Statistics Final – Project

Student Name: Khubim Kumar Chhetri **Student ID:** D00251757

Project Title: **Analysis of Forest Fire and Temperature**

Background Research/ Problem Statement

A forest fire is unplanned and uncontrollable and can occur by lightning or human carelessness in forests, grasslands or shrubbery (Government of Canada, 2020). The vast majority of forest fires are human-caused, however dry climate, hot temperatures, lightning, and volcanic eruption can also lead to their occurrence (National Park Service, 2018). The past decade of climate change has only exacerbated the amount of forest fires, leading to more frequent and extreme occurences.

Wildfire agencies use many variables to indicate an imminent wildfire and the evolution of machine learning has provided us the ability to predict future events by analyzing these variables. Thus, we pose the predictive question: do certain variables allow us to determine if a forest fire has or will occur and if so, how accurate will they be?

Forest fires are a major concern all over the world; each year, millions of hectares are lost. Algeria is one of the countries afflicted by this phenomena, which occurs primarily during the summer time. Forest fire detection and forecasting become critical issues for reducing the disaster's damage. Exploration of new fire detection and forecast systems as alternatives to existing ones becomes a necessity. The goal is to predict whether or not the fire will break out based on weather data.

To support our hypothesis, we used a dataset on Algerian Forest Fires from UCI (Faroudja & Izeboudjen, 2020). The dataset contains a culmination of forest fire observations and data in two regions of Algeria: the Bejaia region and the Sidi Bel-Abbes region. The timeline of this dataset is from June 2012 to September 2012. In this project, we focused on whether certain weather characteristics could predict forest fires in these regions . we used a dataset on **Algerian Forest Fires from UCI**. The dataset contains a culmination of forest fire observations and data in two regions of Algeria: the Bejaia region and the Sidi Bel-Abbes region.

The timeline of this dataset is from June 2012 to September 2012. In this project, we focused on whether certain weather features could predict forest fires in these regions using few Machine Learning algorithms.

The dataset used can be found here from the UCI Machine Learning Repository.

We analyse it by taking into considering the following objectives:

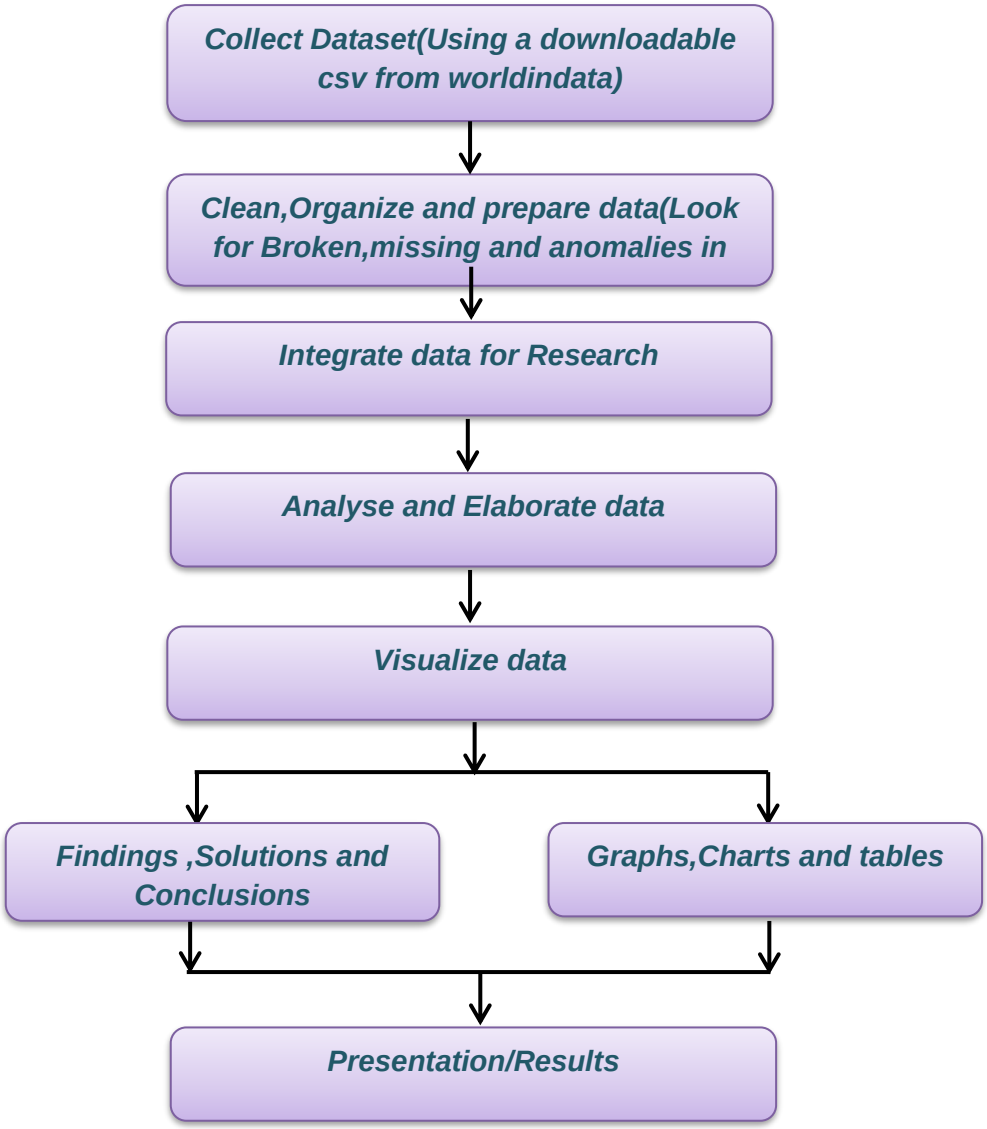
Research questions	Project Goals
Predict the possibility of fire and predict temperature?	Learn more about Forest fire happening and check its impacts over the years
Which Years are going to generate more Temperature coming years?	To detect future fires based on certain Weather report.
Predict the change in atmospheric Humidity over the coming years.	Understand different factors affecting the Fire and check for accuracies to find out the cause?
Check and find out if there is any seasonality in Fire changes?	Gain Hands-on experience with visualization, Hypothesis testing and the Machine learning models with accuracy score

