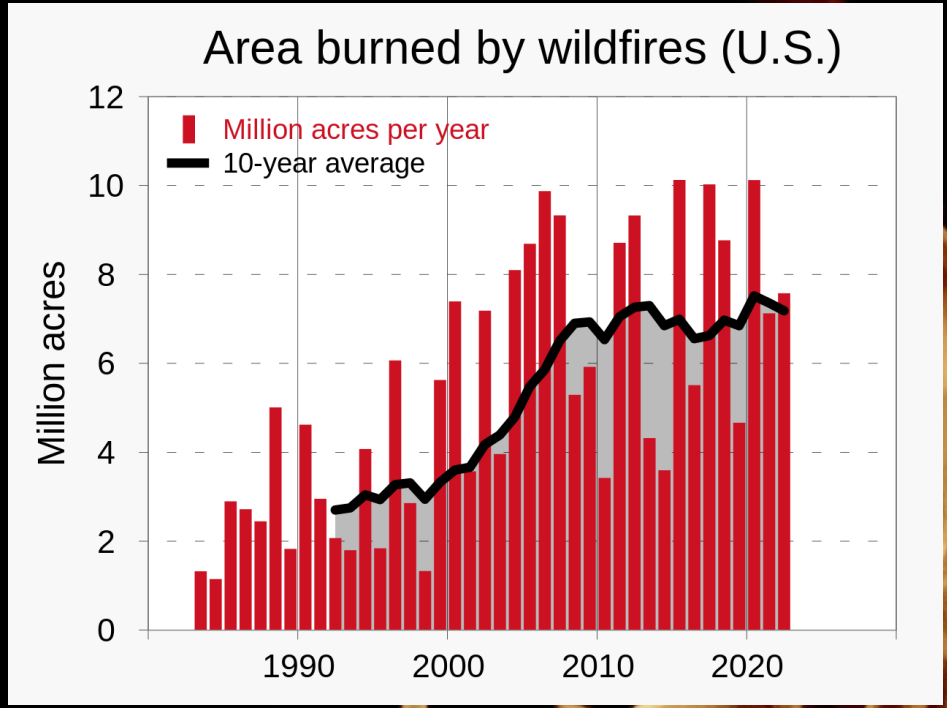
Group 2:
Celina E., Kent R., Pankaj S.,
Sujatha

BACKGROUND

Wildfires are increasing in intensity, severity, size and duration.

Western, mountainous states bear the highest burden.

Increased temperatures and changes in weather patterns will continue contributing to wildfires.

