

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade



Project Report

on

Algerian Forest Fires Dataset EDA Project

Submitted By:

Aditya Patidar (0901AI211004)

Aryan Bhai Patel (0901AI211012)

Faculty Mentor:

Dr.Kritika Bansal

CENTRE FOR ARTIFICIAL INTELLIGENCE

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE

GWALIOR - 474005 (MP) est. 1957

JULY-DEC. 2023

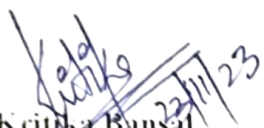
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

CERTIFICATE

This is certified that **Aditya Patidar (0901AI211004), Aryan Bhai Patel (0901AI211012)** has submitted the project report titled **Algerian Forest Fires Dataset EDA Project** under the mentorship of **Dr. Kritika Bansal**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in **Artificial Intelligence and Robotics** from Madhav Institute of Technology and Science, Gwalior.

 22/11/23

Dr. Kritika Bansal
Faculty Mentor
Assistant Professor
Centre for Artificial Intelligence



Dr. R. R. Singh
Coordinator
Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR


(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

DECLARATION

I hereby declare that the work being presented in this project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in **Artificial Intelligence and Robotics** at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr.Kritika Bansal .Assistant Professor .Centre for Artificial Intelligence.**

I declare that I have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.



Aditya Patidar (0901AI211004)



Arjun Bhai Patel (0901AI211012)

III Year,

Centre for Artificial Intelligence

MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE GWALIOR

(A Govt. Aided UGC Autonomous Institute Affiliated to RGPV, Bhopal)

NAAC Accredited with A++ Grade

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology and Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme (based on the AICTE Model Curriculum 2018), approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Artificial Intelligence**, for allowing me to explore this project. I humbly thank **Dr. R. R. Singh**, Coordinator, Centre for Artificial Intelligence, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Kritika Bansal**, Assistant Professor, Centre for Artificial Intelligence, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.



Aditya Patidar (0901AI211004)



Aryan Bhai Patel (0901AI211012)

III Year
Centre for Artificial Intelligence

ABSTRACT

The Algerian Forest Fires EDA project aims to conduct a comprehensive exploration of a dataset related to forest fires in Algeria. This initiative is driven by the necessity to understand the patterns, trends, and contributing factors to forest fires, ultimately contributing to informed decision-making and effective preventive measures. The dataset under examination is expected to contain temporal, meteorological, and geographical information, offering a multifaceted perspective on the occurrence and characteristics of forest fires in the region.

The project follows a structured approach to Exploratory Data Analysis (EDA), encompassing key steps such as loading and overviewing the dataset, generating descriptive statistics, and conducting data cleaning to ensure the quality and reliability of the information. Through visualizations and statistical analyses, the EDA process seeks to uncover insights into the distribution of relevant features, correlations among variables, and the temporal evolution of critical factors like temperature. Outlier detection mechanisms will be employed to identify anomalous values that may impact the overall analysis. The inclusion of a target variable, likely indicating whether a forest fire occurred, necessitates a focused examination of its distribution and relationships with other features. Through this analysis, we aim to provide stakeholders, researchers, and policymakers with a deeper understanding of the dynamics surrounding forest fires in Algeria.

The iterative nature of the EDA process allows for the refinement and adjustment of analyses based on emerging patterns and insights. The project documentation serves as a guide for transparently documenting methodologies, code implementations, and visualizations. The results of this EDA will serve as a foundational step for subsequent phases of analysis, modeling, and decision support, ultimately contributing to effective strategies for forest fire prevention and management in Algeria.

सार

अल्जीरियाई जंगल की आग ईडीए परियोजना का उद्देश्य अल्जीरिया में जंगल की आग से संबंधित डेटासेट की व्यापक खोज करना है। यह पहल जंगल की आग के पैटर्न, रुझान और योगदान करने वाले कारकों को समझने की आवश्यकता से प्रेरित है, जो अंततः सूचित निर्णय लेने और प्रभावी निवारक उपायों में योगदान करती है। परीक्षण के तहत डेटासेट में अस्थायी होने की उम्मीद है। मौसम संबंधी और भौगोलिक जानकारी, क्षेत्र में जंगल की आग की घटनाओं और विशेषताओं पर एक बहुआयामी परिप्रेक्ष्य पेश करती है।