2. Research Plan/Design

I Plan to spend nearly 12 days for collection, cleaning and preparation of data. The data can be collected from various sources from https://archive.ics.uci.edu/ml/dataset where we can get the downloadable csv file, with the clear guidance and inference drawn from the data. With the documentation reports.

Once the data is collected , I can go for initial or preliminary investigation .I will plan to check for misleading , broken data that is need to be fixed or removed for further analysis. After cleaning data and integrating dataset I will compare all dataset together to analyze correctly for the changes in mobility before and during the Pandemic.

Apart from dataset, I may look for articles ,helpful information which are relevant and required for to find the result or findings of the project. Consequently once the dataset are arranged ,the data then analyse and identify for results. The data needs to be displayed and visualized with graphs, plots, hypothesis, conclusions once the statistical results are valid and clear from the analysis.



Aim and Objectives

The paper aims to develop a trustworthy and interpretable Hypothesis amd visualisation that will predict the temperature that will determine the forest fire. Deadly forest fires have become an annual scourge in the North African country,Algeria where climate change is turning large areas into a tinderbox. The identification of the factors that will check for fire/No fire can aid Regions and country in significantly reducing the factors responsible to determine and take precaution.

The objectives of the research are based on the above aim and are as follows:

- To analyze the relationship and visualize patterns of Temperature
- To suggest suitable steps to extract the most value from the data including picking the most significant features
- To find appropriate techniques to determine the factors on the dataset
- To compare the classification or regression to visualise accurate to determine the Classes
- To predict whether or not the fire will break out based on weather data

Significance of the research

The research is contributing to the explanation and interpretation of the prediction of various predictive factors to support decision making and increase the safety measures that are going to minimize the fire. This will help Countries allocate budget and time to the Region that are likely to Burn by running targeted campaigns. This can help the any country the pain points faced by its region and can ultimately help aid in fundamental policy changes that can increase the significance measures to minimize the cause.

Scope of the study

Due to the limitation of the time frame in this research, the scope of the research will be limited to the below points:

- The data for the study has directly been obtained from the authorized source.
- The research will include the development and evaluation of various visualization techiniques and various models will not be considered as a part of this study due to a lack of resources and time
- The study will limit the use of classification and Regression such as logistic regression, MLR,SLR for this study

Research Methodology

Forest fire is a disaster that causes economic and ecological damage and human life threat. Thus predicting such critical environmental issue is essential to mitigate this threat. In this paper, we were able to identify that the telecom industry is an extremely competitive industry where customers have the free will to move across companies if they believe they are getting more value with another service provider. In this step, we will apply Exploratory Data Analysis (EDA) to extract insights from the data set to know which features have contributed more in predicting Forest fire by performing Data Analysis using Pandas and Data visualization using Matplotlib & Seaborn. The fire prediction is based on the meteorological data corresponding to the critical weather elements that influence the forest fire occurrence, namely temperature, relative humidity and wind speed. It is a classification and regression problem. Here we can predict the possibility of fire and predict temperature

It is always a good practice to understand the data first and try to gather as many insights from it.

Data Understanding

There are various data sources used to predict Forest fires are a major concern all over the world; each year, millions of hectares are lost. Algeria is one of the countries afflicted by this phenomena, which occurs primarily during the summer time. Forest fire detection and forecasting become critical issues for reducing the disaster's damage. Exploration of new fire detection and forecast systems as alternatives to existing ones becomes a necessity. The goal is to predict whether or not the fire will

break out based on weather data.The given data consists of multiple factors about the Forest fire regarding Relative Humidity, Wind speed, FFMC, DMC, DC, FWI, Rain, with the Region in a Yes or No format that can be leveraged post-processing. It is presented in a .csv format with different attributes information as metadata.

