Deep Learning for Wildfire Risk Prediction

Caleb Maxie and Koi Cooks

Introduction: Wildfires

Wildfires are becoming more frequent, more intense, and more destructive driven by prolonged droughts, and human activities.

In the face of this growing crisis, early and accurate prediction is critical. Our research explores the power of deep learning and classical machine learning models to forecast wildfire occurrences before they escalate. By training models on weather patterns, seasonal trends, and environmental indicators, we aim to build a system that can assist emergency responders, guide resource deployment, and ultimately help save lives and landscapes.

Through this work, we're not just responding to the wildfire crisis we're working to stay ahead of it.

Ways it can be Useful

Wildfires pose serious threats to people, ecosystems, and economies. Predictive modeling especially using machine learning and deep learning can provide critical insights that help reduce those impacts. Here are several ways our wildfire prediction system can be useful:

Dataset 1 Overview

Provides a comprehensive compilation of weather observations and wildfire data in California from from 1984 to 2025. Features include:

- DATE
- PRECIPITATION
- MAX_TEMP
- MIN_TEMP
- AVG_WIND_SPEED
- FIRE_START_DAY
- YEAR
- TEMP_RANGE
- WIND_TEMP_RATIO
- MONTH
- SEASON
- LAGGED_PRECIPITATION
- LAGGED_AVG_WIND_SPEED

Dataset 2 Overview

CAL FIRE Historical Wildland Fire Perimeters nominated to Living Atlas. The service includes layers that are data subsets symbolized by size and year. Hosted on CAL FIRE AGOL from 1878 to 2025

Features include:

- OBJECTID
- Year
- State
- Agency
- Unit ID
- Fire Name
- Local Incident Number
- Alarm Date
- Containment Date
- Cause

Combine Datasets

We decided to combine the datasets into one that had all the features from the first one and includes fire name, longitude and latitude.

