FOREST FIRE PREDICTION USING MACHINE LEARNING METHODS

Submitted in partial fulfilment of the requirements for the award of the

DIPLOMA

in

DATA SCIENCE

By

ABHIMANYU B (Reg.No:K2375L200001)

KELTRON
A Government of Kerala Undertaking

DIPLOMA IN DATA SCIENCE & AI

KELTRON KNOWLEDGE CENTRE
ATTINGAL, THIRUVANATHAPURAM

MARCH 2024

DIPLOMA IN DATASCIENCE & AI
KELTRON KNOWLEDGE CENTRE

THIRUVANANTHAPURAM

CERTIFICATE

This is to certify that the dissertation stage report entitled,**'Forest fire prediction using Machine learning methods'** submitted by Abhimanyu B in partial fulfilment of the requirement for the award of the Diploma in Data science, keltron knowlede -ge centre.This dissertation stage report in any form has not been submitted to any other institute for any purpose.

Manoj S
Head Of Departmemt

DECLARATION

I am hereby declare that dissertation entitled as **"Forest fire prediction using Machine learning methods"**is an authentic record of the dissertation carried out by me under the guidance and supervision of Mrs. Arathi p raj teacher of keltron knowledge centre, Attingal in partial fulfilment of the requirments for the diploma in Data science and AI .No parts of this dissertation has been presented eariler for any other degree or diploma

ABHIMANYU B

ACKNOWLEDGEMENT

I would like to thank faculties Arathi P Raj,Department of Data science & AI, Keltron knowledge centre and all our friends,family whose continued support and insiration helped us to accomplish this task.

I thank the Almighty,without whose blessings we would have never completed our work.

ABHIMANYU B

ABSTRACT

Forest fire is a disaster that causes economic and ecological damageand human life threat. Thus predicting such critical environmental issue is essential to mitigate this threat.In this paper we propose a decision tree based system for forest fire prediction.The aim being the integration of the decision tree classifier as a part of the smart sensor node architecture that allows fire prediction in automated and intelligent way without requiring human inter vention. The fire prediction is based on the mete -orological data corresponding to the critical weather elements that influ -ence the forest fire occurrence, namely temperature, relative humidity and wind speed.

CONTENTS

1.INTRODUCTION

The most common hazard in forests is forests fire. Forests fires are as old as the forests themselves. They pose a threat not only to the forest wealth but also to the entire regime to fauna and flora seriously disturbing the bio-diversity and the ecology and environment of a region. During summer, when there is no rain for months, the forests become littered with dry senescent leaves and twinges, which could burst into flames ignited by the slightest spark. The Himalayan forests, particularly, Garhwal Himalayas have been burning regularly during the last few summers, with colossal loss of vegetation cover of that region.

1.1Causes of Forest Fire

Forest fires are caused by Natural causes as well as Man made causes

Natural causes - Many forest fires start from natural causes such as lightning which set trees on fire.However, rain extinguishes such fires without causing much damage. High atmospheric temperatures and dryness (low humidity) offer favorable circumstance for a fire to start.

Man made causes - Fire is caused when a source of fire like naked flame, cigarette or bidi, electric spark or any source of ignition comes into contact with inflammable material.

1.2 Classification of Forest Fire

Forest fire can broadly be classified into three categories;
Natural or controlled forest fire.
Forest fires caused by heat generated in the litter and other biomes in summer through carelessness of people (human neglect) and
Forest fires purposely caused by local inhabitants.

1.3 Types of Forest Fire

Surface Fire - A forest fire may burn primarily as a surface fire, spreading along the ground as the surface litter (senescent leaves and twigs and dry grasses etc) on the forest floor and is engulfed by the spreading flames.

Underground Fire - The fires of low intensity, consuming the organic matter beneath and the surface litter of forest floor are sub-grouped as underground fire.In most of the dense forests a thick mantle of organic matter is find on top of the mineral soil.This fire spreads in by consuming such materials.These fires usually spread entirely underground and burn for some meters below the surface.This fire spreads very slowly and in most of the cases it becomes very hard to detect and control such type of fires. They may continue to burn for months and destroy vegetative cover of the soil. The other terminology for this type of fire is Muck fires.

Ground Fire - These fires are fires in the sub surface organic fuels, such as duff layers under forest stands, Arctic tundra or taiga, and organic soils of swamps or bogs.There is no clear distinction between underground and ground fires. The smoldering under ground fires sometime changes into Ground fire.This fire burns root and other material on or beneath the surface i.e. burns the herbaceous growth on forest floor together with the layer of organic matter in various stages of decay. They are more damaging than surface fires , as they can destroy vegetation completely. Ground fires burn underneath the surface by smoldering combustion and are more often ignited by surface fires.

Crown Fire - A crown fire is one in which the crown of trees and shrubs burn, often sustained by a surface fire. A crown fire is particularly very dangerous in a coniferous forest because resinous material given off burning logs burn furiously. On hill slopes, if the fire starts downhill, it spreads up fast as heated air adjacent to a slope tends to flow up the slope spreading flames along with it. If the fire starts uphill, there is less likelihood of it spreading downwards.