Data Set Information:

- The dataset includes 244 instances that regroup a data of two regions of **Algeria**,namely the
- **Bejaia region** located in the **northeast of Algeria** and the **Sidi Bel-abbes region** located in the **northwest of Algeria**.
- 122 instances for each region.
- The period from June 2012 to September 2012.
- The dataset includes 11 attribues and 1 output attribue (class)
- The 244 instances have been classified into **fire** (138 classes) and **not fire** (106 classes) classes.

Attribute Information:

- **Date :** (DD/MM/YYYY) Day, month ('june' to 'september'), year (2012)

Weather data observations

- **Temp :** temperature noon (temperature max) in Celsius degrees: 22 to 42
- **RH :** Relative Humidity in %: 21 to 90
- **Ws :** Wind speed in km/h: 6 to 29
- **Rain:** total day in mm: 0 to 16.8

FWI Components

- **Fine Fuel Moisture Code (FFMC) index from the FWI system:** 28.6 to 92.5. It is the numeric rating in the moisture content of litter and other cured fine fuels.And this indicates the the factors or relative rate of ignition and the flammability of fine fuel.
- **Duff Moisture Code (DMC) index from the FWI system:** 1.1 to 65.9.It is the numeric rating of the average moisture content of loosely compacted organic layers of moderate depth. This indicates fuel consumption in moderate duff layers and medium-size woody material.
- **Drought Code (DC) index from the FWI system:** 7 to 220.4.It is the numeric indicator of the average moisture content of deep, compact organic layers. This can be used as the indicator of seasonal drought effects on forest fuels and the amount of smoldering in deep duff layers and large logs.
- **Initial Spread Index (ISI) index from the FWI system:** 0 to 18.5.It is the numeric indicator of the expected rate of fire spread. It is based on wind speed and FFMC.
- **Buildup Index (BUI) index from the FWI system:** 1.1 to 68.It is the numeric indicator of the total amount of fuel available for combustion. It is mainly based on DMC and the DC.
- **Fire Weather Index (FWI) Index:** 0 to 31.1.It is the numeric indicator of fire intensity. It is based on the ISI and the BUI, and is used as a general index of fire danger throughout the forested

Classes: two classes, namely **Fire** and **not Fire**

The **Canadian Forest Fire Weather Index (FWI) System** comprised of 6 factors that plays important role in the impacts of moisture in the fuel present in the forest and weather conditions on fire behavior.

The first three factors are fuel moisture codes, which represent the numeric ratings of the moisture content of the forest floor and other dead organic matter that are seen in the forest for the fire.So if the moisture content decreases then it plays a major role in fire.

The other 3 factors are fire behavior indices, that can be used as spread rate of fire, the fuel for combustion, and the upfront intensity of the fire; these three values increases as the fire danger increases

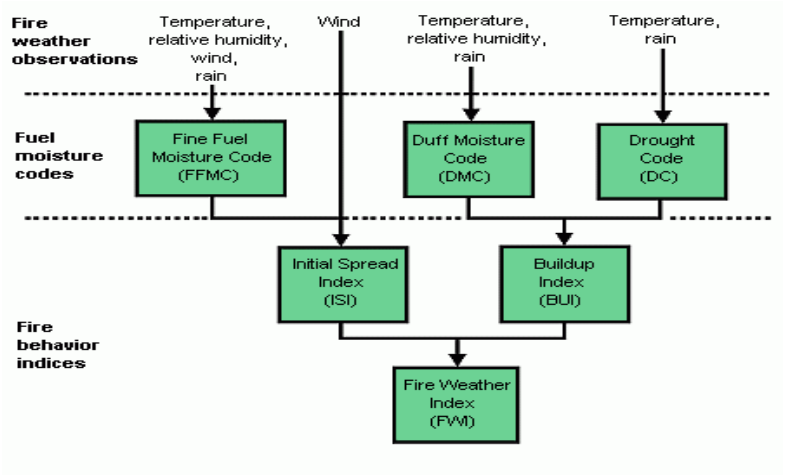


Figure: Natural Resource Canada(Src:cwfis.cfs.nrcan.gc.ca)

Data Selection

There were a few datasets we can choose from when it comes to Regression and classification. The data we have selected is the Algerian forest fire Data. The dataset is at an Class level dataset with binary problem and can be used as Regression problem . The dataset has information that can be leveraged at a Region level to identify Classes likely to Fire /No Fire.

In Bejaïa, the summers are warm, muggy, dry, and mostly clear and the winters are long, cold, wet, and partly cloudy. Over the course of the year, the temperature typically varies from *8°C* to *29°C* and is rarely below *5°C* or above *32°C*.