5/13/2017	0	70	56	8.72	TRUE	2017	14	0.124571	5	Spring	133	APACHE	-1.3E+07	4130789	0.13	8.884286
4/14/2013	0	63	56	6.49	TRUE	2013	7	0.103016	4	Spring	104	APPLE	-1.3E+07	4179005	0	9.461429
4/15/2013	0.06	64	54	9.4	FALSE	2013	10	0.146875	4	Spring	105	APPLE	-1.3E+07	4179005	0.06	8.407143
4/16/2013	0	65	51	13.2	FALSE	2013	14	0.203077	4	Spring	106	APPLE	-1.3E+07	4179005	0.06	8.311429
6/9/2018	0	75	61	8.5	TRUE	2018	14	0.113333	6	Summer	160	APPLE	-1.4E+07	4855123	0	7.861429
6/10/2018	0	73	62	6.93	FALSE	2018	11	0.094932	6	Summer	161	APPLE	-1.4E+07	4855123	0	7.828571
10/8/2014	0	75	64	5.14	TRUE	2014	11	0.068533	10	Fall	281	APPLEGAT	-1.3E+07	4723838	0	5.398571
10/9/2014	0	76	66	6.93	FALSE	2014	10	0.091184	10	Fall	282	APPLEGAT	-1.3E+07	4723838	0	5.557143
10/10/2014	0	73	66	6.49	FALSE	2014	7	0.088904	10	Fall	283	APPLEGAT	-1.3E+07	4723838	0	5.59
5/17/2005	0	69	57	8.5	TRUE	2005	12	0.123188	5	Spring	137	APRICOT	-1.3E+07	4389836	0	7.477143
5/14/2017	0	71	53	8.5	TRUE	2017	18	0.119718	5	Spring	134	AQUADUC	-1.3E+07	4243539	0	9.011429
7/10/1996	0	76	63	8.95	TRUE	1996	13	0.117763	7	Summer	192	AQUADUC	-1.3E+07	4106210	0	8.341429
10/13/2003	0	82	60	5.14	TRUE	2003	22	0.062683	10	Fall	286	AQUEDUC	-1.3E+07	4132262	0	5.847143
10/14/2003	0	74	61	5.82	FALSE	2003	13	0.078649	10	Fall	287	AQUEDUC	-1.3E+07	4132262	0	5.911429
8/13/2005	0	74	65	8.28	TRUE	2005	9	0.111892	8	Summer	225	AQUEDUC	-1.3E+07	4215238	0	8.085714
4/12/2018	0	68	55	14.76	TRUE	2018	13	0.217059	4	Spring	102	AQUEDUC	-1.3E+07	4281375	0	8.628571
4/13/2018	0	78	59	8.28	FALSE	2018	19	0.106154	4	Spring	103	AQUEDUC	-1.3E+07	4281375	0	8.724286
10/5/1985	0.04	88	67	9.84	TRUE	1985	21	0.111818	10	Fall	278	ARCHIBAL	-1.3E+07	4052807	0.04	7.86
10/6/1985	0	78	65	9.62	FALSE	1985	13	0.123333	10	Fall	279	ARCHIBAL	-1.3E+07	4052807	0.04	8.275714
8/12/2011	0	72	62	7.83	TRUE	2011	10	0.10875	8	Summer	224	ARDO	-1.3E+07	4292093	0	6.902857

```python
import pandas as pd
import geopandas as gpd
from shapely.geometry import Point

perim_path = r"California_Fire_Perimeters.shp"
fire_perimeters = gpd.read_file(perim_path)


zenodo_csv_path = r"wildfire_dataset.csv"
df = pd.read_csv(zenodo_csv_path, parse_dates=["DATE"])


fire_perimeters['centroid'] = fire_perimeters.geometry.centroid
fire_perimeters['LONGITUDE'] = fire_perimeters.centroid.x
fire_perimeters['LATITUDE'] = fire_perimeters.centroid.y


fire_locations = fire_perimeters[['FIRE_NAME', 'ALARM_DATE', 'LONGITUDE', 'LATITUDE']].copy()
fire_locations.rename(columns={'ALARM_DATE': 'DATE'}, inplace=True)


df['DATE'] = pd.to_datetime(df['DATE']).astype('datetime64[ns]')
fire_locations['DATE'] = pd.to_datetime(fire_locations['DATE']).astype('datetime64[ns]')


df = df.sort_values('DATE')
fire_locations = fire_locations.sort_values('DATE')


if 'FIRE_NAME' in df.columns:
    merged = pd.merge(df, fire_locations, on=['FIRE_NAME', 'DATE'], how='left')
else:
    merged = pd.merge_asof(df, fire_locations, on='DATE', direction='nearest')


merged['TEMP_RANGE'] = merged['MAX_TEMP'] - merged['MIN_TEMP']
merged['WIND_TEMP_RATIO'] = merged['AVG_WIND_SPEED'] / merged['MAX_TEMP']
merged['MONTH'] = merged['DATE'].dt.month
merged['YEAR'] = merged['DATE'].dt.year
merged['DAY_OF_YEAR'] = merged['DATE'].dt.dayofyear
```

Handling Imbalanced Data

To balance our dataset we used two machine learning techniques called SHAP and SMOTE .

SHAP

SHAP explains ML models by showing each feature's impact on predictions. For wildfires, it highlights key factors like temperature and wind speed, helping stakeholders trust and understand model decisions.

SMOTE

SMOTE fixes class imbalance by creating synthetic wildfire samples. This boosts detection of real fires (recall) while keeping false alarms low (precision), improving model reliability.

