

# SAMHITHA-21BCB7025

## ASSIGNMENT-4

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.linear_model import LogisticRegression
```

```
df = pd.read_csv("Employee-Attrition.csv")
df.head()
```

	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
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]

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

df.describe()

      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]
```

```
df.isnull().any()
```

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

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

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

```

```

corr = df.corr()
corr

```

C:\Users\tejos\AppData\Local\Temp\ipykernel\_18972\2438084875.py:1:  
FutureWarning: The default value of numeric\_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
corr = df.corr()
```

	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

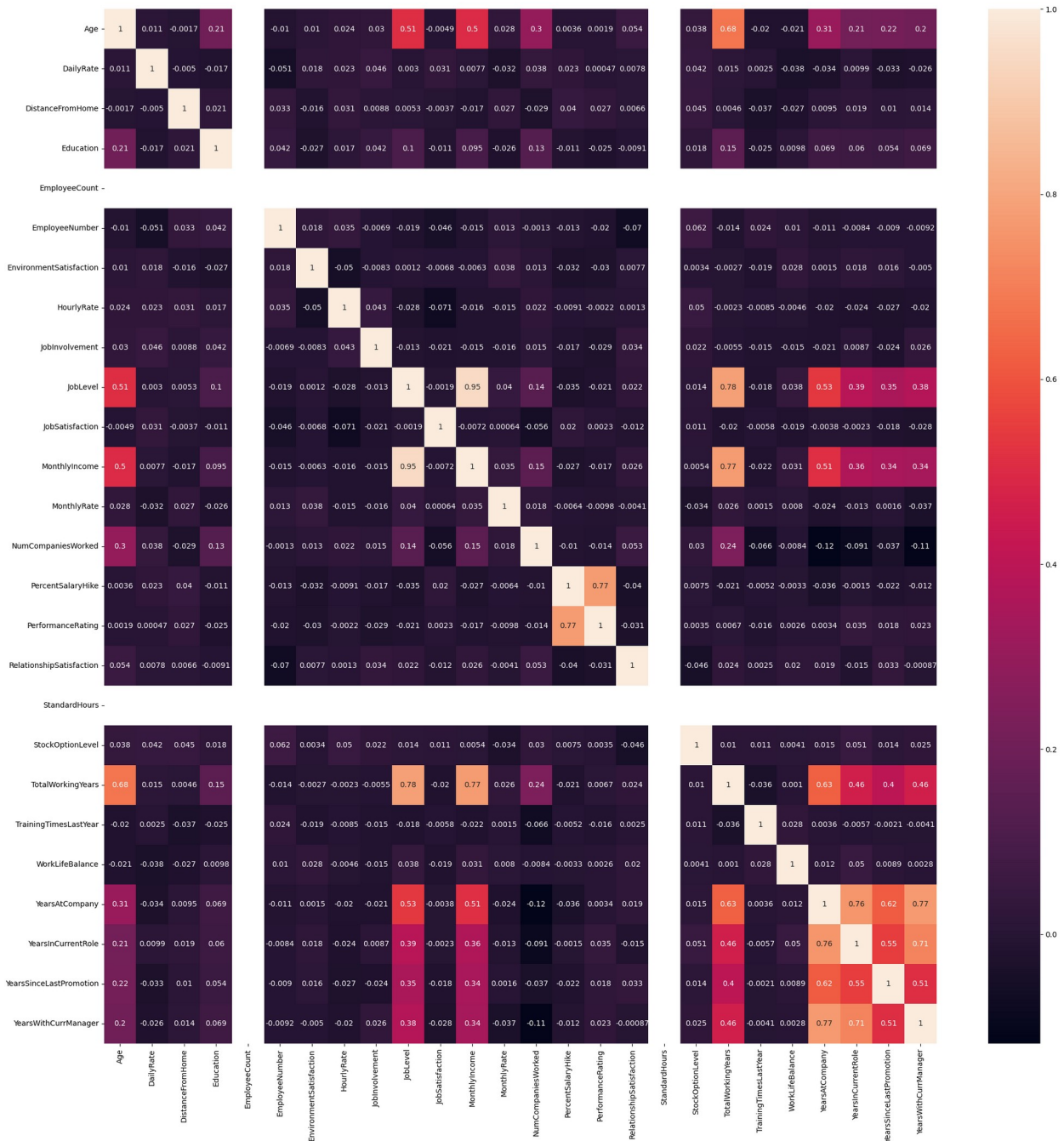
	YearsSinceLastPromotion	
YearsWithCurrManager		
Age	0.216513	
0.202089		
DailyRate	-0.033229	-
0.026363		
DistanceFromHome	0.010029	
0.014406		
Education	0.054254	
0.069065		
EmployeeCount	NaN	
NaN		
EmployeeNumber	-0.009019	-
0.009197		
EnvironmentSatisfaction	0.016194	-
0.004999		
HourlyRate	-0.026716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