In Sidi Bel Abbès, the summers are short, hot, dry, and mostly clear and the winters are long, cold, windy, and partly cloudy. Over the course of the year, the temperature typically varies from *3°C* to *34°C* and is rarely below *-1°C* or above *38°C*. Dataset shows the hazards based on both natural factors, such as humidity, as well as artificial factors, such as litter and climate change

The information obtained from the data can be broken down into four broad categories and is as follows:

- Day,Month,Year are the dates of the observation of the given weather causing Fire
- Fuel moisture codes or the ease of ignition (FFMC,DMC,DC) are the moisture content of the fuel in forest.
- Fire behaviour or Intensity information such as ISI,BUI,FWI
- The given data consists of multiple factors about the Classes in a Fire or No Fire format that can be leveraged post-processing. It is presented in a .csv format with customer attributes information as metadata

Understanding the different segments of the data available will help us profile the various segments and their effects, which will, in turn, be able to accurately flag the set of factors that are indicative of Classes(Fire/No Fire).

Data Visualisation

Data visualisation is an integral part of exploratory data analysis to be able to understand the data. We can use the packages to analyse and understand the data prep. This will help us understand the distribution of the columns, the variance, and the data profile. Comparing the data visually before and after processing will also help us understand datasets that will serve as inputs to the ML operation.We shall visualise a few of the features and the target variables to understand the distribution of the data points.

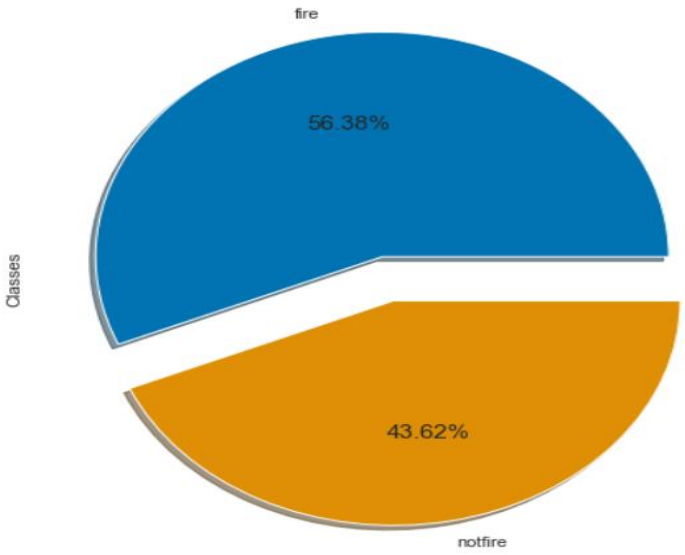


Figure: Pie Chart of Classes

From the Above Pie charts that showcase the % of the response variable, it is noted that the crucial features that results in a target are the Fire as compared to the No Fore . The attributes have the highest variance.

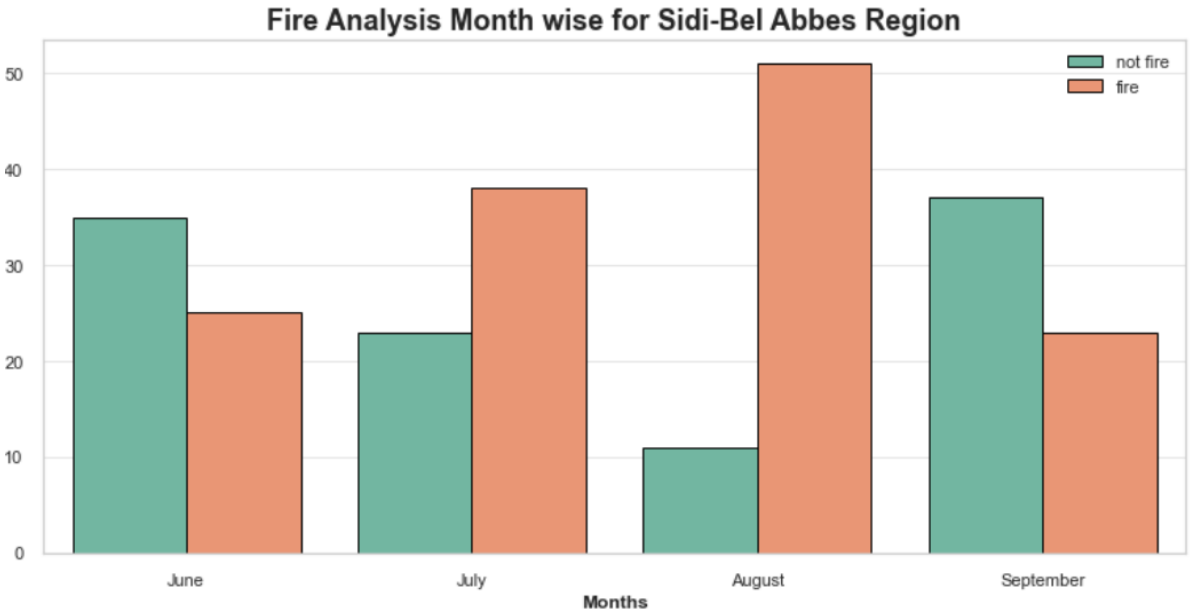


Figure2: Month wise Class Analysis

Its observed that from the above and below Bar chart (Fig-2, Fig-2.1) that July and September had the most number of forest fires for both regions. And from the above plot of months, we can understand few things.Most of the fires happened in August and very high Fires happened in only 3 months - **June, July and August and** Less Fires was on September