Firestorms - Among the forest fires, the fire spreading most rapidly is the firestorm, which is an intense fire over a large area. As the fire burns, heat rises and air rushes in, causing the fire to grow. More air makes the fire spin violently like a storm. Flames fly out from the base and burning ember spew out the top of the fiery twister, starting smaller fires around it. Temperatures inside these storms can reach around 2,000 degrees Fahrenheit.

1.4 Vulnerability

The youngest mountain ranges of Himalayas are the most vulnerable stretches of the world susceptible to forest fires. The forests of Western are more frequently vulnerable to forest fires as compared to those in Eastern Himalayas. This is because forests of Eastern Himalayas grow in high rain density. With large scale expansion of chirr (Pine) forests in many areas of the Himalayas the frequency and intensity of forest fires has increased.

1.5 Preparedness and Mitigation Measures

Forest fires are usually seasonal. They usually start in the dry season and can be prevented by adequate precautions. Successive Five Year Plans have provided funds for forests fighting. During the British period, fire was prevented in the summer through removal of forest litter all along the forest boundary. This was called "Forest Fire Line" This line used to prevent fire breaking into the forest from one compartment to another. The collected litter was burnt in isolation. Generally, the fire spreads only if there is continuous supply of fuel (Dry vegetation) along its path. The best way to control a forest fire is therefore, to prevent it from spreading, which can be done by creating firebreaks in the shape of small clearings of ditches in the forests.

1.6 Precautions

- To keep the source of fire or source of ignition separated from combustible and inflammable material.
- To keep the source of fire under watch and control.
- Not allow combustible or inflammable material to pile up unnecessarily and to stock the same as per procedure recommended for safe storage of such combustible or inflammable material.
- To adopt safe practices in areas near forests viz. factories, coalmines, oil stores, chemical plants and even in household kitchens.
  To incorporate fire reducing and fire fighting techniques and equipment

Source : Department of Home, Himachal Pradesh

2.Literature Review

[1]George E. Sakr, Imad H. Elhajj, George Mitri and Uchechukwu C. Wejinya “Artificial Intelligence for Forest Fire Prediction” 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics Montréal, Canada, July 6 9, 2010: This paper presented a forest fire risk prediction method. The findings show that a small quantity of data can be used to estimate forest fire risk.

[2]Mauro Castelli, Leonardo Vanneschi, and Ales Popovic “Predicting burned areas of forest fires: an artificial intelligence approach” Fire Ecology 2015: They demonstrated a novel intelligent GP-based approach for examining burned areas in this demonstration. The major goal was to create a system that could forecast how much land will be destroyed in the event of a forest fire. The experimental findings revealed that geometric semantic genetic programming outperforms due to the small MAE.

[3]A. Kansal, Y. Singh, N. Kumar and V. Mohindru, "Detection of forest fires using machine learning technique: A perspective" 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015: The use of regression and the division of datasets has been proposed in this paper as a method for detecting fire. The algorithm achieves a low R-squared and a low root mean square error. This method could be used for other calamities in the future. The use of specific transformations may also help to increase the model's efficiency.

[4]L. Yu, N. Wang, and X. Meng "Real-time forest fire detection with Wireless Sensor Networks" in Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on, vol. 2. IEEE, 2005: Ensemble learning is used at all cluster heads in this case. At the base station, SVM, a supervised machine learning technique, is used with a polynomial kernel function. Carbon dioxide, temperature, humidity, and carbon monoxide can all be detected using the sensors that have been installed. Clustered stream generates data in tabular or clustered form. After that, the SVM is used to detect fire.

[5]Guruh Fajar Shidik and Khabib Mustofa “Predicting Size of Forest Fire Using Hybrid Model “ICT-Eurasia 2014: An alternative hybrid model capable of predicting the extent of forest fire has been developed in this study. The algorithm, which includes meteorological and forest weather index variables, has successfully classified the level of burning into three categories: No Burn Area, Light Burn, and Heavy Burn. The proposed model's examination revealed encouraging results in terms of accuracy. of confusion matrix around 97.50% and Kappa 0.961.

[6]Paulo Cortez and Anibal Morais "A Data Mining Approach to Predict Forest Fires using Meteorological Data": They investigate a Data Mining approach for predicting the burned area of forest fires in this paper. The optimal configuration combines an SVM with four meteorological inputs to forecast the burned area of minor fires. Such information is especially valuable for bettering the administration of firefighting resources.

3. PROBLEM STATEMENT AND PROPOSED METHOD

Forest fires are a typical occurrence in the natural world. Forest fire is the most common threat in forests. Forest fires are a major environmental hazard that threatens forest preservation, causing economic and ecological harm as well as human suffering. They endanger not only the forest's wealth, but also the entire ecosystem's animals and vegetation, causing major disruption of a region's biodiversity, ecology, and environment.

The proposed system is to predict Forest Fire using parameters like temperature,fire weather index,drought code,intial spread index etc. And use different machine learning methods like logistic regression, random forest,decision tree,XGboost and also to bulid a jar site to predict forest fire. It will achieve the fire prediction with the best accurate values.