यह परियोजना खोजपूर्ण डेटा विश्लेषण (ईडीए) के लिए एक संरचित दृष्टिकोण का पालन करती है, जिसमें डेटासेट को लोड करना और अवलोकन करना, वर्णनात्मक आंकड़े तैयार करना और सूचना की गुणवत्ता और विश्वसनीयता सुनिश्चित करने के लिए डेटा सफाई का संचालन करना जैसे प्रमुख चरण शामिल हैं। विजुअलाइज़ेशन और सांख्यिकीय विश्लेषण के माध्यम से, ईडीए प्रक्रिया प्रासंगिक सुविधाओं के वितरण, चर के बीच सहसंबंध और तापमान जैसे महत्वपूर्ण कारकों के अस्थायी विकास में अंतर्दृष्टि को उजागर करना चाहती है। समग्र विश्लेषण को प्रभावित करने वाले असामान्य मूल्यों की पहचान करने के लिए बाहरी पहचान तंत्र को नियोजित किया जाएगा। एक लक्ष्य चर को शामिल करने से, संभवतः यह संकेत मिलता है कि क्या जंगल में आग लगी थी, इसके वितरण और अन्य विशेषताओं के साथ संबंधों की एक केंद्रित परीक्षा की आवश्यकता होती है। इस विश्लेषण के माध्यम से, हमारा लक्ष्य हितधारकों, शोधकर्ताओं और नीति निर्माताओं को अल्जीरिया में जंगल की आग के आसपास की गतिशीलता की गहरी समझ प्रदान करना है।

ईडीए प्रक्रिया की पुनरावृत्तीय प्रकृति उभरते पैटर्न और अंतर्दृष्टि के आधार पर विश्लेषण के शोधन और समायोजन की अनुमति देती है। प्रोजेक्ट दस्तावेज़ीकरण पारदर्शी रूप से दस्तावेज़ीकरण पद्धतियों, कोड कार्यान्वयन और विजुअलाइज़ेशन के लिए एक मार्गदर्शक के रूप में कार्य करता है। इस ईडीए के परिणाम विश्लेषण, मॉडलिंग और निर्णय समर्थन के बाद के चरणों के लिए एक मूलभूत कदम के रूप में काम करेंगे, जो अंततः अल्जीरिया में जंगल की आग की रोकथाम और प्रबंधन के लिए प्रभावी रणनीतियों में योगदान देंगे।

TABLE OF CONTENTS

TITLE	PAGE NO.
Chapter 1: Introduction	
1.1 Project Overview.....	VI
1.2 Dataset Information.....	VI
1.3 Objectives.....	VI
Chapter 2: Data Collection and Understanding	
2.1 Data Source.....	VIII
2.2 Tools Used.....	VIII
Chapter 3: Data Exploration	
3.1 Import Libraries.....	IX
3.2 Data Cleaning.....	X
3.3 Exploratory Data Analysis (EDA).....	XIII
3.4 Exploratory Data Analysis (EDA) on all Features.....	XVI
Chapter 4: Conclusion	
References.....	XIX

Chapter 1: INTRODUCTION

Forest fires in Algeria have significant ecological and socio-economic implications. This report outlines the data collection and understanding phase of the Algerian Forest Fires Dataset Project, focusing on the processes involved in acquiring, exploring, and comprehending the dataset.

1.1 Project Overview

This documentation outlines the Exploratory Data Analysis (EDA) phase of the Algerian Forest Fires Dataset Project. The project aims to analyze and gain insights into forest fire occurrences in Algeria through the exploration of a provided dataset. The Exploratory Data Analysis (EDA) project on Algerian forest fires seeks to unravel the complexities surrounding this environmental challenge. The project adopts a multidimensional approach, combining temporal, spatial, and causal analyses to provide a comprehensive understanding of forest fires in Algeria. Below is an overview of the key components and methodologies employed in this EDA.

1.2 Dataset Information

Source: Algerian Forest Fires Dataset

Description: The dataset contains information about forest fires, including date, location, weather conditions, and other relevant variables. To perform an Exploratory Data Analysis (EDA) on Algerian forest fires, you would ideally need a dataset that includes relevant information on various aspects of the fires. Below is a suggested outline for the dataset information you might need:

Temporal Information:

- Date and time of each recorded forest fire incident.
- Duration of the fire (if available).
- Historical records spanning multiple years to analyze temporal trends.
- Geographic coordinates (latitude and longitude) of each fire incident.
- Region or province information to understand the spatial distribution.

1.3 Objective

The EDA phase focuses on understanding the dataset's structure, identifying patterns, and extracting valuable insights to inform further analysis and decision-making. Acquire the Algerian Forest Fires dataset.

