

**A DOCUMENTATION ON**

**FIRE WEATHER INDEX PREDICTOR**



Project Duration  
DEC 2025 – JAN 2026

Project Mentor  
PRAVEEN

PROJECT SUBMITTED BY

Pranjali Kolawale  
**Infosys Springboard Internship Program**

# **INDEX**

Introduction.....	1
Problem Statement.....	1
Objectives.....	1
1. Module 1: Data Collection and Initial Data Preparation	
1.1 Loading Dataset and Explored Structure.....	2
1.2 Identifying Missing Values and Incorrect Values.....	2
1.3 Cleaning Column Names and Formats.....	2
1.4 Handling Missing Values.....	2
1.5 Validating Data After Cleaning.....	3
2. Module 2: Data Exploration and Preprocessing	
2.1 Checking for Duplicates and Data Overview.....	4
2.2 Handling Categorical Features.....	4
2.3 Visualizing Data Distribution.....	4 - 6
2.4 Preparing the Dataset.....	6
2.5 Saving the Processed Data.....	6

## **FIGURES**

Figure 1.....	4
Figure 2.....	5
Figure 3.....	6

# FIRE WEATHER INDEX PREDICTOR

## INTRODUCTION

Wildfires have become more common in recent years and can cause serious damage to forests, animals, houses, and even people. Changes in climate, such as higher temperatures and uneven rainfall, make it easier for fires to start and spread. Because of this, it is important to know the fire danger level in advance so that forest officials and emergency teams can prepare and take action on time.

The Fire Weather Index (FWI) is a measure used to understand how risky the weather conditions are for a wildfire to occur. It depends on factors like temperature, humidity, wind speed, rainfall, and the dryness of forest fuel. Checking all these conditions manually can be slow and may not always give accurate predictions.

To solve this problem, this project aims to build a machine learning model that can predict the FWI using environmental data. By analysing the dataset and creating a predictive system, the project helps in giving early warnings and improving wildfire risk management.

## PROBLEM STATEMENT

Wildfires can spread quickly and cause heavy damage, so it is important to know the fire risk in advance. The Fire Weather Index (FWI) helps measure how likely a fire is to start, but calculating it manually is not always accurate or fast. To solve this, the aim of this project is to use environmental factors like temperature, humidity, wind speed, and rainfall to build a machine learning model that can predict the FWI automatically. This will help in giving early warnings and improving wildfire preparedness.

## OBJECTIVES

- To study the weather factors that affect the Fire Weather Index (FWI).
- To clean and prepare the dataset for accurate analysis.
- To explore the data and understand its patterns.
- To build a machine learning model that can predict FWI.
- To test the model and improve its accuracy.
- To create a simple web app for making FWI predictions.

## MODULE 1: DATA COLLECTION AND INITIAL DATA PREPARATION

### 1.1 Loading Dataset and Explored Structure

Collected the Fire Weather Index (FWI) Dataset from the Kaggle. Then dataset was imported into a Pandas DataFrame. Basic exploration was done by checking the first few rows, the shape of the dataset, column names, and data types. This helped in understanding the kind of information available, such as weather parameters (*Temperature, RH, Ws, Rain*) and FWI-related indices (*FFMC, DMC, DC, ISI, BUI, FWI*), along with the *Classes* column representing fire risk levels.

### 1.2 Identifying Missing Values and Incorrect Values

The dataset was checked for missing entries and inconsistencies. Some columns, particularly *Classes*, *FWI*, and *DC*, contained missing or misformatted values, like spaces within numbers or text-based numeric entries. Detecting these issues early is important to ensure accurate analysis and modeling later.

```
missing_row = df[df.isnull().any(axis=1)]  
print(missing_row)
```

