Assignment_4

September 28, 2023

1 Shaik Mansoor Vali

[1]:

```
#Import the Libraries.
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:

```
#Importing the dataset.
df=pd.read_csv("WA_Fn-UseC_-HR-Employee-Attrition.csv")
```

[3]:

```
df.head()
```

[3]:

```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
2  …                         2            80                 0
3  …                         3            80                 0
4  …                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
```

```
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

[4]:

```
df.shape
```

[4]:

```
(1470, 35)
```

[5]:

```
df.Age.value_counts()
```

[5]:

```
Age
35    78
34    77
36    69
31    69
29    68
32    61
30    60
33    58
38    58
40    57
37    50
27    48
28    48
42    46
39    42
45    41
41    40
26    39
44    33
46    33
43    32
50    30
25    26
24    26
49    24
47    24
55    22
51    19
```

```
53    19
48    19
54    18
52    18
22    16
56    14
23    14
58    14
21    13
20    11
59    10
19     9
18     8
60     5
57     4
Name: count, dtype: int64
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

[7]:
```
df.describe()
```

[7]:
```
               Age    DailyRate  DistanceFromHome    Education  EmployeeCount  \
count  1470.000000  1470.000000       1470.000000  1470.000000         1470.0
mean     36.923810   802.485714          9.192517     2.912925            1.0
std       9.135373   403.509100          8.106864     1.024165            0.0
min      18.000000   102.000000          1.000000     1.000000            1.0
25%      30.000000   465.000000          2.000000     2.000000            1.0
50%      36.000000   802.000000          7.000000     3.000000            1.0
75%      43.000000  1157.000000         14.000000     4.000000            1.0
max      60.000000  1499.000000         29.000000     5.000000            1.0

       EmployeeNumber  EnvironmentSatisfaction   HourlyRate  JobInvolvement  \
count     1470.000000              1470.000000  1470.000000     1470.000000
mean      1024.865306                 2.721769    65.891156        2.729932
std        602.024335                 1.093082    20.329428        0.711561
min          1.000000                 1.000000    30.000000        1.000000
25%        491.250000                 2.000000    48.000000        2.000000
50%       1020.500000                 3.000000    66.000000        3.000000
75%       1555.750000                 4.000000    83.750000        3.000000
max       2068.000000                 4.000000   100.000000        4.000000

          JobLevel  …  RelationshipSatisfaction  StandardHours  \
count  1470.000000  …               1470.000000         1470.0
mean      2.063946  …                  2.712245           80.0
std       1.106940  …                  1.081209            0.0
min       1.000000  …                  1.000000           80.0
25%       1.000000  …                  2.000000           80.0
50%       2.000000  …                  3.000000           80.0
75%       3.000000  …                  4.000000           80.0
max       5.000000  …                  4.000000           80.0

       StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
count       1470.000000        1470.000000            1470.000000
mean           0.793878          11.279592               2.799320
std            0.852077           7.780782               1.289271
min            0.000000           0.000000               0.000000
25%            0.000000           6.000000               2.000000
50%            1.000000          10.000000               3.000000
75%            1.000000          15.000000               3.000000
max            3.000000          40.000000               6.000000

       WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
count      1470.000000     1470.000000         1470.000000
mean          2.761224        7.008163            4.229252
std           0.706476        6.126525            3.623137
min           1.000000        0.000000            0.000000
25%           2.000000        3.000000            2.000000
50%           3.000000        5.000000            3.000000
75%           3.000000        9.000000            7.000000
max           4.000000       40.000000           18.000000

       YearsSinceLastPromotion  YearsWithCurrManager
count              1470.000000           1470.000000
mean                  2.187755              4.123129
std                   3.222430              3.568136
min                   0.000000              0.000000
25%                   0.000000              2.000000
50%                   1.000000              3.000000
75%                   3.000000              7.000000
max                  15.000000             17.000000

[8 rows x 26 columns]
```

[8]:
```
#Checking for Null Values.
df.isnull().any()
```

[8]:
```
Age                         False
Attrition                   False
BusinessTravel              False
DailyRate                   False
Department                  False
DistanceFromHome            False
Education                   False
EducationField              False
EmployeeCount               False
EmployeeNumber              False
EnvironmentSatisfaction     False
Gender                      False
HourlyRate                  False
```

```
     JobInvolvement              False
     JobLevel                    False
     JobRole                     False
     JobSatisfaction             False
     MaritalStatus               False
     MonthlyIncome               False
     MonthlyRate                 False
     NumCompaniesWorked          False
     Over18                      False
     OverTime                    False
     PercentSalaryHike           False
     PerformanceRating           False
     RelationshipSatisfaction    False
     StandardHours               False
     StockOptionLevel            False
     TotalWorkingYears           False
     TrainingTimesLastYear       False
     WorkLifeBalance             False
     YearsAtCompany              False
     YearsInCurrentRole          False
     YearsSinceLastPromotion     False
     YearsWithCurrManager        False
     dtype: bool
```

```
[9]: df.isnull().sum()
```

```
[9]: Age                         0
     Attrition                   0
     BusinessTravel              0
     DailyRate                   0
     Department                  0
     DistanceFromHome            0
     Education                   0
     EducationField              0
     EmployeeCount               0
     EmployeeNumber              0
     EnvironmentSatisfaction     0
     Gender                      0
     HourlyRate                  0
     JobInvolvement              0
     JobLevel                    0
     JobRole                     0
     JobSatisfaction             0
     MaritalStatus               0
     MonthlyIncome               0
     MonthlyRate                 0
     NumCompaniesWorked          0
```

```
Over18                      0
OverTime                    0
PercentSalaryHike           0
PerformanceRating           0
RelationshipSatisfaction    0
StandardHours               0
StockOptionLevel            0
TotalWorkingYears           0
TrainingTimesLastYear       0
WorkLifeBalance             0
YearsAtCompany              0
YearsInCurrentRole          0
YearsSinceLastPromotion     0
YearsWithCurrManager        0
dtype: int64
```

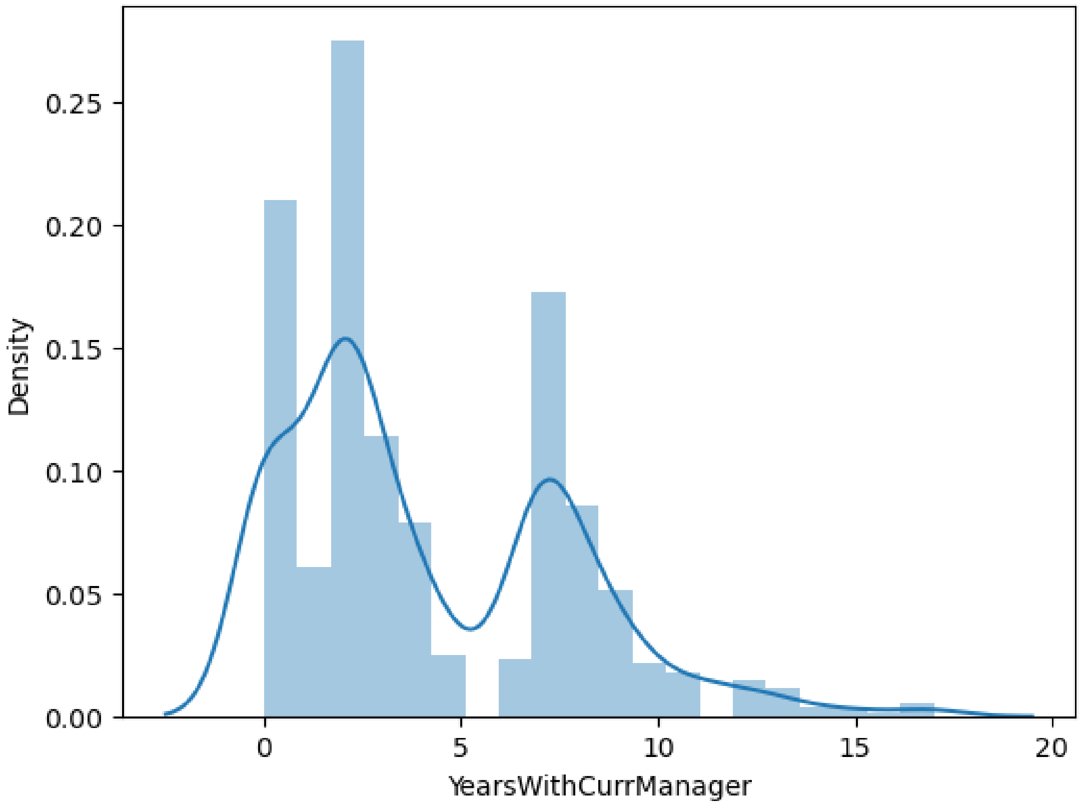
[10]:

```
#Data Visualization.
sns.distplot(df["YearsWithCurrManager"])
```

```
C:\Users\Mansoorvali\AppData\Local\Temp\ipykernel_3476\944934001.py:2:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df["YearsWithCurrManager"])
```

[10]:

```
<Axes: xlabel='YearsWithCurrManager', ylabel='Density'>
```



[11]:
```
num_df= df.select_dtypes(include=['number'])
num_df.corr()
```

[11]:
```
                               Age  DailyRate  DistanceFromHome  Education  \
Age                       1.000000   0.010661         -0.001686   0.208034
DailyRate                 0.010661   1.000000         -0.004985  -0.016806
DistanceFromHome         -0.001686  -0.004985          1.000000   0.021042
Education                 0.208034  -0.016806          0.021042   1.000000
EmployeeCount                  NaN        NaN               NaN        NaN
EmployeeNumber           -0.010145  -0.050990          0.032916   0.042070
EnvironmentSatisfaction   0.010146   0.018355         -0.016075  -0.027128
HourlyRate                0.024287   0.023381          0.031131   0.016775
JobInvolvement            0.029820   0.046135          0.008783   0.042438
JobLevel                  0.509604   0.002966          0.005303   0.101589
JobSatisfaction          -0.004892   0.030571         -0.003669  -0.011296
MonthlyIncome             0.497855   0.007707         -0.017014   0.094961
MonthlyRate               0.028051  -0.032182          0.027473  -0.026084
NumCompaniesWorked        0.299635   0.038153         -0.029251   0.126317
PercentSalaryHike         0.003634   0.022704          0.040235  -0.011111
PerformanceRating         0.001904   0.000473          0.027110  -0.024539
RelationshipSatisfaction  0.053535   0.007846          0.006557  -0.009118
```

```
StandardHours                  NaN        NaN               NaN        NaN
StockOptionLevel          0.037510   0.042143          0.044872   0.018422
TotalWorkingYears         0.680381   0.014515          0.004628   0.148280
TrainingTimesLastYear    -0.019621   0.002453         -0.036942  -0.025100
WorkLifeBalance          -0.021490  -0.037848         -0.026556   0.009819
YearsAtCompany            0.311309  -0.034055          0.009508   0.069114
YearsInCurrentRole        0.212901   0.009932          0.018845   0.060236
YearsSinceLastPromotion   0.216513  -0.033229          0.010029   0.054254
YearsWithCurrManager      0.202089  -0.026363          0.014406   0.069065

                          EmployeeCount  EmployeeNumber  \
Age                                 NaN       -0.010145
DailyRate                           NaN       -0.050990
DistanceFromHome                    NaN        0.032916
Education                           NaN        0.042070
EmployeeCount                       NaN             NaN
EmployeeNumber                      NaN        1.000000
EnvironmentSatisfaction             NaN        0.017621
HourlyRate                          NaN        0.035179
JobInvolvement                      NaN       -0.006888
JobLevel                            NaN       -0.018519
JobSatisfaction                     NaN       -0.046247
MonthlyIncome                       NaN       -0.014829
MonthlyRate                         NaN        0.012648
NumCompaniesWorked                  NaN       -0.001251
PercentSalaryHike                   NaN       -0.012944
PerformanceRating                   NaN       -0.020359
RelationshipSatisfaction            NaN       -0.069861
StandardHours                       NaN             NaN
StockOptionLevel                    NaN        0.062227
TotalWorkingYears                   NaN       -0.014365
TrainingTimesLastYear               NaN        0.023603
WorkLifeBalance                     NaN        0.010309
YearsAtCompany                      NaN       -0.011240
YearsInCurrentRole                  NaN       -0.008416
YearsSinceLastPromotion             NaN       -0.009019
YearsWithCurrManager                NaN       -0.009197

                          EnvironmentSatisfaction  HourlyRate  JobInvolvement  \
Age                                      0.010146    0.024287        0.029820
DailyRate                                0.018355    0.023381        0.046135
DistanceFromHome                        -0.016075    0.031131        0.008783
Education                               -0.027128    0.016775        0.042438
EmployeeCount                                 NaN         NaN             NaN
EmployeeNumber                           0.017621    0.035179       -0.006888
EnvironmentSatisfaction                  1.000000   -0.049857       -0.008278
HourlyRate                              -0.049857    1.000000        0.042861
```