Understand the dataset's structure, variables, and context.

Temporal and Spatial Analysis: Investigate temporal and spatial distribution patterns of forest fires to discern trends and variations.

Causative Factors Identification: Identify natural and human-induced causes of forest fires.

Ecological Impact Assessment: Assess the impact on ecosystems, biodiversity, and soil health.

Chapter 2: DATA COLLECTION AND UNDERSTANDING

2.1 Data Source

I 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 Classification algorithms.

Official Government Databases:- Check government agencies responsible for forestry, environment, or disaster management in Algeria. Look for official reports, databases, or publications related to forest fires. These may include statistics on the frequency, location, and causes of fires.

Meteorological Data:- Obtain historical meteorological data for relevant regions in Algeria. Look for factors like temperature, humidity, wind speed, and precipitation. These variables can contribute to understanding the climatic conditions conducive to forest fires.

Satellite Imagery:- Utilize satellite imagery from platforms like NASA or European Space Agency (ESA). Satellite data can provide visualizations of the extent and spread of forest fires. Tools like Google Earth Engine can be valuable for analyzing such imagery.

Historical Records and Archives:- Access historical records, newspaper archives, or academic publications that may contain information on past forest fire incidents. Local libraries, research institutions, and online databases can be valuable resources.

Government Reports and Policies:- Look for government reports or policies related to forest fire management. These documents can provide insights into mitigation strategies, prevention efforts, and the overall response framework.

Collaboration with Research Institutions:- Reach out to research institutions or universities in Algeria involved in environmental studies, forestry, or climate science. Collaborate to access relevant datasets or research findings.

International Organizations:- Check databases and reports from international organizations like the United Nations, FAO (Food and Agriculture Organization), or regional environmental bodies. These sources may offer a broader perspective on the issue.

2.2 Tools Used

Software used:- Jupyter Notebook (Python)

Chapter 3: DATA EXPLORATION

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

Below are tasks to be performed in EDA:

1. Importing Libraries
2. Data Cleaning for EDA Report
3. Exploratory Data Analysis (EDA) on all Features

3.1 Import Libraries

```
# Import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import seaborn as sns
from warnings import filterwarnings
filterwarnings('ignore')
```

```
# Import dataset
dataset = pd.read_csv('forestfire.csv', delimiter=',')
dataset.head()
```

	day	month	year	Temperature	RH	Ws	Rain	FEMC	DMC	DC	ISI	BUI	FWI	Classes
0	01	08	2012	29	87	13	0	857	34	78	13	34	0.5	no fire
1	02	08	2012	29	87	13	13	844	41	78	1	36	0.4	no fire
2	03	08	2012	28	82	22	13.1	471	26	77	0.3	27	0.1	no fire
3	04	08	2012	28	88	13	2.6	283	13	89	0	17	0	no fire
4	05	08	2012	27	77	13	0	840	3	142	12	39	0.6	no fire

```
# Convert the dataset to dictionary
data = dataset.to_dictorient('records')
```

```
# Import the dataset from the file
# Import the dataset from the file
fb, fire_names, intent_list, data = fb.import_data('forestfire.csv',
                                                    delimiter=',',
                                                    data_format='records',
                                                    usecols=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14])
```

```
# Create a DataFrame object
df = pd.DataFrame(list(data))
df.columns = ['day', 'month', 'year', 'Temperature', 'RH', 'Ws', 'Rain', 'FEMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI', 'Classes']
```

	day	month	year	Temperature	RH	Ws	Rain	FEMC	DMC	DC	ISI	BUI	FWI	Classes
0	01	08	2012	29	87	13	0	857	34	78	13	34	0.5	no fire
1	02	08	2012	29	87	13	13	844	41	78	1	36	0.4	no fire
2	03	08	2012	28	82	22	13.1	471	26	77	0.3	27	0.1	no fire
3	04	08	2012	28	88	13	2.6	283	13	89	0	17	0	no fire
4	05	08	2012	27	77	13	0	840	3	142	12	39	0.6	no fire

241	26	09	2012	30	85	14	0	264	15	44.5	4.5	13.6	3.5	fire
242	27	09	2012	28	87	15	44	411	8.5	3	0.1	3.2	0	no fire
243	28	09	2012	27	87	28	0.8	459	3.6	7.9	0.4	1.4	0.2	no fire
244	29	09	2012	24	84	12	0.1	797	4.3	16.2	1.7	6.1	0.7	no fire
245	30	09	2012	24	84	16	0.2	873	3.8	16.6	1.2	4.3	0.6	no fire