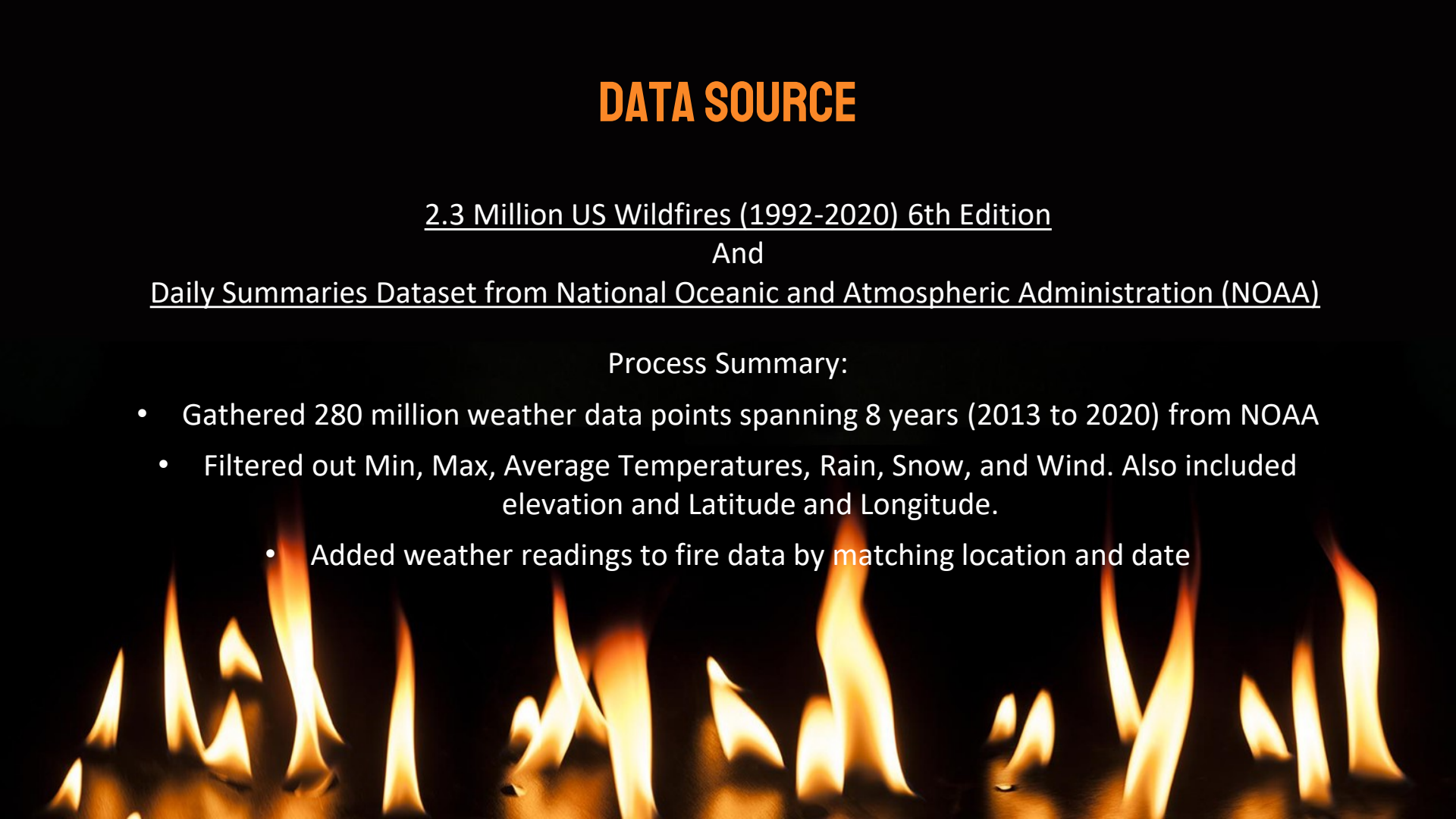


DATA SOURCE

2.3 Million US Wildfires (1992-2020) 6th Edition
And
Daily Summaries Dataset from National Oceanic and Atmospheric Administration (NOAA)

Process Summary:

- Gathered 280 million weather data points spanning 8 years (2013 to 2020) from NOAA
- Filtered out Min, Max, Average Temperatures, Rain, Snow, and Wind. Also included elevation and Latitude and Longitude.
- Added weather readings to fire data by matching location and date



DATA SUMMARY

Final Dataset had 955k weather readings for both Fire and No Fire present

Fire Size	Ave Temp (C)	Max Temp (C)	Min Temp (C)	Precipitation (10ths of mm)	Snow (mm)	Snow Depth (mm)	Wind (10ths meters/second)	Elevation	Days to Contain
No Fire	10.8	16.1	5.5	31.4	2.9	24.3	37.5	380.7	N/A
A	18.7	26.1	11.4	8.7	0.2	0.7	35.0	566.3	0.6
B	17.4	25.0	10.3	8.5	0.1	0.4	35.3	333.3	0.5
C	17.1	25.0	9.7	8.2	0.1	0.2	37.0	361.2	1.1
D	18.8	26.9	10.9	7.9	0.1	0.1	39.6	506.1	3.2
E	19.8	28.0	11.7	6.5	0.0	0.0	40.6	587.0	6.2
F	21.8	30.1	13.4	5.4	0.1	0.0	40.8	674.3	12.4
G	23.5	31.5	14.8	3.9	0.0	0.0	37.4	655.0	31.0