```
JobInvolvement                          -0.008278   0.042861        1.000000
JobLevel                                 0.001212  -0.027853       -0.012630
JobSatisfaction                         -0.006784  -0.071335       -0.021476
MonthlyIncome                           -0.006259  -0.015794       -0.015271
MonthlyRate                              0.037600  -0.015297       -0.016322
NumCompaniesWorked                       0.012594   0.022157        0.015012
PercentSalaryHike                       -0.031701  -0.009062       -0.017205
PerformanceRating                       -0.029548  -0.002172       -0.029071
RelationshipSatisfaction                 0.007665   0.001330        0.034297
StandardHours                                 NaN        NaN             NaN
StockOptionLevel                         0.003432   0.050263        0.021523
TotalWorkingYears                       -0.002693  -0.002334       -0.005533
TrainingTimesLastYear                   -0.019359  -0.008548       -0.015338
WorkLifeBalance                          0.027627  -0.004607       -0.014617
YearsAtCompany                           0.001458  -0.019582       -0.021355
YearsInCurrentRole                       0.018007  -0.024106        0.008717
YearsSinceLastPromotion                  0.016194  -0.026716       -0.024184
YearsWithCurrManager                    -0.004999  -0.020123        0.025976

                          JobLevel  …  RelationshipSatisfaction  \
Age                       0.509604  …                  0.053535
DailyRate                 0.002966  …                  0.007846
DistanceFromHome          0.005303  …                  0.006557
Education                 0.101589  …                 -0.009118
EmployeeCount                  NaN  …                       NaN
EmployeeNumber           -0.018519  …                 -0.069861
EnvironmentSatisfaction   0.001212  …                  0.007665
HourlyRate               -0.027853  …                  0.001330
JobInvolvement           -0.012630  …                  0.034297
JobLevel                  1.000000  …                  0.021642
JobSatisfaction          -0.001944  …                 -0.012454
MonthlyIncome             0.950300  …                  0.025873
MonthlyRate               0.039563  …                 -0.004085
NumCompaniesWorked        0.142501  …                  0.052733
PercentSalaryHike        -0.034730  …                 -0.040490
PerformanceRating        -0.021222  …                 -0.031351
RelationshipSatisfaction  0.021642  …                  1.000000
StandardHours                  NaN  …                       NaN
StockOptionLevel          0.013984  …                 -0.045952
TotalWorkingYears         0.782208  …                  0.024054
TrainingTimesLastYear    -0.018191  …                  0.002497
WorkLifeBalance           0.037818  …                  0.019604
YearsAtCompany            0.534739  …                  0.019367
YearsInCurrentRole        0.389447  …                 -0.015123
YearsSinceLastPromotion   0.353885  …                  0.033493
YearsWithCurrManager      0.375281  …                 -0.000867
```

```
                          StandardHours  StockOptionLevel  TotalWorkingYears  \
Age                                 NaN          0.037510           0.680381
DailyRate                           NaN          0.042143           0.014515
DistanceFromHome                    NaN          0.044872           0.004628
Education                           NaN          0.018422           0.148280
EmployeeCount                       NaN               NaN                NaN
EmployeeNumber                      NaN          0.062227          -0.014365
EnvironmentSatisfaction             NaN          0.003432          -0.002693
HourlyRate                          NaN          0.050263          -0.002334
JobInvolvement                      NaN          0.021523          -0.005533
JobLevel                            NaN          0.013984           0.782208
JobSatisfaction                     NaN          0.010690          -0.020185
MonthlyIncome                       NaN          0.005408           0.772893
MonthlyRate                         NaN         -0.034323           0.026442
NumCompaniesWorked                  NaN          0.030075           0.237639
PercentSalaryHike                   NaN          0.007528          -0.020608
PerformanceRating                   NaN          0.003506           0.006744
RelationshipSatisfaction            NaN         -0.045952           0.024054
StandardHours                       NaN               NaN                NaN
StockOptionLevel                    NaN          1.000000           0.010136
TotalWorkingYears                   NaN          0.010136           1.000000
TrainingTimesLastYear               NaN          0.011274          -0.035662
WorkLifeBalance                     NaN          0.004129           0.001008
YearsAtCompany                      NaN          0.015058           0.628133
YearsInCurrentRole                  NaN          0.050818           0.460365
YearsSinceLastPromotion             NaN          0.014352           0.404858
YearsWithCurrManager                NaN          0.024698           0.459188

                          TrainingTimesLastYear  WorkLifeBalance  \
Age                                   -0.019621        -0.021490
DailyRate                              0.002453        -0.037848
DistanceFromHome                      -0.036942        -0.026556
Education                             -0.025100         0.009819
EmployeeCount                               NaN              NaN
EmployeeNumber                         0.023603         0.010309
EnvironmentSatisfaction               -0.019359         0.027627
HourlyRate                            -0.008548        -0.004607
JobInvolvement                        -0.015338        -0.014617
JobLevel                              -0.018191         0.037818
JobSatisfaction                       -0.005779        -0.019459
MonthlyIncome                         -0.021736         0.030683
MonthlyRate                            0.001467         0.007963
NumCompaniesWorked                    -0.066054        -0.008366
PercentSalaryHike                     -0.005221        -0.003280
PerformanceRating                     -0.015579         0.002572
RelationshipSatisfaction               0.002497         0.019604
StandardHours                               NaN              NaN
```

```
StockOptionLevel                       0.011274         0.004129
TotalWorkingYears                     -0.035662         0.001008
TrainingTimesLastYear                  1.000000         0.028072
WorkLifeBalance                        0.028072         1.000000
YearsAtCompany                         0.003569         0.012089
YearsInCurrentRole                    -0.005738         0.049856
YearsSinceLastPromotion               -0.002067         0.008941
YearsWithCurrManager                  -0.004096         0.002759

                          YearsAtCompany  YearsInCurrentRole  \
Age                             0.311309            0.212901
DailyRate                      -0.034055            0.009932
DistanceFromHome                0.009508            0.018845
Education                       0.069114            0.060236
EmployeeCount                        NaN                 NaN
EmployeeNumber                 -0.011240           -0.008416
EnvironmentSatisfaction         0.001458            0.018007
HourlyRate                     -0.019582           -0.024106
JobInvolvement                 -0.021355            0.008717
JobLevel                        0.534739            0.389447
JobSatisfaction                -0.003803           -0.002305
MonthlyIncome                   0.514285            0.363818
MonthlyRate                    -0.023655           -0.012815
NumCompaniesWorked             -0.118421           -0.090754
PercentSalaryHike              -0.035991           -0.001520
PerformanceRating               0.003435            0.034986
RelationshipSatisfaction        0.019367           -0.015123
StandardHours                        NaN                 NaN
StockOptionLevel                0.015058            0.050818
TotalWorkingYears               0.628133            0.460365
TrainingTimesLastYear           0.003569           -0.005738
WorkLifeBalance                 0.012089            0.049856
YearsAtCompany                  1.000000            0.758754
YearsInCurrentRole              0.758754            1.000000
YearsSinceLastPromotion         0.618409            0.548056
YearsWithCurrManager            0.769212            0.714365

                          YearsSinceLastPromotion  YearsWithCurrManager
Age                                      0.216513              0.202089
DailyRate                               -0.033229             -0.026363
DistanceFromHome                         0.010029              0.014406
Education                                0.054254              0.069065
EmployeeCount                                 NaN                   NaN
EmployeeNumber                          -0.009019             -0.009197
EnvironmentSatisfaction                  0.016194             -0.004999
HourlyRate                              -0.026716             -0.020123
JobInvolvement                          -0.024184              0.025976
```

```
JobLevel                                 0.353885              0.375281
JobSatisfaction                         -0.018214             -0.027656
MonthlyIncome                            0.344978              0.344079
MonthlyRate                              0.001567             -0.036746
NumCompaniesWorked                      -0.036814             -0.110319
PercentSalaryHike                       -0.022154             -0.011985
PerformanceRating                        0.017896              0.022827
RelationshipSatisfaction                 0.033493             -0.000867
StandardHours                                 NaN                   NaN
StockOptionLevel                         0.014352              0.024698
TotalWorkingYears                        0.404858              0.459188
TrainingTimesLastYear                   -0.002067             -0.004096
WorkLifeBalance                          0.008941              0.002759
YearsAtCompany                           0.618409              0.769212
YearsInCurrentRole                       0.548056              0.714365
YearsSinceLastPromotion                  1.000000              0.510224
YearsWithCurrManager                     0.510224              1.000000

[26 rows x 26 columns]
```

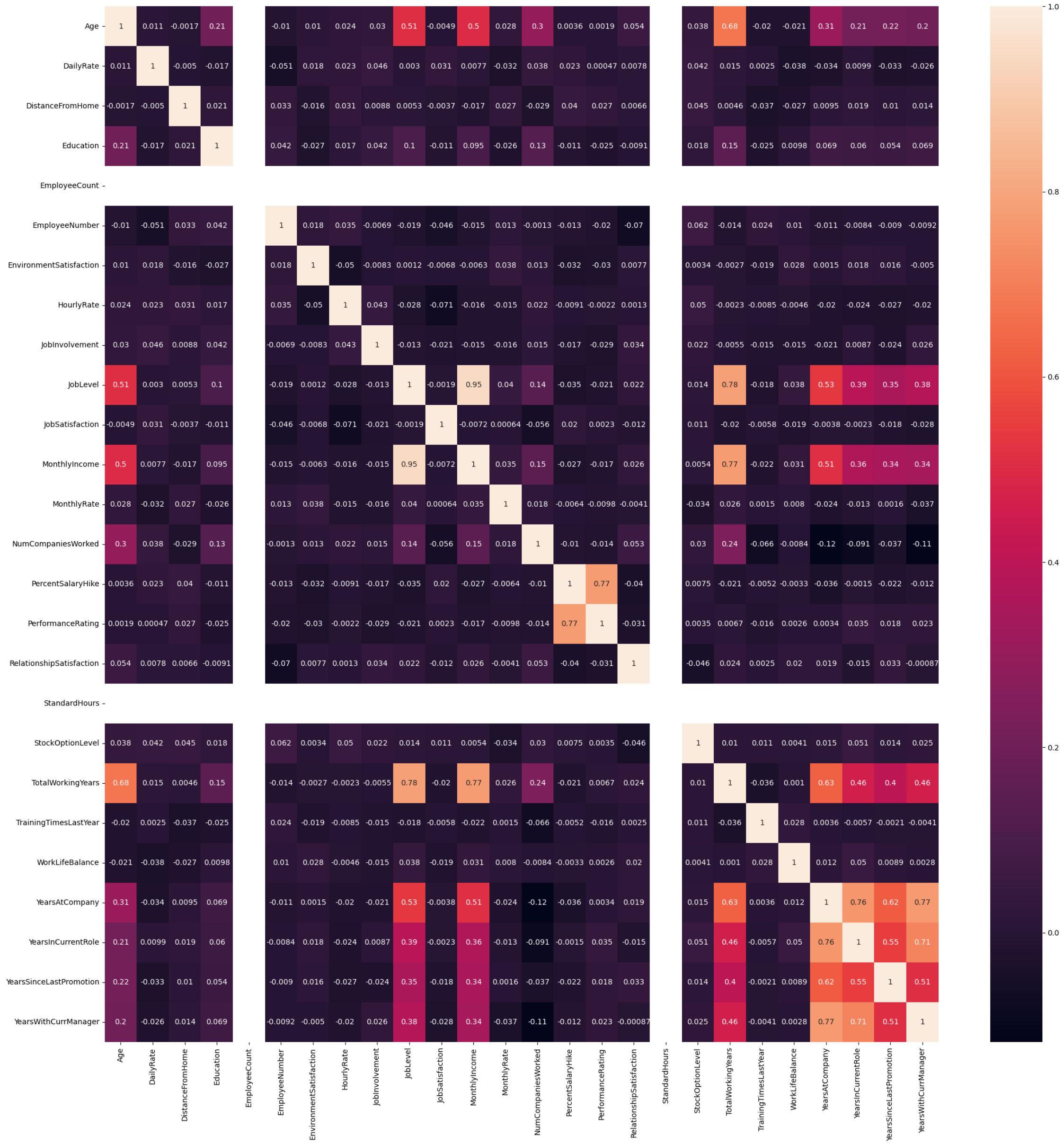
[12]:
```
df.head()
```

[12]:
```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
2  …                         2            80                 0
3  …                         3            80                 0
4  …                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
```

