By Sujen Purty
Qeds (Quantitative Economics and Data Science)

Algerian forest fires project

About My Project

- **1.** This project helps to **predict forest fires early**.
- **2.** It uses **real data** and **machine learning techniques**.
- **3.** The data is collected from two places in Algeria – **Bejaia** and **Sidi Bel-Abbes**.
- **4.** The data includes things like **temperature, wind speed, humidity, and rain**.
- **5.** Our machine learning model checks these values and **predicts if fire will happen or not**.
- **6.** This prediction can help the **forest department take early action**.
- **7.** It helps to **stop big fires and protect the forest**.

Introduction

Forest fires are dangerous and can destroy trees, animals, and homes. If we can predict forest fires early, we can save the forest and reduce damage.

This project is called Algerian Forest Fire Prediction. It uses real weather data and machine learning to check if a fire may happen.

We used data from two areas in Algeria: Bejaia and Sidi Bel-Abbes. The data includes details like temperature, humidity, wind, and rain.

By using this data, our model can predict whether there will be a fire or not. This can help the forest department take quick action and stop big fires before they spread.

Algerian Forest Fires Dataset

✅ **Total entries:** 244 records.
🌍 **Regions covered:**
- Bejaia (northeast Algeria) – 122 records
-**Sidi Bel-Abbes** (northwest Algeria) – 122 records

📅 **Data collection period:**
- From **June 2012 to September 2012**

📊 **Number of features:**

- 11 input features (like temperature, wind, rain, etc.)

- **1 output/class feature** (Fire or No Fire)

🔥 **Fire classification:**

- 138 instances marked as **Fire**

- 106 instances marked as **Not Fire**

Attribute Information:

📅 **Date** – Day, Month (June to September), Year (2012)

🌦️ **Weather Observations:**

🌡️ **Temp – Noon temperature (22°C to 42°C)**

💧 **RH (Relative Humidity) – 21% to 90%**

🌬️ **Ws (Wind Speed) – 6 to 29 km/h**

☔ **Rain – Rainfall in mm (0 to 16.8 mm)**

🔥 **Fire Weather Index (FWI) System Components:**

🔸 **FFMC (Fine Fuel Moisture Code) – 28.6 to 92.5**

🔸 **DMC (Duff Moisture Code) – 1.1 to 65.9**

🔸 **DC (Drought Code) – 7 to 220.4**

🔸 **ISI (Initial Spread Index) – 0 to 18.5**

🔸 **BUI (Build-Up Index) – 1.1 to 68**

🔥 **FWI (Fire Weather Index) – 0 to 31.1**

🟩 **Classes – Two categories: Fire and Not Fire**

By Sujen Purty
Qeds (Quantitative Economics and Data Science)

Algerian forest fires project

1. Algerian forest fires

- **Algerian Forest Fire Prediction Using Machine Learning**

2. Problem Statement

- To develop a predictive model that uses weather and environmental data to **predict the likelihood of forest fires** in two regions of Algeria — **Bejaia** and **Sidi Bel-Abbes** to help in **early detection, resource planning, and disaster prevention**.

2. **Business Use Case**

- Forest fires lead to environmental damage, economic loss, and threats to biodiversity.
- Early predictions help forest departments to take preventive measures like alerts, controlled burns, or patrolling.
- This system can assist in policy making, disaster management, and AI-based alert systems.

4. **Dataset Overview**

Source: UCI Machine Learning Repository – Algerian Forest Fires Dataset.
Records: 244 total entries (122 from Bejaia, 122 from Sidi Bel-Abbes).
Duration: June to September 2012.
Features:

- 11 Input Variables: Temp, RH, Wind, Rain, FFMC, DMC, DC, ISI, BUI, FWI, Region.
- 1 Output: Class (Fire / Not Fire).

5. Tools & Technologies

Category	Tools/Tech
Programming Language	Python
Libraries used	Pandas,Numpy,Scikit-learn,Matplotlib,Seaborn
Deployment	Flask
Version control	Git + GitHub

6. Data Preprocessing

Missing Value Handling: Checked and cleaned.

