A Comparative Study of Forest Fire Prediction using Machine Learning models

Sachin Malego st125171 | Sila N Mahoot st125127

9 October 2024

For the fulfillment project proposal of
AT82.01 Computer Programming for Data Science and Artificial Intelligence course
by Dr. Chantri Polprasert

Agenda

1. Introduction
2. Project Goals
3. Problem Statement
4. Related Works
5. Datasets
6. Methodology
7. Preliminary Results
8. Deployment

Introduction

Why Are Forest Fires a Problem?

Global Tree Cover Loss (2001–2023)

- 2001-2021: The world lost approximately 437 million hectares of tree cover, representing around an **11% decrease since 2000**.

- Annual Average: On average, around **25 million hectares** of tree cover have been lost each year since 2000.

Introduction

Why Are Forest Fires a Problem?

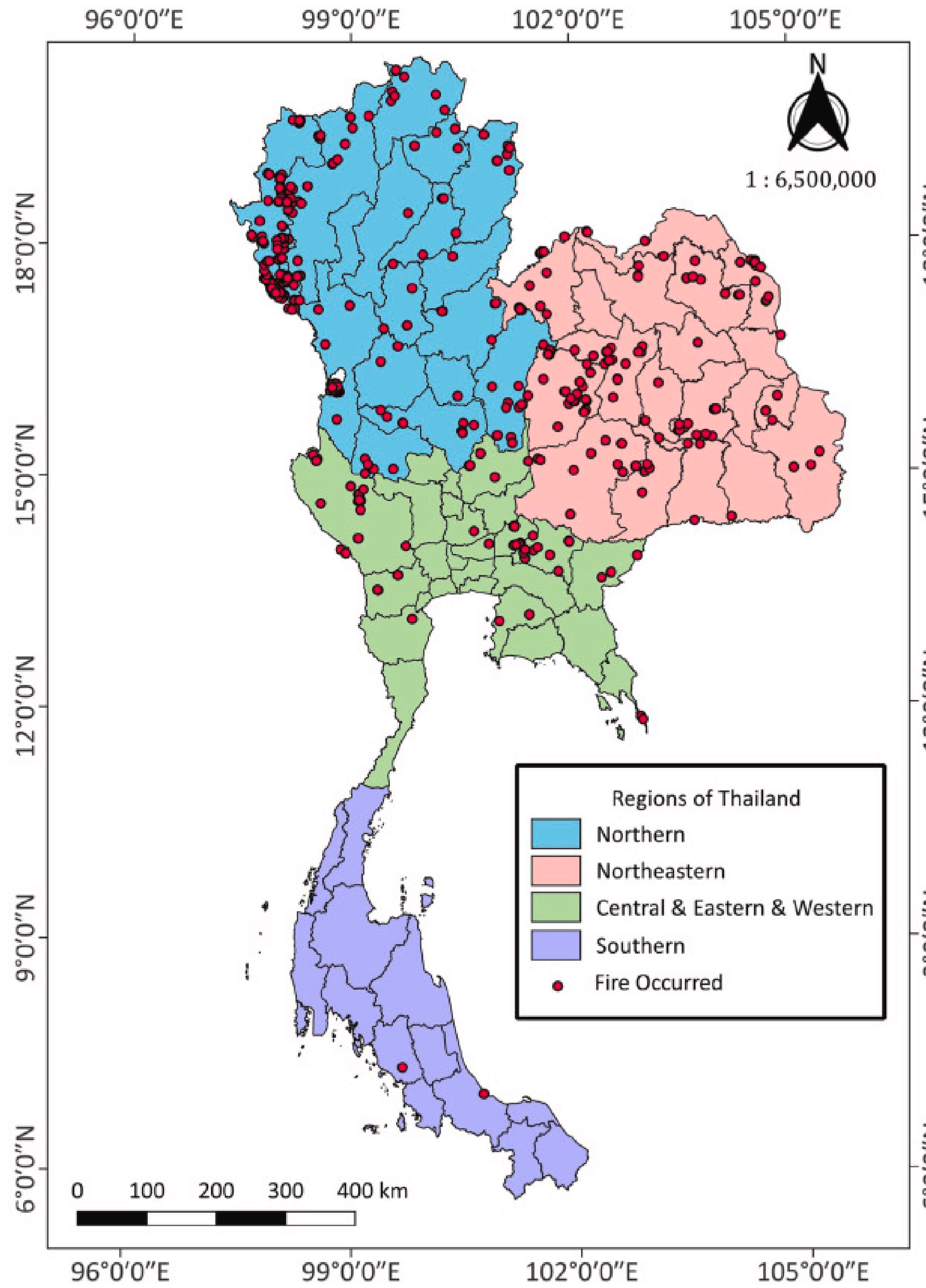
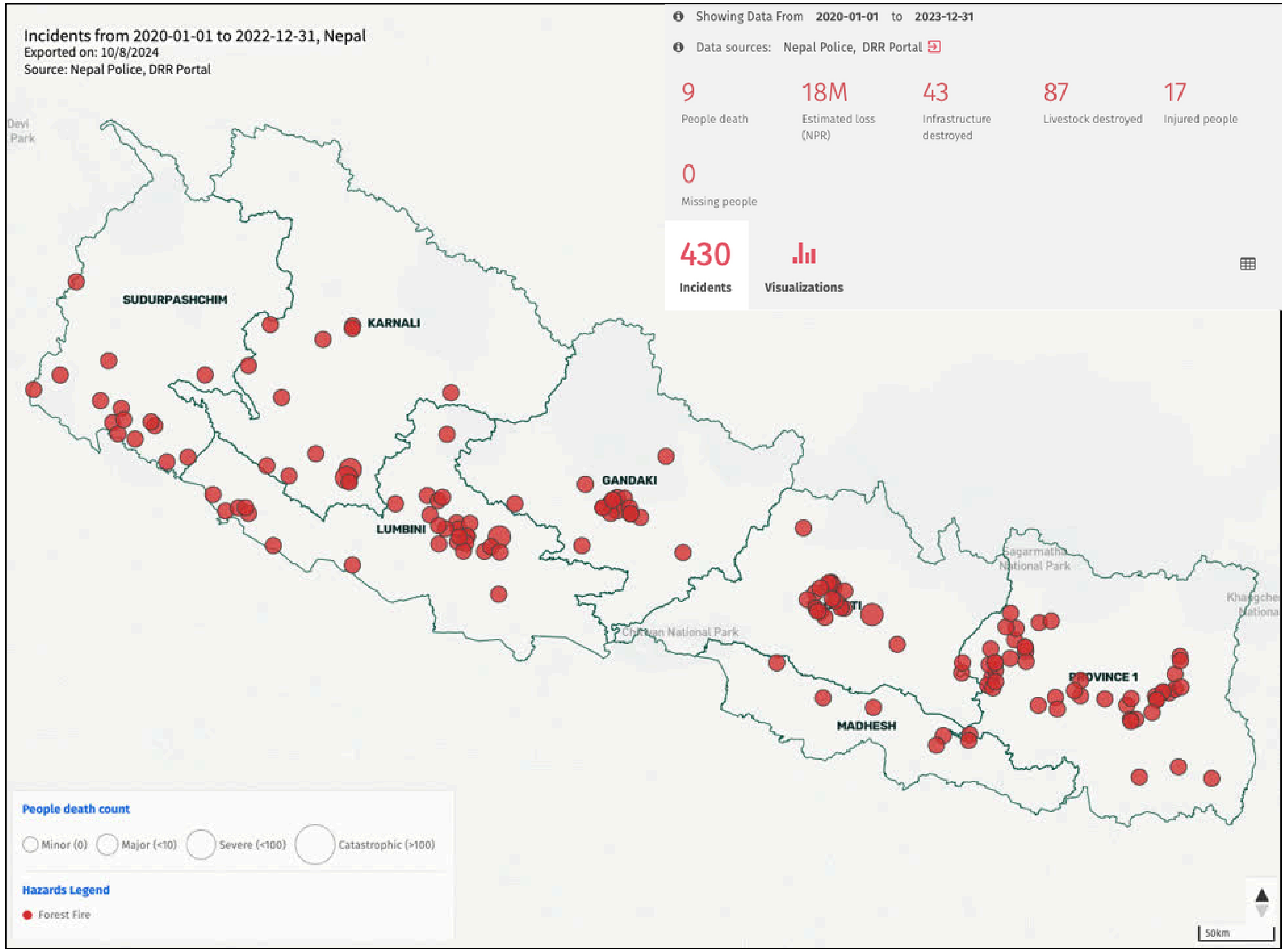


Image Source: https://doi.org/10.1016/j.heliyon.2024.e34021



Source: https://bipadportal.gov.np/incidents/

Impact of Forest Fires

Why Are Forest Fires a Problem?