```
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2

[5 rows x 35 columns]
```

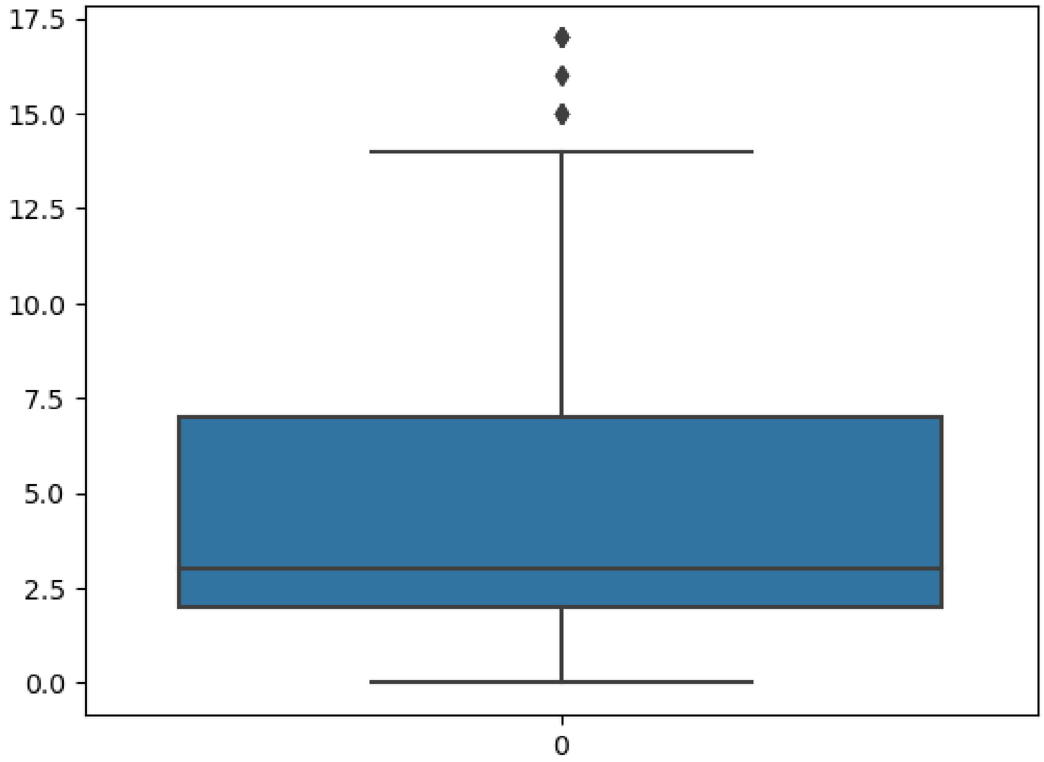
[13]:
```
plt.subplots(figsize = (25,25))
sns.heatmap(num_df.corr(),annot=True)
```

[13]:
```
<Axes: >
```



```
[14]: sns.boxplot(df.YearsWithCurrManager)
```

```
[14]: <Axes: >
```



[15]:
```
from scipy import stats
z_scores = np.abs(stats.zscore(df['YearsWithCurrManager']))
max_threshold=3
outliers = df['YearsWithCurrManager'][z_scores > max_threshold]

# Print and visualize the outliers
print("Outliers detected using Z-Score:")
print(outliers)
```

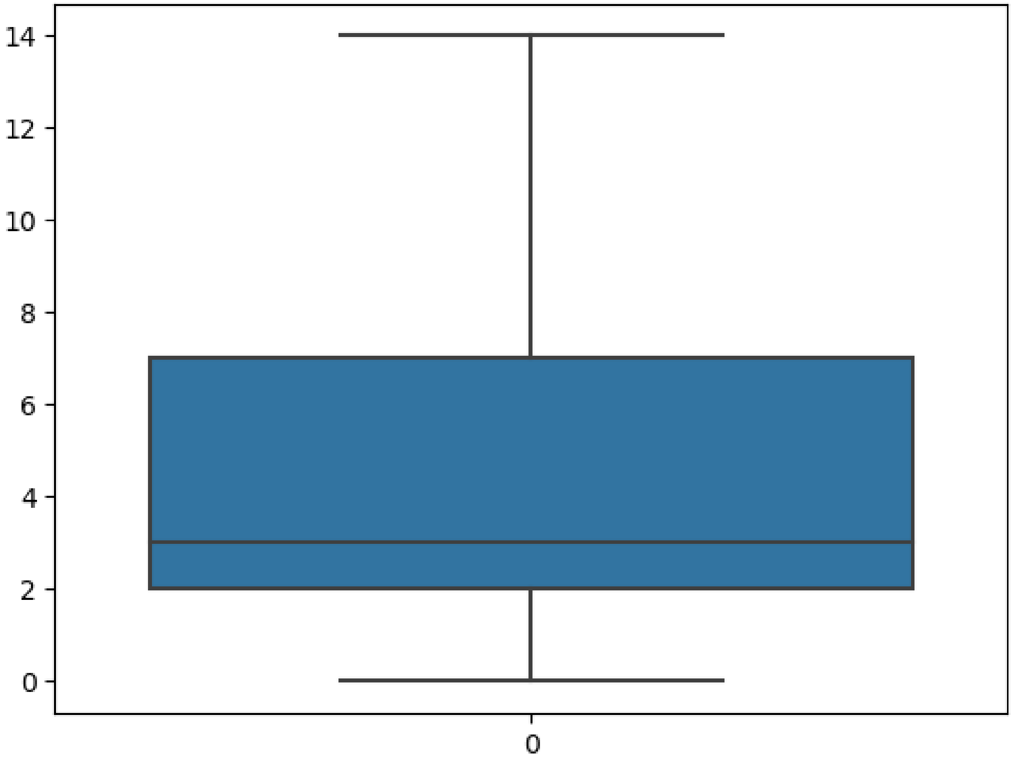
```
Outliers detected using Z-Score:
28      17
123     15
153     15
187     15
231     15
386     17
561     16
616     17
635     15
686     17
875     17
926     17
```

```
1078    17
1348    16
Name: YearsWithCurrManager, dtype: int64
```

```
[16]: q1 = df.YearsWithCurrManager.quantile(0.25)
q3 = df.YearsWithCurrManager.quantile(0.75)
print(q1)
print(q3)
upperlimit = q3+1.5*(q3-q1)
upperlimit
lowerlimit = q1-1.5*(q3-q1)
lowerlimit
num_df.median()
df["YearsWithCurrManager"]=np.
 ↪where(df["YearsWithCurrManager"]>upperlimit,14,df['YearsWithCurrManager'])
sns.boxplot(df.YearsWithCurrManager)
```

```
2.0
7.0
```

```
[16]: <Axes: >
```



[17]:

```
from scipy import stats
z_scores = np.abs(stats.zscore(df['YearsWithCurrManager']))
max_threshold=3
outliers = df['YearsWithCurrManager'][z_scores > max_threshold]

# Print and visualize the outliers
print("Outliers detected using Z-Score:")
print(outliers)
```

```
Outliers detected using Z-Score:
Series([], Name: YearsWithCurrManager, dtype: int64)
```

[18]:

```
df.head()
```

[18]:

```
   Age Attrition     BusinessTravel  DailyRate              Department  \
0   41       Yes      Travel_Rarely       1102                   Sales
1   49        No  Travel_Frequently        279  Research & Development
2   37       Yes      Travel_Rarely       1373  Research & Development
3   33        No  Travel_Frequently       1392  Research & Development
4   27        No      Travel_Rarely        591  Research & Development

   DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
0                 1          2  Life Sciences              1               1
1                 8          1  Life Sciences              1               2
2                 2          2          Other              1               4
3                 3          4  Life Sciences              1               5
4                 2          1        Medical              1               7

   …  RelationshipSatisfaction StandardHours  StockOptionLevel  \
0  …                         1            80                 0
1  …                         4            80                 1
2  …                         2            80                 0
3  …                         3            80                 0
4  …                         4            80                 1

   TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  YearsAtCompany  \
0                  8                      0               1               6
1                 10                      3               3              10
2                  7                      3               3               0
3                  8                      3               3               8
4                  6                      3               3               2

  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                  4                        0                     5
1                  7                        1                     7
2                  0                        0                     0
3                  7                        3                     0
4                  2                        2                     2
```

```
[5 rows x 35 columns]
```

```
[19]: x=df.drop('Attrition',axis=1)
      x.head()
```

```
[19]:    Age     BusinessTravel  DailyRate              Department  \
      0   41      Travel_Rarely       1102                   Sales
      1   49  Travel_Frequently        279  Research & Development
      2   37      Travel_Rarely       1373  Research & Development
      3   33  Travel_Frequently       1392  Research & Development
      4   27      Travel_Rarely        591  Research & Development

         DistanceFromHome  Education EducationField  EmployeeCount  EmployeeNumber  \
      0                 1          2  Life Sciences              1               1
      1                 8          1  Life Sciences              1               2
      2                 2          2          Other              1               4
      3                 3          4  Life Sciences              1               5
      4                 2          1        Medical              1               7

         EnvironmentSatisfaction  … RelationshipSatisfaction  StandardHours  \
      0                        2  …                        1             80
      1                        3  …                        4             80
      2                        4  …                        2             80
      3                        4  …                        3             80
      4                        1  …                        4             80

         StockOptionLevel  TotalWorkingYears TrainingTimesLastYear  WorkLifeBalance  \
      0                 0                  8                     0                1
      1                 1                 10                     3                3
      2                 0                  7                     3                3
      3                 0                  8                     3                3
      4                 1                  6                     3                3

        YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  \
      0              6                   4                        0
      1             10                   7                        1
      2              0                   0                        0
      3              8                   7                        3
      4              2                   2                        2

         YearsWithCurrManager
      0                     5
      1                     7
      2                     0
      3                     0
      4                     2
```

```
[5 rows x 34 columns]
```

[20]:
```
y=df.Attrition
y.head()
```

[20]:
```
0    Yes
1     No
2    Yes
3     No
4     No
Name: Attrition, dtype: object
```

[21]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x.BusinessTravel        =le.fit_transform(x.BusinessTravel        )
x.head()
x.Department        =le.fit_transform(x.Department        )
x.head()
x.EducationField        =le.fit_transform(x.EducationField        )
x.head()
x.Gender=le.fit_transform(x.Gender)
x.head()
x.JobRole        =le.fit_transform(x.JobRole        )
x.head()
x.MaritalStatus        =le.fit_transform(x.MaritalStatus        )
x.head()
x.Over18        =le.fit_transform(x.Over18        )
x.head()
x.OverTime        =le.fit_transform(x.OverTime        )
x.head()
```

[21]:
```
   Age  BusinessTravel  DailyRate  Department  DistanceFromHome  Education  \
0   41               2       1102           2                 1          2
1   49               1        279           1                 8          1
2   37               2       1373           1                 2          2
3   33               1       1392           1                 3          4
4   27               2        591           1                 2          1

   EducationField  EmployeeCount  EmployeeNumber  EnvironmentSatisfaction  \
0               1              1               1                        2
1               1              1               2                        3
2               4              1               4                        4
3               1              1               5                        4
4               3              1               7                        1
```

```
   …  RelationshipSatisfaction  StandardHours  StockOptionLevel  \
0  …                         1             80                 0
1  …                         4             80                 1
2  …                         2             80                 0
3  …                         3             80                 0
4  …                         4             80                 1

   TotalWorkingYears  TrainingTimesLastYear  WorkLifeBalance  YearsAtCompany  \
0                  8                      0                1               6
1                 10                      3                3              10
2                  7                      3                3               0
3                  8                      3                3               8
4                  6                      3                3               2

   YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager
0                   4                        0                     5
1                   7                        1                     7
2                   0                        0                     0
3                   7                        3                     0
4                   2                        2                     2

[5 rows x 34 columns]
```

[22]:

```
df.columns
```

[22]:

```
Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
       'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
       'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
       'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
       'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
       'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
       'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
       'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
       'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
       'YearsWithCurrManager'],
      dtype='object')
```

[23]:

```
#feature scaling
from sklearn.preprocessing import MinMaxScaler
ms=MinMaxScaler()
x_scaled=pd.DataFrame(ms.fit_transform(x),columns=x.columns)
```

