Forest Fire Area Prediction

THE HOPLITES

-mamatha,neeharika,surya,suresh

Introduction

Forests play an important role for supporting the human environment and Forest fires are one among the largest dangers for forest preservation.
Forest fires importantly influence our environment and lives . Negative effects resulting from high levels of burn severity include significant removal of organic matter, deterioration of both soil structure and porosity, considerable loss of nutrients through volatilization, ash entrapment in smoke columns, leaching, and erosion. Also, the release of hazardous chemicals significantly impacts human health and increases the risk of future diseases, wildfire smoke is accompanied by high concentrations of carbon dioxide, which can result
in consequences such as headache, mental confusion, nausea, disorientation, coma, and even death. Even at lower concentrations, the effects of carbon dioxide should not be neglected; individuals with cardiovascular disease may experience chest pain and cardiac arrhythmia. A comprehensive study tracking wildfire firefighter deaths from 1990 to 2006
reported that 21.9 % of their deaths occurred from heart attacks .The ability of accurately predicting the area that may be involved in a forest fire event may help in optimizing fire management efforts. Based on the complexity of the task, powerful computational tools are needed for predicting the amount of area that will be burned during a forest fire.

> *"A house you can rebuild; a bridge you can restring; a washed-out road you can fill in.*
> *But there is nothing you can do about a tree but mourn."*

Objectives Of Research

- *To protect the wildlife and natural resources*
- *To protect the environment from green house gases*
- *Large Fire Management*
- *Resource Acquisation and strategic Deployment*

Problem Statement

Large Area Of Forests are burnt and still we are facing the problem.
Our Project is to predict the area of the forest that will be burned in case of any wild fire

Industry Profile:

This project is helpful to people of fire department
This project is helpful to people of forest department

Review Of Literature:

- *Prediction of Forest Fire Area Better Helps in forest fire management*
- *Management Of Active Fires On the Basis of Behaviour Prediction*
- *A Direct Forest Fire Management is Control Of Each Fire Spread and development on the basis of it's behavior prediction*

Data Collection:

This data set is taken from https://archive.ics.uci.adu/ml/datasets/forestfires
Data is collected from this website https://www.kaggle1
These are different variables and there ranges in the dataset
X - x-axis spatial coordinate
Y - y-axis spatial coordinate
month - month of the year
day - day of the week
FFMC - FFMC index from the FWI system
DMC - DMC index from the FWI system
DC - DC index from the FWI system
ISI - ISI index from the FWI system:
temp - temperature in Celsius degrees
RH - relative humidity in %
wind - wind speed in km/h
rain - outside rain in mm/m2
area - the burned area of the forest (in ha)
FWI represents Fire Weather Index

Fuel Moisture Codes:
The FWI System evaluates fuel moisture content and relative fire behavior using past and present weather effects on ground level fuels.
The moisture codes reflect the net effects of daily moisture gains and losses.

FFMC:-The Fine Fuel Moisture Code (FFMC) is a numerical rating of the moisture content of surface litter and other cured fine fuels. It shows the relative ease of ignition and flammability of fine fuels. The moisture content of fine fuels is very sensitive to the weather. Even a day of rain, or of fine and windy weather, will significantly affect the FFMC rating. The system uses a time lag of two-thirds of a day to accurately measure the moisture content in fine fuels.

DMC:-The Duff Moisture Code (DMC) is a numerical rating of the average moisture content of loosely compacted organic layers of moderate depth. The code indicates the depth that fire will burn in moderate duff layers and medium size woody material. Duff layers take longer than surface fuels to dry out but weather conditions over the past couple of weeks will significantly affect the DMC. The system applies a time lag of 12 days to calculate the DMC

DC:-The Drought Code (DC) is a numerical rating of the moisture content of deep, compact, organic layers. It is a useful indicator of seasonal drought and shows the likelihood of fire involving the deep duff layers and large logs. A long period of dry weather (the system uses 52 days) is needed to dry out these fuels and affect the Drought Code. A DC rating of 200 is high, and 300 or more is extreme indicating that fire will involve deep sub-surface and heavy fuels. Burning off should not be permitted when the DC rating is above 300.

ISI:-The Initial Spread Index (ISI) indicates the rate fire will spread in its early stages. It is calculated from the FFMC rating and the wind factor.
The open-ended ISI scale starts at zero and a rating of 10 indicates high rate of spread shortly after ignition. A rating of 19 or more indicates extremely rapid rate of spread.