Key Difference

- SHAP = Model interpretation
- SMOTE = Data balancing

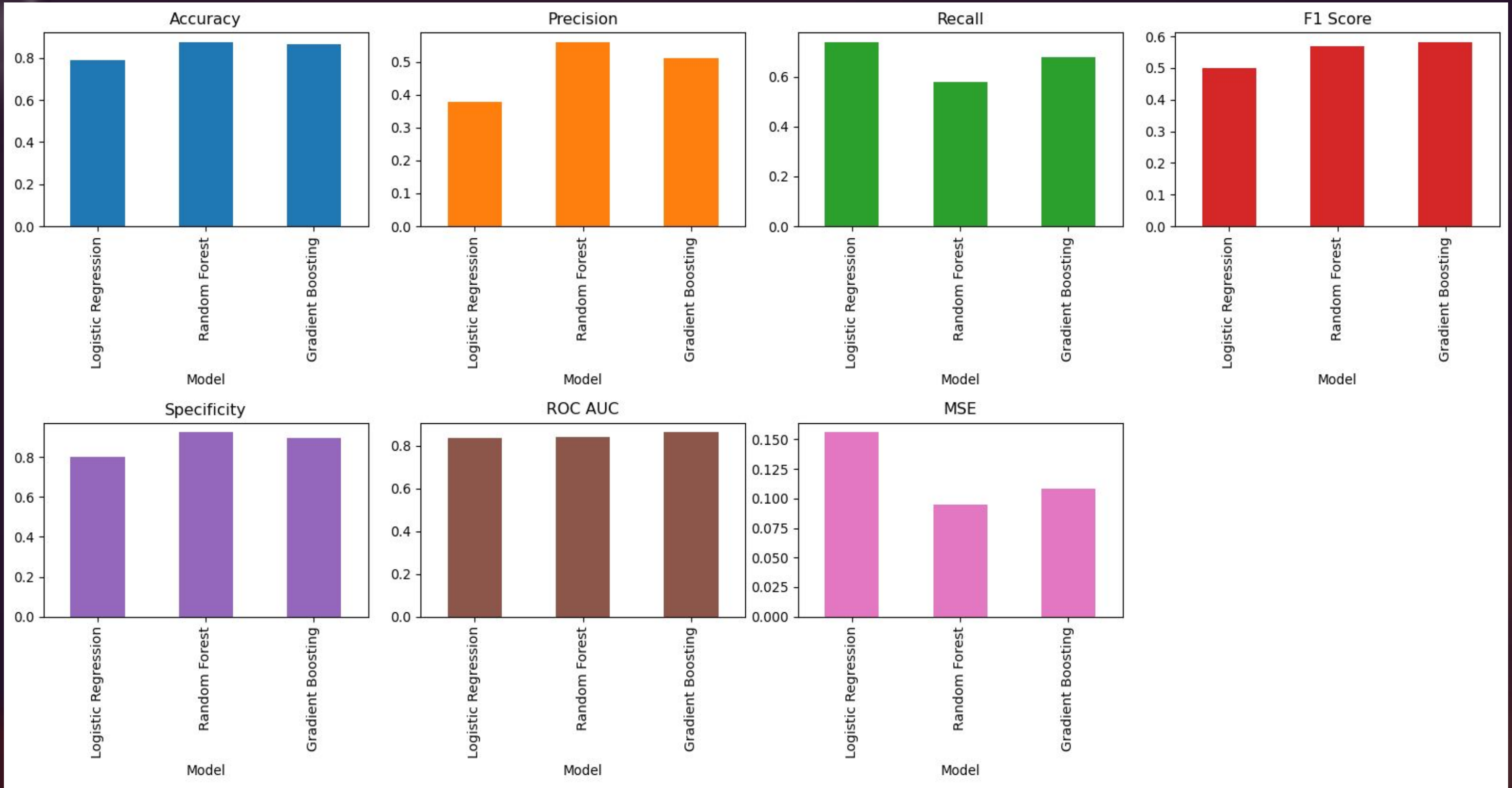
Traditional Machine Learning Models

- Random Forest – combines multiple decision trees to vote on predictions, finding complex patterns while avoiding overfitting. It's ideal for analyzing how weather, seasons, and terrain interact to predict fire risks.

- Gradient Boosting – builds trees one by one, each fixing the last's errors. It's great for catching rare wildfire events in data. The model improves by focusing on past mistakes.