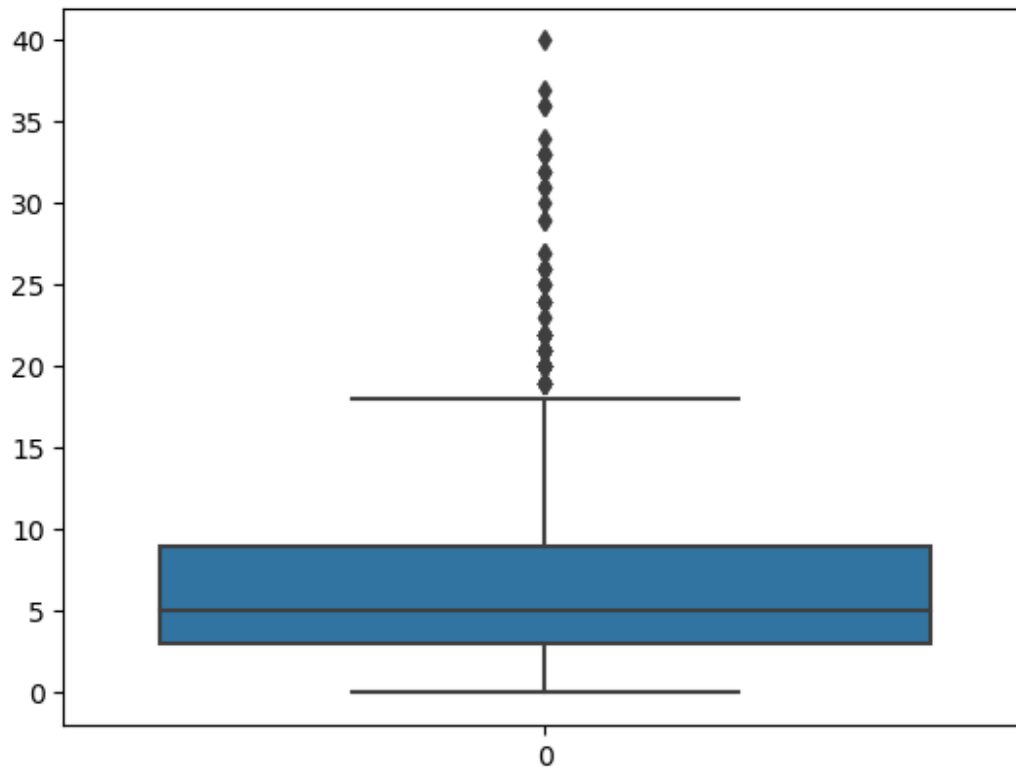
```
plt.subplots(figsize = (25,25))  
sns.heatmap(corr,annot = True)
```

<Axes: >



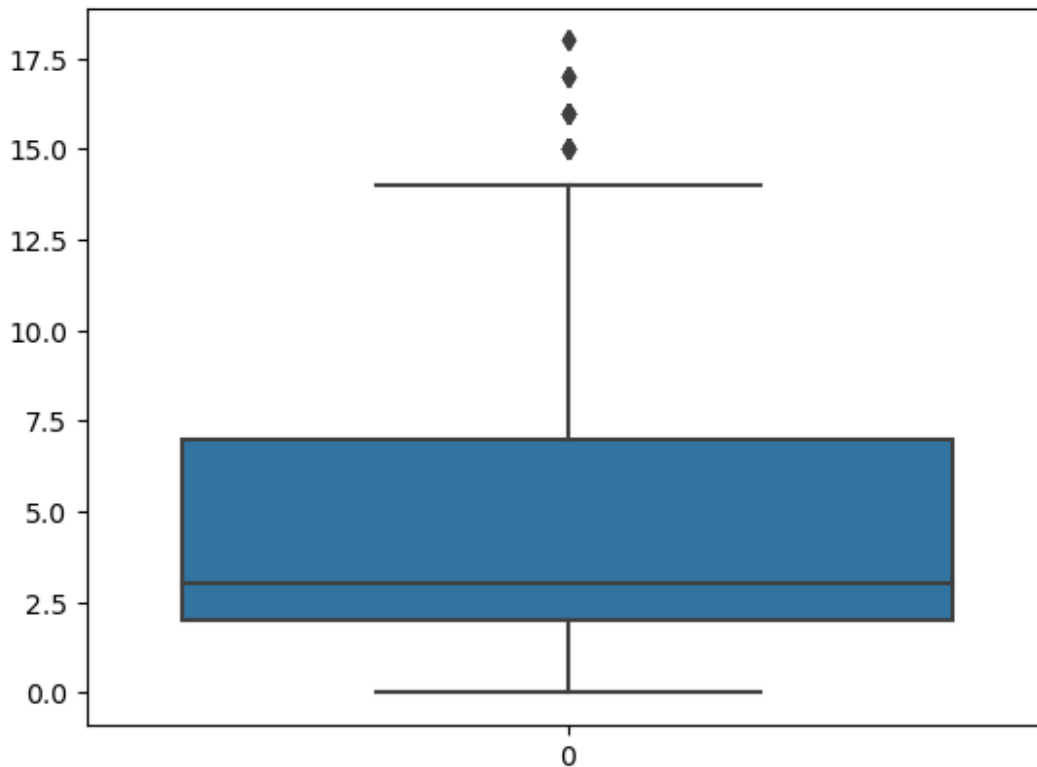
```
sns.boxplot(df.YearsAtCompany)
```

```
<Axes: >
```



```
sns.boxplot(df.YearsInCurrentRole)
```

```
<Axes: >
```



```
q1 = df.YearsInCurrentRole.quantile(0.25)
q3 = df.YearsInCurrentRole.quantile(0.75)
print(q1)
print(q3)
```

```
2.0
7.0
```

```
q3-q1
```

```
5.0
```

```
upperlimit = q3+1.5*(q3-q1)
upperlimit
```

```
14.5
```

```
lowerlimit = q1-1.5*(q3-q1)
lowerlimit
```

```
-5.5
```

```
df.median()
```

```
C:\Users\tejos\AppData\Local\Temp\ipykernel_18972\530051474.py:1:
FutureWarning: The default value of numeric_only in DataFrame.median
is deprecated. In a future version, it will default to False. In
```



addition, specifying 'numeric\_only=None' is deprecated. Select only valid columns or specify the value of numeric\_only to silence this warning.

```
df.median()
```