[24]:

```
x_scaled
```

[24]:

```
           Age  BusinessTravel  DailyRate  Department  DistanceFromHome  \
0     0.547619             1.0   0.715820         1.0          0.000000
1     0.738095             0.5   0.126700         0.5          0.250000
```

```
2     0.452381             1.0   0.909807         0.5          0.035714
3     0.357143             0.5   0.923407         0.5          0.071429
4     0.214286             1.0   0.350036         0.5          0.035714
…        …             …        …         …               …
1465  0.428571             0.5   0.559771         0.5          0.785714
1466  0.500000             1.0   0.365784         0.5          0.178571
1467  0.214286             1.0   0.037938         0.5          0.107143
1468  0.738095             0.5   0.659270         1.0          0.035714
1469  0.380952             1.0   0.376521         0.5          0.250000

      Education  EducationField  EmployeeCount  EmployeeNumber  \
0          0.25             0.2            0.0        0.000000
1          0.00             0.2            0.0        0.000484
2          0.25             0.8            0.0        0.001451
3          0.75             0.2            0.0        0.001935
4          0.00             0.6            0.0        0.002903
…         …             …            …             …
1465       0.25             0.6            0.0        0.996613
1466       0.00             0.6            0.0        0.997097
1467       0.50             0.2            0.0        0.998065
1468       0.50             0.6            0.0        0.998549
1469       0.50             0.6            0.0        1.000000

      EnvironmentSatisfaction  …  RelationshipSatisfaction  StandardHours  \
0                    0.333333  …                  0.000000            0.0
1                    0.666667  …                  1.000000            0.0
2                    1.000000  …                  0.333333            0.0
3                    1.000000  …                  0.666667            0.0
4                    0.000000  …                  1.000000            0.0
…                       …  …                       …            …
1465                 0.666667  …                  0.666667            0.0
1466                 1.000000  …                  0.000000            0.0
1467                 0.333333  …                  0.333333            0.0
1468                 1.000000  …                  1.000000            0.0
1469                 0.333333  …                  0.000000            0.0

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
0             0.000000              0.200               0.000000
1             0.333333              0.250               0.500000
2             0.000000              0.175               0.500000
3             0.000000              0.200               0.500000
4             0.333333              0.150               0.500000
…                …                …                    …
1465          0.333333              0.425               0.500000
1466          0.333333              0.225               0.833333
1467          0.333333              0.150               0.000000
1468          0.000000              0.425               0.500000
```

```
1469           0.000000            0.150            0.500000

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
0            0.000000           0.150            0.222222
1            0.666667           0.250            0.388889
2            0.666667           0.000            0.000000
3            0.666667           0.200            0.388889
4            0.666667           0.050            0.111111
…                 …             …                  …
1465         0.666667           0.125            0.111111
1466         0.666667           0.175            0.388889
1467         0.666667           0.150            0.111111
1468         0.333333           0.225            0.333333
1469         1.000000           0.100            0.166667

      YearsSinceLastPromotion  YearsWithCurrManager
0                    0.000000              0.357143
1                    0.066667              0.500000
2                    0.000000              0.000000
3                    0.200000              0.000000
4                    0.133333              0.142857
…                        …                     …
1465                 0.000000              0.214286
1466                 0.066667              0.500000
1467                 0.000000              0.214286
1468                 0.000000              0.571429
1469                 0.066667              0.142857

[1470 rows x 34 columns]
```

[25]:
```
#Splitting Data into Train and Test.
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_scaled,y,test_size=0.
 ↪2,random_state=0)
```

[26]:
```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

[26]:
```
((1176, 34), (294, 34), (1176,), (294,))
```

[27]:
```
x_train.head()
```

[27]:
```
           Age  BusinessTravel  DailyRate  Department  DistanceFromHome  \
1374  0.952381             1.0   0.360057         1.0          0.714286
1092  0.642857             1.0   0.607015         0.5          0.964286
768   0.523810             1.0   0.141732         1.0          0.892857
569   0.428571             0.0   0.953472         1.0          0.250000
911   0.166667             0.5   0.355762         1.0          0.821429
```

```
      Education  EducationField  EmployeeCount  EmployeeNumber  \
1374       0.50             0.2            0.0        0.937107
1092       0.50             1.0            0.0        0.747460
768        0.50             0.4            0.0        0.515239
569        0.75             0.2            0.0        0.381229
911        0.00             0.2            0.0        0.615385

      EnvironmentSatisfaction  …  RelationshipSatisfaction  StandardHours  \
1374                 1.000000  …                  0.666667            0.0
1092                 1.000000  …                  1.000000            0.0
768                  0.666667  …                  0.333333            0.0
569                  0.000000  …                  0.333333            0.0
911                  0.666667  …                  1.000000            0.0

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
1374          0.333333              0.725               0.333333
1092          0.333333              0.200               0.500000
768           0.333333              0.200               0.500000
569           0.000000              0.250               0.166667
911           0.000000              0.025               0.666667

      WorkLifeBalance  YearsAtCompany  YearsInCurrentRole  \
1374         0.333333           0.025            0.000000
1092         0.666667           0.125            0.222222
768          0.333333           0.175            0.388889
569          0.666667           0.250            0.388889
911          0.666667           0.025            0.000000

      YearsSinceLastPromotion  YearsWithCurrManager
1374                 0.000000              0.000000
1092                 0.000000              0.214286
768                  0.466667              0.357143
569                  0.000000              0.642857
911                  0.066667              0.000000

[5 rows x 34 columns]
```

[28]:

```
from sklearn.linear_model import LogisticRegression
model=LogisticRegression()
```

[29]:

```
model.fit(x_train,y_train)
pred=model.predict(x_test)
pred
```

[29]:

```
array(['No', 'No', 'No', 'No', 'Yes', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'Yes', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No'],
      dtype=object)
```

[30]:

```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y=le.fit_transform(y)
```

[31]:

```
y_test
```

[31]:

```
442      No
1091     No
981     Yes
785      No
1332    Yes
       …
1439     No
481      No
124     Yes
198      No
1229     No
Name: Attrition, Length: 294, dtype: object
```

```
[32]: df
```

```
[32]:       Age Attrition     BusinessTravel  DailyRate              Department  \
      0       41       Yes      Travel_Rarely       1102                   Sales
      1       49        No  Travel_Frequently        279  Research & Development
      2       37       Yes      Travel_Rarely       1373  Research & Development
      3       33        No  Travel_Frequently       1392  Research & Development
      4       27        No      Travel_Rarely        591  Research & Development
      …   …       …                 …        …                     …
      1465   36        No  Travel_Frequently        884  Research & Development
      1466   39        No      Travel_Rarely        613  Research & Development
      1467   27        No      Travel_Rarely        155  Research & Development
      1468   49        No  Travel_Frequently       1023                   Sales
      1469   34        No      Travel_Rarely        628  Research & Development

            DistanceFromHome  Education EducationField  EmployeeCount  \
      0                    1          2  Life Sciences              1
      1                    8          1  Life Sciences              1
      2                    2          2          Other              1
      3                    3          4  Life Sciences              1
      4                    2          1        Medical              1
      …                …        …            …            …
      1465                23          2        Medical              1
      1466                 6          1        Medical              1
      1467                 4          3  Life Sciences              1
      1468                 2          3        Medical              1
      1469                 8          3        Medical              1

            EmployeeNumber  …  RelationshipSatisfaction StandardHours  \
      0                  1  …                         1            80
      1                  2  …                         4            80
      2                  4  …                         2            80
      3                  5  …                         3            80
      4                  7  …                         4            80
      …              …  …                       …           …
      1465            2061  …                         3            80
      1466            2062  …                         1            80
      1467            2064  …                         2            80
      1468            2065  …                         4            80
      1469            2068  …                         1            80

            StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear  \
      0                    0                  8                      0
      1                    1                 10                      3
      2                    0                  7                      3
      3                    0                  8                      3
      4                    1                  6                      3
```

```
…                …                …                    …
1465                 1                 17                      3
1466                 1                  9                      5
1467                 1                  6                      0
1468                 0                 17                      3
1469                 0                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole  \
0                  1               6                  4
1                  3              10                  7
2                  3               0                  0
3                  3               8                  7
4                  3               2                  2
…              …             …                …
1465               3               5                  2
1466               3               7                  7
1467               3               6                  2
1468               2               9                  6
1469               4               4                  3

      YearsSinceLastPromotion  YearsWithCurrManager
0                           0                     5
1                           1                     7
2                           0                     0
3                           3                     0
4                           2                     2
…                       …                   …
1465                        0                     3
1466                        1                     7
1467                        0                     3
1468                        0                     8
1469                        1                     2

[1470 rows x 35 columns]
```

2 Evaluation of classification model

[33]:

```
#Accuracy score
from sklearn.metrics import␣
 ↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[34]:

```
accuracy_score(y_test,pred)
```

[34]:

```
0.8843537414965986
```

[35]:

```
confusion_matrix(y_test,pred)
```

```
[35]: array([[242,   3],
             [ 31,  18]], dtype=int64)
```

```
[36]: pd.crosstab(y_test,pred)
```

```
[36]: col_0       No  Yes
      Attrition
      No         242    3
      Yes         31   18
```

2.0.1 Roc-AUC curve

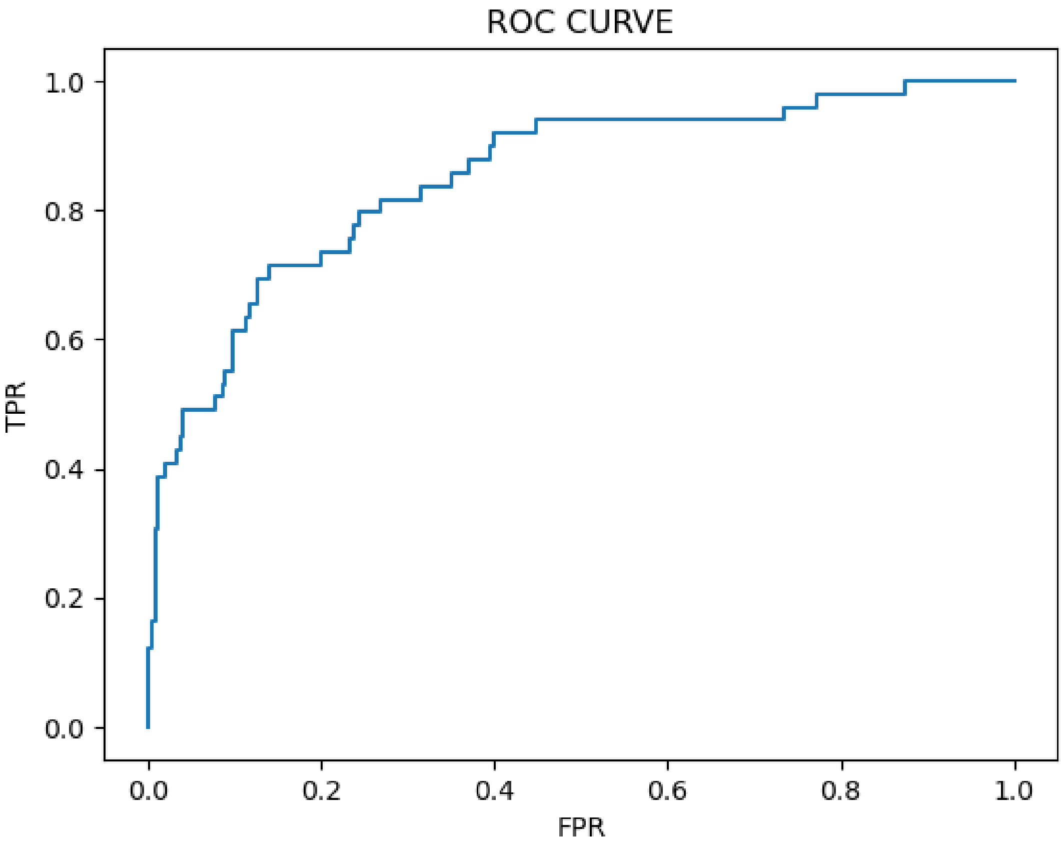
```
[37]: probability=model.predict_proba(x_test)[:,1]
      probability
```

```
[37]: array([0.15843867, 0.20617997, 0.31691729, 0.09672152, 0.63876647,
             0.06205401, 0.61414184, 0.07466397, 0.00797252, 0.39157785,
             0.05281564, 0.33160211, 0.02022395, 0.6671328 , 0.19419683,
             0.0335299 , 0.10954936, 0.17130578, 0.043804  , 0.2241511 ,
             0.23531373, 0.01475346, 0.06562592, 0.05019163, 0.59115162,
             0.44667993, 0.07401303, 0.0449937 , 0.67637047, 0.05859033,
             0.01545736, 0.03386798, 0.07021403, 0.1707141 , 0.07767295,
             0.04154894, 0.08312937, 0.06997437, 0.03567429, 0.05269126,
             0.05742727, 0.02144976, 0.01779053, 0.01301572, 0.02825292,
             0.50162054, 0.41541766, 0.00299378, 0.74315718, 0.51799699,
             0.09708281, 0.48942319, 0.07941138, 0.25720931, 0.66861063,
             0.26482373, 0.01970983, 0.30281497, 0.02858501, 0.16213966,
             0.02040161, 0.2173984 , 0.13768821, 0.03568054, 0.37558052,
             0.03010741, 0.29718154, 0.15832399, 0.10264349, 0.08700774,
             0.0815183 , 0.30943969, 0.08708969, 0.07442596, 0.12300414,
             0.0618342 , 0.04633075, 0.07672219, 0.19834226, 0.03129952,
             0.00857215, 0.02394842, 0.13606932, 0.02587787, 0.03217004,
             0.0821409 , 0.00518749, 0.035308  , 0.03813342, 0.14270872,
             0.26418695, 0.16461435, 0.27401734, 0.24146954, 0.02119787,
             0.17774284, 0.34102562, 0.28338745, 0.06906981, 0.04948532,
             0.24465264, 0.74929682, 0.35691434, 0.01878265, 0.08772637,
             0.03239915, 0.05413857, 0.15215059, 0.07127406, 0.13828798,
             0.09342465, 0.04693869, 0.02494493, 0.15041914, 0.07133392,
             0.03025642, 0.05306455, 0.1165452 , 0.00872431, 0.01229042,
             0.17575238, 0.05005249, 0.09018395, 0.82857166, 0.03066995,
             0.0228189 , 0.00874605, 0.13496234, 0.16593413, 0.05060052,
             0.01520085, 0.29791945, 0.54919611, 0.33581407, 0.0469494 ,
             0.38773566, 0.61348127, 0.14171081, 0.07455884, 0.2409655 ,
             0.09528764, 0.06730943, 0.09797576, 0.20026612, 0.20053142,
             0.03046036, 0.14877431, 0.0036571 , 0.11146887, 0.15912883,
             0.06017571, 0.17964687, 0.06063618, 0.1199213 , 0.03284092,
             0.02688355, 0.06536903, 0.08335812, 0.01464284, 0.01536292,
             0.37701597, 0.01262506, 0.15004068, 0.80530948, 0.11655522,
             0.28461049, 0.17042029, 0.15392139, 0.02756879, 0.00599553,
             0.04142216, 0.09958411, 0.11567269, 0.10448555, 0.01830036,
             0.1444171 , 0.1048541 , 0.10079777, 0.05099176, 0.09183576,
             0.02893646, 0.09754427, 0.00516687, 0.75206394, 0.04227453,
             0.04018918, 0.37563319, 0.04457964, 0.72551665, 0.10583031,
             0.36656526, 0.38293703, 0.32923777, 0.05248015, 0.08216713,
             0.13748888, 0.04309097, 0.01429957, 0.2656631 , 0.06297408,
             0.16075744, 0.15388494, 0.67190498, 0.05834473, 0.28467369,
             0.04694404, 0.46237195, 0.00339026, 0.13927388, 0.02695884,
             0.12707414, 0.17395277, 0.0750947 , 0.10135673, 0.16496216,
             0.02583798, 0.01790826, 0.08850395, 0.02838351, 0.13795992,
             0.08655223, 0.22164621, 0.73379009, 0.17294814, 0.40907888,
             0.01503347, 0.11411826, 0.21412683, 0.32566668, 0.03366086,
             0.04472831, 0.32127248, 0.05442236, 0.0242917 , 0.16228044,
             0.32858438, 0.22879119, 0.00852736, 0.0798162 , 0.01140248,
             0.14102568, 0.29116266, 0.01282151, 0.17118076, 0.04051376,
             0.04165738, 0.42684273, 0.35009936, 0.0366853 , 0.11692325,
             0.37940034, 0.31562415, 0.79587005, 0.05488792, 0.21568794,
             0.06397987, 0.00569145, 0.66085682, 0.35796045, 0.37592133,
             0.3650533 , 0.03568965, 0.21192376, 0.05892118, 0.06428028,
             0.10143977, 0.00796354, 0.2678938 , 0.4288445 , 0.0652538 ,
             0.09309022, 0.01226927, 0.14314823, 0.04989664, 0.02304292,
             0.02508766, 0.06618985, 0.24272596, 0.26663754, 0.1979951 ,
             0.26504226, 0.01648205, 0.15826843, 0.08519882, 0.02669729,
             0.18757572, 0.00768502, 0.27928747, 0.0027473 , 0.02506718,
             0.22608608, 0.72428674, 0.07739605, 0.26575953])
```