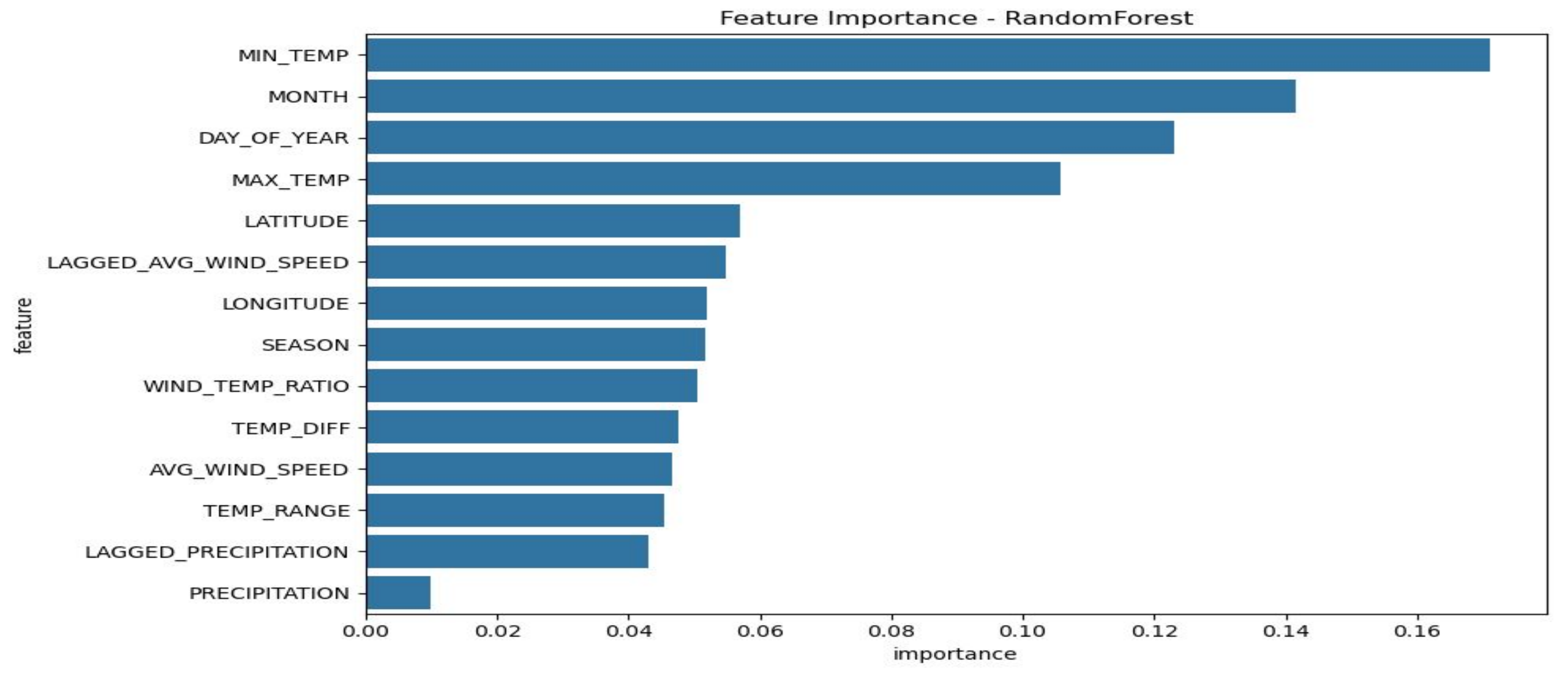
- Logistic Regression – calculates fire probability using simple math. It shows how factors like temperature increase fire chances. Fast and clear, it's a good starting point.