RH:- Relative humidity is the ratio of the partial pressure of water vapor to the equilibrium vapor pressure of water at a given temperature. Relative humidity depends on temperature and the pressure of the system of interest. The same amount of water vapor results in higher relative humidity in cool air than warm air.

4.Methodology

4.1 Exploratory Data Analysis:

Description Of Elements:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
count	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000	517.000000
mean	4.669246	4.299807	7.475822	4.259188	90.644681	110.872340	547.940039	9.021663	18.889168	44.288201	4.017602	0.021663	12.847292
std	2.313778	1.229900	2.275990	2.072929	5.520111	64.046482	248.066192	4.559477	5.806625	16.317469	1.791653	0.295959	63.655818
min	1.000000	2.000000	1.000000	1.000000	18.700000	1.100000	7.900000	0.000000	2.200000	15.000000	0.400000	0.000000	0.000000
25%	3.000000	4.000000	7.000000	2.000000	90.200000	68.600000	437.700000	6.500000	15.500000	33.000000	2.700000	0.000000	0.000000
50%	4.000000	4.000000	8.000000	5.000000	91.600000	108.300000	664.200000	8.400000	19.300000	42.000000	4.000000	0.000000	0.520000
75%	7.000000	5.000000	9.000000	6.000000	92.900000	142.400000	713.900000	10.800000	22.800000	53.000000	4.900000	0.000000	6.570000
max	9.000000	9.000000	12.000000	7.000000	96.200000	291.300000	860.600000	56.100000	33.300000	100.000000	9.400000	6.400000	1090.840000

Here we detect outliers for the variables and replace them with the median of their respective columns using the following code snippet.

Median=df.x.median()
df['X']=df['X'].mask((df['X']<maxvalue),median)

4.1.1. Figures and Tables

The Co-relation between the attributes can be given as

Df.corr(method="pearson")

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
X	1.000000	0.357587	-0.098388	-0.024922	-0.038482	-0.070968	-0.078751	-0.013501	-0.065848	0.058564	0.018362	0.065387	0.063385
Y	0.357587	1.000000	-0.033843	-0.015037	0.039550	0.071541	-0.111427	0.047612	-0.037137	0.033773	-0.016619	0.037700	0.059396
month	-0.098388	-0.033843	1.000000	-0.100779	-0.037123	-0.007462	0.353501	-0.086663	-0.008731	-0.064194	-0.048694	-0.018515	0.026890
day	-0.024922	-0.015037	-0.100779	1.000000	0.051132	0.009082	-0.008877	0.006446	0.037368	0.073047	0.047737	-0.048340	0.023226
FFMC	-0.038482	0.039550	-0.037123	0.051132	1.000000	0.488035	0.300721	0.620735	0.527279	-0.197777	-0.005098	0.113342	0.041474
DMC	-0.070968	0.071541	-0.007462	0.009082	0.488035	1.000000	0.491266	0.456437	0.528986	0.010030	-0.100597	0.092718	0.089956
DC	-0.078751	-0.111427	0.353501	-0.008877	0.300721	0.491266	1.000000	0.197260	0.326258	0.086830	-0.142949	0.029331	0.036577
ISI	-0.013501	0.047612	-0.086663	0.006446	0.620735	0.456437	0.197260	1.000000	0.399622	-0.159826	0.125031	0.088711	0.017701
temp	-0.065848	-0.037137	-0.008731	0.037368	0.527279	0.528986	0.326258	0.399622	1.000000	-0.496490	-0.158561	0.070252	0.101744
RH	0.058564	0.033773	-0.064194	0.073047	-0.197777	0.010030	0.086830	-0.159826	-0.496490	1.000000	0.113697	0.110424	-0.072226
wind	0.018362	-0.016619	-0.048694	0.047737	-0.005098	-0.100597	-0.142949	0.125031	-0.158561	0.113697	1.000000	0.073225	0.013307
rain	0.065387	0.037700	-0.018515	-0.048340	0.113342	0.092718	0.029331	0.088711	0.070252	0.110424	0.073225	1.000000	-0.007366
area	0.063385	0.059396	0.026890	0.023226	0.041474	0.089956	0.036577	0.017701	0.101744	-0.072226	0.013307	-0.007366	1.000000

4.2 Statistical Data and Data Visualization

X - 1 to 9.