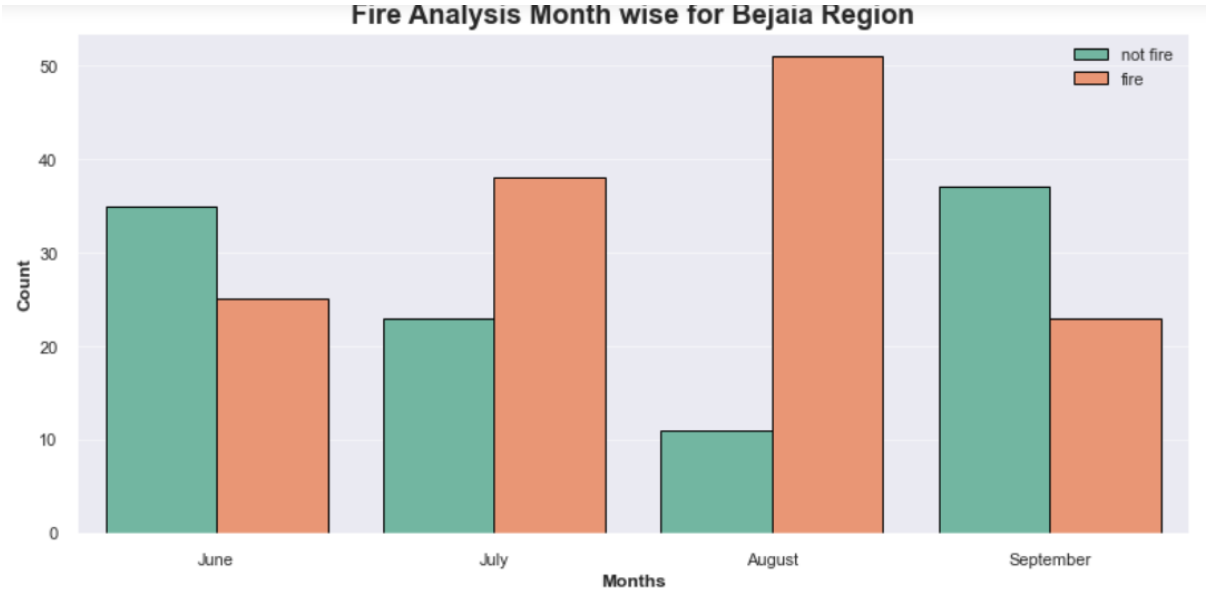


Figure2.1: Month wise Class Analysis

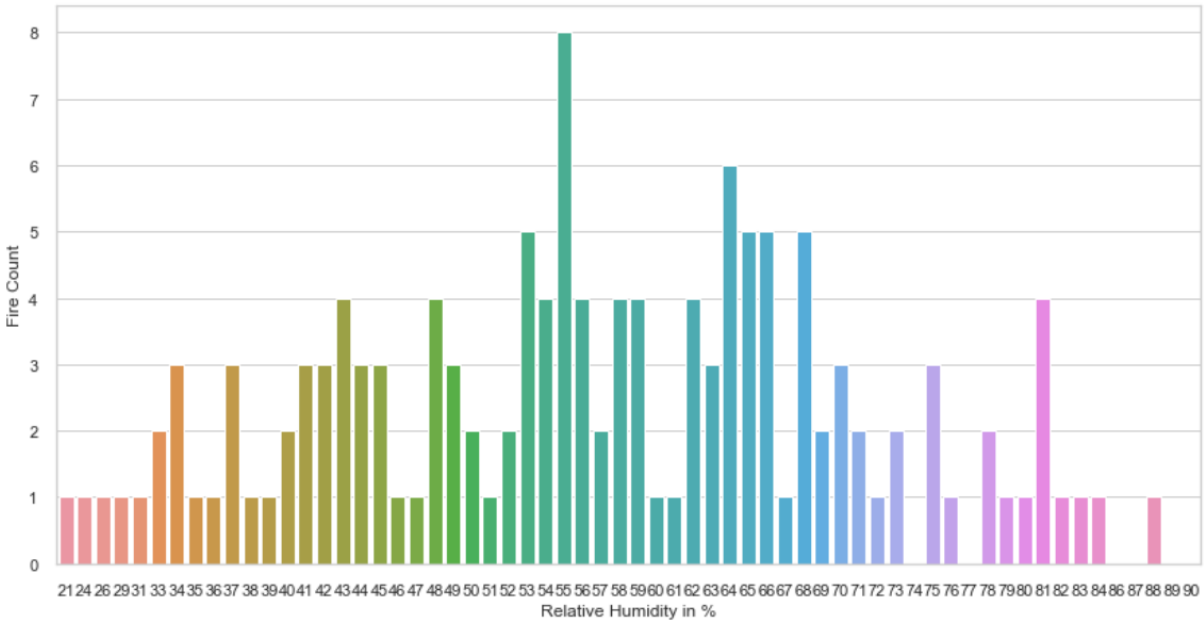


Figure 3: Relative Humidity Vs Fire Count

From the above bar plots ,we see that the RH greater than 53 results in