Precision, Recall, and F1 Score



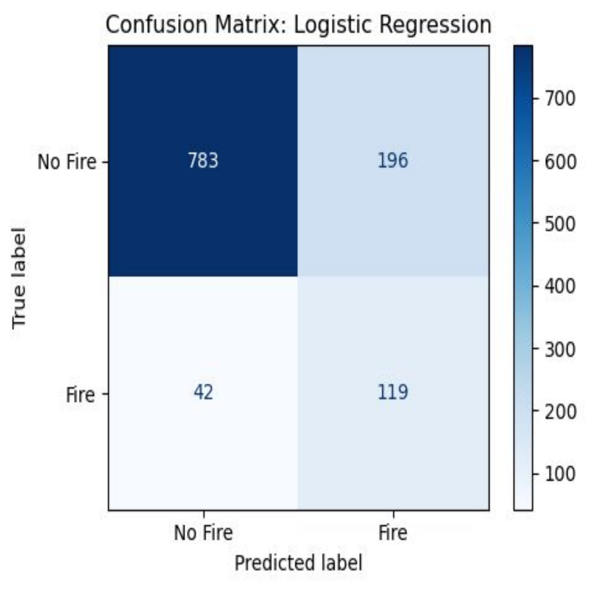
Feature Engineering

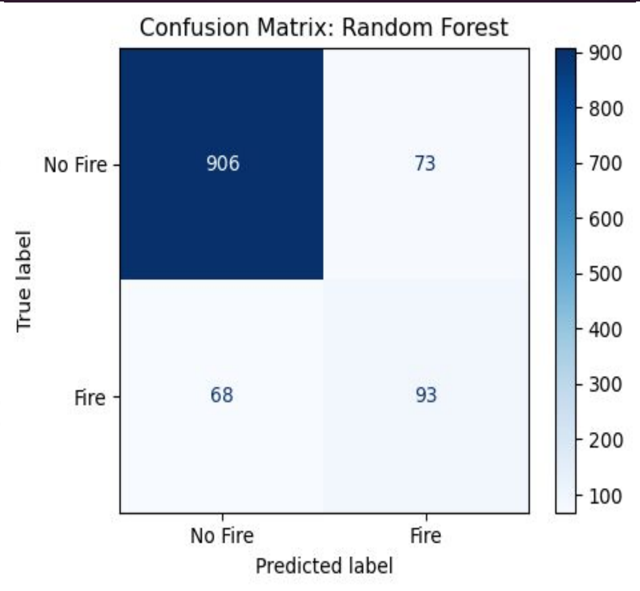
The RandomForest model reveals that minimum temperature (MIN_TEMP) and month (MONTH)are the most critical predictors of wildfire risk, followed closely by day of year (DAY_OF_YEAR) and maximum temperature (MAX_TEMP). Geographic coordinates (LATITUDE, LONGITUDE) and derived metrics like wind-to-temperature ratio (WIND_TEMP_RATIO) also contribute significantly.