TECHNOLOGY

Pandas

SQLite

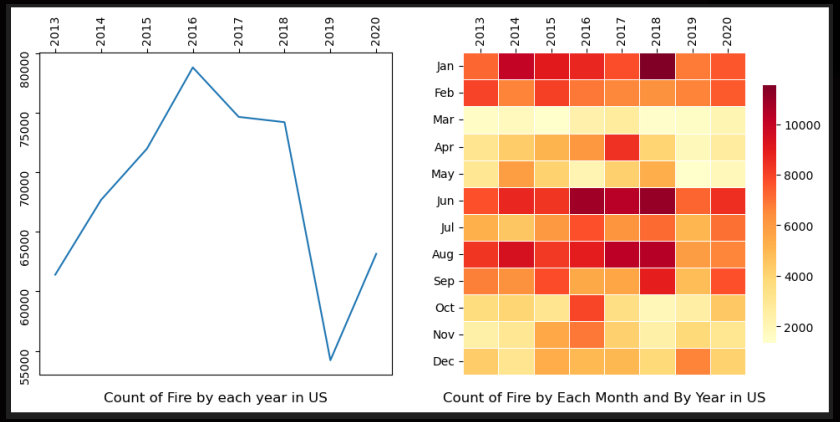
Matplotlib

Tableau

ML Ski-kit learn

Random forest &
Neural Network

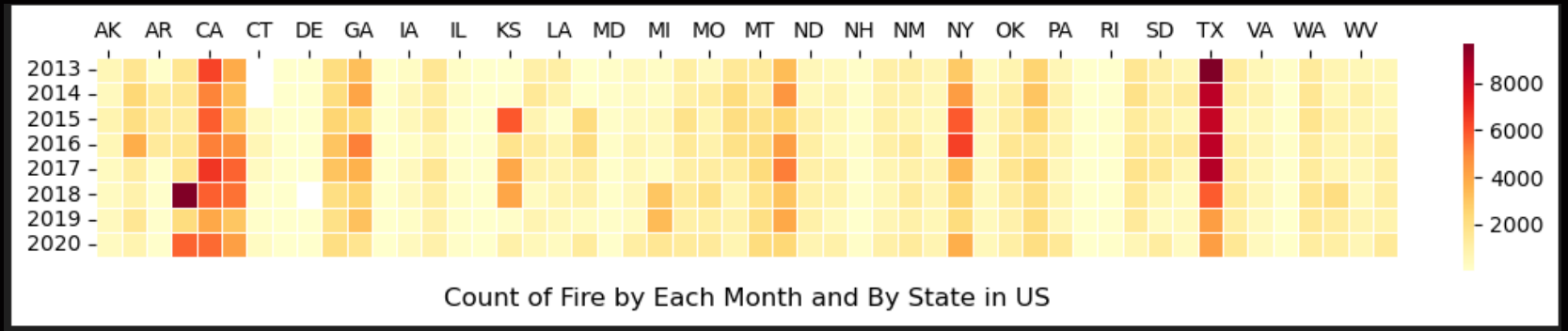
ANALYZING WILDFIRE TREND 2013 TO 2020



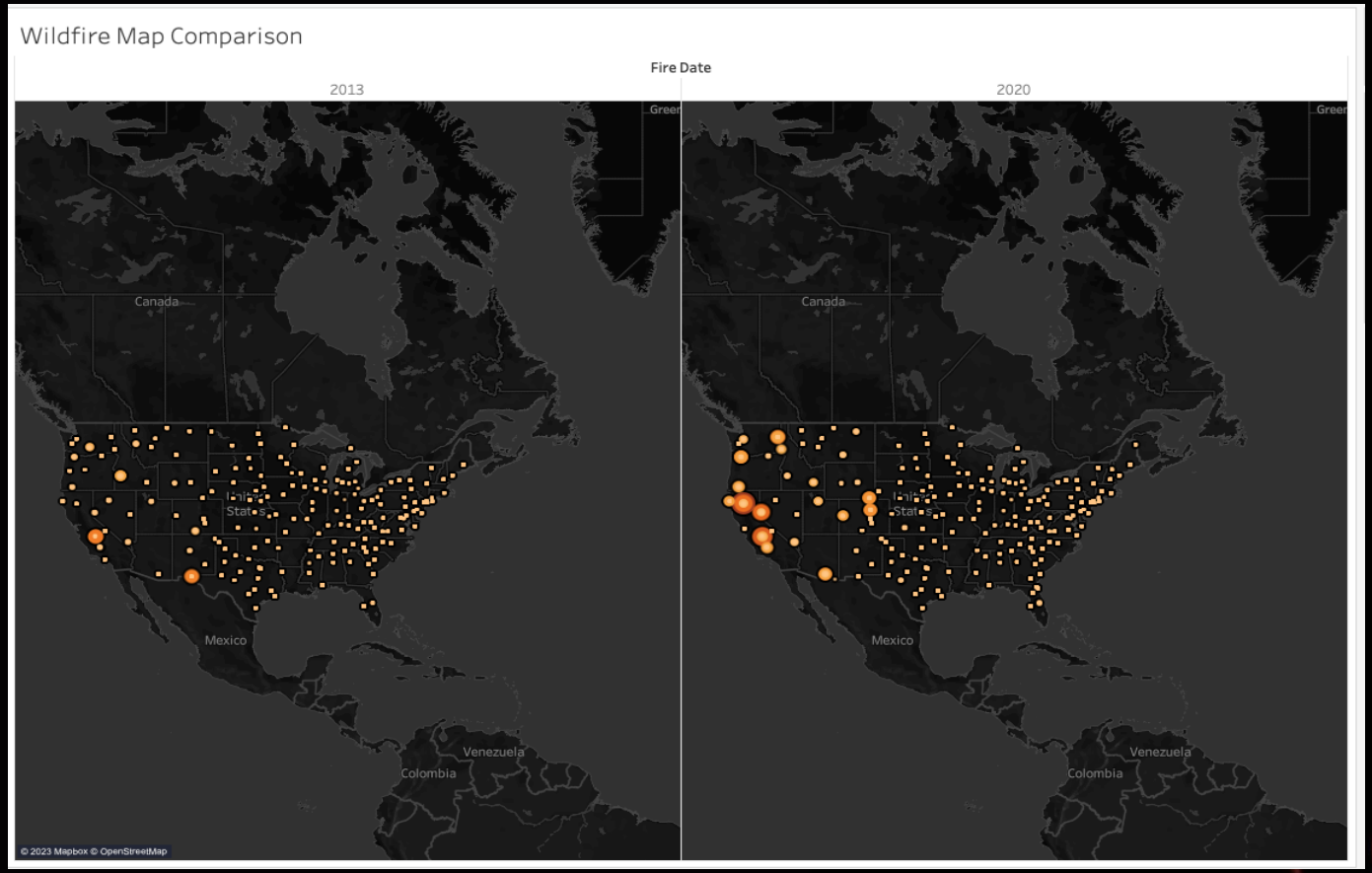
Wildfire

Wildfire seemed to be higher in 2016

California and Texas seems to has more fires compared to any other states



WILDFIRE MAP COMPARISON



In 2013, wildfire incidents in the United States numbered 46,615, consuming a total of 4,307,176 acres. Fast-forward to 2020, the scale of devastation has multiplied, witnessing over double the acreage burned. A staggering 10,274,679 acres were ravaged by wildfires that year, coinciding with a substantial rise in the number of fires, reaching 58,258 incidents.

"Nature is not a place to visit. It is home."
- Gary Snyder

ANALYZING CLIMATE FACTORS AFFECTING WILDFIRE

As of June 1, 2023, around 18,300 wildfires have impacted over half of million (511,000) acres this year.

Wildfires have the potential to harm property, livelihoods, and human health. Fire related threats are increasing, especially as more people live in and around forests, grasslands, and other natural areas. Over the past few decades, the United States has routinely spent more than $1 billion per year to fight wildfires, including $2.3 billion in 2020.*EPA data