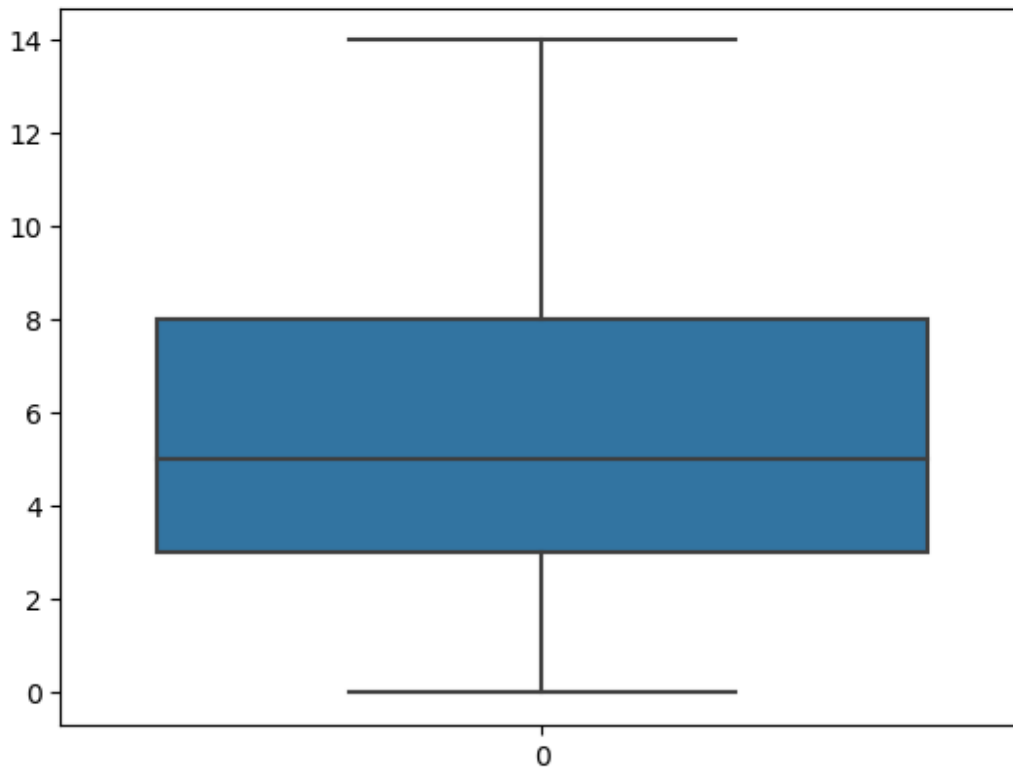
Age	36.0
DailyRate	802.0
DistanceFromHome	7.0
Education	3.0
EmployeeCount	1.0
EmployeeNumber	1020.5
EnvironmentSatisfaction	3.0
HourlyRate	66.0
JobInvolvement	3.0
JobLevel	2.0
JobSatisfaction	3.0
MonthlyIncome	4919.0
MonthlyRate	14235.5
NumCompaniesWorked	2.0
PercentSalaryHike	14.0
PerformanceRating	3.0
RelationshipSatisfaction	3.0
StandardHours	80.0
StockOptionLevel	1.0
TotalWorkingYears	10.0
TrainingTimesLastYear	3.0
WorkLifeBalance	3.0
YearsAtCompany	5.0
YearsInCurrentRole	3.0
YearsSinceLastPromotion	1.0
YearsWithCurrManager	3.0

dtype: float64

```
df["YearsAtCompany"] = np.where(df["YearsAtCompany"] > upperlimit, 5, df["YearsAtCompany"])
```