Impact on natural habitats and biodiversity.

Threats to human lives and properties.

Contribution to air pollution and climate change.

Project Goals - Proposed Solution

Predict and Prevent Forest Fires

1

Develop a **machine learning** model to predict forest fires

2

Provide an **early warning system** to help manage resources

3

Reduce the impact of fires on people and nature

Problem Statements

Challenges in Fire Prediction

- Weather Impact Analysis
- Drought-Induced Fire Risk
- El Niño and Fire Behavior
- Fire Spread Prediction
- Difficulty in accurately predicting fires
- Current systems lack precision for early warnings
- Need for better environmental data handling

Related Works

- Previous studies used machine learning and satellite data.
- Our model includes climate data like El Niño for improved accuracy.

Toward a More Resilient Thailand Developing a Machine Learning-Powered Forest Fire Warning System

doi.org/10.1016/j.heliyon.2024.e34021

Developed a machine learning-powered forest fire warning system using satellite data and gas measurements. The **XGBoost model achieved 99.6% accuracy**.

Predicting Wildfires in Algerian Forests Using Machine Learning Models

doi.org/10.1016/j.heliyon.2023.e18064

Used PCA for reducing data complexity and developed an **ANN for predicting wildfires, achieving an accuracy of 96.7%**. It highlighted key features like relative humidity and drought code. The dataset included various weather features collected from Algeria.

Comparison of Forest Fire Prediction System Using Machine Learning Algorithms

doi.org/10.1109/ICACITE57410.2023.10182818

Compared several machine learning models, including logistic regression and random forest, to predict forest fires. It discussed the strengths and weaknesses of each model. The dataset included temperature, wind speed, and humidity, and the authors suggested **integrating climate patterns like El Niño for better predictions.**

Understanding El Niño

What is El Niño?

- Periodic warming of the Pacific Ocean affects weather worldwide.
- Causes hotter, drier conditions that increase fire risk.

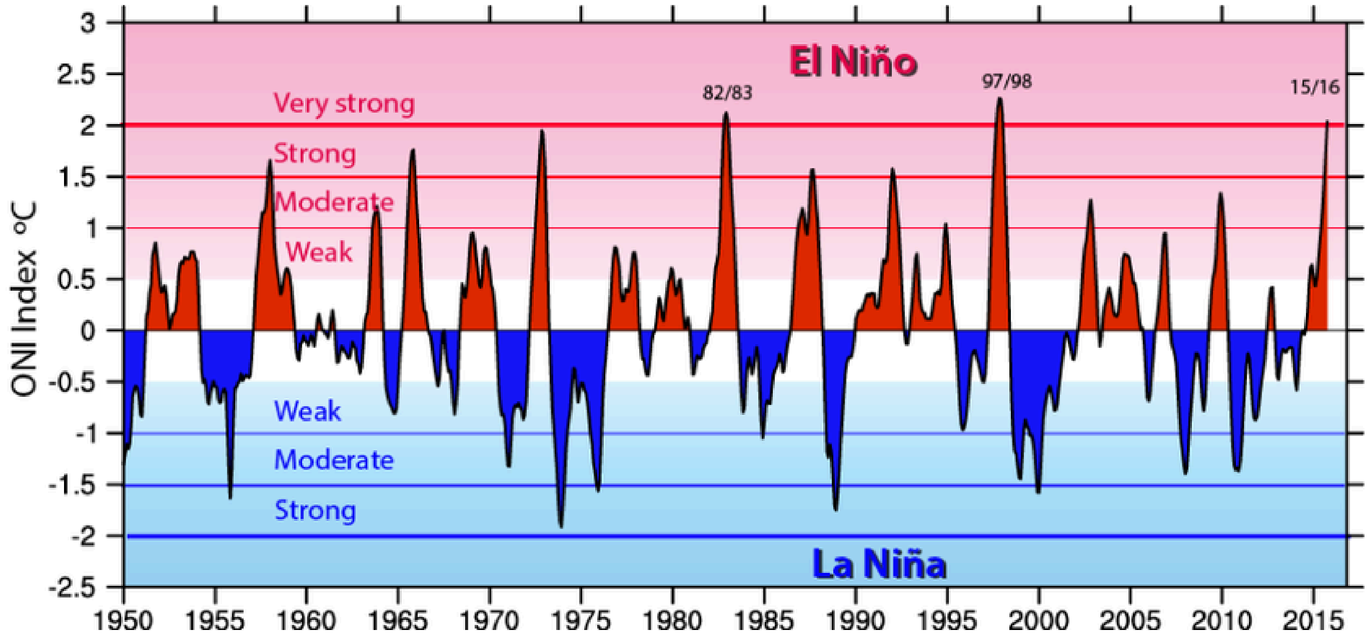


Image Source: climatedataguide.ucar.edu

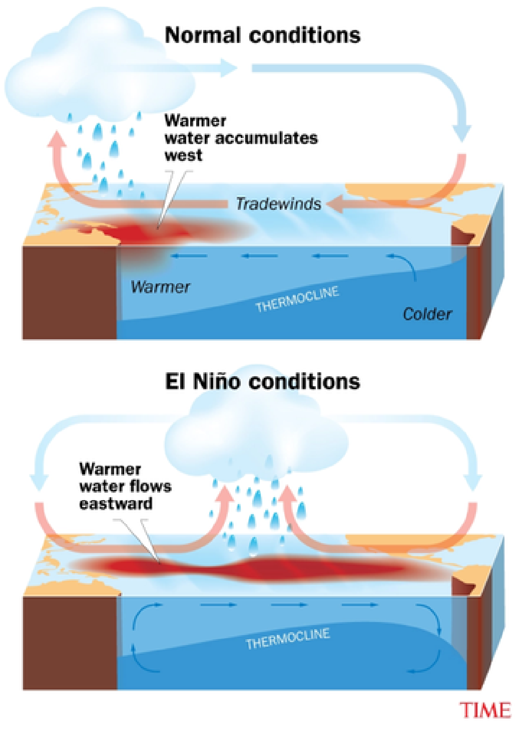
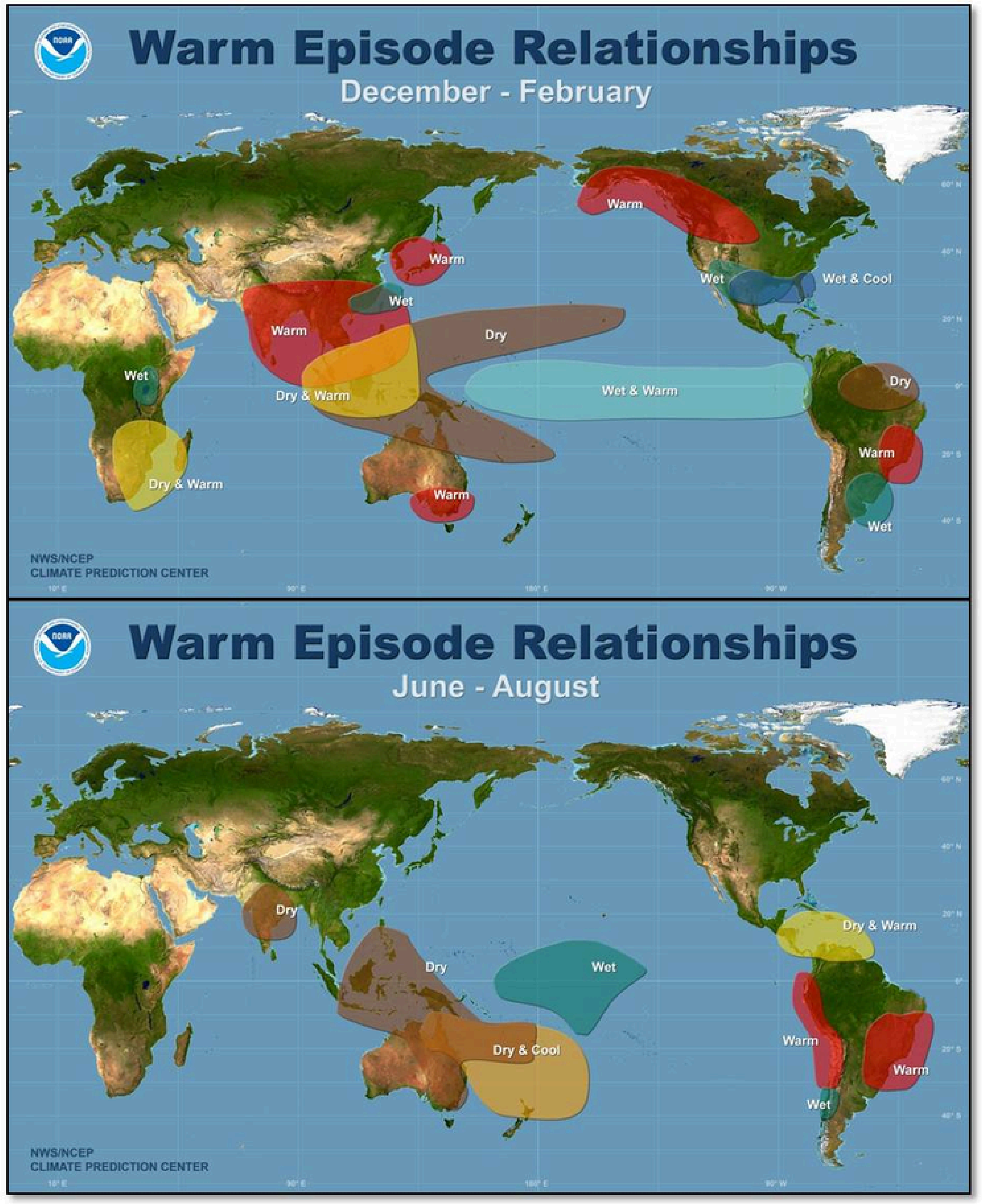


