**Statistics Final** – Project

D00251757

Khubim Kumar Chhetri

**Student Name: Student ID:**

**Predicting Forest Fire and Temperature**

**Project Title:**

1. 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](https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++) 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    To detect future fires based on certain Weather report.   1. Understand different factors affecting the Fire and check for accuracies to find out the cause?      1. Gain Hands-on experience with visualization, Hypothesis testing and the Machine learning models with accuracy score |
| Which Years are going to generate more Temperature coming years?  Predict the change in atmospheric Humidity over the coming years. |
| Check and find out if there is any seasonality in Fire changes? |

1. 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](https://archive.ics.uci.edu/ml/dataset%20) 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.

***Collect Dataset(Using a downloadable csv from worldindata)***

***Clean,Organize and prepare data(Look for Broken,missing and anomalies in dataset)***

***Integrate data for Research***

***Analyse and Elaborate data***

***Visualize data***

***Graphs,Charts and tables***

***Findings ,Solutions and Conclusions***

***Presentation/Results***

**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, 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 meausres 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. The sales team will be able to offer value-adds to the high-risk and high-value customers. 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 customers regarding lifestyle, behaviour in a Yes or No format that can be leveraged post-processing. It is presented in a .csv format with customer 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

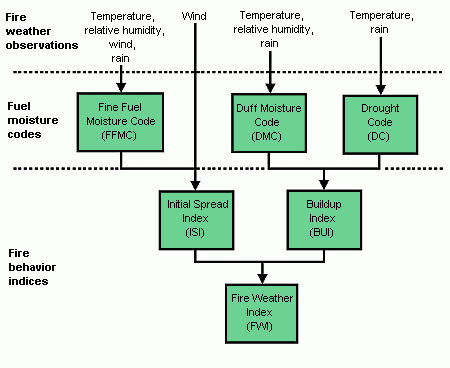


Figure: Natural Resource Canada(Src:cwfis.cfs.nrcan.gc.ca)

1. Special Requirement and deliverable of the project

It is recommended to analyse the data for less than 6 months period .Also we must be aware to analyse and compare weekend data with weekdays.

The deliverables are mobility of visitors with respect to different places during pandemic .As we see the number of cases rising now –then possibility to suggest some solution if required.