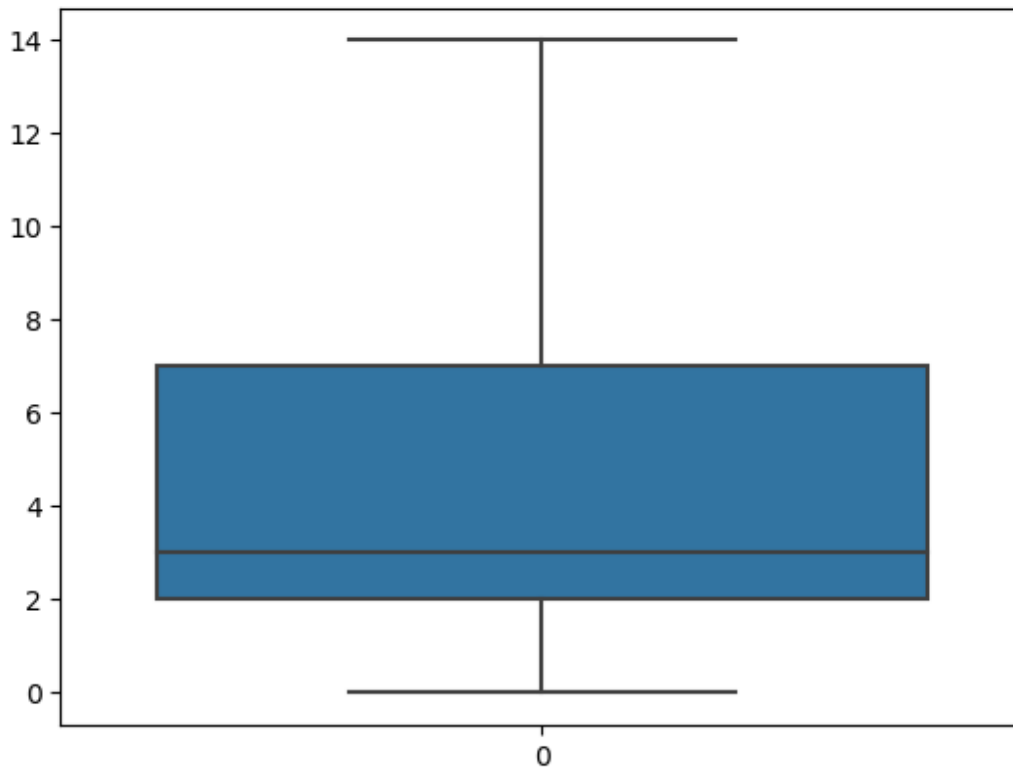
```
sns.boxplot(df.YearsAtCompany)
```

<Axes: >



```
df["YearsInCurrentRole"]=np.where(df["YearsInCurrentRole"]>upperlimit,  
5,df["YearsInCurrentRole"])  
sns.boxplot(df.YearsInCurrentRole)
```

<Axes: >



```
x = df.drop('Attrition',axis=1)
```

```
y = df['Attrition']
```

```
x.head()
```

	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]

type(x)

pandas.core.frame.DataFrame

type(y)

pandas.core.series.Series

```
le = LabelEncoder()
x.BusinessTravel = le.fit_transform(x.BusinessTravel)
x.head()
```

	Age	BusinessTravel	DailyRate	Department
DistanceFromHome \				
0	41	2	1102	Sales
1				
1	49	1	279	Research & Development
8				
2	37	2	1373	Research & Development
2				
3	33	1	1392	Research & Development
3				
4	27	2	591	Research & Development
2				

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

	EnvironmentSatisfaction	... RelationshipSatisfaction
StandardHours \		
0	2	...
80		
1	3	...
80		
2	4	...
80		
3	4	...
80		
4	1	...
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]

```
y = le.fit_transform(y)
y
```

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

```
x.Department = le.fit_transform(x.Department)
x.head()
```

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome
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
0	Life Sciences	1	1
2	...		
1	Life Sciences	1	2
3	...		
2	Other	1	4
4	...		
3	Life Sciences	1	5
4	...		
4	Medical	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
6		1
1	10	3
10		
2	7	3
0		
3	8	3
8		
4	6	3
2		

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

[5 rows x 34 columns]

```
x.EducationField = le.fit_transform(x.EducationField)
x.head()
```

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

EducationField	EmployeeCount	EmployeeNumber
EnvironmentSatisfaction \		
0	1	1
2		
1	1	1
3		2

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]

```
x.Gender = le.fit_transform(x.Gender)
x.head()
```

	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]

```

x.MaritalStatus = le.fit_transform(x.MaritalStatus)
x.Over18 = le.fit_transform(x.Over18)
x.OverTime = le.fit_transform(x.OverTime)
x.JobRole = le.fit_transform(x.JobRole)
x.head()

```

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]

```
ms = MinMaxScaler()
```

```
x_scaled= pd.DataFrame(ms.fit_transform(x),columns=x.columns)
x_scaled.head()
```

	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

	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

	EnvironmentSatisfaction	...	RelationshipSatisfaction
0	0.333333	...	0.000000
1	0.666667	...	1.000000
2	1.000000	...	0.333333
3	1.000000	...	0.666667
4	0.000000	...	1.000000

	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear \
0	0.000000	0.200	0.0
1	0.333333	0.250	0.5
2	0.000000	0.175	0.5
3	0.000000	0.200	0.5
4	0.333333	0.150	0.5

	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole \
--	-----------------	----------------	----------------------

0	0.000000	0.428571	0.285714
1	0.666667	0.714286	0.500000
2	0.666667	0.000000	0.000000
3	0.666667	0.571429	0.500000
4	0.666667	0.142857	0.142857

	YearsSinceLastPromotion	YearsWithCurrManager
0	0.000000	0.294118
1	0.066667	0.411765
2	0.000000	0.000000
3	0.200000	0.000000
4	0.133333	0.117647

[5 rows x 34 columns]

x.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1470 entries, 0 to 1469

Data columns (total 34 columns):

