TITLE: Fire Weather Index Predictor

(A Machine Learning Model To predict Fire Weather Index)

Infosys SpringBoard Virtual Internship Program 6.0

Submitted By

Manyatha N

Under the guidance of Mentor **Praveen**

Project Statement

The increasing frequency and destructiveness of wildfires puts ecosystems, human populations, and financial resources at danger. The Fire Weather Index (FWI), a widely used predictor of wildfire threat, is based on meteorological elements such as temperature, relative humidity, wind speed, rainfall, FFMC, DMC, ISI, and region. However, the substantial reliance of conventional FWI estimates on human procedures and static models limits the speed and accuracy of early-warning systems. The project's objective is to create a machine learning-based FWI prediction model that uses a whole pipeline that includes feature engineering, data preprocessing, StandardScaler-based normalization, and Ridge Regression modeling to assess real-time environmental data. In order to facilitate rapid user input and automatic wildfire-risk prediction, the system is further integrated into a Flask web application, assisting emergency planners, forest departments, and climate researchers in proactive wildfire management.

Expected Outcomes

- A machine learning model based on Ridge Regression that can correctly forecast the Fire Weather Index (FWI).
- A complete preparation pipeline that prepares and normalizes environmental data using StandardScaler.
- A functional Flask web application that lets users enter values and get FWI forecasts in real time.
- An intelligent, data-driven wildfire risk assessment tool for researchers and authorities.

Modules to be Implemented

1. Data Collection
2. Data Exploration & Data Preprocessing
3. Feature Engineering & Scaling
4. Model Training using Ridge Regression
5. Evaluation & Optimization
6. Deployment via Flask App
7. Presentation & Documentation

System Requirements

1. Python
2. Pandas
3. Scikit-learn (Ridge Regression, StandardScaler)
4. Matplotlib / Visualization tools (used in EDA)
5. Flask (for deployment)
6. Pickle (.pkl files) (ridge.pkl, scaler.pkl)
7. Git / GitHub (for final submission)

Milestone 1

Module 1

1. Data Collection

The primary goal of is to gather and prepare the fundamental dataset needed to develop the Fire Weather Index (FWI) prediction model. In this phase, a structured dataset is collected that includes the FWI target variable and important environmental characteristics like temperature, relative humidity, wind speed, rainfall, FFMC, DMC, ISI, and region. After that, the data is examined to make sure the formatting, data types, and consistency are correct. Checking sure each column is appropriately represented and prepared for analysis is part of this. The cleaned dataset is next put into a Pandas DataFrame, which serves as the foundation for all further preprocessing, model construction, and exploration procedures.

Loading and Inspecting the FWI Dataset (Head, Info, and Summary Statistics)

```python
import pandas as pd

file_path = 'FWI_Dataset.csv'
df = pd.read_csv(file_path)

print(df.head())
print(df.info())
print(df.describe())
```

Output

- Check for the attributes in raw dataset
- Drop unwanted / unnecessary attributes (class, DC, Region)

```
   day  month  year  Temperature   RH   Ws  Rain   FFMC  DMC    DC  ISI  BUI  \
0    1      6  2012           29   57   18    0.0  65.7  3.4   7.6  1.3  3.4
1    2      6  2012           29   61   13    1.3  64.4  4.1   7.6  1.0  3.9
2    3      6  2012           26   82   22   13.1  47.1  2.5   7.1  0.3  2.7
3    4      6  2012           25   89   13    2.5  28.6  1.3   6.9  0.0  1.7
4    5      6  2012           27   77   16    0.0  64.8  3.0  14.2  1.2  3.9

   FWI    Classes    Region
0  0.5  not fire     Bejaia
1  0.4  not fire     Bejaia
2  0.1  not fire     Bejaia
3    0  not fire     Bejaia
4  0.5  not fire     Bejaia
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   day          244 non-null    int64
 1   month        244 non-null    int64
 2   year         244 non-null    int64
 3   Temperature  244 non-null    int64
 4    RH          244 non-null    int64
 5    Ws          244 non-null    int64
 6   Rain         244 non-null    float64
 7   FFMC         244 non-null    float64
 8   DMC          244 non-null    float64
 9   DC           244 non-null    object
 10  ISI          244 non-null    float64
 11  BUI          244 non-null    float64
 12  FWI          244 non-null    object
 13  Classes      243 non-null    object
 14  Region       244 non-null    object
dtypes: float64(5), int64(6), object(4)
memory usage: 28.7+ KB
None
```

FIG – 1: Basic information of a dataset

Cleaned_FWI_Dataset:

- Cleaned FWI dataset generated.