high chance of fire .This can also be verified by the following below scatter plots between different variables with respect to the Temperature.(Fig 3 & 3.1)

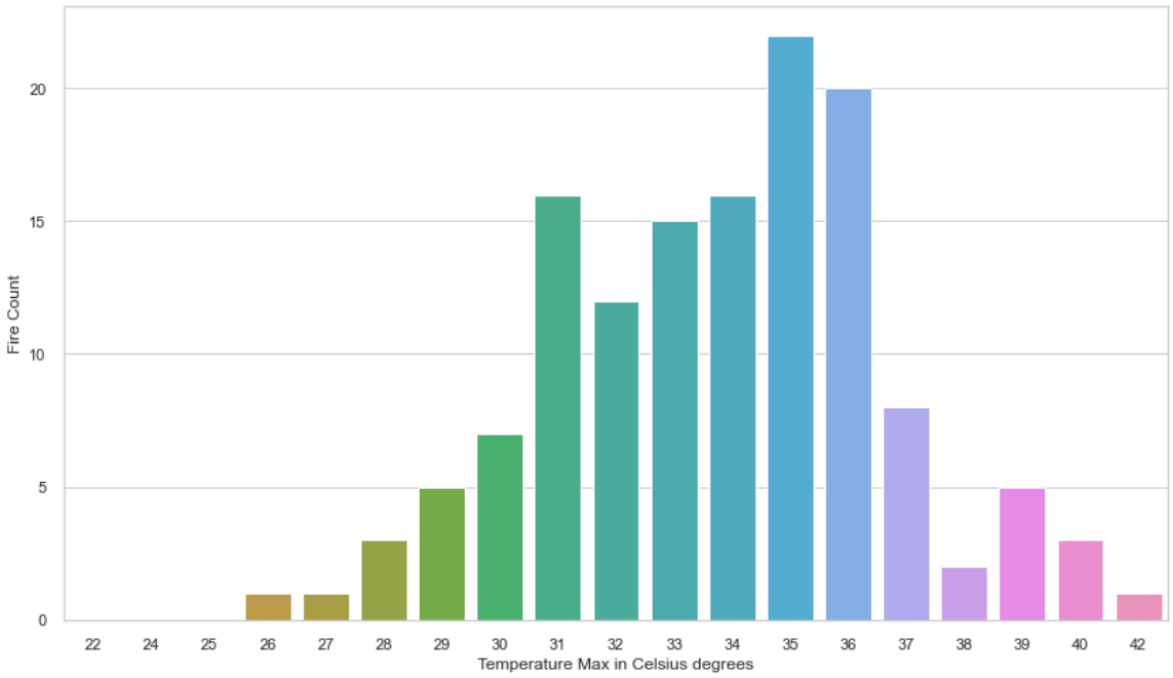
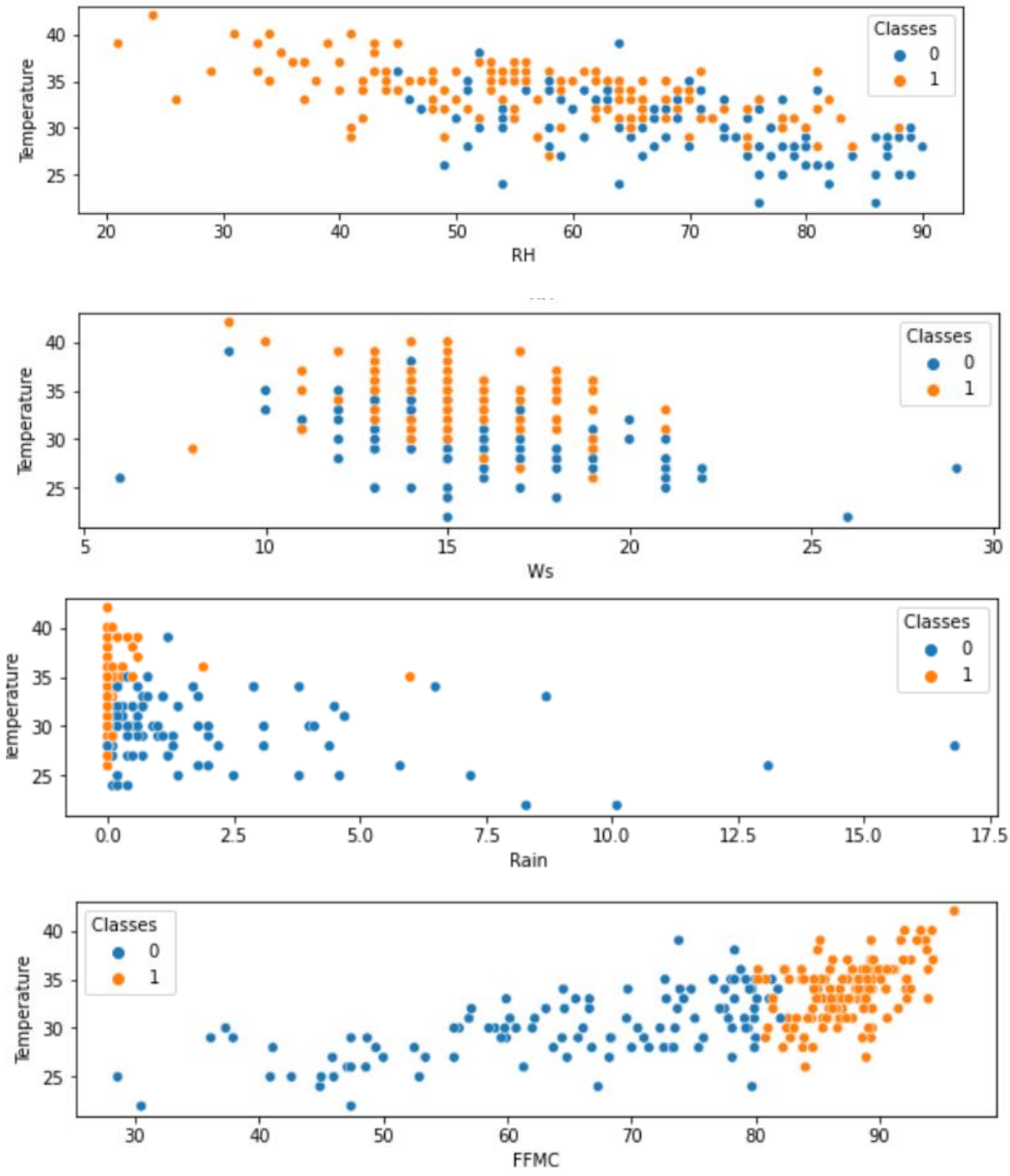