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	Age	1470 non-null	int64
1	BusinessTravel	1470 non-null	int32
2	DailyRate	1470 non-null	int64
3	Department	1470 non-null	int32
4	DistanceFromHome	1470 non-null	int64
5	Education	1470 non-null	int64
6	EducationField	1470 non-null	int32
7	EmployeeCount	1470 non-null	int64
8	EmployeeNumber	1470 non-null	int64
9	EnvironmentSatisfaction	1470 non-null	int64
10	Gender	1470 non-null	int32
11	HourlyRate	1470 non-null	int64
12	JobInvolvement	1470 non-null	int64
13	JobLevel	1470 non-null	int64
14	JobRole	1470 non-null	int32
15	JobSatisfaction	1470 non-null	int64
16	MaritalStatus	1470 non-null	int32
17	MonthlyIncome	1470 non-null	int64
18	MonthlyRate	1470 non-null	int64
19	NumCompaniesWorked	1470 non-null	int64
20	Over18	1470 non-null	int32
21	Overtime	1470 non-null	int32
22	PercentSalaryHike	1470 non-null	int64
23	PerformanceRating	1470 non-null	int64
24	RelationshipSatisfaction	1470 non-null	int64
25	StandardHours	1470 non-null	int64
26	StockOptionLevel	1470 non-null	int64
27	TotalWorkingYears	1470 non-null	int64

[illegible]

```
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, 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,
0,
1, 1, 0, 1, 0, 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, 0, 1, 1, 0, 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, 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, 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, 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,
0, 1, 0, 0, 0, 1, 0, 0])
```

```
from sklearn.metrics import
accuracy_score,confusion_matrix,classification_report
accuracy_score(y_test,pred)
0.8843537414965986
confusion_matrix(y_test,pred)
array([[242,  3],
       [ 31, 18]], dtype=int64)
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.89	0.99	0.93	245
1	0.86	0.37	0.51	49
accuracy			0.88	294
macro avg	0.87	0.68	0.72	294
weighted avg	0.88	0.88	0.86	294

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

```
probability
```

```
array([0.139689 , 0.19203157, 0.31762872, 0.06821401, 0.66363396,  
       0.06679984, 0.6692895 , 0.05667544, 0.00984414, 0.37550747,  
       0.0596486 , 0.32040693, 0.01766806, 0.66462849, 0.20449626,  
       0.0332682 , 0.10081727, 0.17896119, 0.04726071, 0.19875051,  
       0.24840018, 0.02118971, 0.05508412, 0.05624411, 0.59064156,  
       0.39773941, 0.05798602, 0.03316344, 0.69770104, 0.06434571,  
       0.01241986, 0.02760685, 0.08064678, 0.16501468, 0.08024179,  
       0.03300692, 0.08621125, 0.06973968, 0.03323051, 0.05478155,  
       0.07302527, 0.01685177, 0.04834498, 0.01790702, 0.02117587,  
       0.50177468, 0.42499957, 0.00307051, 0.76272065, 0.52732192,  
       0.11573632, 0.48139989, 0.07400635, 0.26197253, 0.68585166,  
       0.28810255, 0.01881209, 0.31829355, 0.02780931, 0.17375126,  
       0.01940688, 0.22571114, 0.15371747, 0.03446668, 0.40839461,  
       0.03541442, 0.24579496, 0.12478435, 0.09620694, 0.0992372 ,  
       0.10888532, 0.30043084, 0.06294337, 0.07266611, 0.12445761,  
       0.06483562, 0.04834911, 0.07949782, 0.20225966, 0.02831225,  
       0.00707532, 0.02164015, 0.14805975, 0.02530552, 0.02966628,  
       0.0762403 , 0.01265841, 0.03303401, 0.03693788, 0.13816936,  
       0.30154937, 0.16993445, 0.28437979, 0.24647952, 0.01774685,  
       0.18495586, 0.35274329, 0.28497481, 0.07620371, 0.04856371,  
       0.25508192, 0.75971301, 0.34218794, 0.01902584, 0.09592795,  
       0.03063844, 0.04903468, 0.16168169, 0.05179291, 0.1244266 ,  
       0.08483899, 0.04678438, 0.02684536, 0.14652185, 0.06109046,  
       0.02971224, 0.04444413, 0.12514182, 0.00707973, 0.01073575,  
       0.16198959, 0.04615528, 0.06739916, 0.82744548, 0.03365619,  
       0.0338438 , 0.0093067 , 0.13406507, 0.17349546, 0.05046873,  
       0.01565673, 0.26448719, 0.57952996, 0.32705869, 0.04297303,  
       0.39117438, 0.58530873, 0.14395331, 0.07955263, 0.25760428,  
       0.11126284, 0.06710735, 0.1009097 , 0.13775765, 0.19696915,  
       0.02455285, 0.16436541, 0.01194276, 0.11728467, 0.1690352 ,  
       0.06262844, 0.15916042, 0.06067153, 0.13343285, 0.03004275,  
       0.02205602, 0.06416646, 0.07854048, 0.01222238, 0.01017505,  
       0.46026173, 0.01474059, 0.16278495, 0.82419141, 0.11115106,  
       0.28572252, 0.15394052, 0.14151269, 0.03102641, 0.00603303,  
       0.03933842, 0.07883446, 0.12642862, 0.10828651, 0.04117055,  
       0.14827404, 0.10452404, 0.13470889, 0.05100001, 0.10385795,  
       0.03033568, 0.09692069, 0.00629245, 0.77914602, 0.04010393,  
       0.03960355, 0.3924741 , 0.04246515, 0.75164761, 0.11066569,  
       0.37636118, 0.39834024, 0.27148738, 0.05231632, 0.08634198,  
       0.14695653, 0.04196124, 0.0125218 , 0.2799495 , 0.05173187,  
       0.14802111, 0.15807771, 0.67948115, 0.06059042, 0.23203306,  
       0.03257483, 0.4824282 , 0.00328 , 0.13866045, 0.02494448,  
       0.11935202, 0.16602654, 0.05667975, 0.10376776, 0.15815346,  
       0.02767992, 0.01908474, 0.07781499, 0.03243343, 0.14721138,  
       0.08482637, 0.23384339, 0.74644943, 0.12150511, 0.39423689,  
       0.0151262 , 0.12000733, 0.23015731, 0.34321134, 0.0366569 ,
```