Image Source: Understanding El Niño: A Climate Phenomenon with Global Implications @ LinkedIn



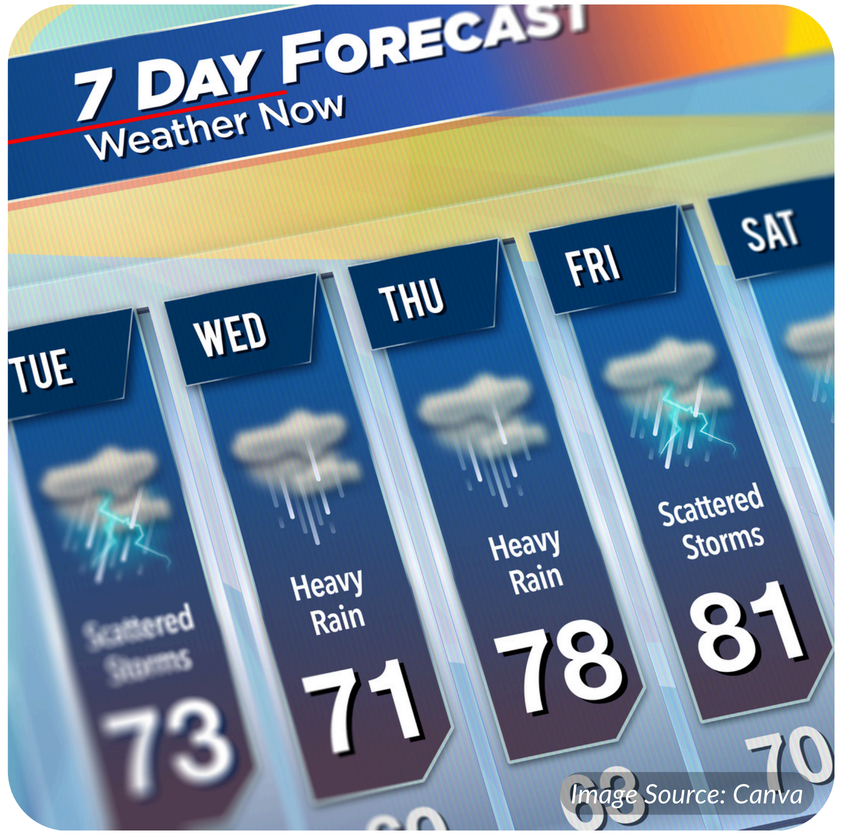
High Resolution Images can be found at:
http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/ENSO-Global-Impacts/