Figure 3.1: Temperature Vs Fire Count

In this below following scatter graphs, we can observe that the different variables such as RH (Relative Humidity) with low value of RH there would be high chance of Fire whereas with the high RH there may be less chance of Fire , ISI(about 2.5), FWI(about 3), BUI (about 10), features showing that after some range there is very high chances of fire .Rain feature shows at 0 there is very high chances of fire.FFMC feature shows that above 80 there is very high chances of fire.(Fig 4)



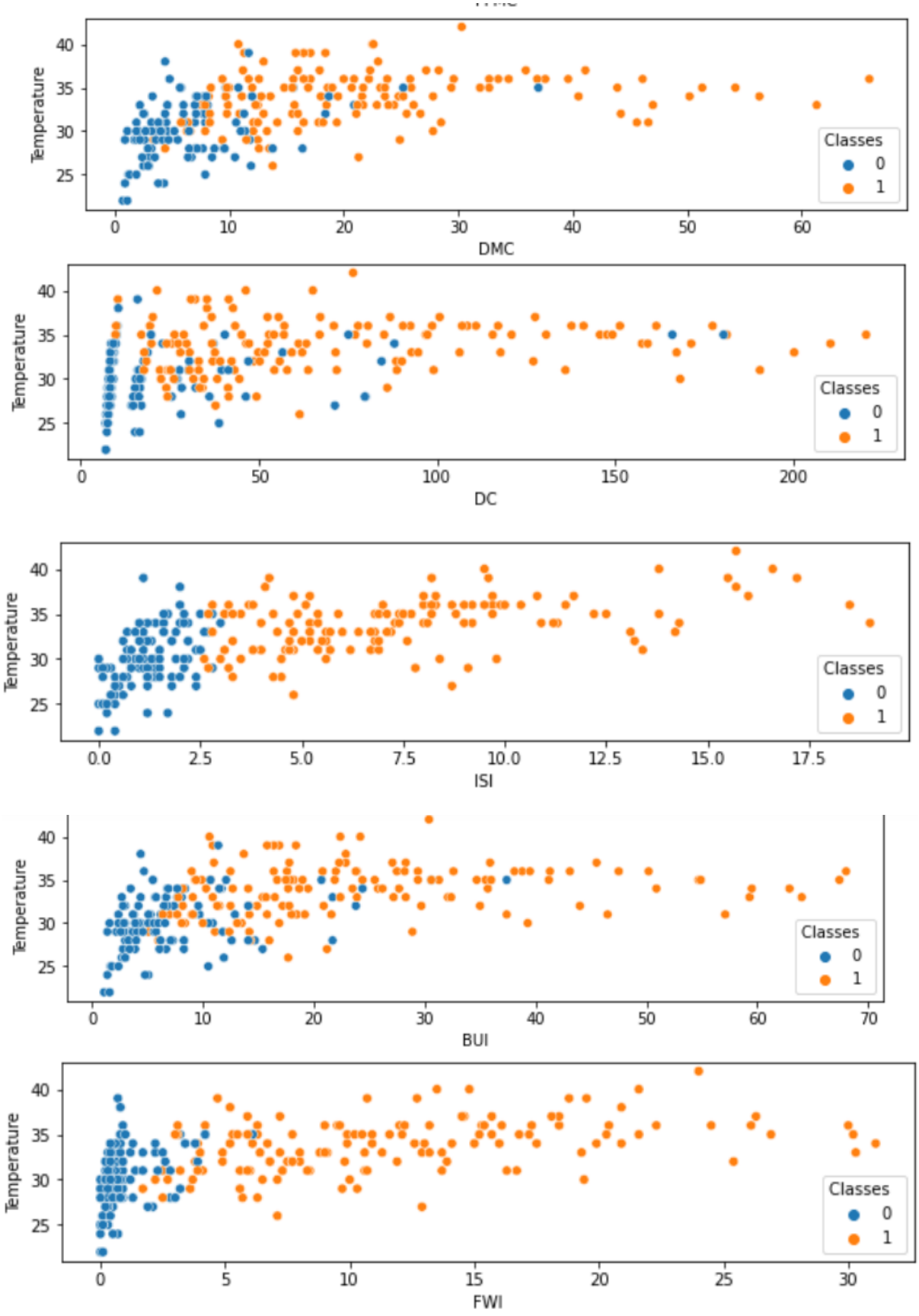


Figure 4: Temperature Vs Features (RH, ISI, FMMC, DMC, DC, Rain, FWI, BUI)

Special Requirement and deliverable of the project

There is No special requirements for the project. The deliverables are. The geographical factors provided in the dataset which may be considered as the responsible for the

Conclusion

In this work, we attempt to predict the factors of forest fires with weather data. Using this kind of data is advantageous, as it can be collected in real time and at a lower cost with respect to other techniques.

We build and evaluate several linear regression models, reaching mediocre results. Then, we solve a simplified version of the problem, namely predicting whether a forest fire is bigger or smaller than a given threshold, by employing a logistic regression model.And we also try to solve the simplified binary classification task with the logistic regression.

Unfortunately, factors are not able to predict the responsible factors correctly to determine the root cause of forest fires in a satisfactory manner using the data at our disposal. This is a difficult regression task, and more advanced techniques may be necessary .

Nonetheless, the linear models are able to give some useful insights on the relationship between predictors and response. They are especially useful to understand what are the variables that influence of forest fires, and to estimate how much they affect the response.