### 1.3 Cleaning Column Names and Formats

Column names were standardized by removing extra spaces. Numeric columns that had text or unwanted spaces, such as *DC*, were cleaned by stripping spaces and converting them into proper numeric types. This step ensures that all calculations and analyses will be accurate and avoids errors during preprocessing.

```
df.columns = df.columns.str.strip()  
  
df['DC'] = df['DC'].astype(str).str.strip()  
  
df['DC'] = df['DC'].str.replace(' ', '')  
  
df['DC'] = df['DC'].astype(float)
```

### 1.4 Handling Missing Values

Different strategies were applied based on column type. For the categorical *Classes* column, missing values were filled with the mode, as it represents the most common fire risk category. For numeric columns like *FWI*, invalid entries were replaced with the mean of the column after conversion to numeric. This ensures that the dataset remains complete and ready for exploratory analysis.

```
df['Classes'] = df['Classes'].fillna(df['Classes'].mode()[0])
```

## 1.5 Validating Data After Cleaning

After performing cleaning and handling missing values, the dataset was reviewed again using info() and sample rows. This confirmed that all corrections were applied properly, column types were consistent, and the dataset was ready for further analysis. This step helps prevent issues in downstream tasks like EDA or modeling.

```
print(df.info())
```

## MODULE 2: DATA EXPLORATION AND PREPROCESSING

After collecting and cleaning the FWI dataset in Module 1, the next step was to explore the data and perform preprocessing to ensure it is ready for modeling. This module focused on understanding the data distribution, identifying patterns, handling categorical features, and preparing the dataset for machine learning.

### 2.1 Checking for Duplicates and Data Overview

The dataset was first examined for duplicate records, which could skew the analysis. Fortunately, no duplicates were found, confirming that each record was unique. Descriptive statistics were generated for numerical features to understand their distribution, central tendency, and variability. This helped in identifying potential anomalies, extreme values, or inconsistencies.

### 2.2 Handling Categorical Features

The dataset contained the categorical feature **Region**, which was converted into numerical values using label encoding. This transformation allowed the model to interpret region information effectively while maintaining the distinction between different regions.

```
df['Region'] = LabelEncoder().fit_transform(df['Region'])
```

### 2.3 Visualizing Data Distribution

Visualization played an important role in understanding the dataset:

- **Boxplots** provided insights into variations and outliers, particularly across regions, helping identify extreme values in feature like Temperature.

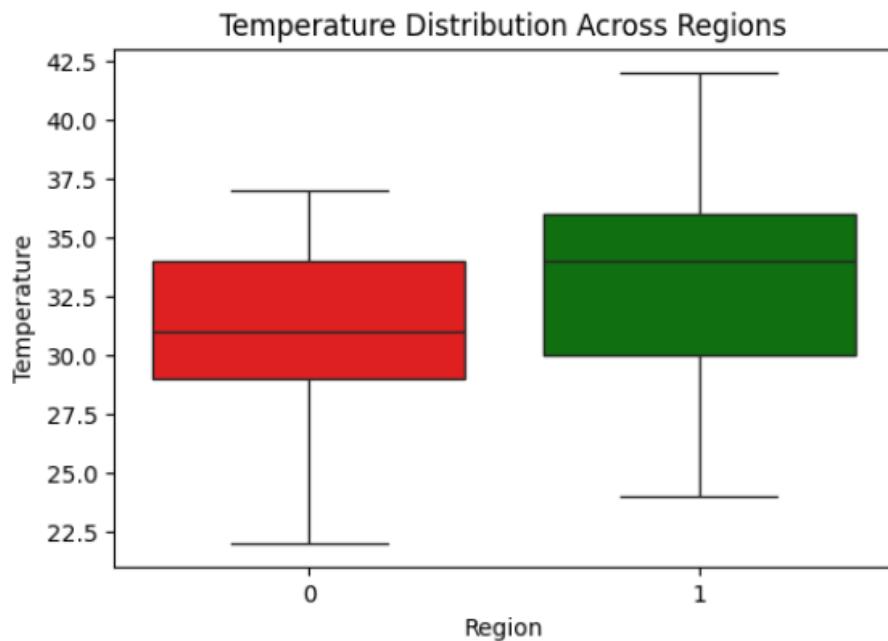


Figure 1

- **Histograms** were used to examine the distribution of numerical features in the data.

