

PROJECT DEVELOPMENT PHASE

DELIVERY OF SPRINT-1

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

```
from sklearn import metrics
from sklearn.metrics import classification_report, confusion_matrix
```

In []:

```
import warnings
warnings.filterwarnings(action="ignore")
%matplotlib inline
pd.set_option("display.max_rows", 1000)
pd.set_option("display.max_columns", 1000)
```

In []:

```
fires = pd.read_csv("forestfires.csv") #reading the dataset
fires.head(15) #show the first 15 instances of dataset
```

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0
5	8	6	aug	sun	92.3	85.3	488.0	14.7	22.2	29	5.4	0.0	0.0
6	8	6	aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0.0	0.0
7	8	6	aug	mon	91.5	145.4	608.2	10.7	8.0	86	2.2	0.0	0.0
8	8	6	sep	tue	91.0	129.5	692.6	7.0	13.1	63	5.4	0.0	0.0
9	7	5	sep	sat	92.5	88.0	698.6	7.1	22.8	40	4.0	0.0	0.0
10	7	5	sep	sat	92.5	88.0	698.6	7.1	17.8	51	7.2	0.0	0.0
11	7	5	sep	sat	92.8	73.2	713.0	22.6	19.3	38	4.0	0.0	0.0
12	6	5	aug	fri	63.5	70.8	665.3	0.8	17.0	72	6.7	0.0	0.0

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area	
13	6	5	sep	mon	90.9	126.5	686.5	7.0	21.3	42	2.2	0.0	0.0
14	6	5	sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0.0	0.0

In []:

```
#show the last 10 instances of dataset
fires.tail(10)
```

Out[]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
507	2	4	aug	fri	91.0	166.9	752.6	7.1	25.9	41	3.6	0.0	0.00
508	1	2	aug	fri	91.0	166.9	752.6	7.1	25.9	41	3.6	0.0	0.00
509	5	4	aug	fri	91.0	166.9	752.6	7.1	21.1	71	7.6	1.4	2.17
510	6	5	aug	fri	91.0	166.9	752.6	7.1	18.2	62	5.4	0.0	0.43
511	8	6	aug	sun	81.6	56.7	665.6	1.9	27.8	35	2.7	0.0	0.00
512	4	3	aug	sun	81.6	56.7	665.6	1.9	27.8	32	2.7	0.0	6.44
513	2	4	aug	sun	81.6	56.7	665.6	1.9	21.9	71	5.8	0.0	54.29
514	7	4	aug	sun	81.6	56.7	665.6	1.9	21.2	70	6.7	0.0	11.16
515	1	4	aug	sat	94.4	146.0	614.7	11.3	25.6	42	4.0	0.0	0.00
516	6	3	nov	tue	79.5	3.0	106.7	1.1	11.8	31	4.5	0.0	0.00

In []:

```
#generate descriptive statistics of each attribute
fires.describe().T
```

Out[]:

	count	mean	std	min	25%	50%	75%	max
X	517.0	4.669246	2.313778	1.0	3.0	4.00	7.00	9.00
Y	517.0	4.299807	1.229900	2.0	4.0	4.00	5.00	9.00
FFMC	517.0	90.644681	5.520111	18.7	90.2	91.60	92.90	96.20
DMC	517.0	110.872340	64.046482	1.1	68.6	108.30	142.40	291.30
DC	517.0	547.940039	248.066192	7.9	437.7	664.20	713.90	860.60

	count	mean	std	min	25%	50%	75%	max
ISI	517.0	9.021663	4.559477	0.0	6.5	8.40	10.80	56.10
temp	517.0	18.889168	5.806625	2.2	15.5	19.30	22.80	33.30
RH	517.0	44.288201	16.317469	15.0	33.0	42.00	53.00	100.00
wind	517.0	4.017602	1.791653	0.4	2.7	4.00	4.90	9.40
rain	517.0	0.021663	0.295959	0.0	0.0	0.00	0.00	6.40
area	517.0	12.847292	63.655818	0.0	0.0	0.52	6.57	1090.84

In []:

```
#given area of land burnt, but we have to predict if there is fire or not so changing values of area to 0 and 1 only
#here 0 represet there is not fire and 1 represent fire, changing all values of area which are greater than 0 to 1
fires['area'].values[fires['area'].values > 0] = 1
```

```
#renaming the area attribute to output for clear understanding
fires = fires.rename(columns={'area': 'output'})
fires.head(5)
```

Out[]:

	X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	output
0	7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
1	7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
2	7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
3	8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
4	8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

In []:

```
#Compute pairwise correlation of columns
fires.corr()
```

Out[]:

	X	Y	FFMC	DMC	DC	ISI	temp	RH	wind	rain	output
X	1.000000	0.539548	-0.021039	0.048384	0.085916	0.006210	-0.051258	0.085223	0.018798	0.065387	0.062491
Y	0.539548	1.000000	-0.046308	0.007782	0.101178	-0.024488	-0.024103	0.062221	-0.020341	0.033234	0.056892

	X	Y	FFM C	DMC	DC	ISI	temp	RH	wind	rain	output
FF MC	- 0.0210 39	- 0.0463 08	1.0000 00	0.3826 19	0.3305 12	0.5318 05	0.4315 32	- 0.3009 95	- 0.0284 85	0.0567 02	0.0738 23
DM C	- 0.0483 84	0.0077 82	0.3826 19	1.0000 00	0.6821 92	0.3051 28	0.4695 94	0.0737 95	- 0.1053 42	0.0747 90	0.0626 72
DC	- 0.0859 16	- 0.1011 78	0.3305 12	0.6821 92	1.0000 00	0.2291 54	0.4962 08	- 0.0391 92	- 0.2034 66	0.0358 61	0.0967 24
ISI	0.0062 10	- 0.0244 88	0.5318 05	0.3051 28	0.2291 54	1.0000 00	0.3942 87	- 0.1325 17	0.1068 26	0.0676 68	0.0356 63
tem p	- 0.0512 58	- 0.0241 03	0.4315 32	0.4695 94	0.4962 08	0.3942 87	1.0000 00	- 0.5273 90	- 0.2271 16	0.0694 91	0.0760 47
RH	0.0852 23	0.0622 21	- 0.3009 95	0.0737 95	- 0.0391 92	- 0.1325 17	- 0.5273 90	1.0000 00	0.0694 10	0.0997 51	- 0.0355 87
wind	0.0187 98	- 0.0203 41	- 0.0284 85	- 0.1053 42	- 0.2034 66	0.1068 26	- 0.2271 16	0.0694 10	1.0000 00	0.0611 19	0.0557 02
rain	0.0653 87	0.0332 34	0.0567 02	0.0747 90	0.0358 61	0.0676 68	0.0694 91	0.0997 51	0.0611 19	1.0000 00	0.0255 50
outp ut	0.0624 91	0.0568 92	0.0738 23	0.0626 72	0.0967 24	0.0356 63	0.0760 47	- 0.0355 87	0.0557 02	0.0255 50	1.0000 00