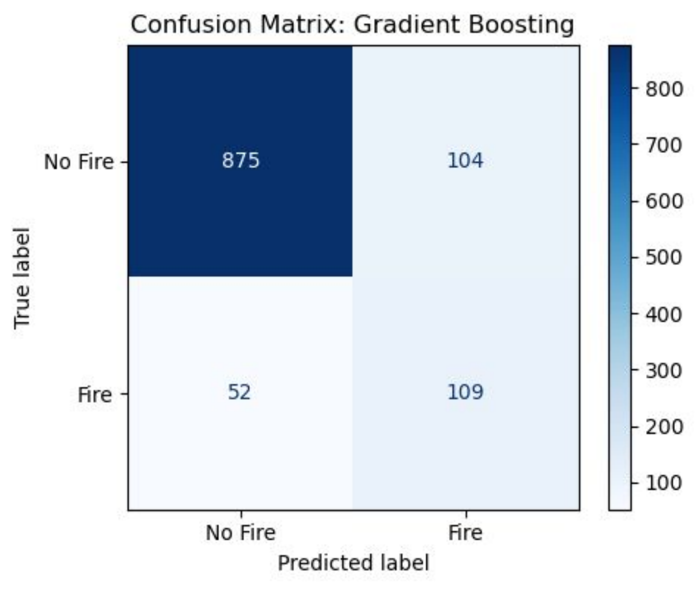


Confusion Matrix

Confusion Matrix of each of the classical machine learning models







Deep Machine Learning Models

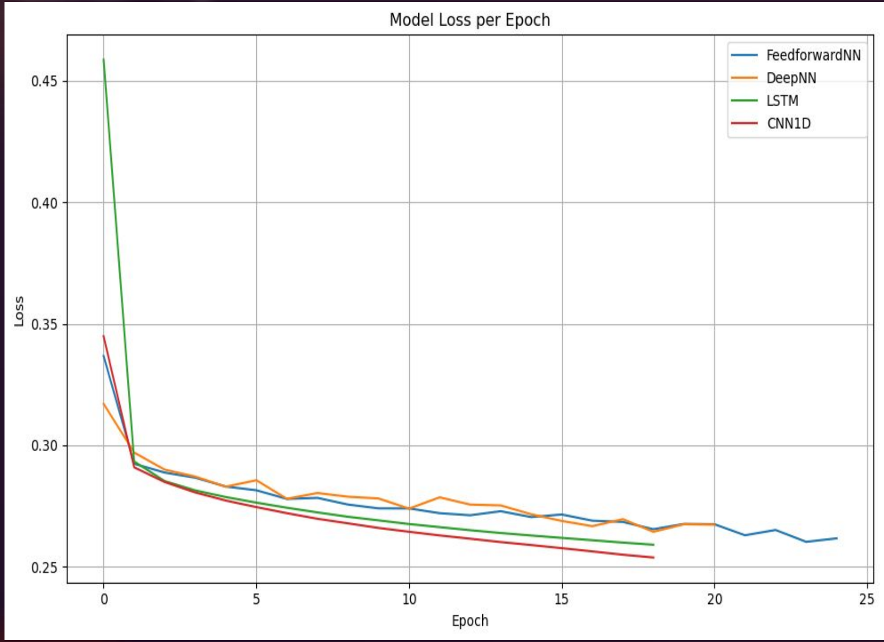
- 1D CNN – Uses sliding filters to scan weather data, detecting local patterns like sudden temperature spikes. Best for short-term fire risk detection. Fast and efficient for time-based patterns.

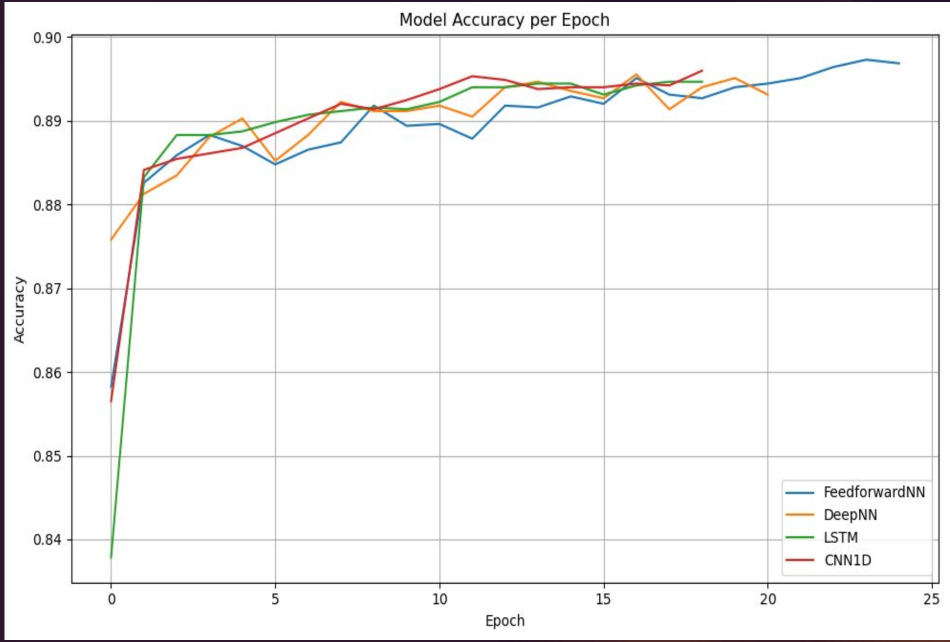
- LSTM – Processes sequences with memory cells, learning long-term trends like drying conditions or seasonal risks. Handles missing data well and retains past dependencies.

- DNN – Applies stacked layers of neurons to model complex relationships in data like weather and terrain. Highly adaptable but requires careful tuning.

- FNN – Passes data one-way through connected layers, analyzing basic fire risk factors. Quick and interpretable but limited in handling complex patterns.