246 rows x 14 columns

```
In [10]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 245 entries, 0 to 244
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  -
 0   day                245 non-null    object
 1   month              245 non-null    object
 2   year               245 non-null    object
 3   Temperature        245 non-null    object
 4   RH                 245 non-null    object
 5   WS                 245 non-null    object
 6   Rain              245 non-null    object
 7   FFMQ              245 non-null    object
 8   DMC               245 non-null    object
 9   DC               245 non-null    object
10   ISI              245 non-null    object
11   BUI              245 non-null    object
12   FWI              245 non-null    object
13   Classes          244 non-null    object
dtypes: object(13)
memory usage: 27.0+ KB
```

3.2 Data Cleaning

```
In [11]: df = df[df['Region'] != '']
In [12]: df.info()
```

```

      day month year Temperature  RH  WS  Rain  FFMQ  DMC  DC  ISI  BUI  FWI  Classes
122  Sidi-Bel-Abbes Region Dataset  float64  float64  float64  float64  float64  float64  float64  float64  float64  float64  float64  float64  float64  float64
167  14      07  2012           27      27      75      0.0      319      129      1459      103      10.4      416      float64

```

The dataset is converted into two sets based on Region from 122th index. We can make a new column based on the Region.

- 1 "Bejaia Region Dataset"
- 2 "Sidi-Bel-Abbes Region Dataset"

Add new column with region

```
In [13]: df.loc[:, 'Region'] =
df.loc[:, 'Region'] =
df[['122:167']] = df[['Region']] astype('int')
```

```
In [14]: df.isnull().sum()
```

```

day                0
month              1
year              1
Temperature        1
RH                 1
WS                 1
Rain              1
FFMQ              1
DMC               1
DC               1
ISI              1
BUI              1
FWI              1
Classes           2
Region            0
dtype: int64

```

```
In [15]: # Remove null or na values from
df = df.dropna()
df.shape
```

```
In [16]: (244, 15)
```


df = df[df['year'] > 1997]

```

day month year Temperature RH Ws Rain FFMC DMC DC ISI BUI FWI Classes Region
122 1 1 month 1998 Temperature RH Ws Rain FFMC DMC DC ISI BUI FWI Classes Region

```

Null and other unwanted rows are removed

```

df = df[df['year'] > 1997]
df = df[df['year'] < 2012]

```

Some column names were misspaced so we can fix it

```

df.columns = df.columns.str.strip()
df.columns

Index: day, month, year, Temperature, RH, Ws, Rain, FFMC, DMC, DC, ISI, BUI, FWI, Classes, Region

```

Looks like the values were misspaced so we can use strip() method

```

df.columns = df.columns.str.strip()
df.columns

Index: day, month, year, Temperature, RH, Ws, Rain, FFMC, DMC, DC, ISI, BUI, FWI, Classes, Region

```

Changing the required columns as integer data type

```

df[['day', 'month', 'year', 'Classes', 'Region']] = df[['day', 'month', 'year', 'Classes', 'Region']].astype(int)

```

Changing the other columns to Float data type

```

df[['Temperature', 'RH', 'Ws', 'Rain', 'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI']] = df[['Temperature', 'RH', 'Ws', 'Rain', 'FFMC', 'DMC', 'DC', 'ISI', 'BUI', 'FWI']].astype(float)

```

df.info()

```

Out[10]:
day      int64 143 non-null
month    int64 143 non-null
year      int64 143 non-null
Temperature  float64 143 non-null
RH        float64 143 non-null
Ws        float64 143 non-null
Rain      float64 143 non-null
FFMC      float64 143 non-null
DMC      float64 143 non-null
DC        float64 143 non-null
ISI       float64 143 non-null
BUI       float64 143 non-null
FWI       float64 143 non-null
Classes   int64 143 non-null
Region    int64 143 non-null
dtypes: float64(10), int64(6)