	day	month	year	Temperature	RH	Ws	\
count	244.000000	244.000000	244.0	244.000000	244.000000	244.000000	
mean	15.754098	7.500000	2012.0	32.172131	61.938525	15.504098	
std	8.825059	1.112961	0.0	3.633843	14.884200	2.810178	
min	1.000000	6.000000	2012.0	22.000000	21.000000	6.000000	
25%	8.000000	7.000000	2012.0	30.000000	52.000000	14.000000	
50%	16.000000	7.500000	2012.0	32.000000	63.000000	15.000000	
75%	23.000000	8.000000	2012.0	35.000000	73.250000	17.000000	
max	31.000000	9.000000	2012.0	42.000000	90.000000	29.000000	

	Rain	FFMC	DMC	ISI	BUI
count	244.000000	244.000000	244.000000	244.000000	244.000000
mean	0.760656	77.887705	14.673361	4.774180	16.664754
std	1.999406	14.337571	12.368039	4.175318	14.204824
min	0.000000	28.600000	0.700000	0.000000	1.100000
25%	0.000000	72.075000	5.800000	1.400000	6.000000
50%	0.000000	83.500000	11.300000	3.500000	12.250000
75%	0.500000	88.300000	20.750000	7.300000	22.525000
max	16.800000	96.000000	65.900000	19.000000	68.000000

FIG – 2: Descriptive Statistics of the Collected Environmental Dataset

Module 2

2. Data Preprocessing

Data preprocessing is a crucial step that prepares the collected dataset for reliable model training. At this point, the dataset is carefully examined for null or missing values, which are treated properly to maintain data integrity. To avoid skewed model behavior, boxplots and statistical criteria are used to identify outliers. Histograms and density plots are used to assess the feature distributions in order to spot trends and problems with data quality. Relationships between variables are explored using correlation matrices and scatterplots to understand how environmental features influence the Fire Weather Index (FWI). Additionally, label encoding or mapping techniques are used to encode categorical data like region so that machine learning algorithms may use them. After cleaning, transforming, and validating all features, the processed dataset is saved for use in the next stages of feature engineering and modeling.

Data Preprocessing: Missing-value handling, outlier treatment, encoding, and scaling

- **Libraries and modules used from python**

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import LabelEncoder
```

- **Missing-value handling**

```
numerical_cols = ['Temperature', 'Relative Humidity', 'Wind Speed', 'Rain', 'FFMC', 'DMC', 'ISI', 'FWI']
for col in numerical_cols:
    if col in df.columns:
        df[col] = pd.to_numeric(df[col].astype(str).str.strip(), errors='coerce')
        df[col].fillna(df[col].median(), inplace=True)
```

```
day             0
month           0
year            0
Temperature     0
 RH             0
 Ws             0
Rain            0
FFMC            0
DMC             0
DC              0
ISI             0
BUI             0
FWI             0
Classes         1
Region          0
dtype: int64
day                   0
month                 0
year                  0
Temperature           0
Relative Humidity     0
Wind Speed            0
Rain                  0
FFMC                  0
DMC                   0
DC                    0
ISI                   0
BUI                   0
FWI                   0
Classes               1
Region                0
dtype: int64
```

FIG-3: Missing Values Summary of the Dataset

- **Histogram statistics**