Y - 2 to 9.

month - 'jan' to 'dec'

day - 'mon' to 'sun'

FFMC - 18.7 to 96.20

DMC - 1.1 to 291.3

DC - 7.9 to 860.6

ISI - 0.0 to 56.10

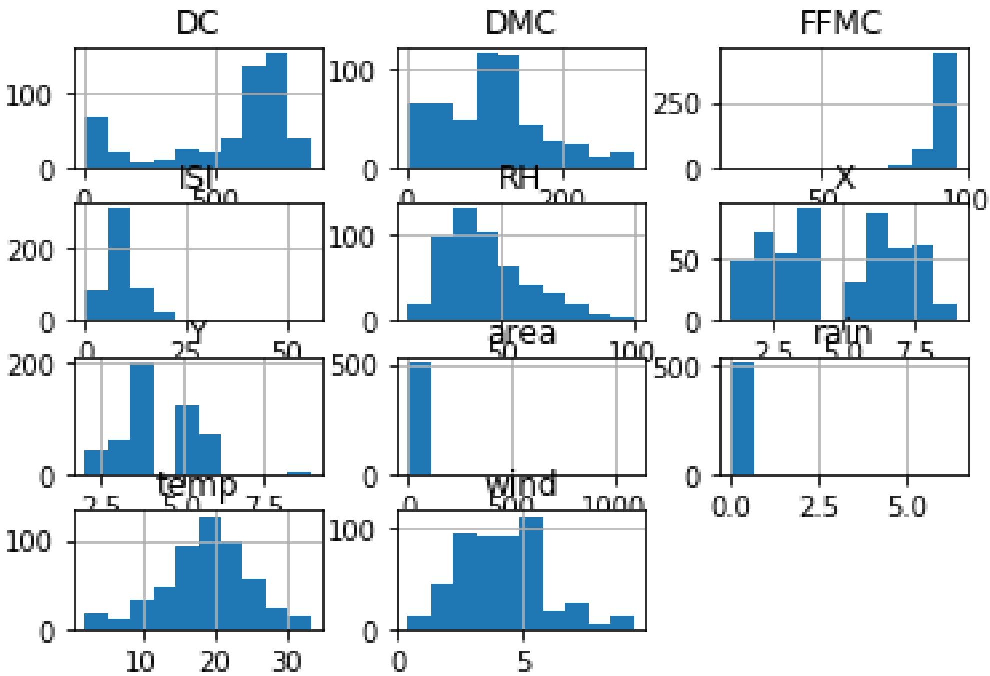
temp - 2.2 to 33.30

RH - relative humidity in %: 15.0 to 100

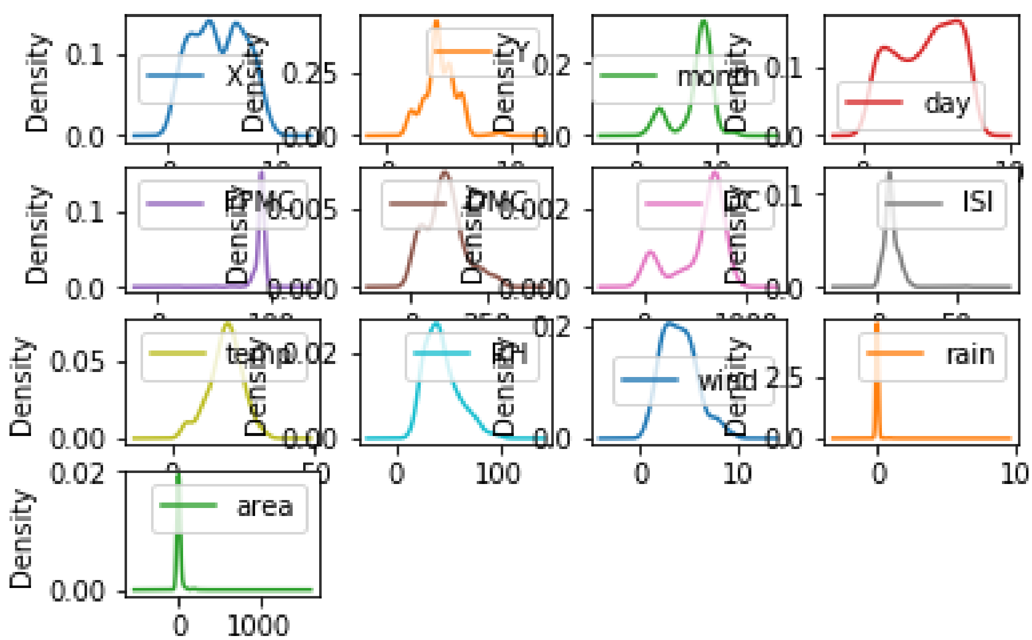
wind - wind speed in km/h: 0.40 to 9.40

rain - outside rain in mm/m2 : 0.0 to 6.4

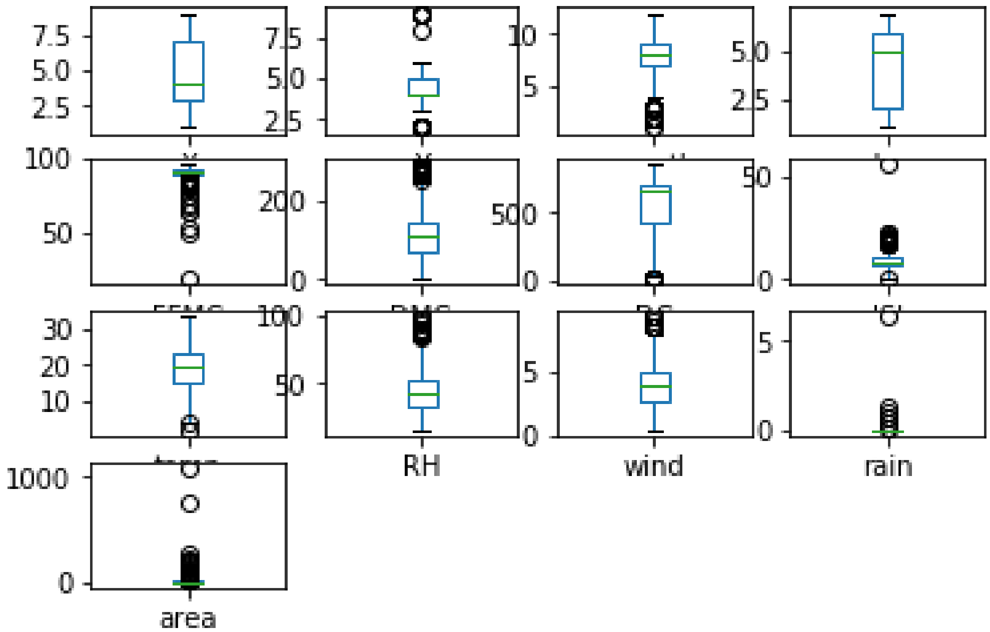
area - the burned area of the forest (in ha): 0.00 to 1090.8 *df.hist()*



df.plot(kind='density', subplots=True, layout=(4,4), sharex=False, sharey=False)



```
df.plot(kind='box', subplots=True, layout=(4,4), sharex=False,sharey=False)
```



Data Modeling Using Supervised ML Techniques:

Here we use Decision Tree Regression which best suits the predicition with an accuracy of 0.999

Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce me0…aningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

Here since we have the dependent feature area as a continuous variable this regression is very suitable.

```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X, y)
```

Here we fit the dataset with decision tree regressor using the above code snippet

```
y_pred = regressor.predict(X_test)
```

Now , the predicition can be done using the above line of code.

```
from sklearn import metrics
print('MAE',metrics.mean_absolute_error(y_test,y_pred))
print('MSE',metrics.mean_squared_error(y_test,y_pred))
print('RMSE',np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
print('r^2',metrics.r2_score(y_test,y_pred))
```

Now we have metric measures as
MAE 0.12163461538461538
MSE 0.5912389423076924
RMSE 0.7689206345961151
r^2 0.9999741805536381

Findings and Suggestions:

The Large Area Of Forest Fire Burns is Mainly because of Rising Temparatures day for day .
Large Forest Area Fire Issues are due to lack of rainfall .
We can also include other features such as cloudiness and hemisphere of location in the dataset to predict more accurately.

Conclusion:

Incidents of Forest Fires in size, frequency and intensity are controlled by weather   variables

References:

https://www.kaggle.com/elikplem/predict -the-burned-area-of-forest-fires
https://www.geeksforgeeks.org/python-decision-tree-regression-using-sklearn
https://www.kaggle1
https://archive.ics.uci.adu/ml/datasets/forestfires