```

df.describe()

	count	mean	std	min	25%	50%	75%	max
day	143	15.761538	8.742792	1.0	8.00	16.0	23.00	31.0
month	143	8.000000	3.141593	1.0	7.00	8.0	8.00	9.0
year	143	2012.000000	0.000000	2012.0	2012.00	2012.0	2012.00	2012.0
Temperature	143	51.172693	10.64089	22.0	30.00	52.0	57.00	62.0
RH	143	62.841192	14.828198	21.0	52.50	69.0	73.50	90.0
Ws	143	15.403627	4.811305	5.0	14.00	15.0	17.00	29.0
Rain	143	5.562947	2.503267	0.0	0.00	3.0	6.50	16.0
FFMC	143	48.821987	12.347661	25.0	31.00	45.0	48.00	65.0
DMC	143	12.660058	12.300640	0.0	0.00	11.0	20.00	45.0
DC	143	47.450004	47.537606	0.0	14.00	50.0	59.00	120.0
ISI	143	4.742187	4.174294	0.0	0.40	3.0	7.20	19.0
BUI	143	15.000000	14.211421	0.0	0.00	12.0	24.00	39.0
FWI	143	35.001991	15.440568	0.0	0.00	4.0	13.00	31.0
Region	143	14.000000	0.000000	0.0	0.00	0.0	0.00	14.0

Data Cleaning in target variable

```
In [56]: # check if there are any NA and strip values
df.Classes.value_counts()
```

```
Out[56]:
fire      131
not fire  101
fire       4
fire       2
not fire   2
not fire   1
not fire   1
not fire   1
Name: Classes, dtype: int64
```

Looks like the values were misspelled so we can use `str.strip()` method

```
In [57]: # Strip misspelled values
df.Classes = df.Classes.str.strip()
```

```
In [58]: df.Classes.value_counts()
```

```
Out[58]:
fire      137
not fire  105
Name: Classes, dtype: int64
```

Region 1 (Bejla Region)

```
In [59]: df[1:122]
```

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	Iai	BUI	FWI	Classes	Region
0	1	6	2012	25.57	18	0.0	69.7	3.2	7.6	1.3	3.4	0.5	not fire		
1	2	6	2012	25.71	13	0.3	62.1	2.1	7.6	1	3.6	0.4	not fire		
2	3	6	2012	26.82	22	13.1	47.1	2.6	7.1	0.3	2.7	0.1	not fire		
3	4	6	2012	25.85	13	2.3	48.1	1.5	6.5	0.0	3.7	0.0	not fire		
4	5	6	2012	27.77	10	0.0	64.8	3.0	14.2	1.2	3.8	0.5	not fire		
...															
117	16	6	2012	31.54	11	0.0	82.0	6.0	10	2.5	6.2	1.7	not fire		
118	17	6	2012	31.66	11	0.0	85.7	8.3	14.6	4.5	6.1	2.1	not fire		
119	18	6	2012	32.47	14	0.7	77.5	7.1	8.2	1.8	6.8	0.6	not fire		
120	19	6	2012	26.89	16	1.8	47.4	2.9	7.7	0.3	3.0	0.1	not fire		
121	20	6	2012	25.78	14	1.4	45.0	1.9	7.5	0.2	2.4	0.1	not fire		

122 rows * 15 columns

Region 2 (Sidi-Bei Abdee Region)

```
In [64]: df[122:]
```

	day	month	year	Temperature	RH	Ws	Rain	FFMC	DMC	DC	Iai	BUI	FWI	Classes	Region
122	1	6	2012	32.71	12	0.7	57.1	2.5	9.2	0.6	2.6	0.2	not fire	2	
123	2	6	2012	30.73	13	4.0	53.7	2.7	7.8	0.6	2.9	0.2	not fire	2	
124	3	6	2012	29.80	14	2.0	48.7	2.2	7.6	0.3	2.6	0.1	not fire	2	
125	4	6	2012	30.64	14	0.0	79.4	5.1	10.4	2.2	5.6	1.0	not fire	2	
126	5	6	2012	32.60	14	0.2	77.1	6.0	17.6	1.8	6.5	0.0	not fire	2	
...															
238	26	6	2012	30.65	14	0.0	85.4	16.0	44.5	4.5	18.9	6.5	fire	2	
239	27	6	2012	26.87	10	4.4	41.1	6.5	8.0	0.1	6.2	0.0	not fire	2	
240	28	6	2012	27.87	29	0.5	45.5	3.5	7.9	0.4	3.4	0.2	not fire	2	
241	29	6	2012	24.64	16	0.1	79.7	4.3	1.2	1.7	5.1	1.1	not fire	2	
242	30	6	2012	24.64	16	0.2	67.9	2.8	10.5	1.2	4.8	0.1	not fire	2	