References

Alexander, M.E.; St. John, P.; Thorburn, R.W.; Simons,P.; MacMillan, A. 2002. CD-ROM based training course Understanding the Fire Weather Index(FWI) System now available! In D.X. Viegas, ed.Forest fire research and wildland fire safety:
(Canadian Wildland Fire Information System | Canadian Forest Fire Weather Index (FWI) System n.d.)

(UCI Machine Learning Repository: Algerian Forest Fires Dataset Data Set n.d.)

Turner , J . A . 1 9 7 0 . Calibrat ion of the Fire Weather Index in British Columbia. Progr . Rep. 1. Can. Fore s t . S erv. , Pacific Forest Res . Centre ,Victoria.
(Sidi Bel Abbès Climate, Weather By Month, Average Temperature (Algeria) - Weather Spark n.d.)
Algeria climate: average weather, temperature, precipitation, when to go. Available from: https://www.climatestotravel.com/climate/algeria [accessed 29 August 2022].

Algeria Forest Fires (2021-08-11).

Canadian Wildland Fire Information System | Canadian Forest Fire Weather Index (FWI) System. Available from: https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi

Sidi Bel Abbès Climate, Weather By Month, Average Temperature (Algeria) - Weather Spark. Available from: https://weatherspark.com/y/42385/Average-Weather-in-Sidi-Bel-Abb%C3%A8s-Algeria-Year-Round

UCI Machine Learning Repository: Algerian Forest Fires Dataset Data Set. Available from: https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++

(Algeria Forest Fires (2021-08-11) n.d.) https://data.jrc.ec.europa.eu/dataset/452fc022-5ff1-40a9-b604-62fb276c0cfb (Algeria climate: average weather, temperature, precipitation, when to go n.d.)

CODE

Codelink-
https://github.com/khubim/Stat_Project/blob/main/Stat_Research_Programming_cross_module_Project_2022.ipynb