- Data storage: Data storage refers to the use of recording media to retain data using computers or other devices. The most prevalent forms of data storage are file storage, block storage and object storage, with each being ideal for different purposes.
- Data preprocessing: Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.
- Data analytics: Data analytics is the science of analyzing raw data to make conclusions about that information.
- Model training: Model training in machine language is the process of feeding an ML algorithm with data to help identify and learn good values for all attributes involved. There are several types of machine learning models, of which the most common ones are supervised and unsupervised learning. The models used here are random forest, linear regression, decision tree and lasso regression.
- Model evaluation: Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Accuracy is the model evaluation method used.
- Forest fire prediction: It plays an essential role in decision making at global, regional levels.Using the different machine learning models.

4. MATERIALS AND METHODS

4.1 Algorithms

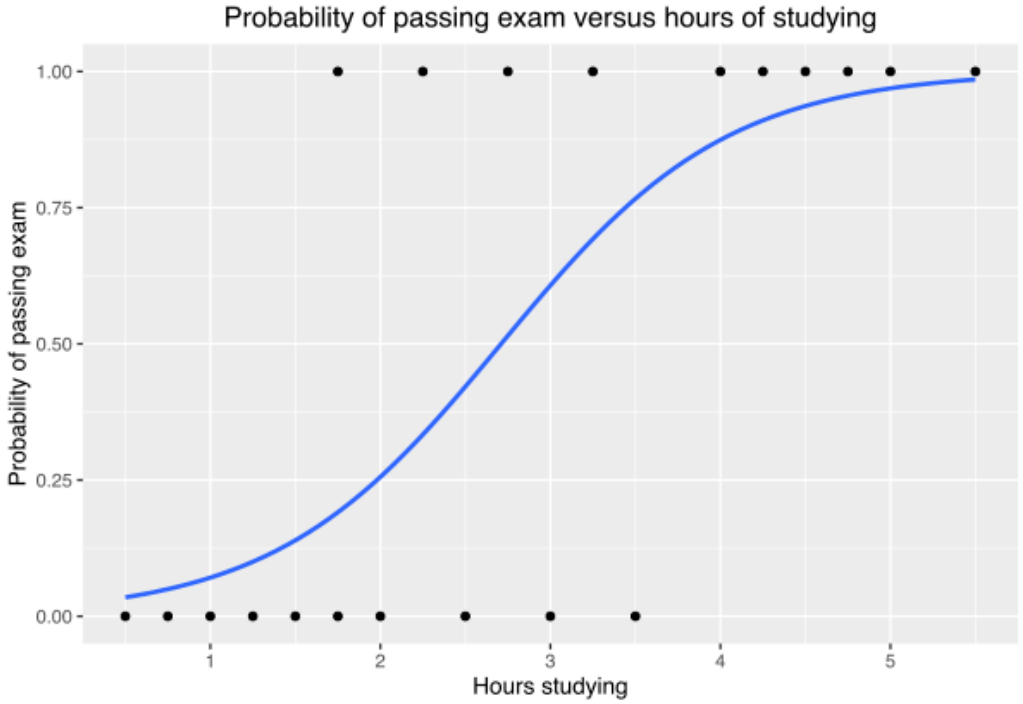
Machine learning (ML) is basically the field of computer science that uses a computer system that can understand data in a way that a human can. Simply put, ML is a type of artificial intelligence that extracts patterns from raw data using an algorithm or method. The main goal of ML is to allow computer systems to learn from experience without being explicitly programmed or without human intervention.

The strategy focuses on predicting forest fire using ML methods. The machine learning methods are :

4.1.1 Logistic regression

Logistic regression is a supervised machine learning algorithm mainly used for binary classification where we use a logistic function, also known as a sigmoid function that takes input as independent variables and produces a probability value between 0 and 1.

For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 it belongs to Class 0. It's referred to as regression because it is the extension of linear regression but is mainly used for classification problems. The difference between linear regression and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not



4.1.2 Random forest

Random Forest is a well-known supervised machine learning method. It is based on ensemble learning, which is a method of integrating numerous classifiers to solve a complicated issue and increase the model's performance. Random Forest is a classifier that aggregates the results of a number of decision trees on distinct subsets of a dataset to boost the dataset's projected accuracy.

Rather than relying on a single decision tree, the random forest aggregates forecasts from each tree and predicts the final output based on the majority of projections.Random forest is a classification technique that uses numerous decision trees to classify data. When creating each individual tree, it employs bagging and feature randomization in order to generate an uncorrelated forest of trees whose aggregate prediction is more accurate than that of any one tree.