121 rows * 15 columns

3.3 Exploratory Data Analysis (EDA)

```
In [250]: # Drop the 'id' column
df1 = df1.drop(['id'], axis=1)

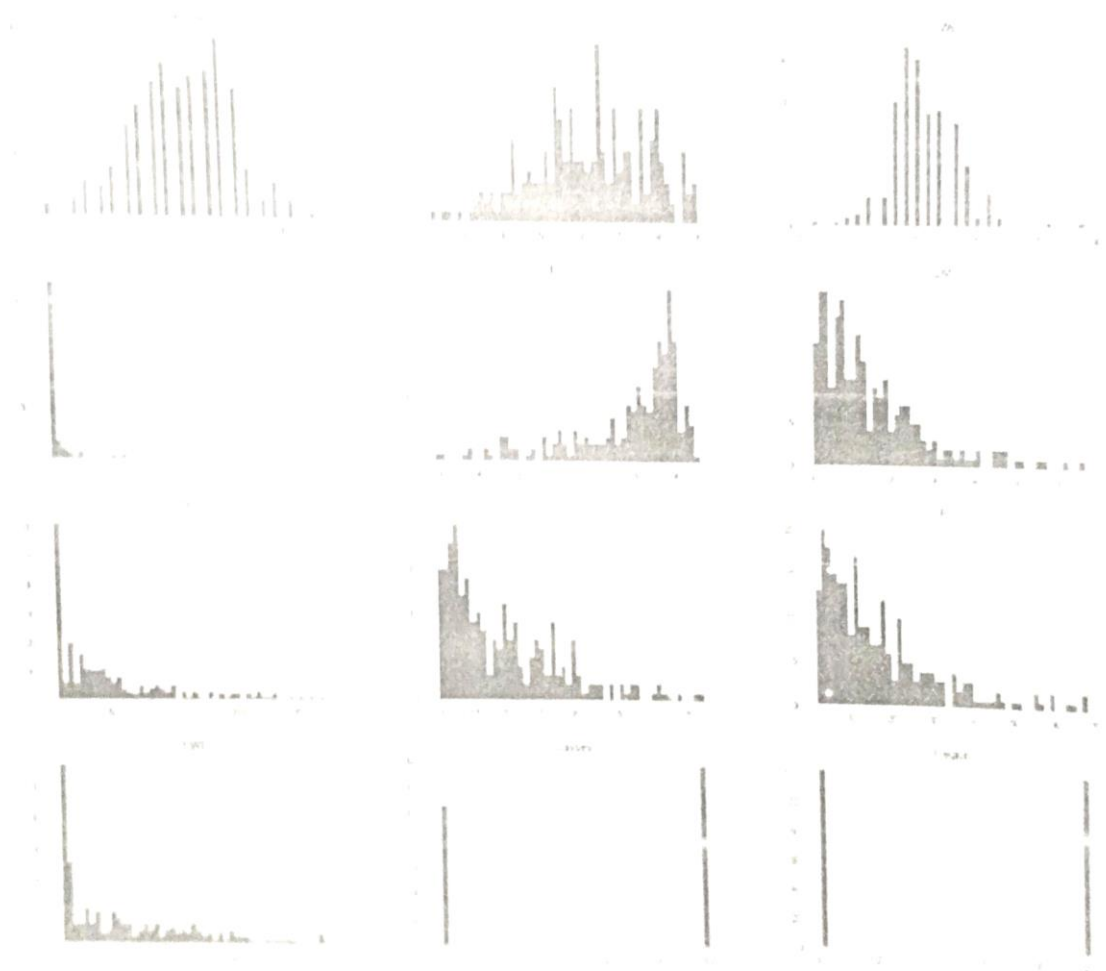
In [251]: # Filter out the 'id' column
df1['classes'] = np.where(df1['classes'] == 0, 1, 2)

In [252]: # Check the counts
df1['classes'].value_counts()

Out[252]: 1    137
          0    106
          Name: classes, dtype: int64
```