```

0.04065181, 0.32726154, 0.05440085, 0.01691351, 0.16887658,
0.36840505, 0.24385847, 0.00747131, 0.08292112, 0.00903689,
0.13802347, 0.27622546, 0.01151549, 0.16852845, 0.03704734,
0.03618358, 0.43278276, 0.37149071, 0.03522335, 0.10998048,
0.39416705, 0.31640703, 0.80951128, 0.04966195, 0.19133522,
0.07099553, 0.005368 , 0.68465104, 0.37811123, 0.367362 ,
0.36756812, 0.03401359, 0.26248562, 0.05986855, 0.06345854,
0.10722762, 0.00694377, 0.23640487, 0.46213154, 0.07122532,
0.09022254, 0.01262613, 0.15422746, 0.04787988, 0.01993732,
0.02368743, 0.06355887, 0.25322214, 0.26058801, 0.20169548,
0.28111899, 0.01582808, 0.162111 , 0.09351842, 0.02701473,
0.19563321, 0.00809015, 0.23963531, 0.00733077, 0.02078013,
0.20737464, 0.73363449, 0.0803108 , 0.28551904])

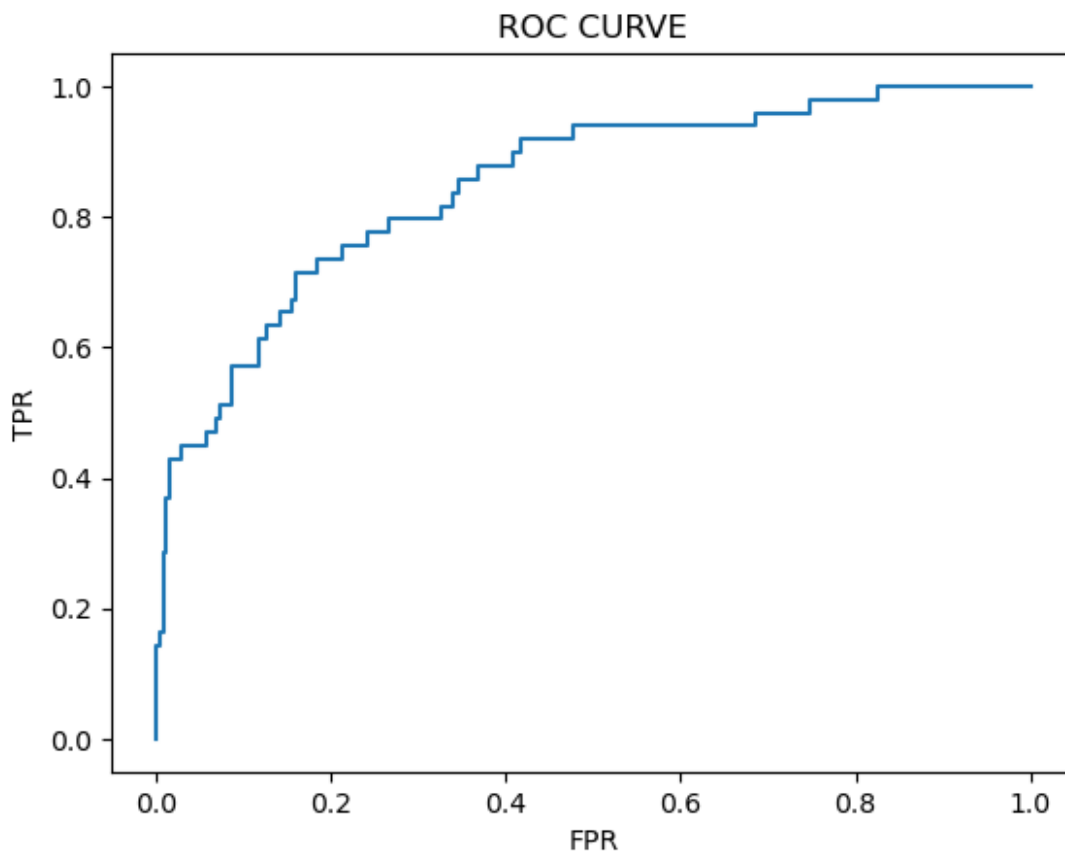
```

```

from sklearn.metrics import roc_curve
fpr,tpr,threshsholds = roc_curve(y_test,probability)

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

```





# Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
dt = DecisionTreeClassifier()

y_train
array([0, 0, 0, ..., 0, 0, 0])

dt.fit(x_train,y_train)

DecisionTreeClassifier()

pred = dt.predict(x_test)

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

y_test
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, 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, 1, 1, 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,
1,
    1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,
1,
    0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 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,
    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, 1, 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, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
0,
    0, 1, 0, 0, 0, 1, 0, 0])

```

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

```
0.7721088435374149
```

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

```
array([[208,  37],
       [ 30,  19]], dtype=int64)
```

```
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.87	0.85	0.86	245
1	0.34	0.39	0.36	49
accuracy			0.77	294
macro avg	0.61	0.62	0.61	294
weighted avg	0.78	0.77	0.78	294