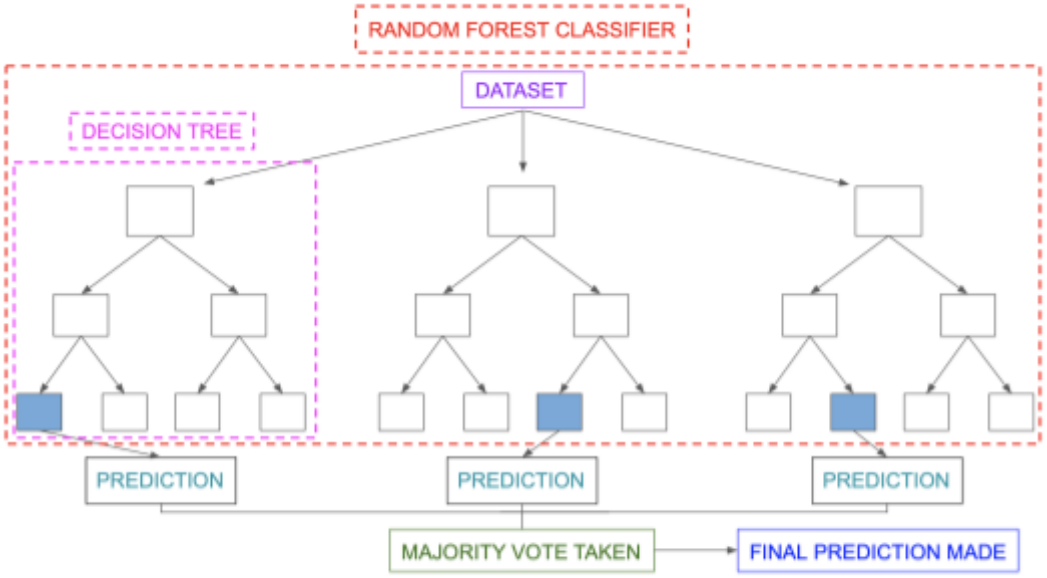


Figure 4.2: Random forest

4.1.3 Decision tree

Decision Tree is a supervised learning strategy that can handle classification and regression problems, however it is most typically used to address classification problems. A Decision tree has two nodes: the Decision Node and the Leaf Node. Decision nodes are used to make any decision and have numerous branches, whereas Leaf nodes are the outcome of such decisions and do not have any more branches. In this tree-structured classifier, internal nodes hold dataset attributes, branches indicate decision rules, and each leaf node offers the conclusion.Instances are classified using decision trees by sorting them along the tree from the root to a leaf node that provides the classification. Starting with the root node of the tree and checking the attribute provided by this node, an instance is categorised by travelling down the tree branch according to the value of the attribute. After then, the procedure is repeated for the new node-rooted subtree.

4.2 Tools

- Python: An interpreted, object-oriented programming language with dynamic semantics. Its high-level, built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components together. Python and EDA can be used together to identify missing values in a data set, which is important so you can decide how to handle missing values for machine learning.
- Pandas**:** Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The 34 name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010.
- NumPy: NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high level mathematical functions to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Number array into Numeric, with extensive modifications. NumPy is open-source software and has many contributors.
- Scikit-learn: Scikit-learn (formerly scikits.learn and also known as sklearn)

is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, *k*-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. Scikit-learn is largely written in Python and uses NumPy extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written in Cython to improve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM; logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR. In such cases, extending these methods with Python may not be possible. Scikit-learn integrates well with many other Python libraries, such as Matplotlib and plotly for plotting, NumPy for array vectorization, Pandas data frames, SciPy, and many more.

- Seaborn: Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures. Seaborn helps you explore and understand your data. Its plotting functions operate on data frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.
- Matplotlib: Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open-source alternative to MATLAB.

4.3 About the dataset

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.

Dataset columns:

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

Duff Moisture Code (DMC) index from the FWI system: 1.1 to 65.9

Drought Code (DC) index from the FWI system: 7 to 220.4

Initial Spread Index (ISI) index from the FWI system: 0 to 18.5

Buildup Index (BUI) index from the FWI system: 1.1 to 68

Fire Weather Index (FWI) Index: 0 to 31.1

Classes: two classes, namely Fire and not Fire

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29	57	18	0	65.7	3.4	7.6	1.3	3.4	0.5	not fire
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1	3.9	0.4	not fire
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0	1.7	0	not fire
4	5	6	2012	27	77	16	0	64.8	3	14.2	1.2	3.9	0.5	not fire

5. IMPLEMENTATION AND DEPLOYMENT

5.1 Import libraries and loading the dataset

Import the required libraries such as numpy, pandas, seaborn, matplotlib,...

Then using pandas load the dataset into a dataframe.

```
## IMPORT LIBRARIES

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.read_csv("C:/Users/bijua/Downloads/Datasets/Algerian_forest_fires_dataset.csv")
df.head(5)
```

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	ISI	BUI	FWI	Classes
0	1	6	2012	29	57	18	0	65.7	3.4	7.6	1.3	3.4	0.5	not fire
1	2	6	2012	29	61	13	1.3	64.4	4.1	7.6	1	3.9	0.4	not fire
2	3	6	2012	26	82	22	13.1	47.1	2.5	7.1	0.3	2.7	0.1	not fire
3	4	6	2012	25	89	13	2.5	28.6	1.3	6.9	0	1.7	0	not fire
4	5	6	2012	27	77	16	0	64.8	3	14.2	1.2	3.9	0.5	not fire

5.2 Exploratory data analysis (EDA)

Exploratory Data Analysis is the critical process of using summary statistics and graphical representations to conduct early investigations on data in order to uncover patterns, spot anomalies, test hypotheses, and confirm assumptions. The primary goal of EDA is to assist in the analysis of data right before making any assumptions.It can aid in the detection of noticeable errors, as well as a better understanding of data patterns, the detection of outliers or unusual events, and the discovery of interesting relationships between variables.In this stage some basic visualization is also carried out.