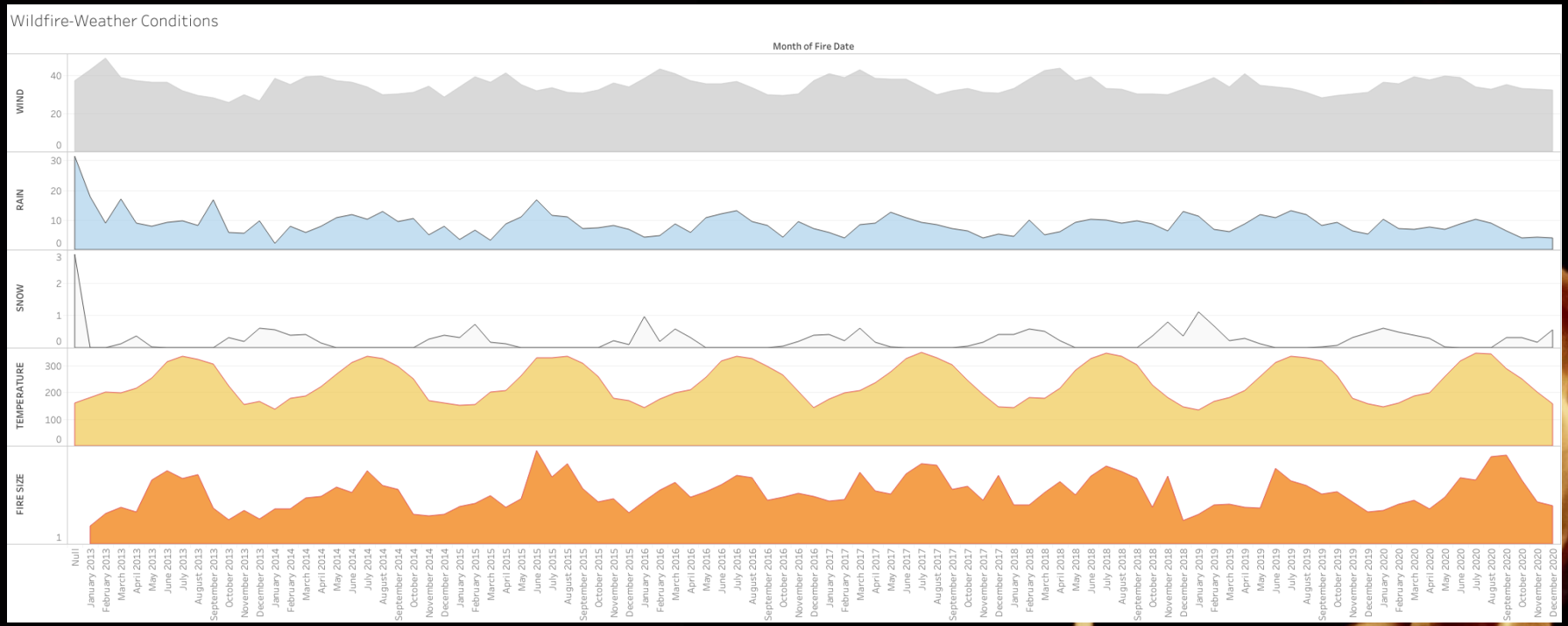
ANALYZING CLIMATE TREND 2013 TO 2020

SUMMARY

with higher temperature and higher wind fire count has significantly increased.

It is evident that climate has impact on wildfire

WILDFIRE CLIMATE CONNECTION



Climate change, including increased heat, extended drought, and a thirsty atmosphere, has been a key driver in increasing the risk and extent of wildfires in the western United States during the last two decades. Wildfires require the alignment of a number of factors, including temperature, humidity, and the lack of moisture in fuels, such as trees, shrubs, grasses, and forest debris. All these factors have strong direct or indirect ties to climate variability and climate change.

RANDOM FOREST MODEL

- Used Random Forest Classifier to predict the occurrence of a fire
- Columns specific to Fire like Fire Size, Fire Classification etc. were dropped and remaining data was standardized and split into Test and Train datasets
- Random Forest Classifier was created with n_estimators = 500 and random_state = 78 and used for the fire occurrence prediction

RANDOM FOREST - MODEL METRICS

Following is the confusion matrix of the model:

	Predicted 0	Predicted 1
Actual 0	79411	13909
Actual 1	10972	134493

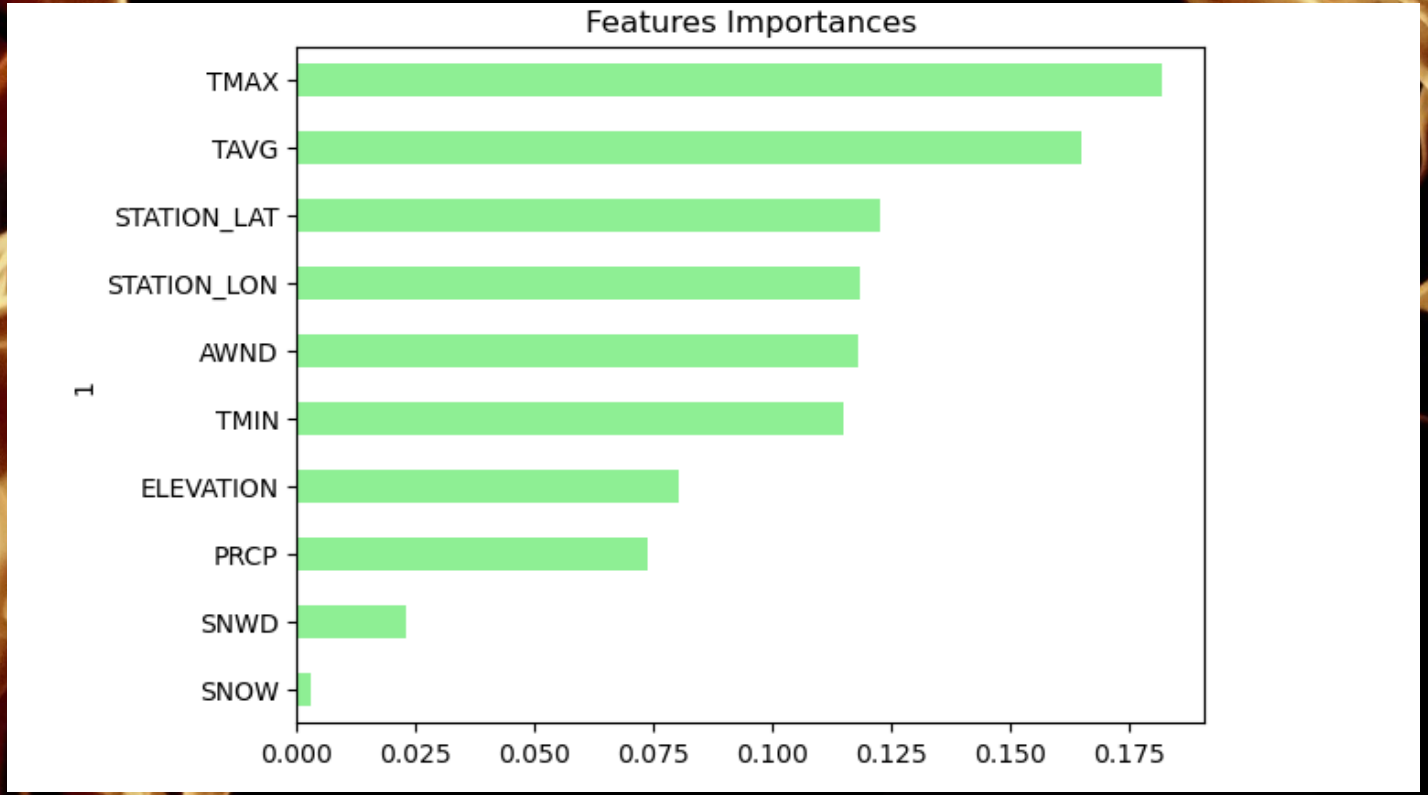
```
Accuracy Score : 0.8958016625834956
Classification Report
              precision    recall  f1-score   support

           0       0.88      0.85      0.86     93320
           1       0.91      0.92      0.92    145465

    accuracy                           0.90    238785
   macro avg       0.89      0.89      0.89    238785
weighted avg       0.90      0.90      0.90    238785
```

- The model does a pretty good job for both positive and negative case with an overall accuracy, recall and f1-score of about 90%
- The scores are slightly better for positive cases probably because of the volume of data of positive cases being higher than the negative ones

RANDOM FOREST – FEATURES IMPORTANCES



RANDOM FOREST - OPTIMIZATIONS

1) Our initial dataset used for the model had the columns populated only for positive (Fire occurred) cases. As a result, it was predicting the Fire occurrence with 100% accuracy. This obviously was a flawed dataset for our purpose.
2) In the next iteration, we dropped the positive case – only columns. The data set had weather columns - Temperature, Precipitation, latitude, longitude and elevation. The period of the dataset was 1 year.
3) With this dataset, we had the following confusion matrix:

	Predicted 0	Predicted 1
Actual 0	88647	8127
Actual 1	7840	10384

Accuracy Score : 0.8611541070279483
Classification Report

	precision	recall	f1-score	support
0	0.92	0.92	0.92	96774
1	0.56	0.57	0.57	18224
accuracy			0.86	114998
macro avg	0.74	0.74	0.74	114998
weighted avg	0.86	0.86	0.86	114998

RANDOM FOREST - OPTIMIZATIONS

- This dataset gave an overall accuracy of 86%. However, the accuracy for positive cases was way lower (56%) than negative cases(92%).