```

probab = dt.predict_proba(x_test)[: ,1]
probab

```

```

array([0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 1., 0., 0.,
0.,
    0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0.,
0.,
    0.,

```

```

0., 0., 0., 0., 1., 0., 0., 0., 0., 1., 0., 0., 0., 0., 1., 1.,
0., 1., 1., 0., 0., 0., 0., 0., 0., 0., 0., 1., 0., 1., 0., 0.,
0., 0., 0., 0., 0., 0., 0., 1., 1., 0., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 0., 0., 0., 0.,
0., 1., 0., 0., 0., 0., 1., 0., 0., 0., 1., 0., 0., 0., 0., 0.,
0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 0., 0., 1., 1., 1., 0.,
0., 0., 1., 0., 0., 0., 0., 0., 1., 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., 0., 0., 1., 0., 1., 1., 0., 0., 0., 0., 1., 0., 0.,
1., 1., 0., 1., 0., 1., 0., 0., 0., 0., 0., 0., 0., 0., 0., 1.,
0., 1., 0., 0., 0., 0., 0., 0., 0., 1., 0., 0., 0., 0., 0., 1.,
0., 0., 1., 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., 1., 0.,
0., 0., 1., 0., 1., 0., 0., 1., 0., 0., 1., 0., 1., 0., 0., 0.,
0., 0., 0., 0., 1., 0., 0., 1., 0., 0., 1., 1., 0., 1., 0., 0.,
0., 0., 0., 0., 0.]

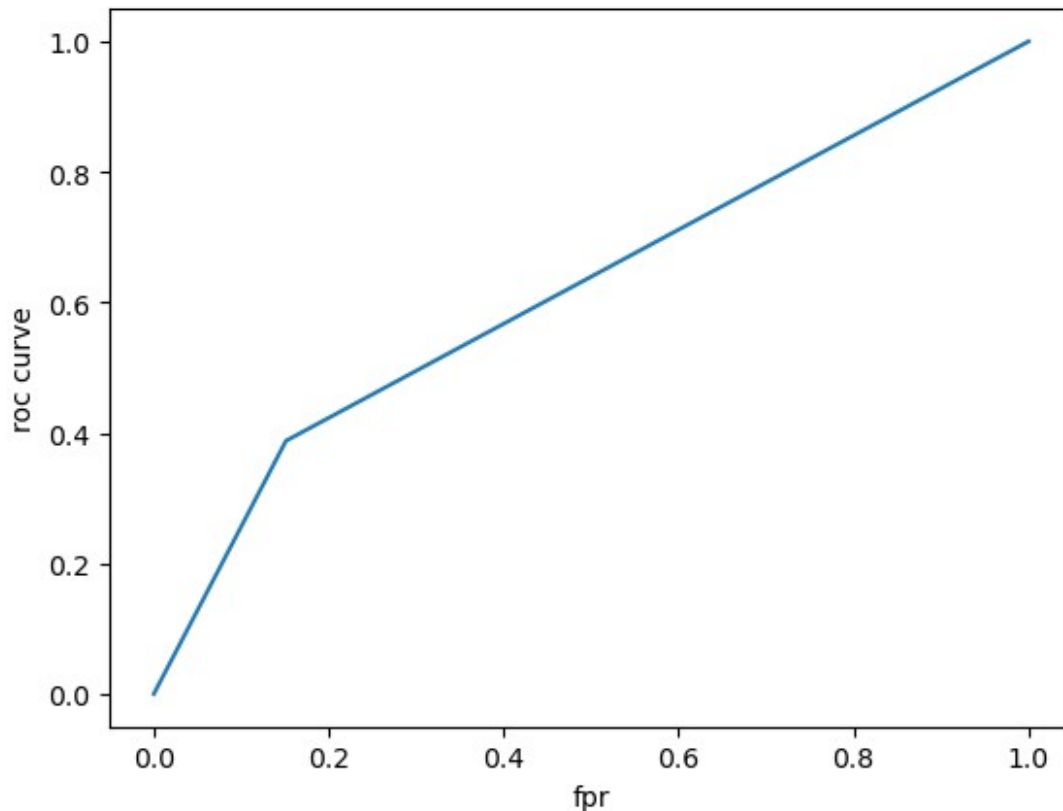
```

```
fpr,tpr,threshholds = roc_curve(y_test,probab)
```

```

plt.plot(fpr,tpr)
plt.xlabel('fpr')
plt.ylabel('roc curve')
plt.show()

```



```
from sklearn import tree
plt.figure(figsize = (25,25))
tree.plot_tree(dt,filled= True)

[Text(0.330265035118525, 0.9705882352941176, 'x[27] <= 0.038\ngini =
0.269\nsamples = 1176\nvalue = [988, 188]'),
Text(0.08428446005267778, 0.9117647058823529, 'x[16] <= 0.75\ngini =
0.5\nsamples = 78\nvalue = [39, 39]'),
Text(0.05267778753292362, 0.8529411764705882, 'x[4] <= 0.554\ngini =
0.426\nsamples = 39\nvalue = [27, 12]'),
Text(0.03511852502194908, 0.7941176470588235, 'x[15] <= 0.167\ngini =
0.312\nsamples = 31\nvalue = [25, 6]'),
Text(0.021071115013169446, 0.7352941176470589, 'x[16] <= 0.25\ngini =
0.49\nsamples = 7\nvalue = [3, 4]'),
Text(0.014047410008779631, 0.6764705882352942, 'x[21] <= 0.5\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.007023705004389816, 0.6176470588235294, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
Text(0.021071115013169446, 0.6176470588235294, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.028094820017559263, 0.6764705882352942, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
Text(0.04916593503072871, 0.7352941176470589, 'x[19] <= 0.056\ngini =
0.153\nsamples = 24\nvalue = [22, 2]'),
```