Describe: The describe() function in pandas is very handy in getting various summary statistics.This function returns the count, mean, standard deviation, minimum and maximum values and the quantiles of the data.

Info: It is also a good practice to know the columns and their corresponding data types,along with finding whether they contain null values or not.For that info() function is used.

Shape: To get the shape of the data using the shape.

Data visualization: Data Visualization is the process of analyzing data in the form of graphs or maps, making it a lot easier to understand the trends or patterns in the data.

Handling Missing Values: Missing Data is a very big problem in real-life scenarios. Missing Data can also refer to as NA(Not Available) values in pandas. There are several useful functions for detecting, removing, and replacing null values in Pandas DataFrame. They are dropna and fillna.

```
df.isnull().sum()
```

```
day            1
month          2
year           2
Temperature    2
 RH            2
 Ws            2
Rain           2
FFMC           2
DMC            2
DC             2
ISI            2
BUI            2
FWI            2
Classes        3
dtype: int64
```

5.3 Data preprocessing

Data preparation is the process of transforming raw data into an understandable format.Because we can’t work with raw data, this is a crucial phase in the data

mining process.Before using machine learning or data mining algorithms, the data quality should be checked.To achieve reliable findings, the techniques are typically used at the earliest phases of the machine learning and AI development pipeline.

Rescale Data: When our data is composed of attributes with varying scales, many machine learning algorithms can benefit from rescaling the attributes to all have the same scale. This is useful for optimization algorithms used in the core of machine learning algorithms like gradient descent.

Train Test Split: Which is one of the important steps in Machine Learning. It is very important because your model needs to be evaluated before it has been deployed. And that evaluation needs to be done on unseen data because when it is deployed, all incoming data is unseen.

```
In [17]: from sklearn.model_selection import train_test_split
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 0)
```

Figure 5.3: Train test split

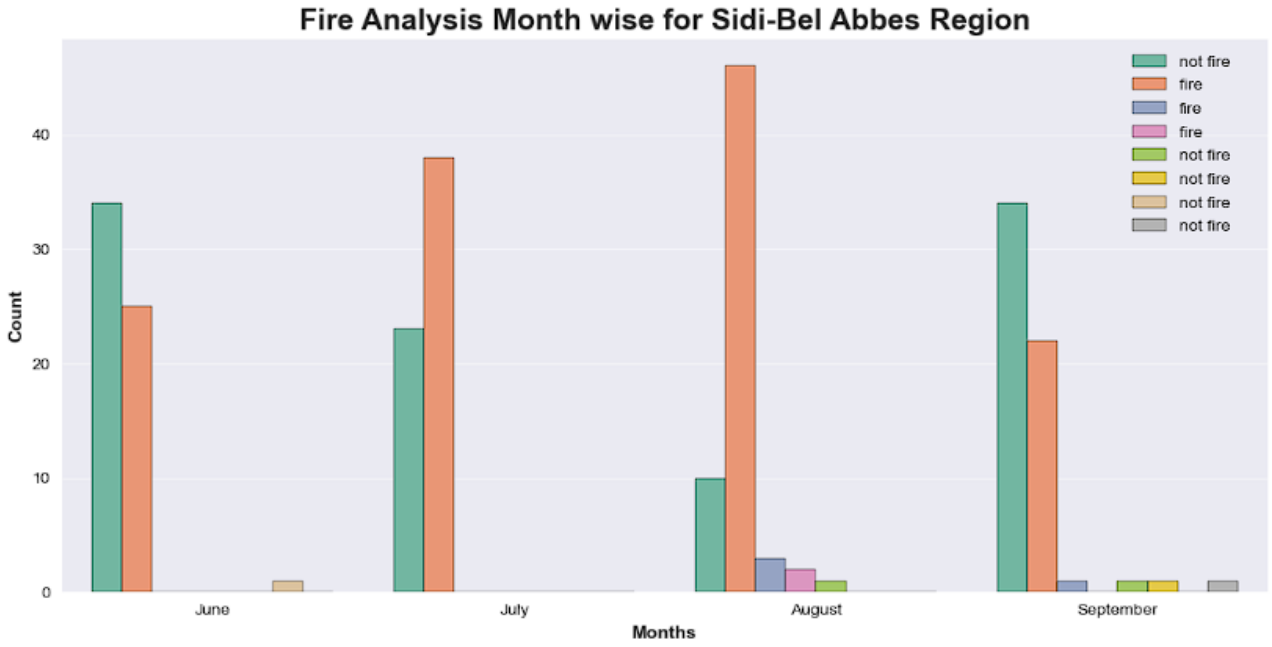
Here the X and y as arguments in train_test_split, which splits X and y such that there is 30% testing data and 70% training data successfully split between X_train, X_test, y_train, and y_test.