How the Deep Learning Models Progressed

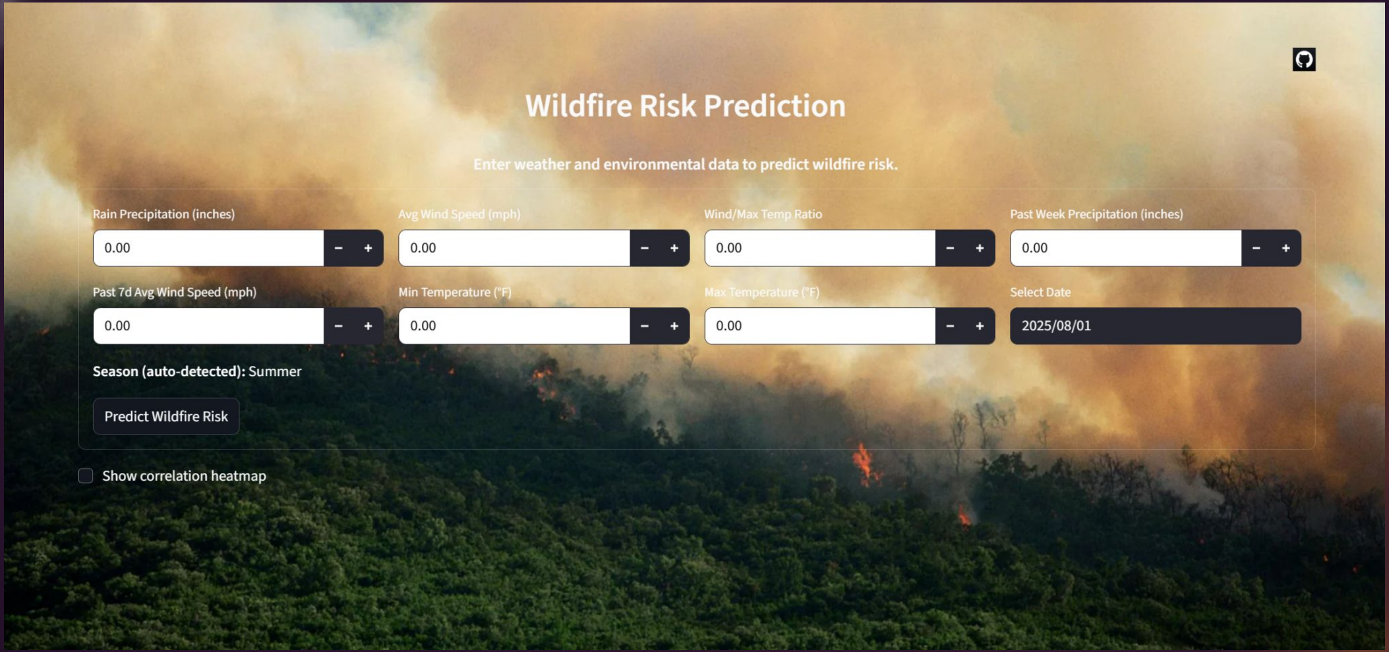




Streamlit Web App

Web App Features

We decided to build the app user interface so that it can be easier for users to understand and use.



Limitations and Future Work

- These models show promise for future applications in real-time fire monitoring systems and environmental alert systems.
- Adding vegetation indices like NDVI or satellite imagery would further contextualize the data and capture fire-prone environmental conditions.

References

Zenodo, “Wildfire Risk Modeling Dataset,” Zenodo, Online dataset, 2025. [Online]. Available: https://zenodo.org/records/14712845

CAL FIRE, “California Fire Perimeters (All),” California Department of Forestry and Fire Protection, GIS Data Portal, Online dataset, 2025. [Online]. Available: https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-fire-perimeters-all/explore?location=37.096038%2C-119.269051%2C6.19

Thank You and Acknowledgements

NSF U.S. National Science Foundation

MS CC

FISK UNIVERSITY
ESTABLISHED 1866

MEHARRY®
MEDICAL COLLEGE

ANY QUESTIONS