Fire Hazard Analysis in Bejaia and Sidi-Bel Abbes Regions

1. Introduction

The comprehensive dataset provides detailed meteorological and environmental data collected over several months in the Bejaia and Sidi-Bel Abbes regions. It includes essential variables such as temperature, relative humidity, wind speed, rainfall, and various fire danger indices like FFMC - Fine Fuel Moisture Code DMC - Duff Moisture Code, DC - Drought Code ISI - Initial Spread Index BUI - Buildup Index FWI - Fire Weather Index.

1. Problem Statement and Objectives:

The dataset covers the period from June to September in 2012, allowing for a seasonal analysis of fire hazards

Understanding the factors influencing the occurrence of fire incidents is crucial for effective fire management and prevention.

The objective of analysing this dataset is to make a predictive model that can predict whether a fire will occur or not based on the various meteorological and environmental factors so that this dataset can be leveraged to assess and mitigate the risk of wildfires in these regions, aiding in proactive fire management and prevention strategies.

1. Methodology:

* Data Preparation and Exploration:

The first step was opening the dataset and then dividing it into two different files as there were two different datasets in one file and then converting them to csv files.

The next step was cleaning the data involved by removing any null and/or unwanted and redundant values from the dataset. Ensuring all data is appropriate for analysis.

The next step involved finding out the basic details of the chosen datset like the shape to figure out the numbers of columns and rows in it and then using the info() functions to find out the column names and the dtype of these values in these columns.

Then we use the describe function to generate descriptive statistics of the dataframe . It provides information such as count, mean, standard deviation, minimum, and maximum values.

* Data Exploration

We used various functions to identify which meteorological and environmental factors have more or less of an effect on whether forest fires will occur or not.

The first step was importing matplotlib and setting the plotting style and font size.

Then the next step was to show the covariance and correlation of the various parameters with the feature Classes which contains the information of whether a fire occurred or not.

Then using lists to count and divide subsets of Classes-Fire and Not Fire-according to different categories depending on the parameters used.

Then using these lists I made bar graphs and plotted line graphs to emphasise the different between the values and labels used in the bar graphs.

* Data Modeling:

I built a predictive model using machine learning algorithms for logistic regression. I selected appropriate algorithms based on the nature of the dataset and then trained the model for which I had to scale the model down and then test it out to ensure that the model was predicting the data correctly.

* Model Validation:

Then I took user input in which the user can enter meteorological and environmental data they have collected into the model and get a prediction of whether a forest fire will occur or not.

1. Result and snapshots

We found out the mean of temperature, RH, WS, DMC, DC, FFMC, ISI,BUI,FWI.

The mean temperature is 31.88.

The mean Relative Humidity is 67.644

The mean Wind Speed is 16

The mean Rain is 0.845

The mean Fine Fuel Moisture Code is 74.679

The mean Duff Moisture Code is 12.4728

The mean Drought Code is 53.516

The mean Initial Spread Index is 3.242

The mean  Buildup Index is 15.450

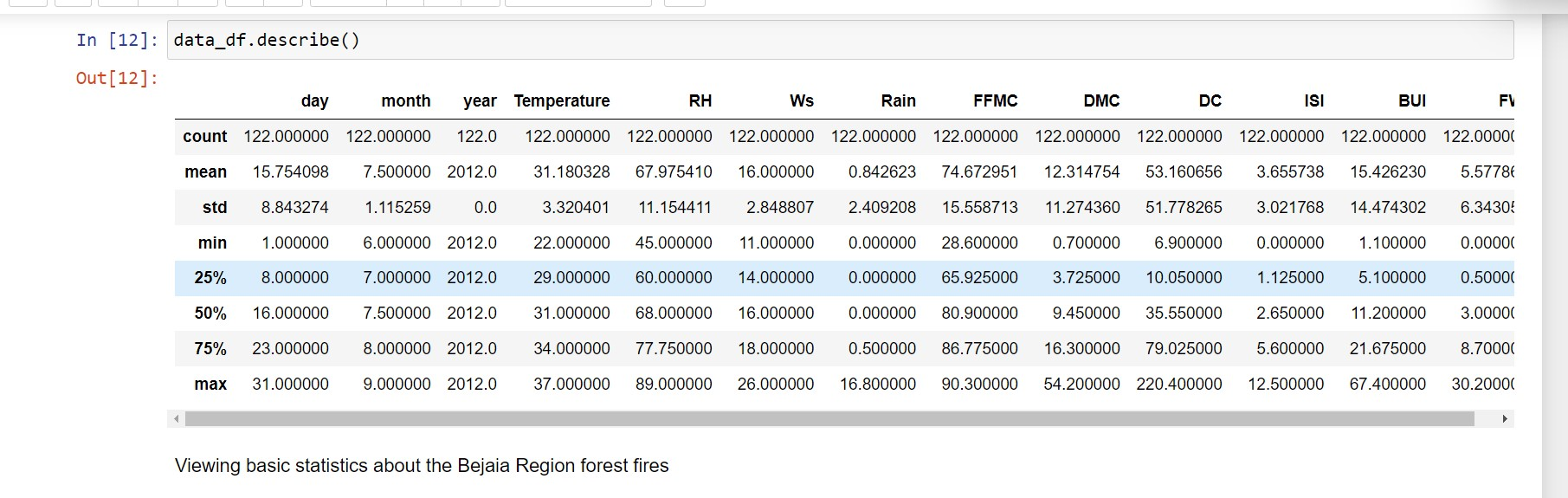
The mean Fire Weather Index is 5.537

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



A white rectangular object with a white background

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A graph with a red line

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Conclusion

Based on the analysis, we can see that the model predicts the occurrence of a forest fire correctly 96% of the time.

Analysing the relationship between the various parameters using various bar graphs and functions like covariance and correlation suggests that some parameters are more important than others when predicting the occurrence of forest fires.

1. Limitations & Future Work

* Limitations:

The dataset is limited to a specific time frame (June to September 2012) so that doesn’t provide us with a very large time frame for building a better prediction model.

Accuracy is 96% only so 4% of the time erroneous judgements are being made about whether or not a forest fire will occur.

* Future Work:

Extend the analysis to multiple years for a more comprehensive understanding.

Include additional environmental variables for a more nuanced analysis.