5.4 Data visualization

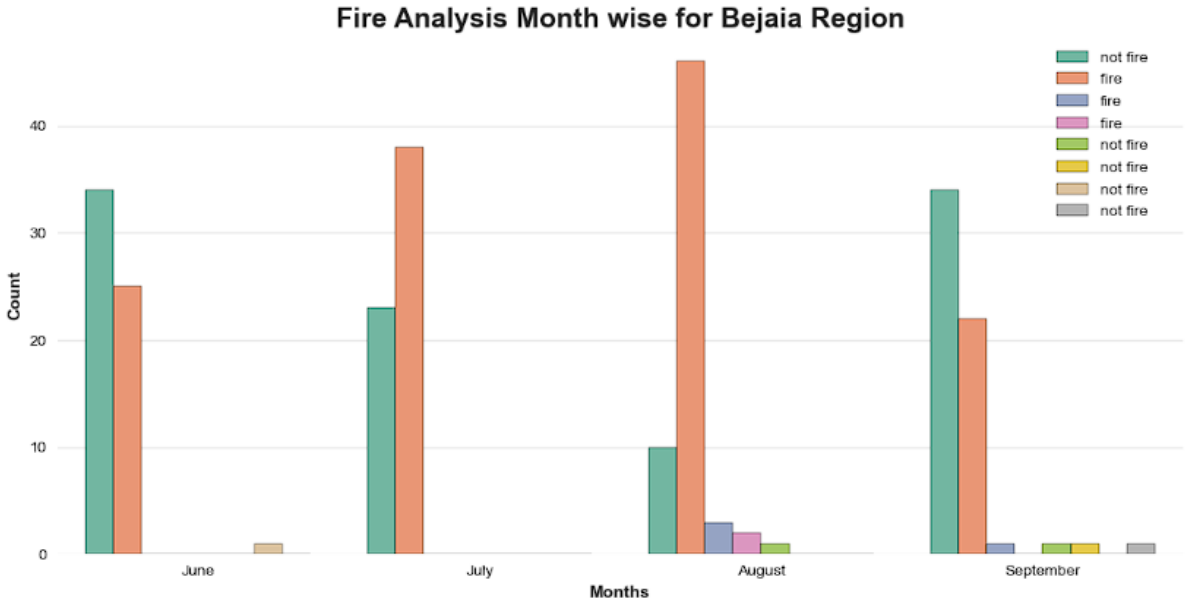
Data Visualization is the process of analyzing data in the form of graphs or maps, making it a lot easier to understand the trends or patterns in the data.

MONTHLY ANALYISIS

```
# Fire Analysis Month wise for Sidi-Bel Abbes Region
dftemp= df.loc[df['Region']== 2]
plt.subplots(figsize=(13,6))
sns.set_style('whitegrid')
sns.countplot(x='month',hue='Classes',data= df,ec = 'black', palette= 'Set2')
plt.title('Fire Analysis Month wise for Sidi-Bel Abbes Region', fontsize=18, weight='bold')
plt.ylabel('Count', weight = 'bold')
plt.xlabel('Months', weight= 'bold')
plt.legend(loc='upper right')
plt.xticks(np.arange(4), ['June','July', 'August', 'September',])
plt.grid(alpha = 0.5,axis = 'y')
plt.show()
```




```
# Fire Analysis Month wise for Bejaia Region
dftemp= df.loc[df['Region']== 1]
plt.subplots(figsize=(13,6))
sns.set_style('whitegrid')
sns.countplot(x='month',hue='Classes',data= df,ec = 'black', palette= 'Set2')
plt.title('Fire Analysis Month wise for Bejaia Region', fontsize=18, weight='bold')
plt.ylabel('Count', weight = 'bold')
plt.xlabel('Months', weight= 'bold')
plt.legend(loc='upper right')
plt.xticks(np.arange(4), ['June','July', 'August', 'September',])
plt.grid(alpha = 0.5,axis = 'y')
plt.show()
```



5.5 Model

To determine which machine learning method is most appropriate for the given model,various algorithums Random Forest, Logistic Regression,Decision Tree, XGboost are applied.

- Random Forest

```
rfc = RandomForestClassifier()
rfc.fit(X_train,y_train)
rfc_pred = rfc.predict(X_test)
score = accuracy_score(y_test,rfc_pred)
cr = classification_report(y_test,rfc_pred)

print("Random Forest")
print ("Accuracy Score value:",score)
print (cr)
```

```
Random Forest
Accuracy Score value: 0.9836065573770492
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        26
           1       1.00      0.97      0.99        35

    accuracy                           0.98        61
   macro avg       0.98      0.99      0.98        61
weighted avg       0.98      0.98      0.98        61
```

● Logistic regression

```
# Create logistic regression model
model = LogisticRegression()
model.fit(X_train, y_train)
```

```
LogisticRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
# Make predictions on the test set
y_pred = model.predict(X_test)
y_pred
```

```
array([1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1,
       0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1])
```

```
# check accuracy and cofusion matrix
score = accuracy_score(y_test, y_pred)
cr = classification_report(y_test,y_pred)

print("Logistic Regression")
print ("Accuracy Score value:",score)
print (cr)
```

```
Logistic Regression
Accuracy Score value: 0.9508196721311475
              precision    recall  f1-score   support

           0       0.90      1.00      0.95        26
           1       1.00      0.91      0.96        35

    accuracy                           0.95        61
   macro avg       0.95      0.96      0.95        61
weighted avg       0.96      0.95      0.95        61
```