[38]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y_test=le.fit_transform(y_test)
```

[39]:
```
 # roc_curve
fpr,tpr,threshsholds = roc_curve(y_test,probability)
```

[40]:
```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show()
```



2.0.2 DecisionTreeClassifier

[41]:
```
from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()
```

[42]:
```
dtc.fit(x_train,y_train)
```

[42]:
```
DecisionTreeClassifier()
```

[43]:
```
pred=dtc.predict(x_test)
```

[44]:
```
pred
```

[44]:
```
array(['No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'No',
       'No', 'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes',
       'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'Yes', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No',
       'No', 'No', 'No', 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'No', 'No',
       'No', 'Yes', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'Yes', 'Yes',
       'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No',
       'Yes', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'Yes', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No'], dtype=object)
```

[45]:

```
y_test
```

[45]:

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

[46]:

```
df
```

[46]:

```
      Age Attrition     BusinessTravel  DailyRate              Department  \
0      41       Yes      Travel_Rarely       1102                   Sales
1      49        No  Travel_Frequently        279  Research & Development
2      37       Yes      Travel_Rarely       1373  Research & Development
3      33        No  Travel_Frequently       1392  Research & Development
4      27        No      Travel_Rarely        591  Research & Development
```

```
…   …       …                …        …                     …
1465   36        No  Travel_Frequently        884  Research & Development
1466   39        No      Travel_Rarely        613  Research & Development
1467   27        No      Travel_Rarely        155  Research & Development
1468   49        No  Travel_Frequently       1023                   Sales
1469   34        No      Travel_Rarely        628  Research & Development

      DistanceFromHome  Education EducationField  EmployeeCount   \
0                    1          2  Life Sciences              1
1                    8          1  Life Sciences              1
2                    2          2          Other              1
3                    3          4  Life Sciences              1
4                    2          1        Medical              1
…                …        …            …            …
1465                23          2        Medical              1
1466                 6          1        Medical              1
1467                 4          3  Life Sciences              1
1468                 2          3        Medical              1
1469                 8          3        Medical              1

      EmployeeNumber  …  RelationshipSatisfaction StandardHours   \
0                  1  …                         1            80
1                  2  …                         4            80
2                  4  …                         2            80
3                  5  …                         3            80
4                  7  …                         4            80
…              …  …                       …           …
1465            2061  …                         3            80
1466            2062  …                         1            80
1467            2064  …                         2            80
1468            2065  …                         4            80
1469            2068  …                         1            80

      StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear   \
0                    0                  8                      0
1                    1                 10                      3
2                    0                  7                      3
3                    0                  8                      3
4                    1                  6                      3
…                …                …                    …
1465                 1                 17                      3
1466                 1                  9                      5
1467                 1                  6                      0
1468                 0                 17                      3
1469                 0                  6                      3

     WorkLifeBalance  YearsAtCompany YearsInCurrentRole   \
0                    1             6                   4
1                    3            10                   7
2                    3             0                   0
3                    3             8                   7
4                    3             2                   2
…               …             …                 …
1465                 3             5                   2
1466                 3             7                   7
1467                 3             6                   2
1468                 2             9                   6
1469                 4             4                   3

      YearsSinceLastPromotion  YearsWithCurrManager
0                           0                     5
1                           1                     7
2                           0                     0
3                           3                     0
4                           2                     2
…                       …                   …
1465                        0                     3
1466                        1                     7
1467                        0                     3
1468                        0                     8
1469                        1                     2

[1470 rows x 35 columns]
```

3 Evaluation of classification model

[47]:
```
#Accuracy score
from sklearn.metrics import␣
 ↪accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

[48]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y=le.fit_transform(y)
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
pred=le.fit_transform(pred)
```

[49]:
```
y_test
```

[49]:
```
array([0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 1, 0, 0])
```

```
[50]: accuracy_score(y_test,pred)
```

```
[50]: 0.7551020408163265
```

```
[51]: confusion_matrix(y_test,pred)
```

```
[51]: array([[206,  39],
             [ 33,  16]], dtype=int64)
```

```
[52]: pd.crosstab(y_test,pred)
```

```
[52]: col_0    0   1
      row_0
      0      206  39
      1       33  16
```

```
[53]: print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.86      0.84      0.85       245
           1       0.29      0.33      0.31        49

    accuracy                           0.76       294
   macro avg       0.58      0.58      0.58       294
weighted avg       0.77      0.76      0.76       294
```

3.0.1 Roc-AUC curve

```
[54]: probability=dtc.predict_proba(x_test)[:,1]
```

```
[55]: probability
```

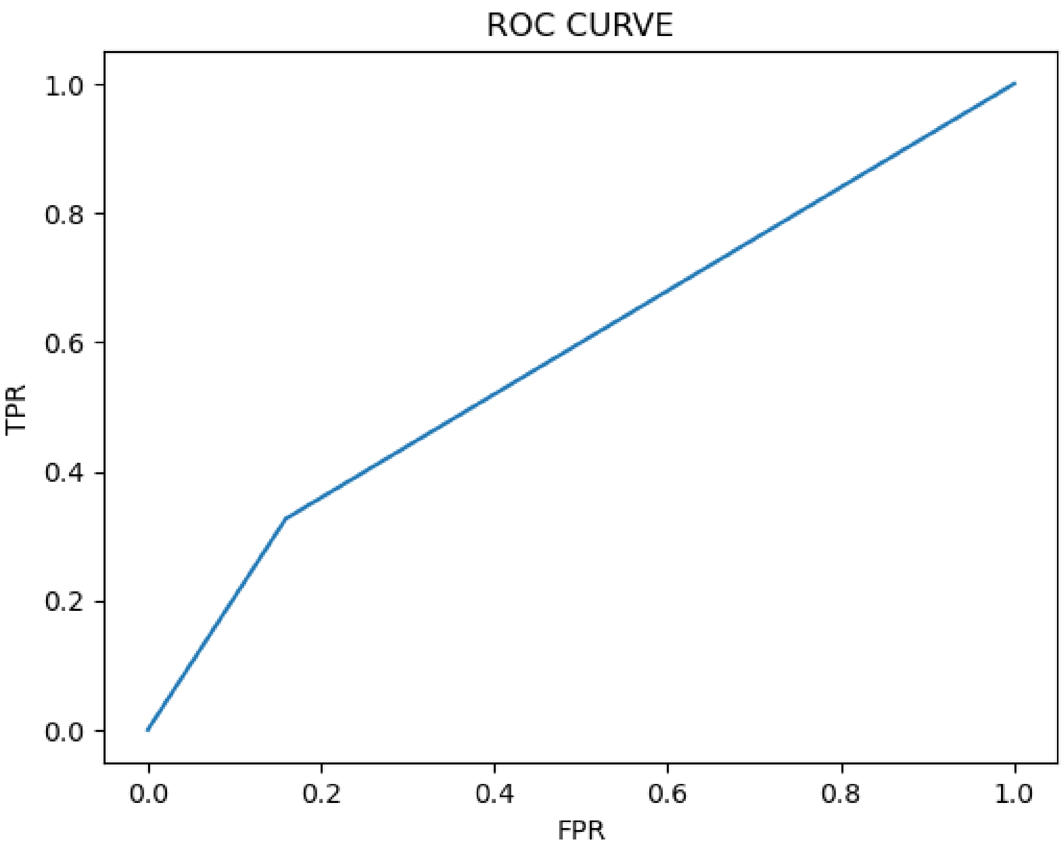
```
[55]: array([0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
             0., 0., 1., 1., 0., 0., 0., 1., 0., 0., 0., 0., 0., 1., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 1., 1.,
       1., 0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 1., 0., 1., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 1., 1., 0., 0., 0., 1., 0., 0., 1., 0.,
       0., 1., 0., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0., 1., 1., 1., 1., 0.,
       0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 1., 0., 1., 1., 0., 0., 0., 0., 1., 0., 0., 1.,
       0., 0., 1., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0.,
       1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 1., 0.,
       0., 1., 0., 0., 0., 0., 1., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0.,
       0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
       0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1., 0., 0., 0., 0.,
       0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 1., 0., 0., 0.,
       0., 0., 0., 0., 0.])
```

[56]:

```
fpr,tpr,thresholds =  roc_curve(y_test,probability)
```

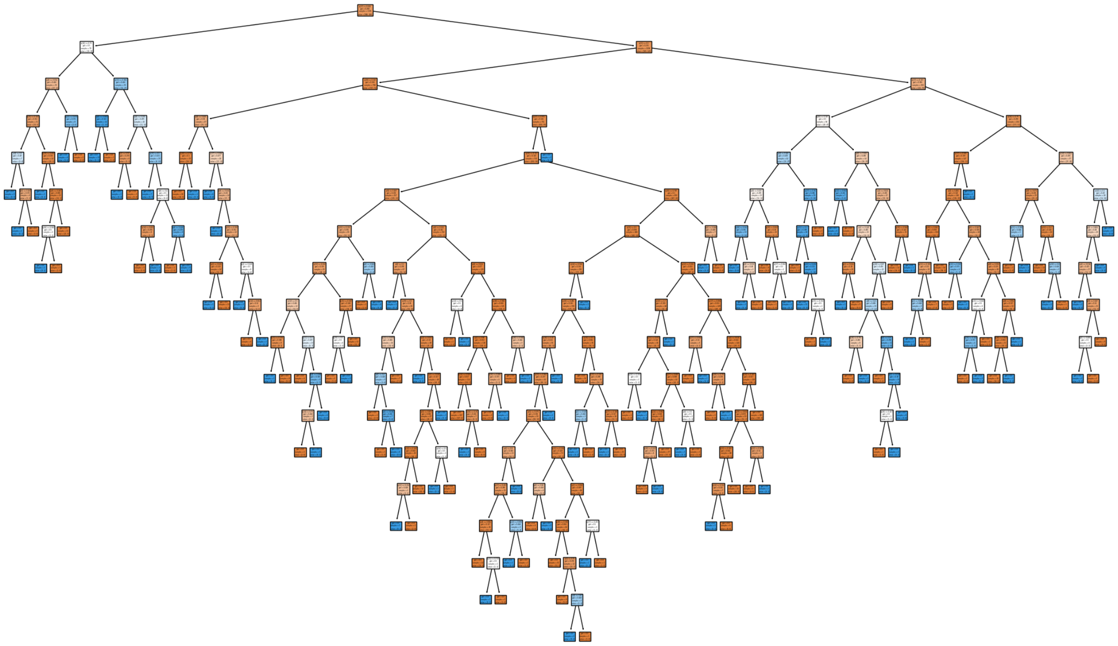
[57]:

```
plt.plot(fpr,tpr)
plt.xlabel('FPR')
plt.ylabel('TPR')
plt.title('ROC CURVE')
plt.show()
```



[58]:
```
from sklearn import tree
plt.figure(figsize=(25,15))
tree.plot_tree(dtc,filled=True)
```

[58]:
```
[Text(0.32591391509433965, 0.9722222222222222, 'x[27] <= 0.038\ngini =
0.269\nsamples = 1176\nvalue = [988, 188]'),
 Text(0.07547169811320754, 0.9166666666666666, 'x[16] <= 0.75\ngini =
0.5\nsamples = 78\nvalue = [39, 39]'),
 Text(0.044596912521440824, 0.8611111111111112, 'x[4] <= 0.554\ngini =
0.426\nsamples = 39\nvalue = [27, 12]'),
 Text(0.0274442538593482, 0.8055555555555556, 'x[15] <= 0.167\ngini =
0.312\nsamples = 31\nvalue = [25, 6]'),
 Text(0.0137221269296741, 0.75, 'x[9] <= 0.5\ngini = 0.49\nsamples = 7\nvalue =
[3, 4]'),
 Text(0.00686106346483705, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.02058319039451115, 0.6944444444444444, 'x[0] <= 0.167\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.0137221269296741, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.0274442538593482, 0.6388888888888888, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.0411663807890223, 0.75, 'x[19] <= 0.056\ngini = 0.153\nsamples =
24\nvalue = [22, 2]'),
 Text(0.03430531732418525, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.048027444253859346, 0.6944444444444444, 'x[9] <= 0.167\ngini =
0.083\nsamples = 23\nvalue = [22, 1]'),
 Text(0.0411663807890223, 0.6388888888888888, 'x[28] <= 0.583\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.03430531732418525, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.048027444253859346, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue
= [1, 0]'),
 Text(0.0548885077186964, 0.6388888888888888, 'gini = 0.0\nsamples = 21\nvalue =
[21, 0]'),
 Text(0.06174957118353345, 0.8055555555555556, 'x[22] <= 0.679\ngini =
0.375\nsamples = 8\nvalue = [2, 6]'),
 Text(0.0548885077186964, 0.75, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
 Text(0.0686106346483705, 0.75, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
 Text(0.10634648370497427, 0.8611111111111112, 'x[11] <= 0.364\ngini =
0.426\nsamples = 39\nvalue = [12, 27]'),
 Text(0.08919382504288165, 0.8055555555555556, 'x[0] <= 0.369\ngini =
0.133\nsamples = 14\nvalue = [1, 13]'),
 Text(0.0823327615780446, 0.75, 'gini = 0.0\nsamples = 13\nvalue = [0, 13]'),
 Text(0.09605488850771869, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.1234991423670669, 0.8055555555555556, 'x[8] <= 0.105\ngini =
0.493\nsamples = 25\nvalue = [11, 14]'),
 Text(0.1097770154373928, 0.75, 'x[1] <= 0.75\ngini = 0.278\nsamples = 6\nvalue
= [5, 1]'),
 Text(0.10291595197255575, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.11663807890222985, 0.6944444444444444, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
 Text(0.137221269296741, 0.75, 'x[15] <= 0.5\ngini = 0.432\nsamples = 19\nvalue
= [6, 13]'),
 Text(0.13036020583190394, 0.6944444444444444, 'gini = 0.0\nsamples = 7\nvalue =
[0, 7]'),
 Text(0.14408233276157806, 0.6944444444444444, 'x[6] <= 0.4\ngini = 0.5\nsamples
= 12\nvalue = [6, 6]'),
 Text(0.13036020583190394, 0.6388888888888888, 'x[14] <= 0.875\ngini =
0.278\nsamples = 6\nvalue = [5, 1]'),
 Text(0.1234991423670669, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
 Text(0.137221269296741, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.15780445969125215, 0.6388888888888888, 'x[8] <= 0.249\ngini =
0.278\nsamples = 6\nvalue = [1, 5]'),
 Text(0.1509433962264151, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.1646655231560892, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue =
[0, 5]'),
 Text(0.5763561320754716, 0.9166666666666666, 'x[21] <= 0.5\ngini =
0.235\nsamples = 1098\nvalue = [949, 149]'),
 Text(0.33024228130360206, 0.8611111111111112, 'x[29] <= 0.167\ngini =
0.162\nsamples = 798\nvalue = [727, 71]'),
 Text(0.1783876500857633, 0.8055555555555556, 'x[8] <= 0.445\ngini =
0.38\nsamples = 47\nvalue = [35, 12]'),
 Text(0.1646655231560892, 0.75, 'x[16] <= 0.75\ngini = 0.1\nsamples = 19\nvalue
= [18, 1]'),
 Text(0.15780445969125215, 0.6944444444444444, 'gini = 0.0\nsamples = 18\nvalue
= [18, 0]'),
 Text(0.17152658662092624, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.19210977701543738, 0.75, 'x[17] <= 0.094\ngini = 0.477\nsamples =
28\nvalue = [17, 11]'),
 Text(0.18524871355060035, 0.6944444444444444, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
 Text(0.19897084048027444, 0.6944444444444444, 'x[8] <= 0.524\ngini =
0.413\nsamples = 24\nvalue = [17, 7]'),
 Text(0.19210977701543738, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.2058319039451115, 0.6388888888888888, 'x[33] <= 0.393\ngini =
0.351\nsamples = 22\nvalue = [17, 5]'),
 Text(0.19210977701543738, 0.5833333333333334, 'x[2] <= 0.025\ngini =
0.133\nsamples = 14\nvalue = [13, 1]'),
 Text(0.18524871355060035, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.19897084048027444, 0.5277777777777778, 'gini = 0.0\nsamples = 13\nvalue
= [13, 0]'),
 Text(0.2195540308747856, 0.5833333333333334, 'x[2] <= 0.329\ngini =
0.5\nsamples = 8\nvalue = [4, 4]'),
 Text(0.21269296740994853, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.22641509433962265, 0.5277777777777778, 'x[14] <= 0.812\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(0.2195540308747856, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
 Text(0.2332761578044597, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.4820969125214408, 0.8055555555555556, 'x[27] <= 0.975\ngini =
0.145\nsamples = 751\nvalue = [692, 59]'),
 Text(0.47523584905660377, 0.75, 'x[30] <= 0.113\ngini = 0.143\nsamples =
750\nvalue = [692, 58]'),
 Text(0.34948542024013723, 0.6944444444444444, 'x[9] <= 0.167\ngini =
0.218\nsamples = 257\nvalue = [225, 32]'),
 Text(0.307032590051458, 0.6388888888888888, 'x[33] <= 0.179\ngini =
0.355\nsamples = 65\nvalue = [50, 15]'),
 Text(0.2847341337907376, 0.5833333333333334, 'x[33] <= 0.036\ngini =
0.303\nsamples = 59\nvalue = [48, 11]'),
 Text(0.2607204116638079, 0.5277777777777778, 'x[12] <= 0.5\ngini =
0.463\nsamples = 22\nvalue = [14, 8]'),
 Text(0.2469982847341338, 0.4722222222222222, 'x[11] <= 0.179\ngini =
0.198\nsamples = 9\nvalue = [8, 1]'),
 Text(0.24013722126929674, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.2538593481989708, 0.4166666666666667, 'gini = 0.0\nsamples = 8\nvalue =
[8, 0]'),
 Text(0.274442538593482, 0.4722222222222222, 'x[11] <= 0.4\ngini =
0.497\nsamples = 13\nvalue = [6, 7]'),
 Text(0.26758147512864494, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
 Text(0.28130360205831906, 0.4166666666666667, 'x[4] <= 0.286\ngini =
0.346\nsamples = 9\nvalue = [2, 7]'),
 Text(0.274442538593482, 0.3611111111111111, 'x[14] <= 0.75\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.26758147512864494, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.28130360205831906, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.2881646655231561, 0.3611111111111111, 'gini = 0.0\nsamples = 6\nvalue =
[0, 6]'),
 Text(0.30874785591766724, 0.5277777777777778, 'x[15] <= 0.167\ngini =
0.149\nsamples = 37\nvalue = [34, 3]'),
 Text(0.3018867924528302, 0.4722222222222222, 'x[29] <= 0.5\ngini = 0.5\nsamples
= 6\nvalue = [3, 3]'),
 Text(0.2950257289879931, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.30874785591766724, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.3156089193825043, 0.4722222222222222, 'gini = 0.0\nsamples = 31\nvalue =
[31, 0]'),
 Text(0.3293310463121784, 0.5833333333333334, 'x[8] <= 0.065\ngini =
0.444\nsamples = 6\nvalue = [2, 4]'),
 Text(0.32246998284734135, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.3361921097770154, 0.5277777777777778, 'gini = 0.0\nsamples = 4\nvalue =
[0, 4]'),
 Text(0.39193825042881647, 0.6388888888888888, 'x[0] <= 0.321\ngini =
0.161\nsamples = 192\nvalue = [175, 17]'),
 Text(0.3567753001715266, 0.5833333333333334, 'x[6] <= 0.1\ngini =
0.294\nsamples = 67\nvalue = [55, 12]'),
 Text(0.34991423670668953, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.36363636363636365, 0.5277777777777778, 'x[29] <= 0.5\ngini =
0.26\nsamples = 65\nvalue = [55, 10]'),
 Text(0.346483704974271, 0.4722222222222222, 'x[11] <= 0.679\ngini =
0.469\nsamples = 16\nvalue = [10, 6]'),
 Text(0.33962264150943394, 0.4166666666666667, 'x[6] <= 0.4\ngini =
0.444\nsamples = 9\nvalue = [3, 6]'),
 Text(0.33276157804459694, 0.3611111111111111, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.346483704974271, 0.3611111111111111, 'x[4] <= 0.018\ngini =
0.245\nsamples = 7\nvalue = [1, 6]'),
 Text(0.33962264150943394, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.35334476843910806, 0.3055555555555556, 'gini = 0.0\nsamples = 6\nvalue =
[0, 6]'),
 Text(0.35334476843910806, 0.4166666666666667, 'gini = 0.0\nsamples = 7\nvalue =
[7, 0]'),
 Text(0.38078902229845624, 0.4722222222222222, 'x[2] <= 0.037\ngini =
0.15\nsamples = 49\nvalue = [45, 4]'),
 Text(0.37392795883361923, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.3876500857632933, 0.4166666666666667, 'x[2] <= 0.938\ngini =
0.117\nsamples = 48\nvalue = [45, 3]'),
 Text(0.38078902229845624, 0.3611111111111111, 'x[5] <= 0.875\ngini =
0.081\nsamples = 47\nvalue = [45, 2]'),
 Text(0.3670668953687822, 0.3055555555555556, 'x[12] <= 0.167\ngini =
0.043\nsamples = 45\nvalue = [44, 1]'),
 Text(0.3602058319039451, 0.25, 'x[22] <= 0.214\ngini = 0.444\nsamples =
3\nvalue = [2, 1]'),
 Text(0.35334476843910806, 0.19444444444444445, 'gini = 0.0\nsamples = 1\nvalue
= [0, 1]'),
 Text(0.3670668953687822, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.37392795883361923, 0.25, 'gini = 0.0\nsamples = 42\nvalue = [42, 0]'),
 Text(0.39451114922813035, 0.3055555555555556, 'x[24] <= 0.5\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.3876500857632933, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.4013722126929674, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.39451114922813035, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.42710120068610635, 0.5833333333333334, 'x[8] <= 0.022\ngini =
0.077\nsamples = 125\nvalue = [120, 5]'),
 Text(0.40823327615780447, 0.5277777777777778, 'x[14] <= 0.5\ngini =
0.5\nsamples = 4\nvalue = [2, 2]'),
 Text(0.4013722126929674, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.41509433962264153, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.44596912521440824, 0.5277777777777778, 'x[18] <= 0.968\ngini =
0.048\nsamples = 121\nvalue = [118, 3]'),
 Text(0.4288164665523156, 0.4722222222222222, 'x[2] <= 0.98\ngini =
0.033\nsamples = 118\nvalue = [116, 2]'),
 Text(0.41509433962264153, 0.4166666666666667, 'x[14] <= 0.938\ngini =
0.017\nsamples = 114\nvalue = [113, 1]'),
 Text(0.40823327615780447, 0.3611111111111111, 'gini = 0.0\nsamples = 107\nvalue
= [107, 0]'),
 Text(0.4219554030874786, 0.3611111111111111, 'x[16] <= 0.25\ngini =
0.245\nsamples = 7\nvalue = [6, 1]'),
 Text(0.41509433962264153, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.4288164665523156, 0.3055555555555556, 'gini = 0.0\nsamples = 6\nvalue =
[6, 0]'),
 Text(0.4425385934819897, 0.4166666666666667, 'x[30] <= 0.088\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.43567753001715265, 0.3611111111111111, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.44939965694682676, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.4631217838765009, 0.4722222222222222, 'x[28] <= 0.583\ngini =
0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(0.4562607204116638, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.4699828473413379, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6009862778730704, 0.6944444444444444, 'x[30] <= 0.787\ngini =
0.1\nsamples = 493\nvalue = [467, 26]'),
 Text(0.5656089193825042, 0.6388888888888888, 'x[15] <= 0.5\ngini =
0.094\nsamples = 486\nvalue = [462, 24]'),
 Text(0.5154373927958834, 0.5833333333333334, 'x[14] <= 0.938\ngini =
0.154\nsamples = 191\nvalue = [175, 16]'),
 Text(0.5085763293310464, 0.5277777777777778, 'x[18] <= 0.481\ngini =
0.145\nsamples = 190\nvalue = [175, 15]'),
 Text(0.49056603773584906, 0.4722222222222222, 'x[18] <= 0.47\ngini =
0.221\nsamples = 95\nvalue = [83, 12]'),
 Text(0.483704974271012, 0.4166666666666667, 'x[33] <= 0.964\ngini =
0.207\nsamples = 94\nvalue = [83, 11]'),
 Text(0.47684391080617494, 0.3611111111111111, 'x[5] <= 0.375\ngini =
0.192\nsamples = 93\nvalue = [83, 10]'),
 Text(0.45454545454545453, 0.3055555555555556, 'x[6] <= 0.9\ngini =
0.363\nsamples = 21\nvalue = [16, 5]'),
 Text(0.44768439108061747, 0.25, 'x[17] <= 0.413\ngini = 0.266\nsamples =
19\nvalue = [16, 3]'),
 Text(0.4339622641509434, 0.19444444444444445, 'x[4] <= 0.982\ngini =
0.117\nsamples = 16\nvalue = [15, 1]'),
 Text(0.42710120068610635, 0.1388888888888889, 'gini = 0.0\nsamples = 14\nvalue
= [14, 0]'),
 Text(0.44082332761578047, 0.1388888888888889, 'x[11] <= 0.729\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.4339622641509434, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.44768439108061747, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue
= [1, 0]'),
 Text(0.4614065180102916, 0.19444444444444445, 'x[22] <= 0.893\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.45454545454545453, 0.1388888888888889, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.46826758147512865, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.4614065180102916, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.49914236706689535, 0.3055555555555556, 'x[31] <= 0.139\ngini =
0.129\nsamples = 72\nvalue = [67, 5]'),
 Text(0.48198970840480276, 0.25, 'x[8] <= 0.68\ngini = 0.444\nsamples = 6\nvalue
= [4, 2]'),
 Text(0.4751286449399657, 0.19444444444444445, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
 Text(0.4888507718696398, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.516295025728988, 0.25, 'x[2] <= 0.958\ngini = 0.087\nsamples = 66\nvalue
= [63, 3]'),
 Text(0.5025728987993139, 0.19444444444444445, 'x[28] <= 0.583\ngini =
0.061\nsamples = 64\nvalue = [62, 2]'),
 Text(0.4957118353344768, 0.1388888888888889, 'gini = 0.0\nsamples = 52\nvalue =
[52, 0]'),
 Text(0.5094339622641509, 0.1388888888888889, 'x[3] <= 0.75\ngini =
0.278\nsamples = 12\nvalue = [10, 2]'),
 Text(0.5025728987993139, 0.08333333333333333, 'gini = 0.0\nsamples = 9\nvalue =