Datasets

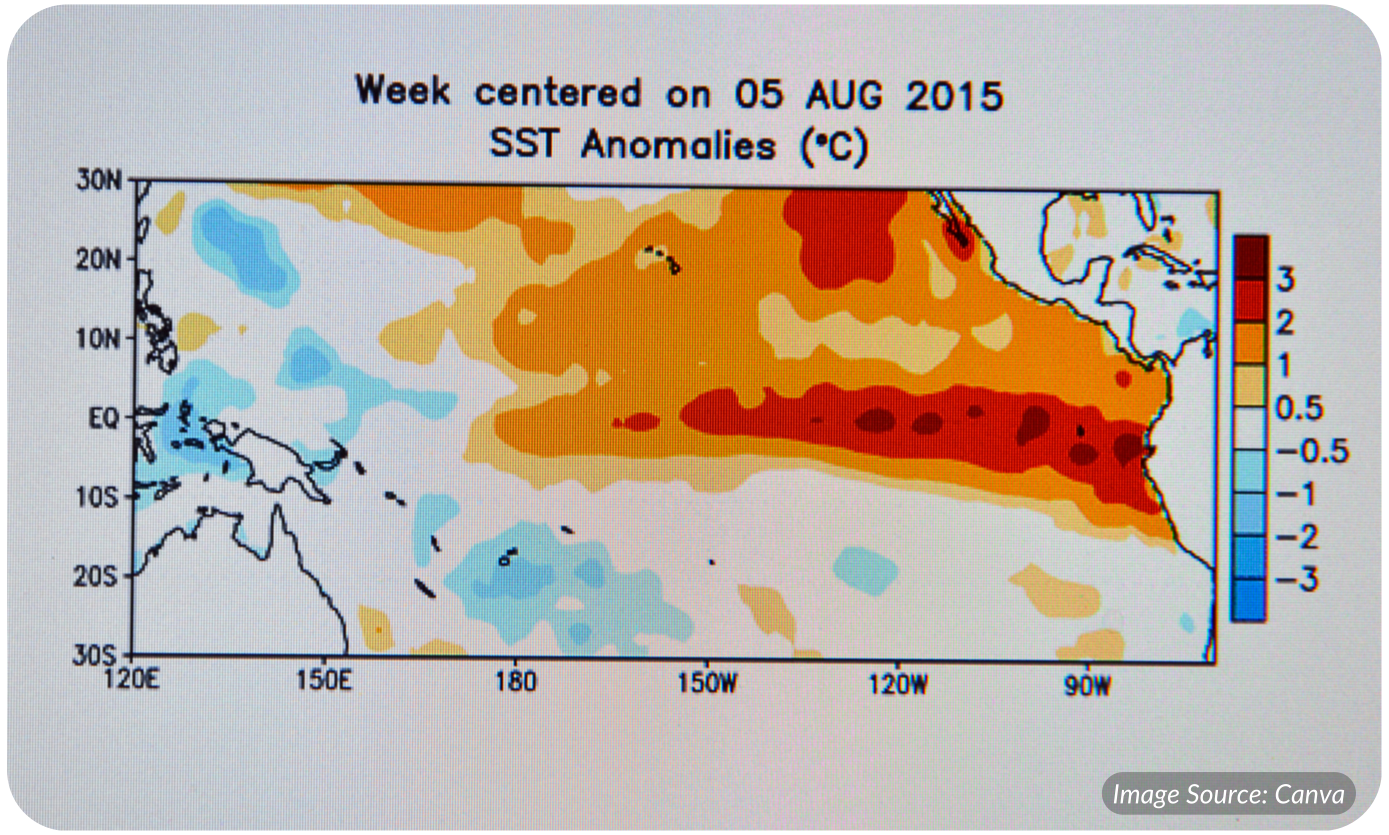
Data Sources Used

Forest Fire Occurrences
in Algeria (2012) and Portugal (2017

Historical Weather Data
from Meteostat & Weather Underground



Sea Surface Temperature
from Climate Prediction Center



Datasets

Data Sources Used

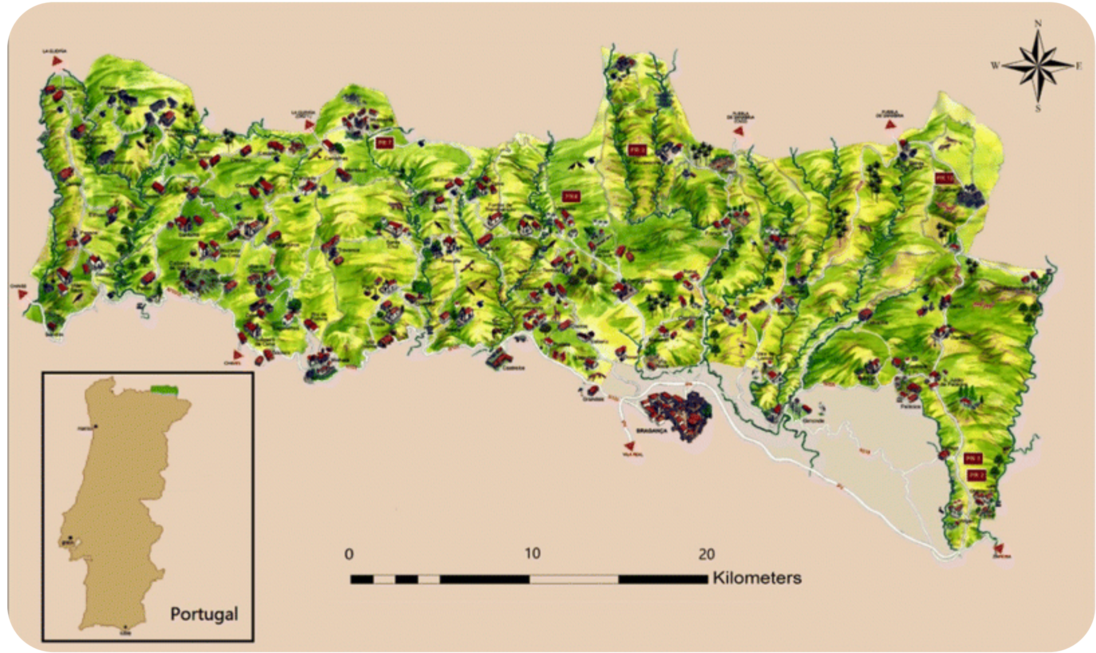
Forest Fire Occurrences in Algeria (2012) and Portugal (2017)

Variables	Description
X	*X*-axis spatial coordinate (from 1 to 9)
Y	*Y*-axis spatial coordinate (from 1 to 9)
Month	Month of the year (from "January" to "December")
Day	Day of the week (from "Monday" to "Sunday")
FFMC	FFMC code from the FWI system (from 18.7 to 96.20)
DMC	DMC code from the FWI system (from 1.1 to 291.3)
DC	DC code from the FWI system (from 7.9 to 860.6)
ISI	ISI code from the FWI system (from 0 to 56.10)
Temp	Temperature in degrees Celsius (from 2.2 to 33.30)
RH	Relative humidity in percentage (from 15.0 to 100)
Wind	Wind speed in km/h (from 0.40 to 9.40)
Rain	Outside rain in mm/m^2 (from 0.0 to 6.40)
Area	Total burned area of the forest (in *ha*) (from 0.00 to 1090.84)

#	Column	Non-Null Count	Dtype
0	day	243 non-null	int32
1	month	243 non-null	int32
2	year	243 non-null	int32
3	Temperature	243 non-null	int32
4	RH	243 non-null	int32
5	Ws	243 non-null	int32
6	Rain	243 non-null	float64
7	FFMC	243 non-null	float64
8	DMC	243 non-null	float64
9	DC	243 non-null	float64
10	ISI	243 non-null	float64
11	BUI	243 non-null	float64
12	FWI	243 non-null	float64
13	Classes	243 non-null	object
14	Region	243 non-null	int32

Features: Date, Temperature, RH, WindSpeed, Rain, Precipitation, FFMC, DMC, DC, FWI, Burn Area

Data Quality Checklist: Completeness, Accuracy, Documentation, Anomaly Detection, others



Datasets

Data Sources Used

Historical Weather Data
from Meteostat & Weather Underground

date	temp_min	temp_avg	temp_max	dew_min	dew_avg	dew_max	hum_min	hum_avg	hum_max	wind_speed_min
2012-01-01	18.888889	11.000000	7.222222	10.000000	7.555556	2.777778	94	80.6	45	22.53076
2012-01-02	20.000000	13.111111	7.777778	12.222222	8.333333	6.111111	93	74.0	49	22.53076
2012-01-03	16.111111	13.222222	8.888889	10.000000	8.777778	7.777778	100	76.7	59	19.31208
2012-01-04	17.222222	11.111111	7.222222	10.000000	7.666667	6.111111	94	80.6	52	22.53076
2012-01-05	18.888889	12.055556	8.888889	12.777778	9.055556	7.222222	94	83.2	52	14.48406

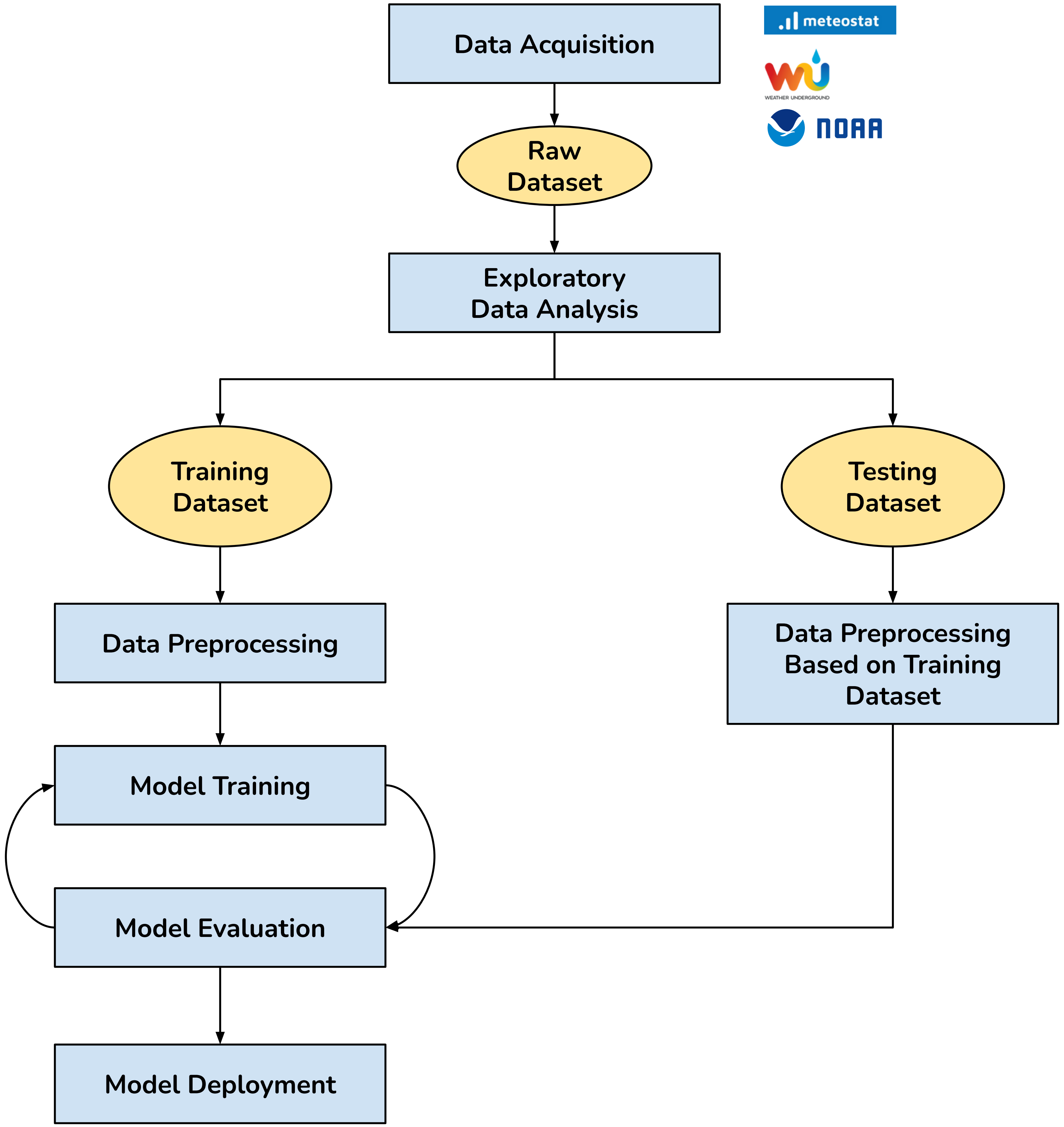
Sea Surface Temperature
from Climate Prediction Center