● Decision tree

```
dtc = DecisionTreeClassifier()
dtc.fit(X_train,y_train)
dtc_pred = dtc.predict(X_test)
score = accuracy_score(y_test,dtc_pred)
cr = classification_report(y_test,dtc_pred)

print("Accuracy:",score)
print(cr)
```

```
Accuracy: 0.9836065573770492
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        26
           1       1.00      0.97      0.99        35

    accuracy                           0.98        61
   macro avg       0.98      0.99      0.98        61
weighted avg       0.98      0.98      0.98        61
```

- XGboost

```
# Import KNeighborsClassifier to Train from SKlearn
xgb = XGBClassifier()
xgb.fit(X_train,y_train)
xgb_pred = xgb.predict(X_test)
score = accuracy_score(y_test, xgb_pred)
cr = classification_report(y_test, xgb_pred)

print("XGboost Classifier")
print ("Accuracy Score value:",score)
print (cr)
```

```
XGboost Classifier
Accuracy Score value: 0.9672131147540983
              precision    recall  f1-score   support

           0       0.93      1.00      0.96        26
           1       1.00      0.94      0.97        35

    accuracy                           0.97        61
   macro avg       0.96      0.97      0.97        61
weighted avg       0.97      0.97      0.97        61
```

The accuracy of 4 regression models are given below:

Random forest: 98.36

Logistic regression: 95.08

Decision tree: 93.4

XGboost: 96.72

5.6 Deployment

To dump the model I used the Pickle python module.

```
import pandas as pd
import numpy as np
import streamlit as st
import pickle
import sklearn

model = pickle.load(open('model.sav', 'rb'))
```

After the dumping of the model , I created a function named predict. That function takes the inputs from the user and predicts.

Then the output will print on the web page.

```python
import pandas as pd
import numpy as np
import streamlit as st
import pickle
import sklearn

model = pickle.load(open('model.sav', 'rb'))

st.title('Forest fire Prediction')
st.sidebar.header('Forest Fire Data')

# FUNCTION
def user_report():

  FFMC = st.sidebar.slider('FFMC', 0.0,100.0 )
  ISI = st.sidebar.slider('ISI', 0.0,100.0 )
  FWI = st.sidebar.slider('FWI', 0.0,100.0)



  user_report_data = {

      'FFMC': FFMC,
      'ISI':ISI,
      'FWI':FWI,
  }
  report_data = pd.DataFrame(user_report_data, index=[0])
  return report_data

user_data = user_report()
st.header('Forest Fire Data')
st.write(user_data)

clas = model.predict(user_data)
if clas != 1:
  result="Not Fire"
else:
  result="Fire"
st.subheader('Fire Prediction')
st.subheader(result)
```