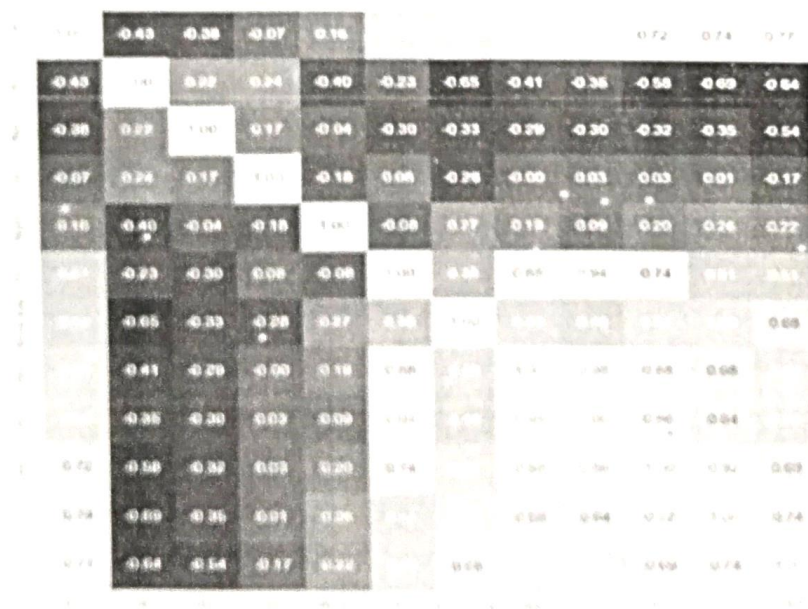
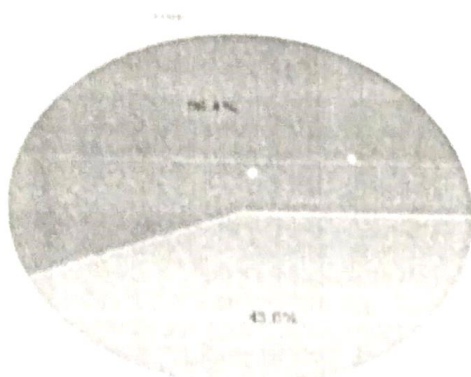
Visualization of Distribution

```
In [254]: # Create a figure with 4 subplots
plt.figure(figsize=(12, 10))
df1.hist()
plt.show()
```



Conclusions

First Part of the Lesson



xiv

BY the number of fire incidents



Monthwise Fire Analysis

```

# Create a bar chart showing the number of fire incidents by month
plt.figure(figsize=(10,6))
plt.title('Monthwise Fire Analysis')
plt.xlabel('Month')
plt.ylabel('Number of Incidents')
plt.grid(True)
plt.show()

```

Fire Analysis Month wise for Bejira Region

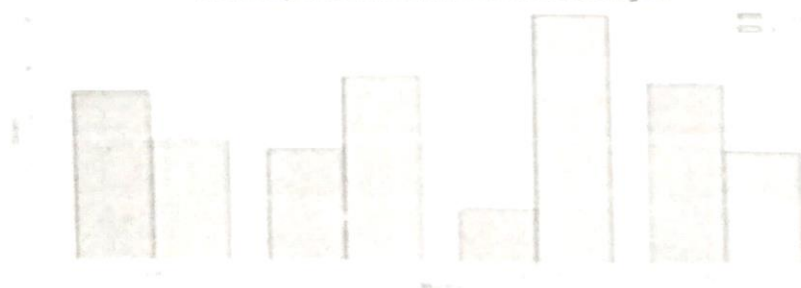


```

# Create a bar chart showing the number of fire incidents by month
plt.figure(figsize=(10,6))
plt.title('Monthwise Fire Analysis')
plt.xlabel('Month')
plt.ylabel('Number of Incidents')
plt.grid(True)
plt.show()

```

Fire Analysis Month wise for Sidi Bel Abbas Region



- By observing the number of fire incidents in both regions, it can be seen that the number of incidents is higher in the month of September.
- Most of the fire incidents in the region are caused by the use of fire in the month of September.
- The number of incidents is higher in the month of September.