	nino12_sst	nono12_ssta	nino3_sst	nino3_ssta	nino34_sst	nino34_ssta	nino4_sst	nino4_ssta
1981-09-02	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-03	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-04	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-05	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-06	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-07	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-08	0.186275	0.022222	0.333333	0.030303	0.421053	0.066667	0.512195	0.15
1981-09-09	0.137255	0.133333	0.317460	0.060606	0.421053	0.066667	0.536585	0.10
1981-09-10	0.137255	0.133333	0.317460	0.060606	0.421053	0.066667	0.536585	0.10
1981-09-11	0.137255	0.133333	0.317460	0.060606	0.421053	0.066667	0.536585	0.10

Threshold for SST/SSTA (NINO 3.4): The most recent three-month average for the area is computed, and if the region is more than **0.5 °C** (**0.9 °F**) above (or below) normal for that period, then an El Niño (or La Niña) is considered in progress.

Methodology

How We Build Our Model

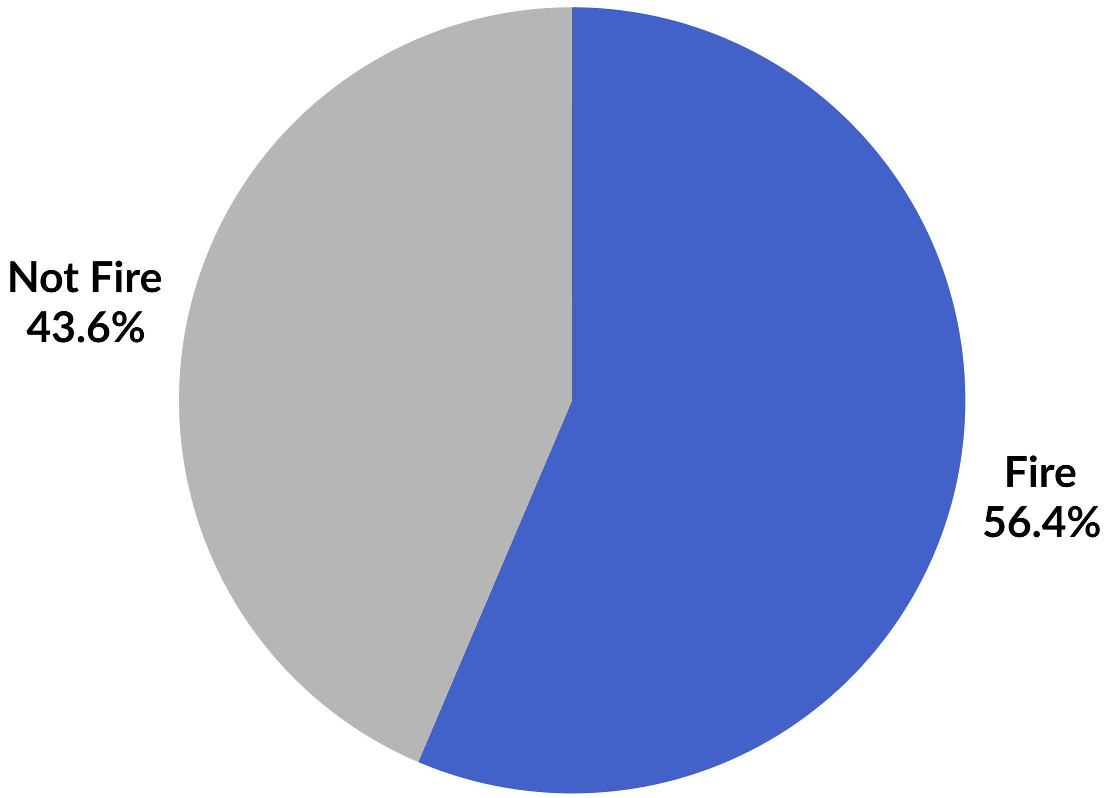


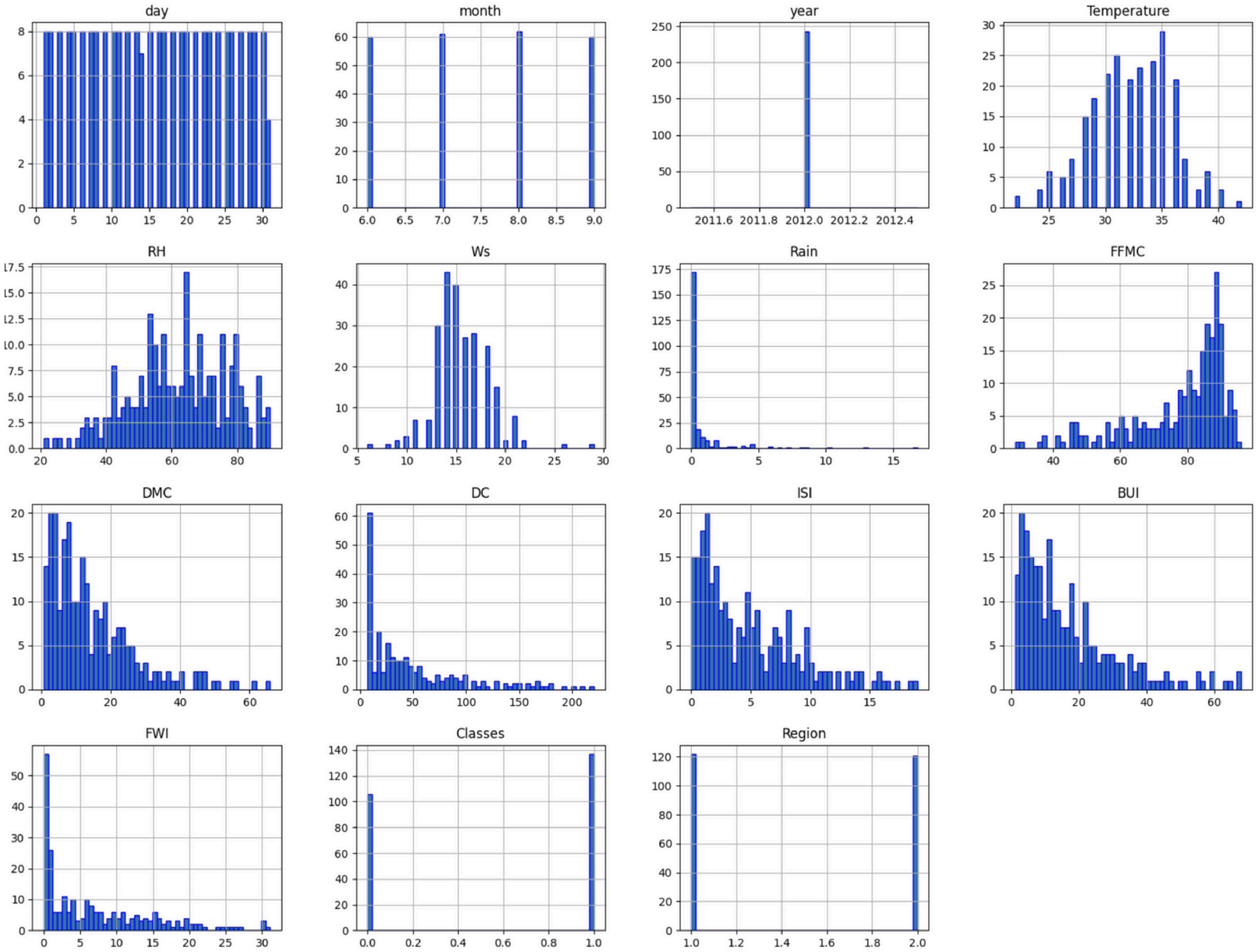
Missing Values
Data Scaling
Data Imbalance