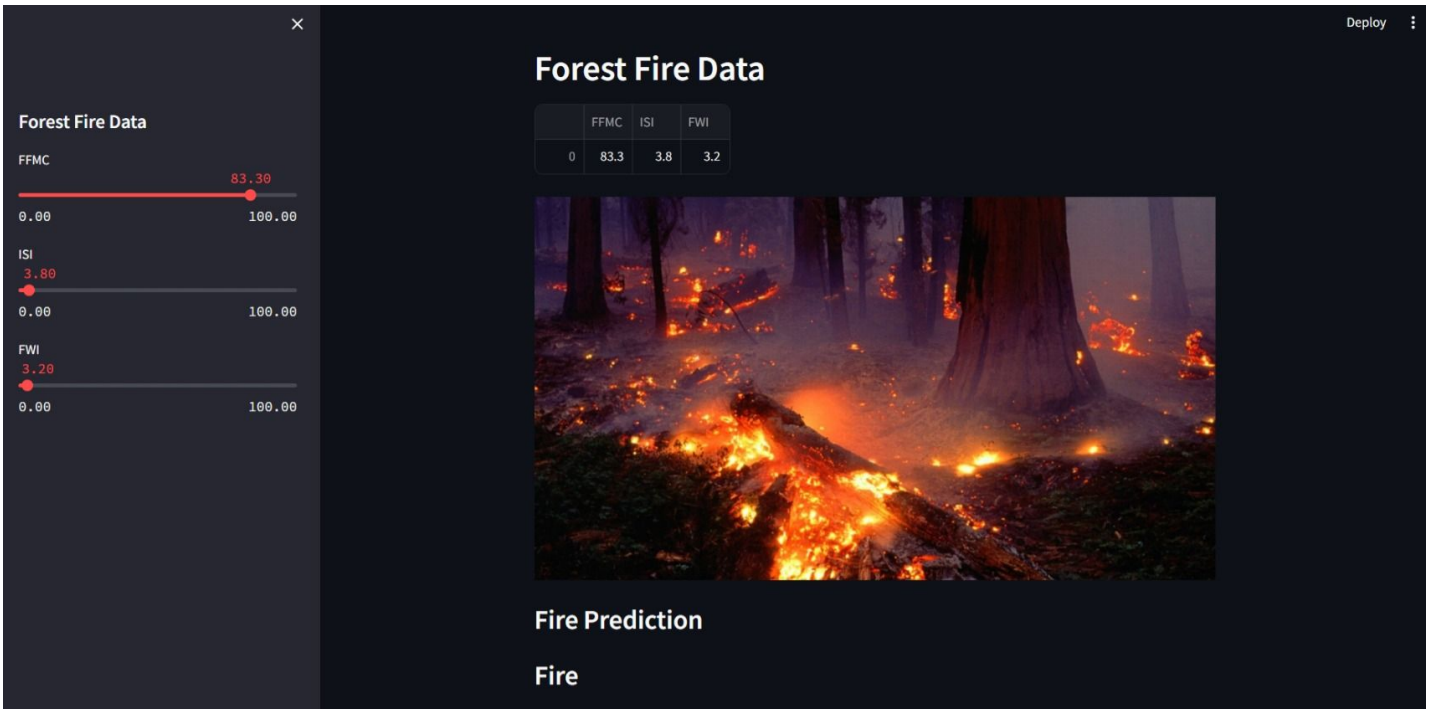
6. RESULTS AND DISCUSSIONS

6.1 Forest Fire prediction

Machine learning involves showing a large volume of data to a machine so that it can learn and make predictions, find patterns, or classify data. The three machine learning types are supervised, unsupervised, and reinforcement learning. Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc. Which is one type of supervised machine learning model. I have used 4 machine learning models for fire prediction. They are Random forest, Logistic regression, Decision tree and XGboost.

Among the 4 models Random forest and XGboost have highest accuracy 98.36 and 96.72 respectively. Logistic regression and Decision have lowest accuracy.

6.2 Web page layout



7.CONCLUSION

Forest fires are a typical occurrence within the flora and fauna. Every year, legion hectares of forest are destroyed round the world. This resulted in significant environmental harm in addition because the loss of irreplaceable human lives. Forest fires are a serious environmental hazard that threatens forest preservation, causing economic and ecological harm likewise as human suffering. Quick fire detection and reaction are efficient methods for decreasing fire damage.Various studies are conducted in try to improve early fire prediction and detection systems, which aid within the development of fireside response methods. It signifies that early caution detection isprovided. The accurate prediction of results supported defined parameters is expounded to early caution detection.

the We completed our research and constructed a system that forecasts proportions of fires that may occur are supported by the weather data provided by the user, like temperature, oxygen, and humidity.

REFERENCES

1. Benkheira, A.: Ministère de l'agriculture et de développement rurale et de la pèche, direction des forets Les feux de forêts en Algérie analyse et perspectives. Technical report (2018)
2. San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., de Rigo, D., Ferrari, D., Maianti, P., Vivancos, T.A., Costa, H., Lana, F.: Advance EFFIS report on Forest Fires in Europe, Middle East and North Africa (2017)
3. Maksimović, M., Vujović, V.: Comparative analysis of data mining techniques applied to wireless sensor network data for fire detection. J. Inf. Technol. Appl. (JITA) 3(2), 65–773 (2013)
4. Giuntini, F.T., Beder, D.M., Ueyama, J.: Exploiting self-organization and fault tolerance in wireless sensor networks: a case study on wildfire detection application. Int. J. Distrib. Sens. Netw. 13(4), 1–16 (2017)
5. Bahrepour, M., Van der Zwaag, B.J., Meratnia, N., Havinga, P.: Fire data analysis and feature reduction using computational intelligence methods. In: Phillips-Wren, G., Jain, L. C., Nakamatsu, K., Howlett, R.J. (eds.) Advances in Intelligent Decision Technologies. Smart Innovation, Systems and Technologies, vol. 4, pp. 289–298. Springer, Heidelberg (2010)
6. Saoudi, M., Bounceur, A., Euler, R., Kechadi, T.: Data mining techniques applied to wireless sensor networks for early forest fire detection. In: International Conference on Internet of Things and Cloud Computing (ICC), Cambridge, United Kingdom (2016)
7. Hefeeda, M., Bagheri, M.: Wireless sensor networks for early detection of forest. In: The IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Pisa, Italy, pp. 1–6 (2007)
8. Liu, Y., Gu, Y., Chen, G., Ji, Y., Li, J.: A novel accurate forest fire detection system using wireless sensor networks. In: IEEE Seventh International Conference on Mobile Ad-Hoc and Sensor Networks (MSN), New York, pp. 52–59 (2011)
9. Han, J., Kamber, M., Pei, J.: Data Mining Concepts and Techniques, 3rd edn. Elsevier, Amsterdam (2012)
10. Stojanova, D., Panov, P., Kobler, A., Džeroski, S., Taškova, K.: Learning to predict forest f
ires with different data mining techniques. In: The 9th International Multi Conference Information Society, Jubljana, Slovinia (2006)
11. Karouni, A., Daya, B., Chauvet, P.: Applying decision tree algorithm and neural networks to predict forest fires in Lebanon. J. Theor. Appl. Inf. Technol. 63(2), 282–291 (2014)
12. Cortez, P., Morais, A.: Data mining approach to predict forest fires using meteorological data. In: Neves, J., Santos, M.F., Machado, J. (eds.) New Trends in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence (EPIA), Guimaraes, Portugal, pp. 512 523 (2007)