A picture containing font, white, symbol, graphics

Description automatically generated

«Algerian Forest Fires Dataset Data Set».

 Alice Anne-Catherine J Destrait (230AIB016)

Professor : Alla Anohina-Naumeca

Report

Practical Assignment 2

Digitalization impact

21/04/2023

Contents

[Introduction 3](#_Toc135383303)

[Exploring the data 4](#_Toc135383304)

[Scatter plots 4](#_Toc135383305)

[References 5](#_Toc135383306)

# Introduction

The dataset used in this practical assignment has been obtained via following link: [UCI Machine Learning Repository: Algerian Forest Fires Dataset Data Set](https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++)

The authors, Faroudja Abid & Nouma Izeboudjen, have developed a dataset called “Algerian Forest Fires Dataset Data Set”. The problem domain of this dataset includes forest fire prediction and fire management strategies. Furthermore, there is no information about the licensing aspects of the dataset. The dataset was obtained as, “We have used the meteorological observations for the period of summer of 2012, from June to September since the ﬁre occurrence is high on this period and the 2012 is the year where the recorded ﬁre occurrence is the highest from 2007 to 2018.” (F. Abid & N. Izeboudjen, 2020). The dataset contains 243 instances and 14 features. The features are divided into two roles namely numerical (day, month, year, temperature, rain, RH, Ws, FFMC, DMC, DC, ISI, BUI and FWI) and categorical (classes) see Table 1. The classes have two values: no fire given the value 0 (106 instances) and fire given the value 1 (137 instances) see Figure 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Meaning | Value type | Range of values |
| Day | Day of the observation | Numerical | 01 to 31 |
| Month | Month of the observation | Numerical | 06 to 09 |
| Year | Year of the observation | Numerical | 2012 |
| Temperature | Temperature noon | Numerical | 22 to 42 |
| Rain | Rain of the day in mm | Numerical | 0 to 16.8 |
| RH | Relative humidity in % | Numerical | 21 to 90 |
| Ws | Wind speed in km/h | Numerical | 6 to 29 |
| FFMC | Fine fuel moisture code index from the FWI system | Numerical | 28.6 to 92.5 |
| DMC | Duff moisture code index from the FWI system | Numerical | 1.1 to 65.9 |
| DC | Drought code index from the FWI system | Numerical | 7 to 220.4 |
| ISI | Initial spread index from the FWI system | Numerical | 0 to 18.5 |
| BUI | Buildup index from the FWI system | Numerical | 1.1 to 68 |
| FWI | Fire weather index | Numerical | 0 to 31.1 |
| Classes | Fire or no fire | Categorical | 0 or 1 |

Table 1: Features with explanations.

A picture containing text, screenshot, number

Description automatically generatedA picture containing text, screenshot

Description automatically generated

Figure 1: Snippet of the structure of the data

# Exploring the data

## A graph with red and blue dots Description automatically generated with low confidenceScatter plots

A picture containing text, screenshot, plot, line

Description automatically generatedFigure 2: FFMC – ISI scatter plot Figure 3 : FWI – ISI scatter plot

The first scatter plot I decided to analyse is the fine fuel moisture code and initial spread index. The graph shows an exponential relationship between the two variables see figure 2. Furthermore, the class separability is clearly seen. The no fire incidents have a mostly linear relationship with a slightly positive slope. The fire incidents clearly have a steeper slope. This shows a relatively stronger association between the two variables. The second scatterplot (see figure 2) has a linear relationship. The graph is interesting because of the class distribution. The non-fire events are clustered together in the lower left corner while the fire incidents mostly have higher values.

## Histogram

## Distributions

A picture containing text, plot, diagram, line

Description automatically generatedA graph with red and blue lines

Description automatically generated with low confidence

Figure 6: FFMC distribution Figure 7: BUI distribution

# References

[(PDF) Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm (researchgate.net)](https://www.researchgate.net/publication/339062373_Predicting_Forest_Fire_in_Algeria_Using_Data_Mining_Techniques_Case_Study_of_the_Decision_Tree_Algorithm)

[UCI Machine Learning Repository: Algerian Forest Fires Dataset Data Set](https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++)