Encoding:

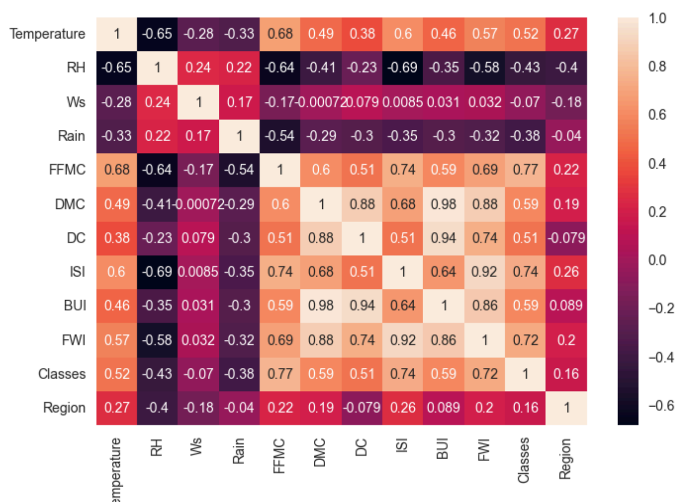
- **`Region`**: 0 = Bejaia, 1 = Sidi Bel-Abbes
- **`Class`**: 0 = Not Fire, 1 = Fire

Feature Scaling: StandardScaler used to normalize numerical features.

Train-Test Split: 80-20 split.

7. **Exploratory Data Analysis (EDA)**

Correlation Heatmap: Identified strong relations (e.g., Temp & Fire).



Color Legend:

Color bar on the right shows values from **-1 to +1**:

+1 (lightest color): Perfect positive correlation.
0 (medium shade): No correlation.
-1 (darkest color): Perfect negative correlation.

Here:

The correlation between `Temperature` and `RH` (Relative Humidity) is **-0.65**, which means:

- As **temperature increases**, **RH tends to decrease** (negative correlation).
- `FWI` (Fire Weather Index) is **strongly positively correlated** with `ISI`, `BUI`, `DC`, `DMC`, etc. (all above 0.8).
- **`Rain` has low or negative correlation with most variables, especially `FFMC` (-0.54).**

Key Observations:

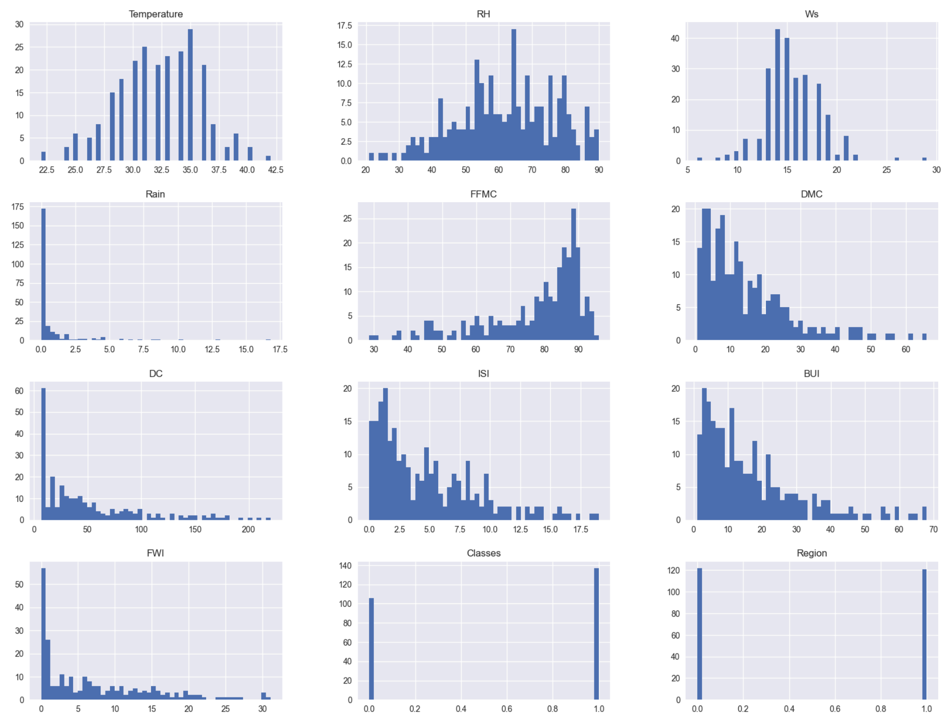
🔥 **Temperature is:**
- Negatively correlated with `RH` (-0.65).
- Positively correlated with fire indices like `FWI`, `DMC`, `ISI`, and `FFMC`.

💧 **Rain has:**
- Negative correlation with fire indices (`FFMC`, `DMC`, `DC`, etc.).
- This makes sense because more rain → less fire risk.

📈 **Fire indices (`FWI`, `ISI`, `BUI`, `DC`, `DMC`) are:**
- Highly positively correlated with each other, showing they measure related conditions for fire risk.