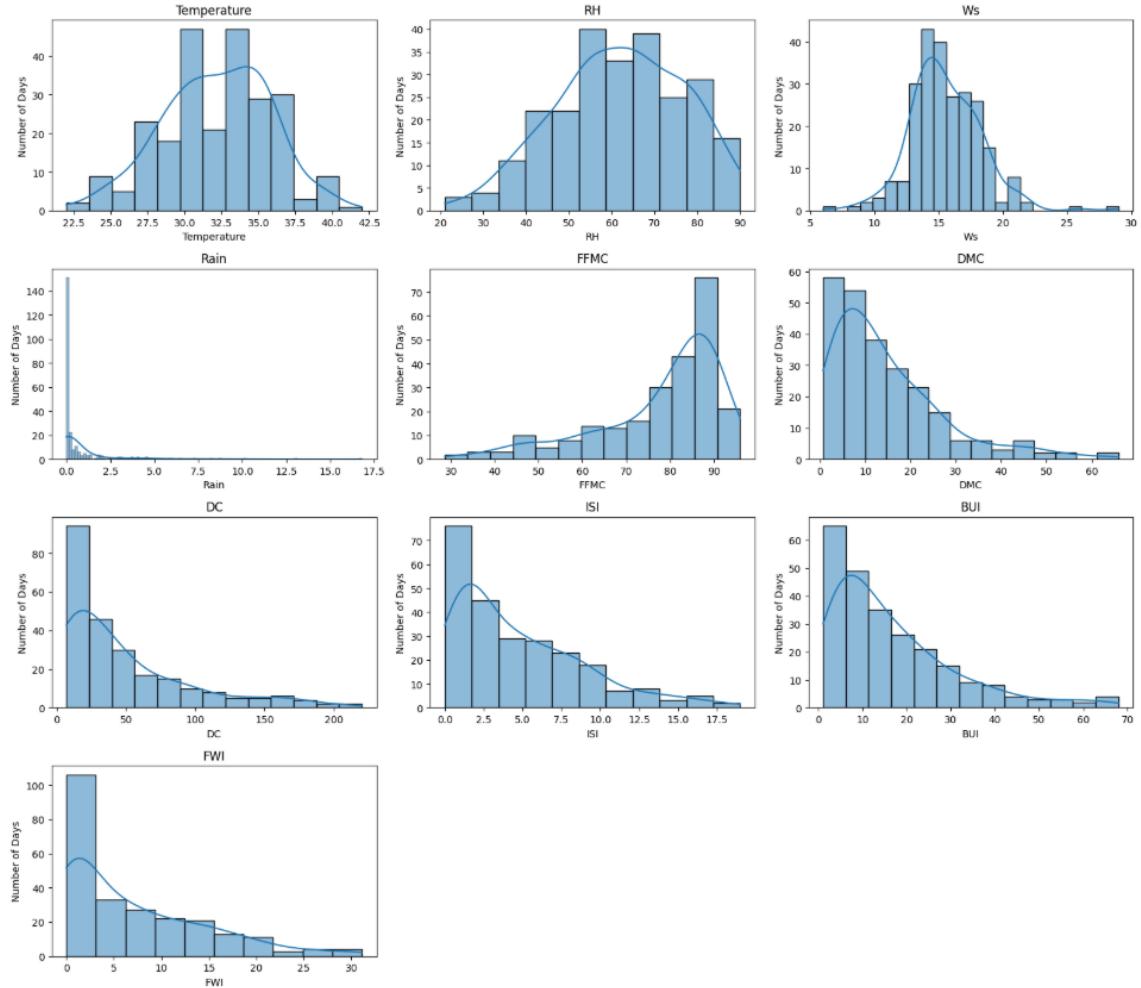


Figure 2

This figure shows histograms with smooth curves for all the main variables in the dataset, such as Temperature, Humidity, Wind Speed, Rainfall, and different fire-risk indexes (FFMC, DMC, DC, ISI, BUI, FWI). These plots help us understand how the data is spread out. Most variables like Rain, DC, ISI, BUI, and FWI are right-skewed, which means high values occur only on a few days. Temperature and RH appear more balanced, showing regular day-to-day patterns. Overall, the figure gives a clear idea of the typical value ranges, variations, and the occasional extreme conditions that can affect fire weather.