Random Forest
Gradient Boosting
Extreme Gradient Boosting

Preliminary Results

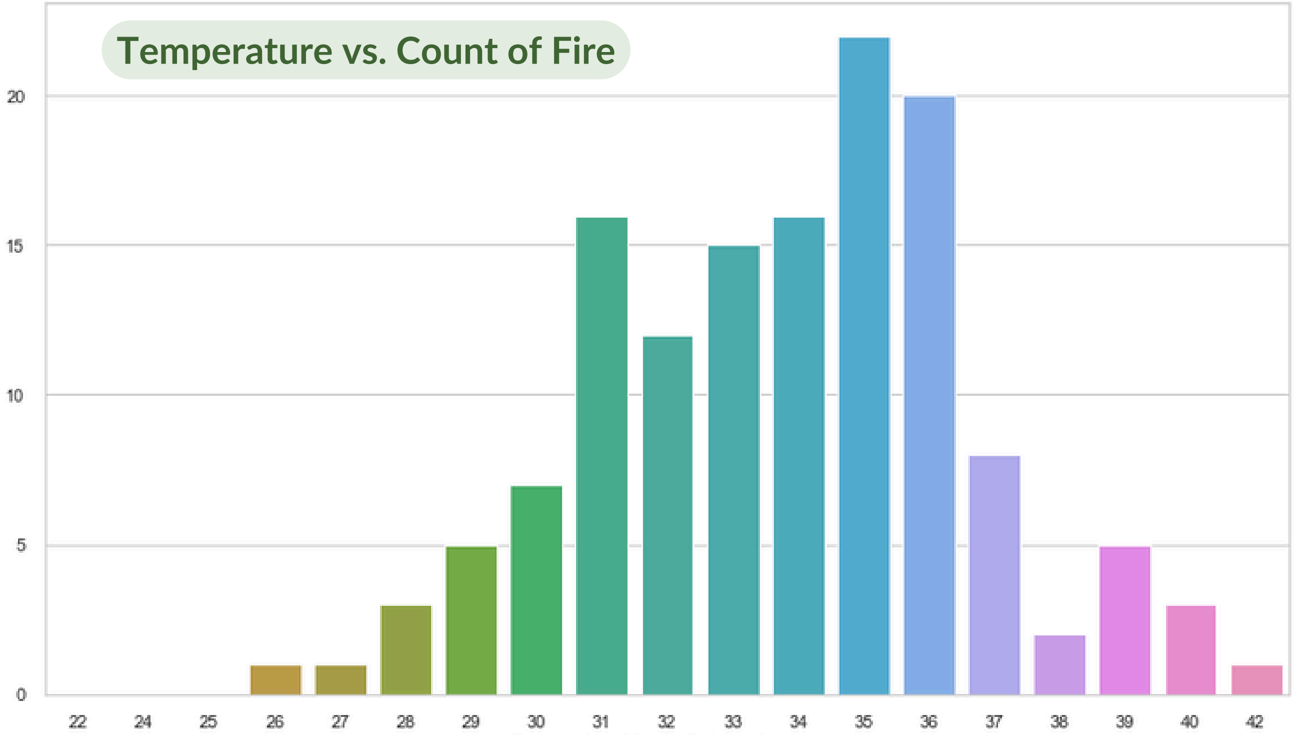
Exploring the Data

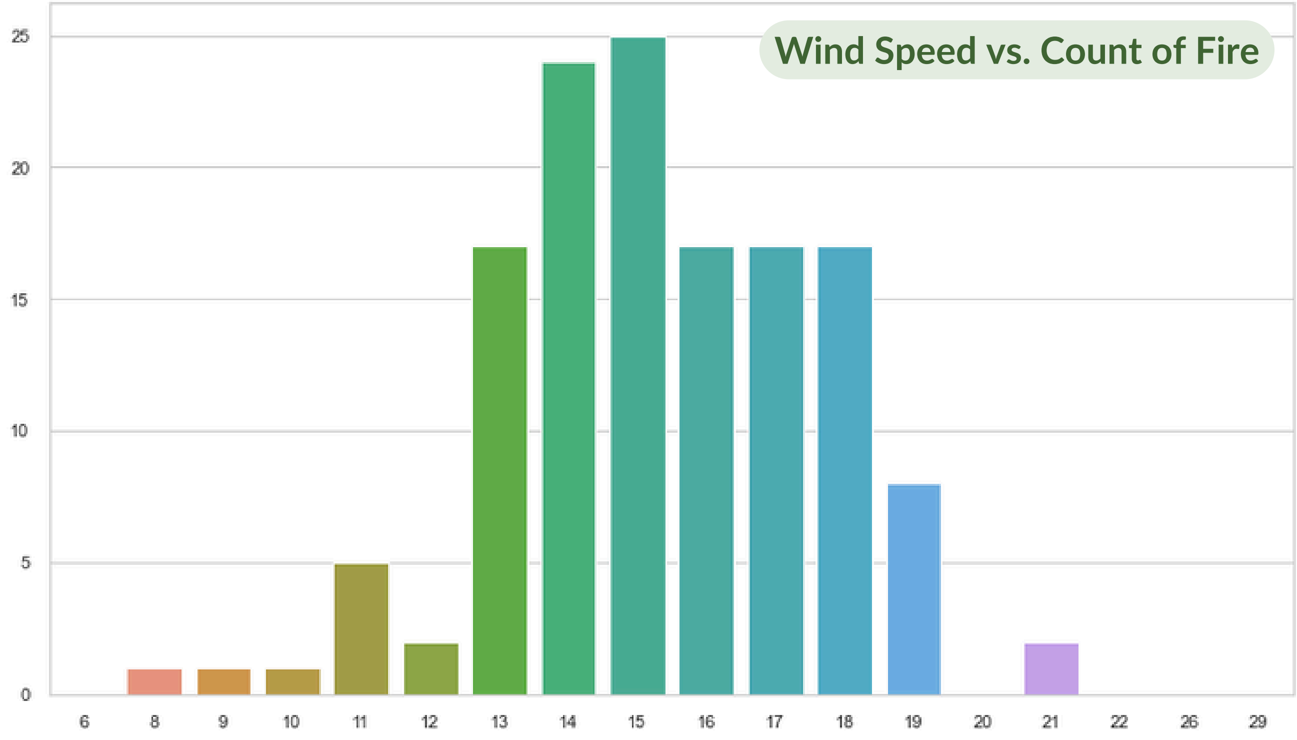


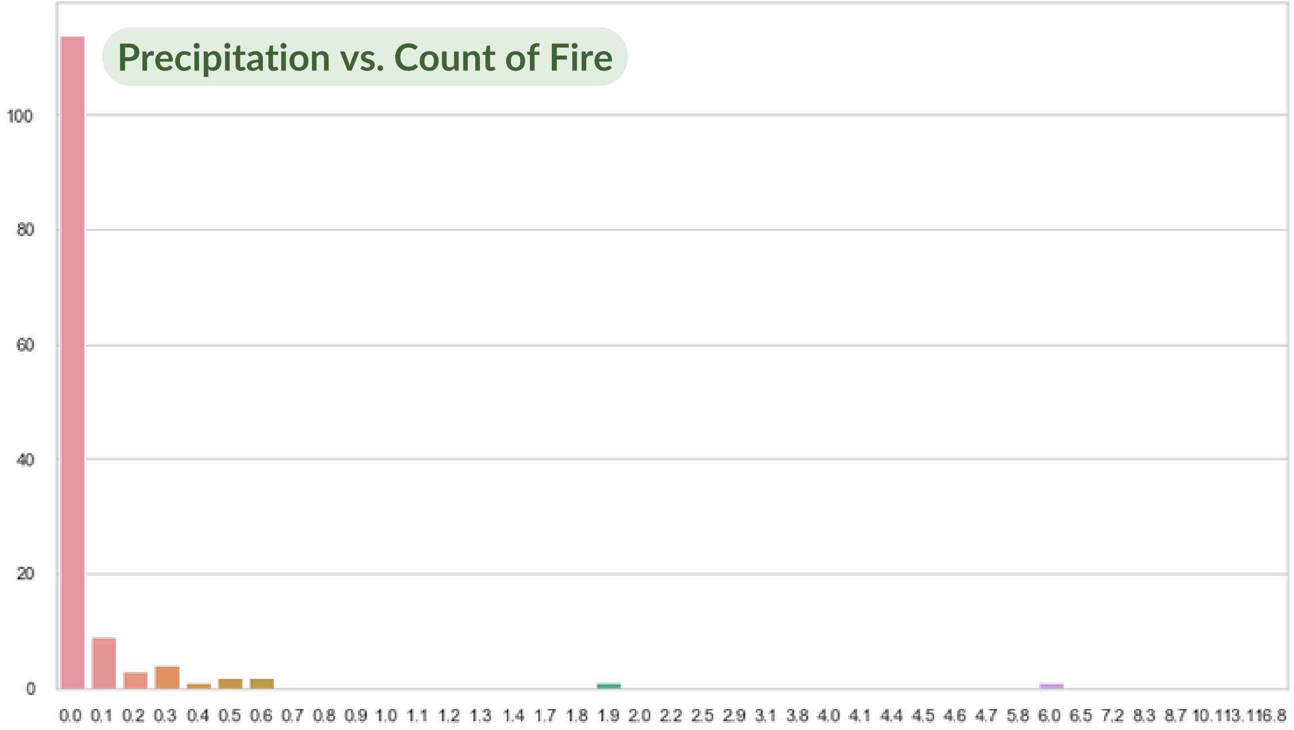