🗺️ **Region and Classes have weak correlation with most variables, suggesting they may be categorical or not strongly related to continuous variables.**

Distribution Plots: Showed imbalance in class distribution.



Continuous Features (Fire Weather & Environmental Factors)

1. Temperature

- Fairly normal distribution, slightly right-skewed.
- Most values fall between **27–35°C**, peaking around **30–32°C**.

2. RH (Relative Humidity)

- Slightly right-skewed, with a wide spread.
- Most common values are between **40–70%**, but some are as high as **90%**.

3. Ws (Wind Speed)

- Somewhat normal, centered around 12–16 km/h.
- Few extreme values above 25.

4. Rain

- Highly right-skewed — nearly all values are zero or very low.
- Indicates rain was rare during data collection.

5. FFMC (Fine Fuel Moisture Code)

- Bimodal distribution peaks around 70 and 90.
- Suggests varying levels of surface dryness.

6. DMC, DC, BUI (Fire weather codes)

- All are right-skewed, concentrated at low values.
- DMC (Duff Moisture Code), DC (Drought Code), and BUI (Build-Up Index) often start
- low, with few high outliers.

7. ISI (Initial Spread Index)

- Right-skewed, mostly values between 0–5.
- Indicates fire spread rate is typically low.

8. FWI (Fire Weather Index)

- Also right-skewed, with many low values and few high-risk fire days.

Categorical/Binary Features

9. Classes

- Values are **either 0 or 1**.
- Likely a **binary classification label**, e.g., fire occurrence: 0 = No Fire, 1 = Fire.
- Class imbalance is visible: more 0s than 1s.

10. Region

- Also binary: **two distinct groups** (likely Region 0 and Region 1).
- Each group appears approximately equal in size.
-

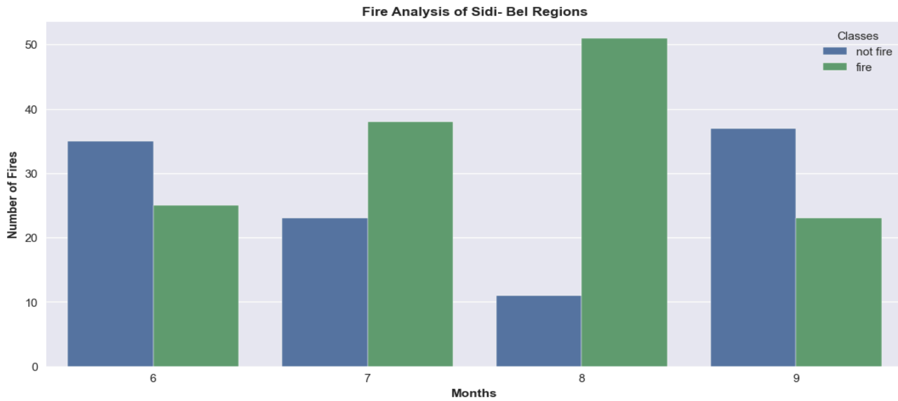
How to Read the Plot and Decide What to Apply

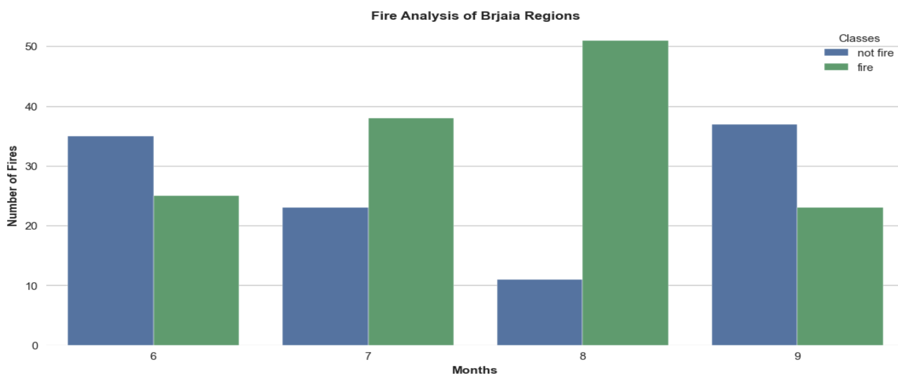
What You See in the Plot	What It Means	What to Apply	why	
🔼 Data is right-skewed (long tail to the right)	Extreme high values, non-normal	✅ Log or Box-Cox transformation	Normalize distribution	
🔽 Data is left-skewed (long tail to the left)	Extreme low values	✅ Square or cube transformation	Normalize distribution	
📈 Single central peak, bell-shaped	Normal-like distribution	✅ Standardization (z-score)	Prepare for linear models	
🔀 Two or more peaks (bimodal/multimodal)	Data comes from multiple groups	⚠️ Check for clustering or split data	May need separate modeling	
🔂 Very narrow range, same value repeats often	Low variability	⚠️ May drop feature or scale	Might not help model	
Two bars only, like `0` and `1`	Binary/categorical feature	✅ Leave as-is or encode	Depends on model input format	
⚠️ Huge spikes at one value (e.g. 0)	Potential imbalance or missing data	⚠️ Check for missing/outliers	Might need imputation or removal	