```
for col in numerical_cols:
    if col in df.columns:
        plt.figure()
        sns.histplot(df[col], kde=True)
        plt.title(f'Distribution of {col}')
        plt.show()
```

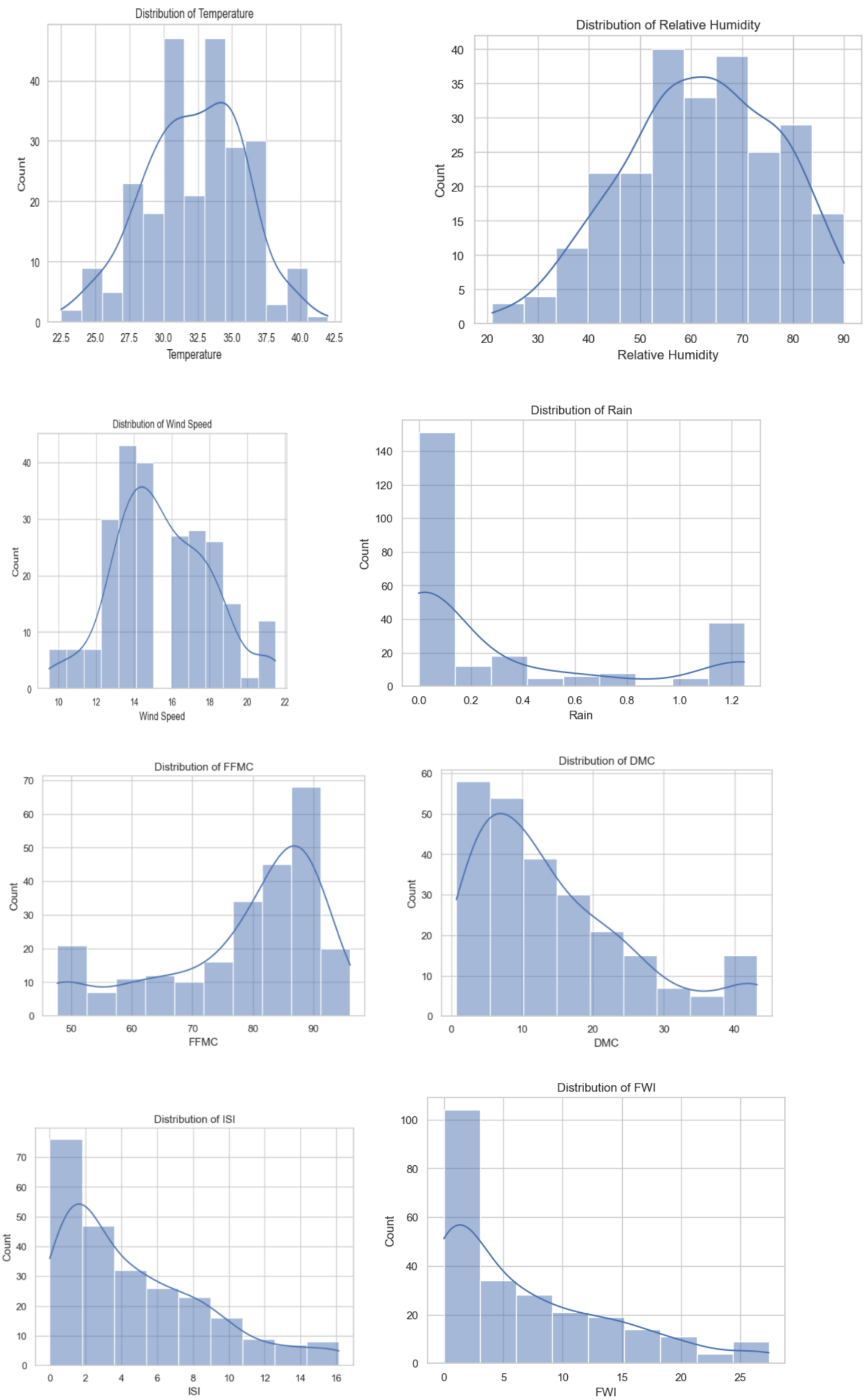

FIG-4: Histograms of statistics

- **Correlation matrix**

```
plt.figure(figsize=(10, 8))
corr_cols = [c for c in numerical_cols if c in df.columns]
corr_matrix = df[corr_cols].corr()
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