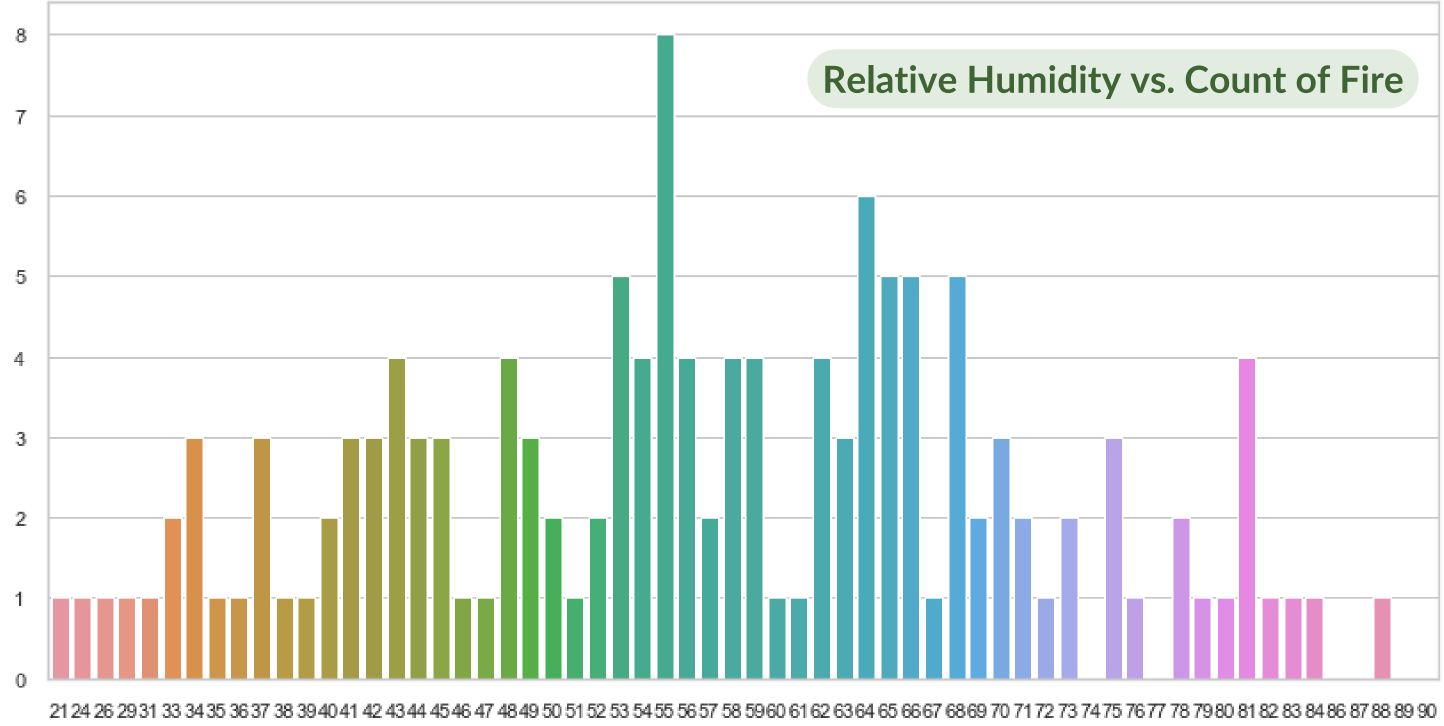
Preliminary Results

Exploring the Data



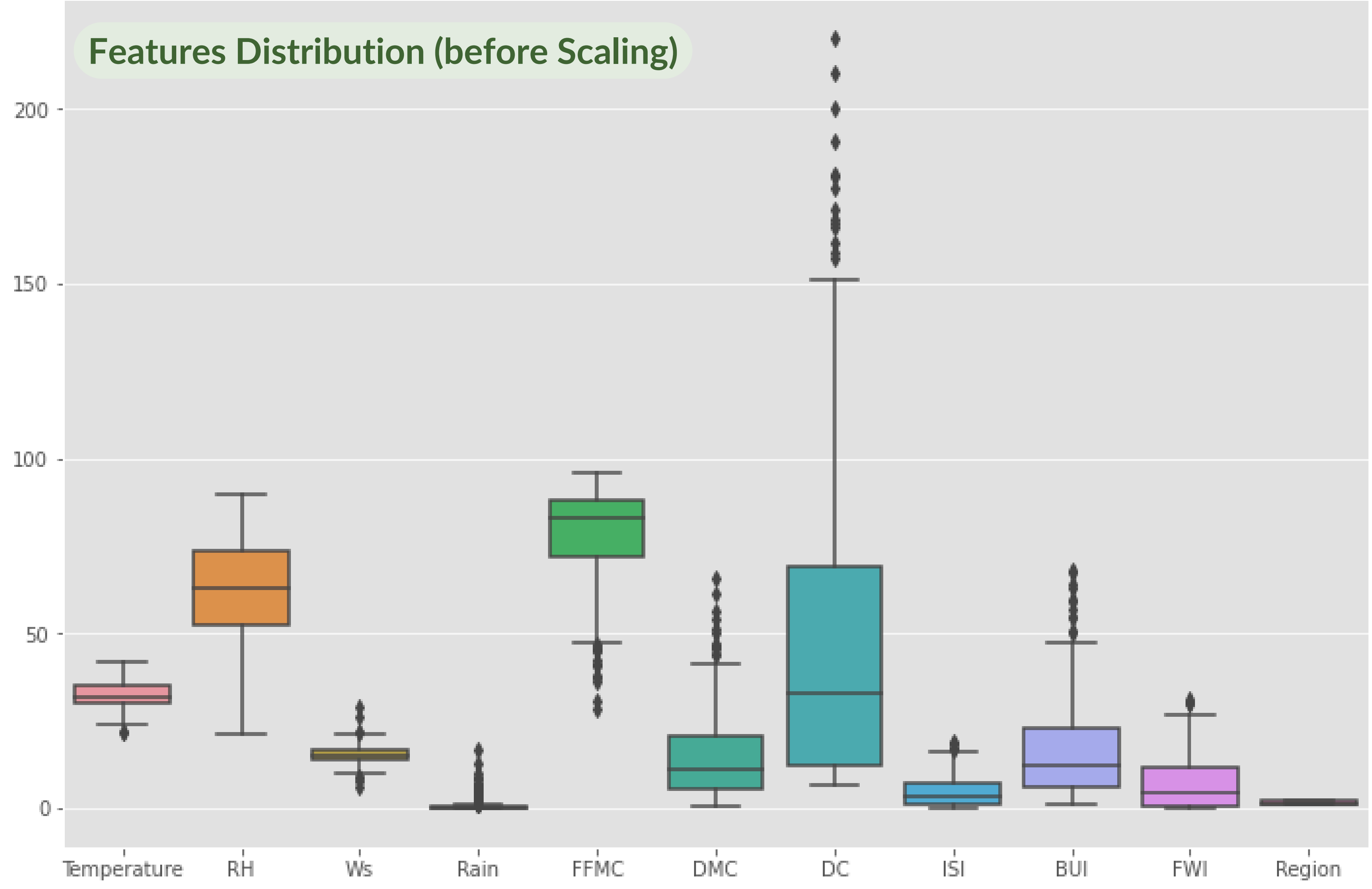


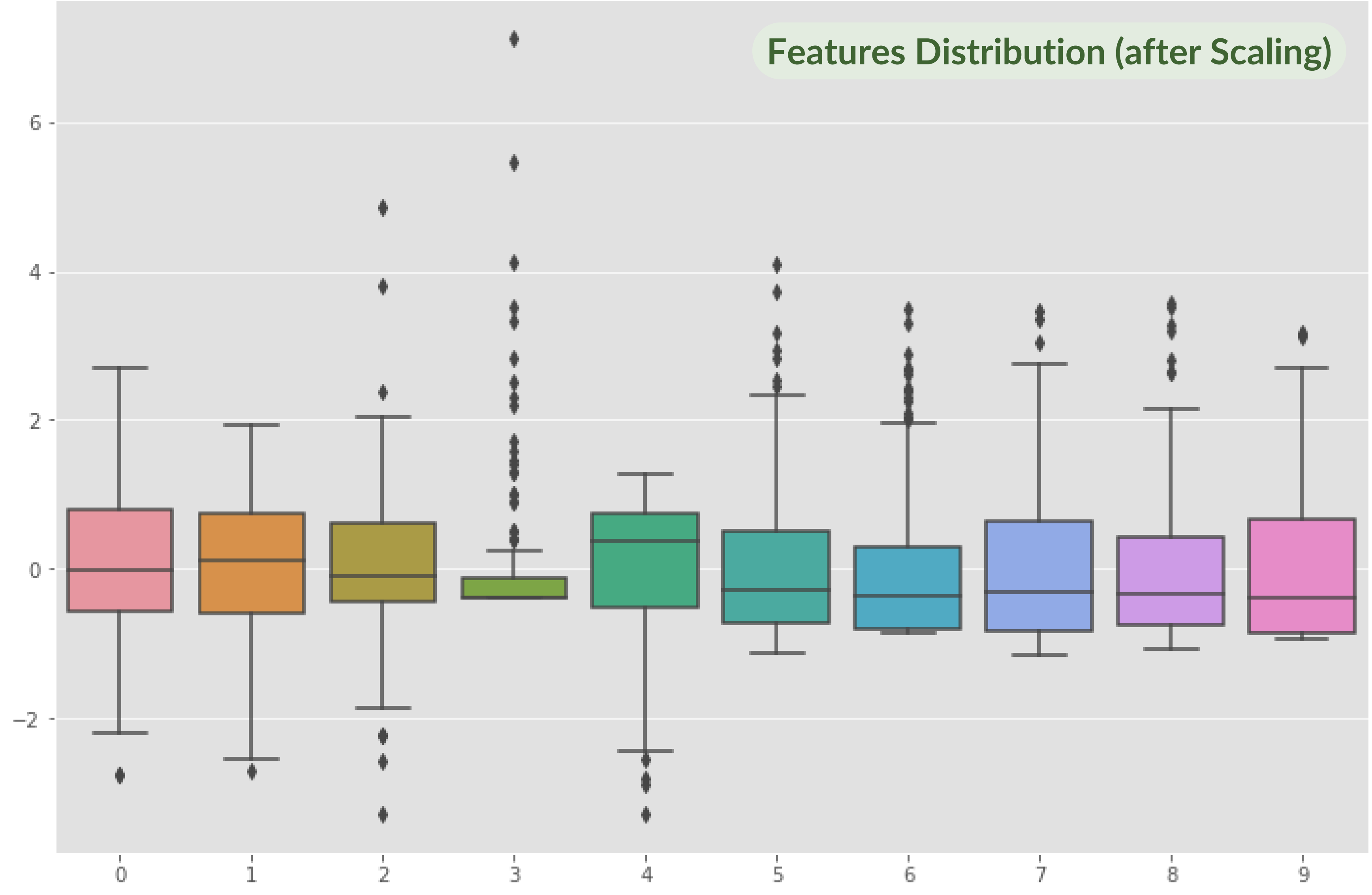




Preliminary Results