3.4 Exploratory Data Analysis (EDA) on all Features

```

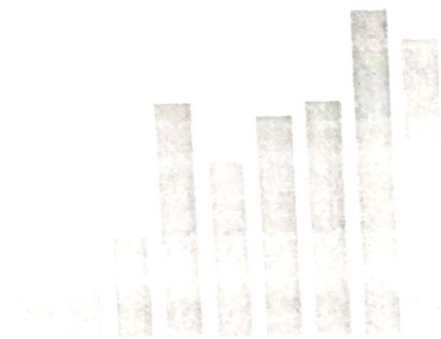
# Feature importance
from sklearn.inspection import permutation_importance

# Feature importance for the model
perm_importance = permutation_importance(model, X_train, y_train, n_repeats=100)

# Sort features by importance
feature_importance = perm_importance['p_values'].argsort()

# Print feature importance
for feature in feature_importance:
    print(f"Feature: {feature}, Importance: {perm_importance['p_values'][feature]}")

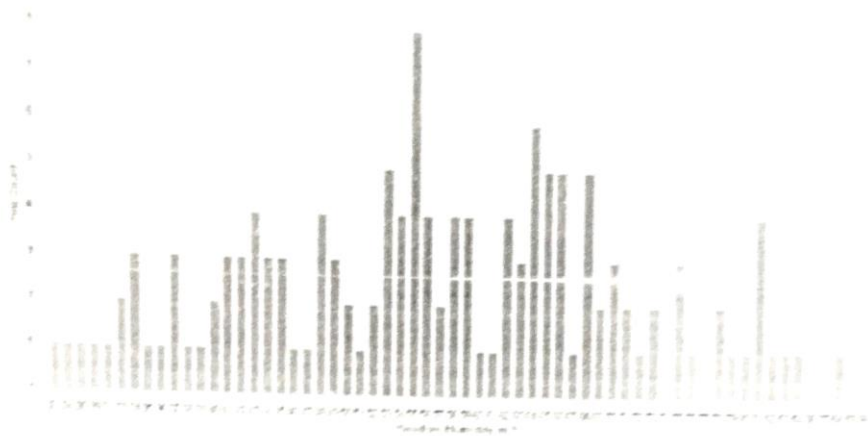
```



Fast-Fuel Moisture Coefficient (FFMC) index




```
in [20]: torolots(100, 100000, 100000, 100000)
```



```
fig = plt.figure(figsize=(10, 8))
ax = diffopt.optimize()
ax.set_title('Random Forest Classifier')
plt.show()
```

2001 / 23 - 2002 /

```

[10]: diff = df['diff'] + df['drop'] + df['drop2'] + df['drop3'] + df['drop4'] + df['drop5']
for feature in diff:
    plt.plot(df[feature], df['drop'], label = feature)
    plt.legend(loc='best')
    plt.title(feature)
    plt.show()

```

Chapter 4: CONCLUSION

The EDA phase provides a foundational understanding of the Algerian Forest Fires Dataset. These insights pave the way for further analysis and modeling. Subsequent project phases may involve feature engineering, statistical tests, and advanced visualizations to address specific research questions or predictive tasks. The data collection and understanding phase provided essential insights into the Algerian Forest Fires dataset. Challenges related to missing values or data anomalies were addressed, and a preliminary exploration of key features was conducted. This phase lays the groundwork for the subsequent Exploratory Data Analysis (EDA), where more in-depth insights will be derived to inform strategic decision-making regarding forest fire prevention and management in Algeria.

In conclusion, the Exploratory Data Analysis (EDA) on Algerian forest fires provides valuable insights into the patterns and factors influencing these environmental incidents. Through a thorough examination of relevant data, we have uncovered the multifaceted nature of the issue, encompassing both natural and human-induced causes. The impact on ecosystems, economies, and local communities underscores the urgency of effective management and prevention strategies.

The examination of challenges and barriers highlights the need for comprehensive solutions, taking into account resource limitations, changing climate conditions, and human behavior. By understanding these complexities, stakeholders can better tailor interventions to address the root causes of forest fires in Algeria. Furthermore, international collaboration emerges as a crucial component in the fight against forest fires, emphasizing the interconnectedness of environmental challenges. Learning from case studies and leveraging data-driven insights, we can pave the way for informed decision-making, fostering a more resilient and sustainable approach to forest fire management.

As we navigate the future, the EDA serves as a foundation for ongoing research and action. By integrating data-driven strategies, fostering community engagement, and staying vigilant against emerging challenges, we can work towards a more secure and sustainable coexistence with Algeria's precious forests.

References

- [1].Algerian National Meteorological Office. (2019). "Meteorological Data for Forest Fire Analysis."
[<https://archive.ics.uci.edu/dataset/547/algerian+forest+fires+dataset>]
- [2].Kaggle. (2012). "Algerian Forest Fires Dataset."Kaggle.
[<https://archive.ics.uci.edu/ml/datasets/Algerian+Forest+Fires+Dataset++>]
- [3].McKinney, W., & Others. (2017). "Pandas: a foundational Python library for data analysis and statistics."
Python for High-Performance and Scientific Computing, 14.
- [4].Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment." Computing in Science & Engineering,
9(3), 90-95. [<http://dx.doi.org/10.1109/MCSE.2007.55>]
- [5].Python Software Foundation. "Python Language Reference, version X.X."
[<https://docs.python.org/3/reference/>]