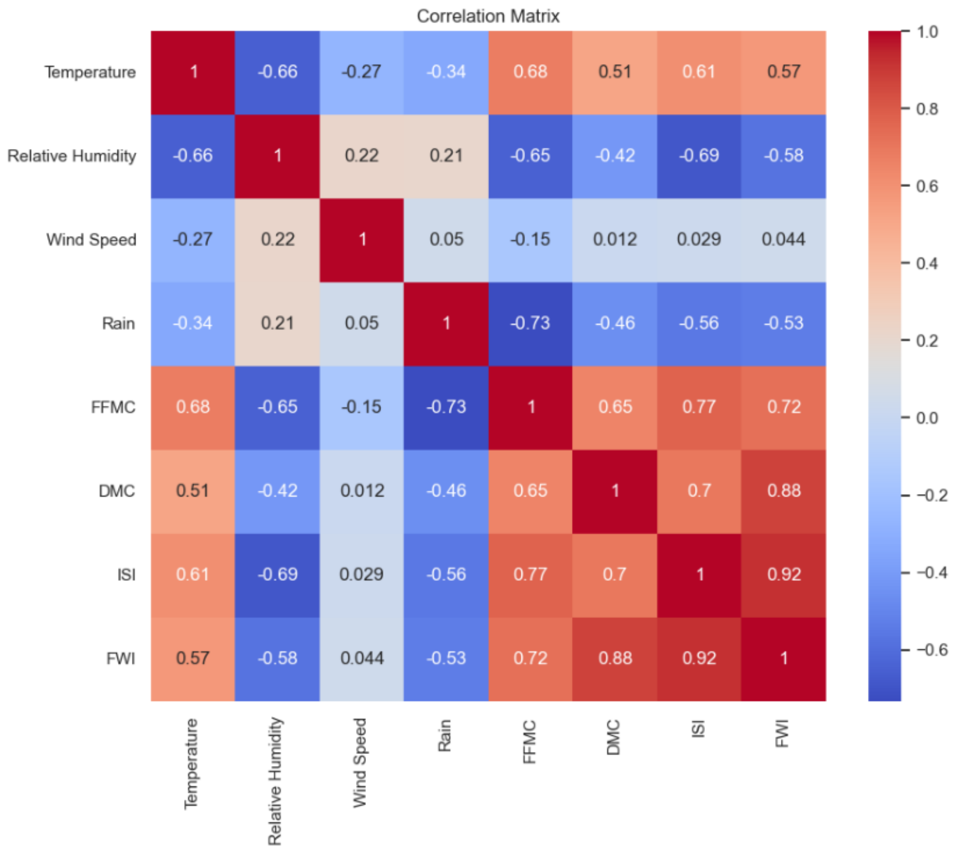

FIG-5: Correlation matrix

- **Scatter plot of FWI**

```
if {'Temperature', 'FWI'}.issubset(df.columns):
    plt.figure()
    sns.scatterplot(x='Temperature', y='FWI', data=df)
    plt.title('Temperature vs FWI')
    plt.show()
```

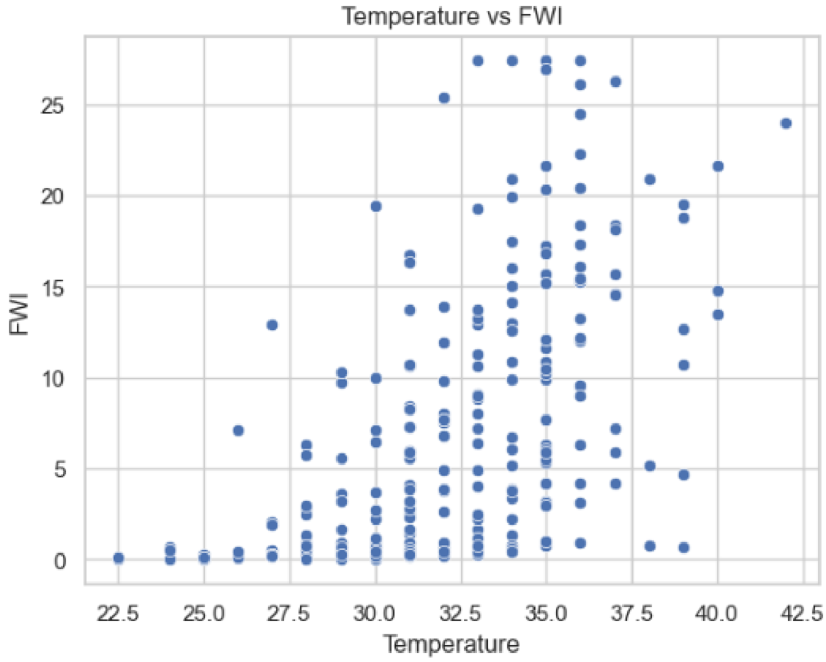

FIG-6: Scatter plot for features of FWI

- **Preview after preprocessing**

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   day                244 non-null    int64  
 1   month              244 non-null    int64  
 2   year               244 non-null    int64  
 3   Temperature        244 non-null    float64
 4   Relative Humidity  244 non-null    float64
 5   Wind Speed         244 non-null    float64
 6   Rain               244 non-null    float64
 7   FFMC               244 non-null    float64
 8   DMC                244 non-null    float64
 9   DC                 244 non-null    object 
 10  ISI                244 non-null    float64
 11  BUI                244 non-null    float64
 12  FWI                244 non-null    float64
 13  Classes            243 non-null    object 
 14  Region_encoded     244 non-null    int32  
dtypes: float64(9), int32(1), int64(3), object(2)
memory usage: 27.8+ KB
None
   day  month  year  Temperature  Relative Humidity  Wind Speed  Rain  \
0    1      6  2012         29.0               57.0        18.0  0.00   
1    2      6  2012         29.0               61.0        13.0  1.25   
2    3      6  2012         26.0               82.0        21.5  1.25   
3    4      6  2012         25.0               89.0        13.0  1.25   
4    5      6  2012         27.0               77.0        16.0  0.00   

      FFMC  DMC    DC  ISI  BUI  FWI      Classes  Region_encoded  
0  65.7000  3.4   7.6  1.3  3.4  0.5  not fire                  0  
1  64.4000  4.1   7.6  1.0  3.9  0.4  not fire                  0  
2  47.7375  2.5   7.1  0.3  2.7  0.1  not fire                  0  
3  47.7375  1.3   6.9  0.0  1.7  0.0  not fire                  0  
4  64.8000  3.0  14.2  1.2  3.9  0.5  not fire                  0  
```

FIG-7: Dataset Information and Preview After Preprocessing