Exploring the Data



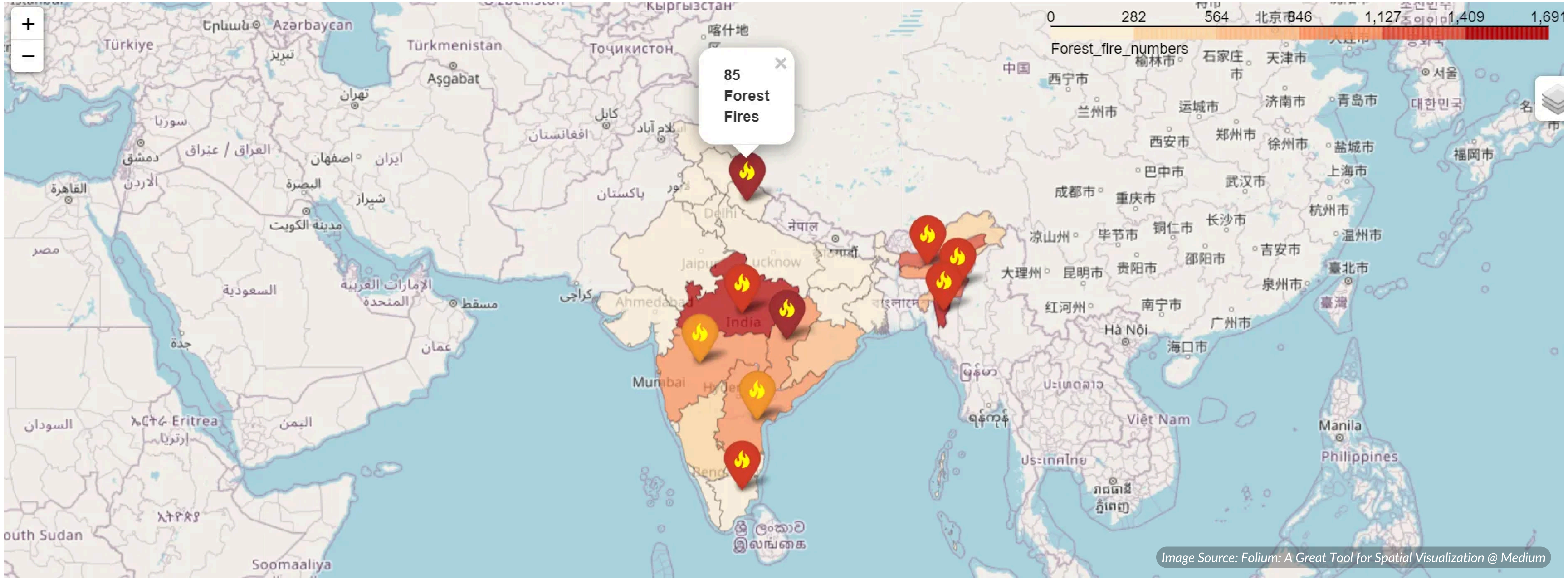


Preliminary Results

Exploring the Data

Deployment

Deploying the Model



Python

Flask

Folium

Azure Web App

Azure Machine Learning

Thank You!