[9, 0]'),
 Text(0.516295025728988, 0.08333333333333333, 'x[8] <= 0.87\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.5094339622641509, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue
= [0, 2]'),
 Text(0.5231560891938251, 0.027777777777777776, 'gini = 0.0\nsamples = 1\nvalue
= [1, 0]'),
 Text(0.5300171526586621, 0.19444444444444445, 'x[17] <= 0.413\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.5231560891938251, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.5368782161234992, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.49056603773584906, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.4974271012006861, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.5265866209262435, 0.4722222222222222, 'x[19] <= 0.5\ngini =
0.061\nsamples = 95\nvalue = [92, 3]'),
 Text(0.5197255574614065, 0.4166666666666667, 'gini = 0.0\nsamples = 76\nvalue =
[76, 0]'),
 Text(0.5334476843910806, 0.4166666666666667, 'x[33] <= 0.107\ngini =
0.266\nsamples = 19\nvalue = [16, 3]'),
 Text(0.5197255574614065, 0.3611111111111111, 'x[2] <= 0.547\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.5128644939965694, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.5265866209262435, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.5471698113207547, 0.3611111111111111, 'x[17] <= 0.108\ngini =
0.117\nsamples = 16\nvalue = [15, 1]'),
 Text(0.5403087478559176, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.5540308747855918, 0.3055555555555556, 'gini = 0.0\nsamples = 15\nvalue =
[15, 0]'),
 Text(0.5222984562607204, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6157804459691252, 0.5833333333333334, 'x[22] <= 0.036\ngini =
0.053\nsamples = 295\nvalue = [287, 8]'),
 Text(0.5917667238421955, 0.5277777777777778, 'x[32] <= 0.7\ngini =
0.159\nsamples = 46\nvalue = [42, 4]'),
 Text(0.5849056603773585, 0.4722222222222222, 'x[11] <= 0.071\ngini =
0.124\nsamples = 45\nvalue = [42, 3]'),
 Text(0.5677530017152659, 0.4166666666666667, 'x[24] <= 0.333\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.5608919382504288, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.5746140651801029, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6020583190394511, 0.4166666666666667, 'x[27] <= 0.688\ngini =
0.089\nsamples = 43\nvalue = [41, 2]'),
 Text(0.5883361921097771, 0.3611111111111111, 'x[14] <= 0.062\ngini =
0.048\nsamples = 41\nvalue = [40, 1]'),
 Text(0.58147512864494, 0.3055555555555556, 'x[30] <= 0.237\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
 Text(0.5746140651801029, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
 Text(0.5883361921097771, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.5951972555746141, 0.3055555555555556, 'gini = 0.0\nsamples = 37\nvalue =
[37, 0]'),
 Text(0.6157804459691252, 0.3611111111111111, 'x[6] <= 0.5\ngini = 0.5\nsamples
= 2\nvalue = [1, 1]'),
 Text(0.6089193825042881, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6226415094339622, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.5986277873070326, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6397941680960549, 0.5277777777777778, 'x[17] <= 0.056\ngini =
0.032\nsamples = 249\nvalue = [245, 4]'),
 Text(0.6226415094339622, 0.4722222222222222, 'x[17] <= 0.054\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(0.6157804459691252, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue =
[4, 0]'),
 Text(0.6295025728987993, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6569468267581475, 0.4722222222222222, 'x[2] <= 0.015\ngini =
0.024\nsamples = 244\nvalue = [241, 3]'),
 Text(0.6432246998284734, 0.4166666666666667, 'x[18] <= 0.715\ngini =
0.278\nsamples = 6\nvalue = [5, 1]'),
 Text(0.6363636363636364, 0.3611111111111111, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
 Text(0.6500857632933105, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6706689536878216, 0.4166666666666667, 'x[24] <= 0.167\ngini =
0.017\nsamples = 238\nvalue = [236, 2]'),
 Text(0.6638078902229846, 0.3611111111111111, 'x[29] <= 0.833\ngini =
0.073\nsamples = 53\nvalue = [51, 2]'),
 Text(0.6500857632933105, 0.3055555555555556, 'x[33] <= 0.107\ngini =
0.041\nsamples = 48\nvalue = [47, 1]'),
 Text(0.6432246998284734, 0.25, 'x[14] <= 0.312\ngini = 0.245\nsamples =
7\nvalue = [6, 1]'),
 Text(0.6363636363636364, 0.19444444444444445, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.6500857632933105, 0.19444444444444445, 'gini = 0.0\nsamples = 6\nvalue =
[6, 0]'),
 Text(0.6569468267581475, 0.25, 'gini = 0.0\nsamples = 41\nvalue = [41, 0]'),
 Text(0.6775300171526587, 0.3055555555555556, 'x[22] <= 0.357\ngini =
0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(0.6706689536878216, 0.25, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
 Text(0.6843910806174958, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.6775300171526587, 0.3611111111111111, 'gini = 0.0\nsamples = 185\nvalue
= [185, 0]'),
 Text(0.6363636363636364, 0.6388888888888888, 'x[2] <= 0.366\ngini =
0.408\nsamples = 7\nvalue = [5, 2]'),
 Text(0.6295025728987993, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.6432246998284734, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
 Text(0.4889579759862779, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.8224699828473413, 0.8611111111111112, 'x[17] <= 0.157\ngini =
0.385\nsamples = 300\nvalue = [222, 78]'),
 Text(0.7367066895368782, 0.8055555555555556, 'x[26] <= 0.167\ngini =
0.5\nsamples = 96\nvalue = [49, 47]'),
 Text(0.7015437392795884, 0.75, 'x[4] <= 0.161\ngini = 0.459\nsamples =
42\nvalue = [15, 27]'),
 Text(0.6775300171526587, 0.6944444444444444, 'x[8] <= 0.415\ngini =
0.499\nsamples = 23\nvalue = [12, 11]'),
 Text(0.6638078902229846, 0.6388888888888888, 'x[18] <= 0.561\ngini =
0.355\nsamples = 13\nvalue = [3, 10]'),
 Text(0.6569468267581475, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue =
[0, 8]'),
 Text(0.6706689536878216, 0.5833333333333334, 'x[9] <= 0.333\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
 Text(0.6638078902229846, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.6775300171526587, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.6912521440823327, 0.6388888888888888, 'x[29] <= 0.833\ngini =
0.18\nsamples = 10\nvalue = [9, 1]'),
 Text(0.6843910806174958, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue =
[8, 0]'),
 Text(0.6981132075471698, 0.5833333333333334, 'x[27] <= 0.063\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.6912521440823327, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.7049742710120068, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.725557461406518, 0.6944444444444444, 'x[27] <= 0.35\ngini =
0.266\nsamples = 19\nvalue = [3, 16]'),
 Text(0.7186963979416809, 0.6388888888888888, 'x[11] <= 0.2\ngini =
0.198\nsamples = 18\nvalue = [2, 16]'),
 Text(0.7118353344768439, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.725557461406518, 0.5833333333333334, 'x[32] <= 0.433\ngini =
0.111\nsamples = 17\nvalue = [1, 16]'),
 Text(0.7186963979416809, 0.5277777777777778, 'gini = 0.0\nsamples = 15\nvalue =
[0, 15]'),
 Text(0.7324185248713551, 0.5277777777777778, 'x[15] <= 0.333\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.725557461406518, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.7392795883361921, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.7324185248713551, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.7718696397941681, 0.75, 'x[0] <= 0.202\ngini = 0.466\nsamples =
54\nvalue = [34, 20]'),
 Text(0.7530017152658662, 0.6944444444444444, 'x[12] <= 0.833\ngini =
0.245\nsamples = 7\nvalue = [1, 6]'),
 Text(0.7461406518010292, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue =
[0, 6]'),
 Text(0.7598627787307033, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.79073756432247, 0.6944444444444444, 'x[2] <= 0.622\ngini =
0.418\nsamples = 47\nvalue = [33, 14]'),
 Text(0.7735849056603774, 0.6388888888888888, 'x[2] <= 0.145\ngini =
0.482\nsamples = 32\nvalue = [19, 13]'),
 Text(0.7598627787307033, 0.5833333333333334, 'x[2] <= 0.024\ngini =
0.18\nsamples = 10\nvalue = [9, 1]'),
 Text(0.7530017152658662, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.7667238421955404, 0.5277777777777778, 'gini = 0.0\nsamples = 9\nvalue =
[9, 0]'),
 Text(0.7873070325900514, 0.5833333333333334, 'x[18] <= 0.87\ngini =
0.496\nsamples = 22\nvalue = [10, 12]'),
 Text(0.7804459691252144, 0.5277777777777778, 'x[8] <= 0.41\ngini =
0.465\nsamples = 19\nvalue = [7, 12]'),
 Text(0.7667238421955404, 0.4722222222222222, 'x[18] <= 0.715\ngini =
0.469\nsamples = 8\nvalue = [5, 3]'),
 Text(0.7598627787307033, 0.4166666666666667, 'gini = 0.0\nsamples = 5\nvalue =
[5, 0]'),
 Text(0.7735849056603774, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.7941680960548885, 0.4722222222222222, 'x[0] <= 0.25\ngini =
0.298\nsamples = 11\nvalue = [2, 9]'),
 Text(0.7873070325900514, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.8010291595197255, 0.4166666666666667, 'x[4] <= 0.018\ngini =
0.18\nsamples = 10\nvalue = [1, 9]'),
 Text(0.7941680960548885, 0.3611111111111111, 'x[28] <= 0.417\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(0.7873070325900514, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.8010291595197255, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.8078902229845626, 0.3611111111111111, 'gini = 0.0\nsamples = 8\nvalue =
[0, 8]'),
 Text(0.7941680960548885, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.8078902229845626, 0.6388888888888888, 'x[19] <= 0.944\ngini =
0.124\nsamples = 15\nvalue = [14, 1]'),
 Text(0.8010291595197255, 0.5833333333333334, 'gini = 0.0\nsamples = 14\nvalue =
[14, 0]'),
 Text(0.8147512864493996, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.9082332761578045, 0.8055555555555556, 'x[16] <= 0.75\ngini =
0.258\nsamples = 204\nvalue = [173, 31]'),
 Text(0.8610634648370498, 0.75, 'x[17] <= 0.992\ngini = 0.138\nsamples =
147\nvalue = [136, 11]'),
 Text(0.8542024013722127, 0.6944444444444444, 'x[4] <= 0.482\ngini =
0.128\nsamples = 146\nvalue = [136, 10]'),
 Text(0.8353344768439108, 0.6388888888888888, 'x[30] <= 0.063\ngini =
0.038\nsamples = 104\nvalue = [102, 2]'),
 Text(0.8284734133790738, 0.5833333333333334, 'x[11] <= 0.193\ngini =
0.32\nsamples = 10\nvalue = [8, 2]'),
 Text(0.8216123499142367, 0.5277777777777778, 'x[2] <= 0.814\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.8147512864493996, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.8284734133790738, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.8353344768439108, 0.5277777777777778, 'gini = 0.0\nsamples = 7\nvalue =
[7, 0]'),
 Text(0.8421955403087479, 0.5833333333333334, 'gini = 0.0\nsamples = 94\nvalue =
[94, 0]'),
 Text(0.8730703259005146, 0.6388888888888888, 'x[9] <= 0.167\ngini =
0.308\nsamples = 42\nvalue = [34, 8]'),
 Text(0.855917667238422, 0.5833333333333334, 'x[11] <= 0.307\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
 Text(0.8490566037735849, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.8627787307032591, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.8902229845626072, 0.5833333333333334, 'x[0] <= 0.393\ngini =
0.229\nsamples = 38\nvalue = [33, 5]'),
 Text(0.8765008576329331, 0.5277777777777778, 'x[11] <= 0.643\ngini =
0.5\nsamples = 6\nvalue = [3, 3]'),
 Text(0.869639794168096, 0.4722222222222222, 'x[19] <= 0.222\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
 Text(0.8627787307032591, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.8765008576329331, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.8833619210977701, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue =
[2, 0]'),
 Text(0.9039451114922813, 0.5277777777777778, 'x[8] <= 0.992\ngini =
0.117\nsamples = 32\nvalue = [30, 2]'),
 Text(0.8970840480274442, 0.4722222222222222, 'x[28] <= 0.917\ngini =
0.062\nsamples = 31\nvalue = [30, 1]'),
 Text(0.8902229845626072, 0.4166666666666667, 'gini = 0.0\nsamples = 30\nvalue =
[30, 0]'),
 Text(0.9039451114922813, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.9108061749571184, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.8679245283018868, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue =
[0, 1]'),
 Text(0.9554030874785592, 0.75, 'x[14] <= 0.812\ngini = 0.456\nsamples =
57\nvalue = [37, 20]'),
 Text(0.9245283018867925, 0.6944444444444444, 'x[8] <= 0.071\ngini =
0.238\nsamples = 29\nvalue = [25, 4]'),
 Text(0.9108061749571184, 0.6388888888888888, 'x[11] <= 0.393\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.9039451114922813, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.9176672384219554, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.9382504288164666, 0.6388888888888888, 'x[32] <= 0.4\ngini =
0.142\nsamples = 26\nvalue = [24, 2]'),
 Text(0.9313893653516295, 0.5833333333333334, 'gini = 0.0\nsamples = 23\nvalue =
[23, 0]'),
 Text(0.9451114922813036, 0.5833333333333334, 'x[30] <= 0.8\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
 Text(0.9382504288164666, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.9519725557461407, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue =
[1, 0]'),
 Text(0.9862778730703259, 0.6944444444444444, 'x[32] <= 0.1\ngini =
0.49\nsamples = 28\nvalue = [12, 16]'),
 Text(0.9794168096054888, 0.6388888888888888, 'x[4] <= 0.804\ngini =
0.48\nsamples = 20\nvalue = [12, 8]'),
 Text(0.9725557461406518, 0.5833333333333334, 'x[4] <= 0.018\ngini =
0.415\nsamples = 17\nvalue = [12, 5]'),
 Text(0.9656946826758147, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue =
[0, 2]'),
 Text(0.9794168096054888, 0.5277777777777778, 'x[9] <= 0.167\ngini =
0.32\nsamples = 15\nvalue = [12, 3]'),
 Text(0.9725557461406518, 0.4722222222222222, 'x[0] <= 0.405\ngini =
0.5\nsamples = 6\nvalue = [3, 3]'),
 Text(0.9656946826758147, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.9794168096054888, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue =
[3, 0]'),
 Text(0.9862778730703259, 0.4722222222222222, 'gini = 0.0\nsamples = 9\nvalue =
[9, 0]'),
 Text(0.9862778730703259, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue =
[0, 3]'),
 Text(0.9931389365351629, 0.6388888888888888, 'gini = 0.0\nsamples = 8\nvalue =
[0, 8]')]
```



[59]:

```python
from sklearn.model_selection import GridSearchCV
parameter={
 'criterion':['gini','entropy'],
  'splitter':['best','random'],
  'max_depth':[1,2,3,4,5],
  'max_features':['auto', 'sqrt', 'log2']

}
```

[60]:

```python
grid_search=GridSearchCV(estimator=dtc,param_grid=parameter,cv=5,scoring="accuracy")
```

[61]:

```python
grid_search.fit(x_train,y_train)
```

```
D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_validation.py:425: FitFailedWarning:
100 fits failed out of a total of 300.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.
```

```
Below are more details about the failures:
--------------------------------------------------------------------------------
100 fits failed with the following error:
Traceback (most recent call last):
  File "D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "D:\programming languages\anaconda\Lib\site-packages\sklearn\base.py",
line 1144, in wrapper
    estimator._validate_params()
  File "D:\programming languages\anaconda\Lib\site-packages\sklearn\base.py",
line 637, in _validate_params
    validate_parameter_constraints(
  File "D:\programming languages\anaconda\Lib\site-
packages\sklearn\utils\_param_validation.py", line 95, in
validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features'
parameter of DecisionTreeClassifier must be an int in the range [1, inf), a
float in the range (0.0, 1.0], a str among {'log2', 'sqrt'} or None. Got 'auto'
instead.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_search.py:976: UserWarning: One or more of the
test scores are non-finite: [       nan        nan 0.84013704 0.84013704
0.84013704 0.84013704
        nan        nan 0.83503065 0.84183916 0.84524342 0.83843491
        nan        nan 0.83673639 0.8409881  0.84014064 0.84183916
        nan        nan 0.83335016 0.83588893 0.8367436  0.84099171
        nan        nan 0.83672917 0.84609088 0.84013343 0.84438875
        nan        nan 0.84013704 0.84013704 0.84013704 0.84013704
        nan        nan 0.8409881  0.84013704 0.83758384 0.84013704
        nan        nan 0.83843491 0.84269023 0.82908402 0.84183916
        nan        nan 0.83502344 0.84183916 0.84182474 0.84097007
        nan        nan 0.83334295 0.83928597 0.83161918 0.84014425]
  warnings.warn(
```

```
[61]: GridSearchCV(cv=5, estimator=DecisionTreeClassifier(),
                   param_grid={'criterion': ['gini', 'entropy'],
                               'max_depth': [1, 2, 3, 4, 5],
                               'max_features': ['auto', 'sqrt', 'log2'],
                               'splitter': ['best', 'random']},
                   scoring='accuracy')
```

```
[62]: grid_search.best_params_
```

[62]:
```
{'criterion': 'gini',
 'max_depth': 5,
 'max_features': 'sqrt',
 'splitter': 'random'}
```

[63]:
```
dtc_cv=DecisionTreeClassifier(criterion= 'entropy',
 max_depth=3,
 max_features='sqrt',
 splitter='best')
dtc_cv.fit(x_train,y_train)
```

[63]:
```
DecisionTreeClassifier(criterion='entropy', max_depth=3, max_features='sqrt')
```

[64]:
```
pred=dtc_cv.predict(x_test)
```

[65]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y=le.fit_transform(y)
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
pred=le.fit_transform(pred)
```

[66]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.84      0.99      0.91       245
           1       0.33      0.02      0.04        49

    accuracy                           0.83       294
   macro avg       0.58      0.51      0.47       294
weighted avg       0.75      0.83      0.76       294
```

RandomForestClassifier

[67]:
```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
```

[68]:
```
forest_params = [{'max_depth': list(range(10, 15)), 'max_features':␣
 ↪list(range(0,14))}]
```

[69]:
```
rfc_cv= GridSearchCV(rfc,param_grid=forest_params,cv=10,scoring="accuracy")
```

[70]:
```
rfc_cv.fit(x_train,y_train)
```

```
D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_validation.py:425: FitFailedWarning:
50 fits failed out of a total of 700.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
50 fits failed with the following error:
Traceback (most recent call last):
  File "D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "D:\programming languages\anaconda\Lib\site-packages\sklearn\base.py",
line 1144, in wrapper
    estimator._validate_params()
  File "D:\programming languages\anaconda\Lib\site-packages\sklearn\base.py",
line 637, in _validate_params
    validate_parameter_constraints(
  File "D:\programming languages\anaconda\Lib\site-
packages\sklearn\utils\_param_validation.py", line 95, in
validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features'
parameter of RandomForestClassifier must be an int in the range [1, inf), a
float in the range (0.0, 1.0], a str among {'log2', 'sqrt'} or None. Got 0
instead.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
D:\programming languages\anaconda\Lib\site-
packages\sklearn\model_selection\_search.py:976: UserWarning: One or more of the
test scores are non-finite: [       nan 0.84779806 0.85459945 0.85797479
0.85545415 0.85884398
 0.85797479 0.86054614 0.85966247 0.86138635 0.8596842  0.85712009
 0.86306678 0.85880052        nan 0.8460959  0.85715631 0.85459221
 0.86055338 0.86394321 0.86054614 0.85798204 0.86138635 0.85885122
 0.85882949 0.86223381 0.86308851 0.85882949        nan 0.84863827
 0.85969868 0.85885122 0.85712734 0.85969144 0.85798204 0.8596842
 0.85798204 0.85967695 0.85541069 0.85885122 0.85627264 0.86221208
        nan 0.84947849 0.85458496 0.85458496 0.86055338 0.86053165
 0.8605389  0.86137911 0.86308127 0.8596842  0.86308127 0.85882949
 0.85373026 0.85712009        nan 0.84865276 0.8528828  0.85715631
 0.86053165 0.85458496 0.86392148 0.86140084 0.86477618 0.85712009
 0.85796755 0.85543242 0.85796031 0.85625815]
  warnings.warn(
```

[70]:
```
GridSearchCV(cv=10, estimator=RandomForestClassifier(),
             param_grid=[{'max_depth': [10, 11, 12, 13, 14],
                          'max_features': [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,
                                           12, 13]}],
             scoring='accuracy')
```

[71]:
```
pred=rfc_cv.predict(x_test)
```

[72]:
```
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
y=le.fit_transform(y)
#label encoding
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
pred=le.fit_transform(pred)
```

[73]:
```
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.86      0.99      0.92       245
           1       0.80      0.16      0.27        49

    accuracy                           0.85       294
   macro avg       0.83      0.58      0.59       294
weighted avg       0.85      0.85      0.81       294
```

[74]:
```
rfc_cv.best_params_
```

[74]:
```
{'max_depth': 14, 'max_features': 8}
```

[ ]:
```

```