- In the final iteration, we introduced 7 years more worth of data. In addition, additional parameters like Wind and Snow were introduced. This final dataset gave us an accuracy of 90%

NEURAL NETWORK MODEL

MODEL I

39%
accuracy

Preprocessing :

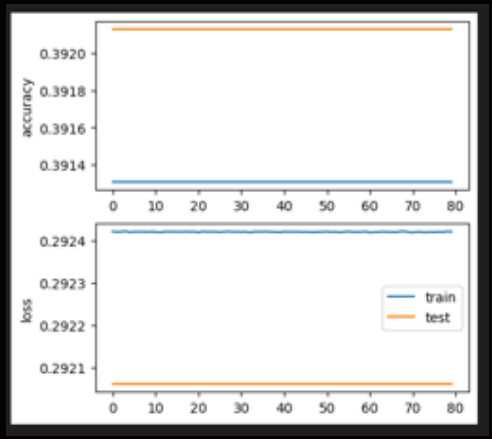
- MinMaxScalar as activation function for scaling each feature
- Train the model using fit

Model

- 4 hidden layer
- Higher neurons units(ranging between 1 to 5)
- Relu as activation function
- Output layer with one neuron and “Sigmod” as activation function

Model compilation

- Loss function: “mean_squared_logarithmic_error”
- Optimizer: ‘adam’
- Metrics : accuracy



NEURAL NETWORK MODEL

MODEL 2

85%
accuracy

Preprocessing :

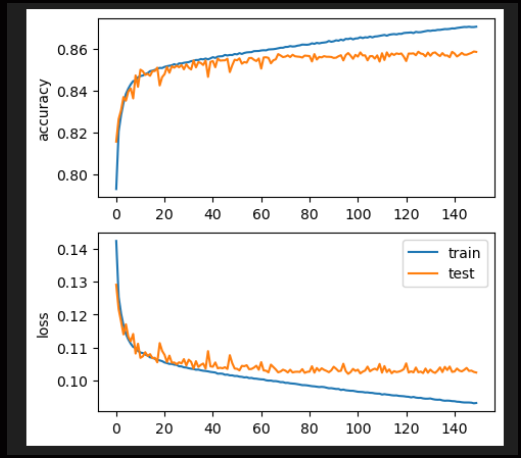
- MinMaxScalar as activation function for scaling each feature
- Train the model using fit

Model

- 5 hidden layer
- Higher neurons units(ranging between 60 to 600)
- Relu as activation function
- Output layer with one neuron and “Sigmod” as activation function

Model compilation

- Loss function: “mean_squared_error”
- Optimizer: ‘adam’
- Metrics : accuracy



NEURAL NETWORK MODEL

MODEL 3

86%
accuracy

Preprocessing :

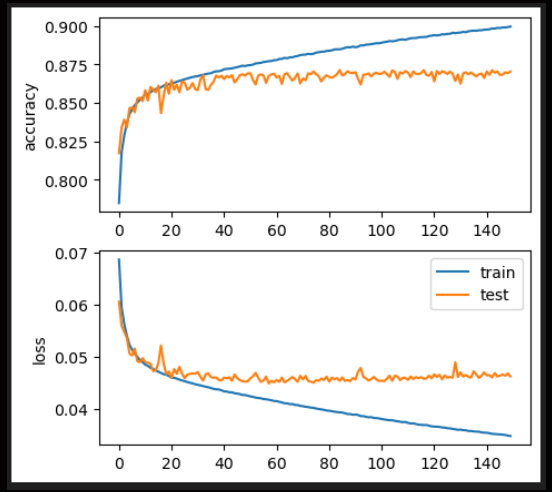
- MinMaxScalar as activation function for scaling each feature
- Train the model using fit

Model

- 5 hidden layer
- Higher neurons units(ranging between 100 to 1100)
- Relu as activation function
- Output layer with one neuron and “Sigmod” as activation function

Model compilation

- Loss function: “mean_squared_logarithmic_error”
- Optimizer: ‘adam’
- Metrics : accuracy



SUMMARY

Using the Machine learning and Neural network, optimized tensor flow model reached accuracy of 86%.
Based on the optimization approach it is recommended

* To have higher neuron units and minimum of 5 hidden layers column
* Using mean_squared_logarithmic_error would significantly reduce loss
* Have a minimum of 8 features for evaluation as reduced input features results in lower accuracy.

Overall, this model performed pretty good and would be produce a reliable fire prediction based on climate factor.

QUESTIONS ?

THANKS!

QUESTION ?