Decision of My Plot

feature	What we see	What to Do
Rain	Right-skewed, many zeros	Log-transform or binarize
FWI, ISI	Right-skewed	**Log-transform**
Temperature	**Normal distribution**	**Standardize**
RH	**Slightly right-skewed**	**Standardize or leave as-is**
FFMC	**Bimodal**	**Check for region-wise pattern**
Classes	**Binary (0,1), imbalanced**	**Resample or class weighting**
Region	**Binary (0,1)**	**Leave or one-hot encode**

Region-wise Analysis: Fire frequency higher in certain areas.





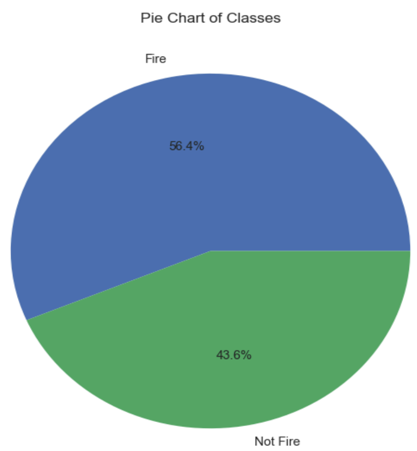
Monthly Fire Trend Observation

It is observed that August and September had the highest number of forest fires across both regions. From the monthly distribution plot, we can understand the following:

- 🔥 August recorded the most forest fires, indicating peak fire season.
- 📈 A high concentration of fires occurred during June, July, and August, suggesting these are the critical summer months for fire activity.
- 🔽 September still had fire incidents, but fewer compared to August.
- 🧊 Very few or no fires occurred in the other months, likely due to cooler or wetter conditions.

Seasonality: Higher fire risk in July & August.

Pie Chart



Observation from Pie Chart of Classes

Fire cases: 56.4% of the total dataset
Not Fire cases: 43.6%

Interpretation:

- The data is relatively balanced, though there are slightly more 'Fire' cases.
- No significant class imbalance, so:
    - we may not need resampling techniques (like SMOTE or undersampling).
    - However, we should still monitor performance metrics like precision, recall, and F1-score, especially if misclassifying a "Fire" case is risky.

How WeCan Use This in Reporting:

"The dataset has a fairly balanced class distribution with 56.4% fire incidents and 43.6% non-fire instances. This balance allows for effective supervised classification without severe bias toward one class."

8. **Model Building**

- **Algorithms Tested:**
- Logistic Regression
- Random Forest Classifier
- K-Nearest Neighbors

- **Hyperparameter Tuning:** GridSearchCV used for optimization.

- **Best Model:** Random Forest (highest accuracy and F1-score).

9. **Model Evaluation**

Metrics Used:

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix

Performance: ~95% Accuracy with Random Forest.

10. Web Application (Flask Interface)

Input Form: Users enter values for temperature, humidity, wind, etc.
Output: Predicts "🔥 Fire" or "✅ No Fire"
UI Features:
- Background theme with forest/fire image.
- Centered inputs and animated results.

11. Insights

Key Drivers of Fire: High temperature, low humidity, high FFMC and ISI.
Preventive Zones: Sidi Bel-Abbes showed more high-risk entries.
Weather Pattern: No rainfall days are more prone to fire.

12. **Benefits**

- Helps forest officers and disaster teams make timely decisions.
- Reduces damage to wildlife, environment, and property.
- Can be extended to mobile apps or IoT-based alerts.

13. Limitations

- Only 4 months of data (June–Sept 2012).
- Works only for Algerian regions.
- Requires daily weather data feed for real-time predictions.

14. Future Scope

- **Expand to other regions or countries.**
- **Add satellite image analysis.**
- **Real-time weather API integration for live prediction.**

15. References

- UCI Machine Learning Repository
- Forest Fire Index (FWI) documentation
- Scikit-learn Documentation