- **Correlation heatmaps** were generated to study relationships between features and the target variable FWI. Features showing strong correlation with FWI were noted as important for model input.

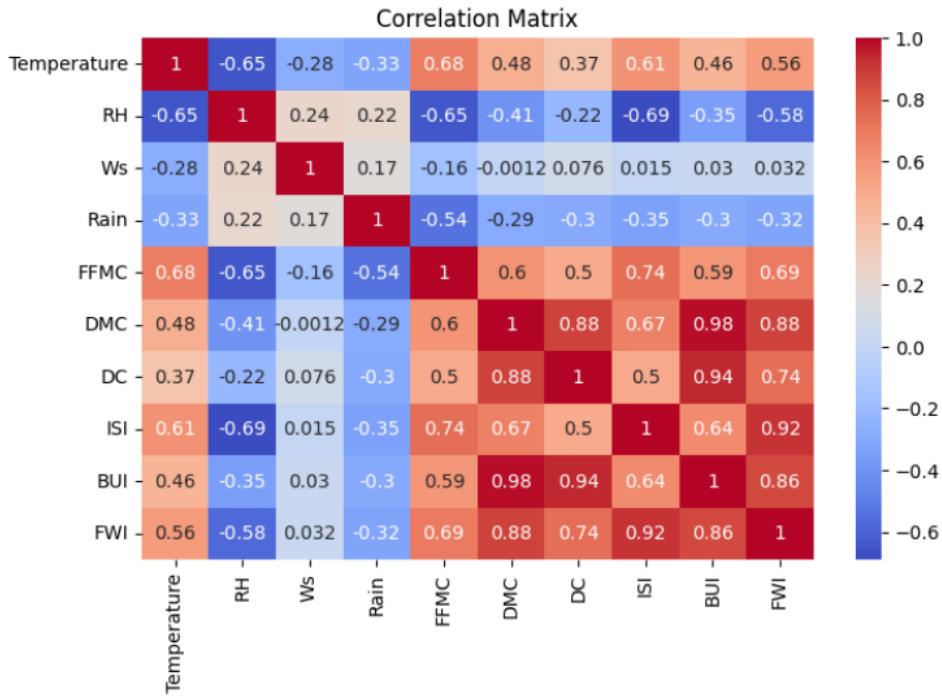


Figure 3

This correlation matrix shows how the main weather and fire-risk variables are related to each other. Fire-related indexes like DMC, DC, ISI, BUI, and FWI have strong positive correlations, meaning they increase together. Temperature also shows positive links with some fire indexes. In contrast, Relative Humidity has strong negative correlations with Temperature and several fire-risk variables. Rainfall has weak negative relationships overall. This figure helps identify which factors influence fire-weather conditions the most.

## 2.4 Preparing the Dataset

After visual exploration, unnecessary columns, such as **Classes**, were dropped since they were not relevant for FWI prediction. All numerical features were checked for consistency and corrected if needed, ensuring that the dataset was clean and formatted correctly for subsequent modeling.

```
df = df.drop('Classes', axis=1)
```

## 2.5 Saving the Processed Dataset

Finally, the fully cleaned and preprocessed dataset was saved as a new file. This ensured that the dataset could be easily used in the next modules for feature engineering, scaling, and model training.

```
df.to_csv("./datasets/FWI_Dataset_Cleaned.csv", index=False)
```