```
Text(0.04214223002633889, 0.6764705882352942, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.056189640035118525, 0.6764705882352942, 'x[9] <= 0.167\ngini = 0.083\nsamples = 23\nvalue = [22, 1]'),
Text(0.04916593503072871, 0.6176470588235294, 'x[28] <= 0.583\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.04214223002633889, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.056189640035118525, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.06321334503950835, 0.6176470588235294, 'gini = 0.0\nsamples = 21\nvalue = [21, 0]'),
Text(0.07023705004389816, 0.7941176470588235, 'x[8] <= 0.385\ngini = 0.375\nsamples = 8\nvalue = [2, 6]'),
Text(0.06321334503950835, 0.7352941176470589, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.07726075504828797, 0.7352941176470589, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.11589113257243196, 0.8529411764705882, 'x[11] <= 0.364\ngini = 0.426\nsamples = 39\nvalue = [12, 27]'),
Text(0.09833187006145742, 0.7941176470588235, 'x[0] <= 0.369\ngini = 0.133\nsamples = 14\nvalue = [1, 13]'),
Text(0.0913081650570676, 0.7352941176470589, 'gini = 0.0\nsamples = 13\nvalue = [0, 13]'),
Text(0.10535557506584724, 0.7352941176470589, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.1334503950834065, 0.7941176470588235, 'x[8] <= 0.105\ngini = 0.493\nsamples = 25\nvalue = [11, 14]'),
Text(0.11940298507462686, 0.7352941176470589, 'x[22] <= 0.464\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.11237928007023705, 0.6764705882352942, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.1264266900790167, 0.6764705882352942, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.14749780509218613, 0.7352941176470589, 'x[15] <= 0.5\ngini = 0.432\nsamples = 19\nvalue = [6, 13]'),
Text(0.14047410008779632, 0.6764705882352942, 'gini = 0.0\nsamples = 7\nvalue = [0, 7]'),
Text(0.15452151009657594, 0.6764705882352942, 'x[6] <= 0.4\ngini = 0.5\nsamples = 12\nvalue = [6, 6]'),
Text(0.14047410008779632, 0.6176470588235294, 'x[14] <= 0.875\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.1334503950834065, 0.5588235294117647, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.14749780509218613, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.16856892010535557, 0.6176470588235294, 'x[8] <= 0.249\ngini = 0.278\nsamples = 6\nvalue = [1, 5]'),
Text(0.16154521510096576, 0.5588235294117647, 'gini = 0.0\nsamples =
```

```
1\nvalue = [1, 0]'),
Text(0.17559262510974538, 0.5588235294117647, 'gini = 0.0\nsamples =
5\nvalue = [0, 5]'),
Text(0.5762456101843723, 0.9117647058823529, 'x[21] <= 0.5\ngini =
0.235\nsamples = 1098\nvalue = [949, 149]'),
Text(0.3377414398595259, 0.8529411764705882, 'x[29] <= 0.167\ngini =
0.162\nsamples = 798\nvalue = [727, 71]'),
Text(0.18964003511852504, 0.7941176470588235, 'x[8] <= 0.445\ngini =
0.38\nsamples = 47\nvalue = [35, 12]'),
Text(0.17559262510974538, 0.7352941176470589, 'x[16] <= 0.75\ngini =
0.1\nsamples = 19\nvalue = [18, 1]'),
Text(0.16856892010535557, 0.6764705882352942, 'gini = 0.0\nsamples =
18\nvalue = [18, 0]'),
Text(0.1826163301141352, 0.6764705882352942, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.20368744512730466, 0.7352941176470589, 'x[17] <= 0.094\ngini =
0.477\nsamples = 28\nvalue = [17, 11]'),
Text(0.19666374012291485, 0.6764705882352942, 'gini = 0.0\nsamples =
4\nvalue = [0, 4]'),
Text(0.21071115013169447, 0.6764705882352942, 'x[8] <= 0.524\ngini =
0.413\nsamples = 24\nvalue = [17, 7]'),
Text(0.20368744512730466, 0.6176470588235294, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
Text(0.2177348551360843, 0.6176470588235294, 'x[33] <= 0.324\ngini =
0.351\nsamples = 22\nvalue = [17, 5]'),
Text(0.20368744512730466, 0.5588235294117647, 'x[2] <= 0.025\ngini =
0.133\nsamples = 14\nvalue = [13, 1]'),
Text(0.19666374012291485, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
Text(0.21071115013169447, 0.5, 'gini = 0.0\nsamples = 13\nvalue =
[13, 0]'),
Text(0.2317822651448639, 0.5588235294117647, 'x[2] <= 0.329\ngini =
0.5\nsamples = 8\nvalue = [4, 4]'),
Text(0.2247585601404741, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
Text(0.23880597014925373, 0.5, 'x[12] <= 0.333\ngini = 0.32\nsamples
= 5\nvalue = [4, 1]'),
Text(0.2317822651448639, 0.4411764705882353, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.24582967515364354, 0.4411764705882353, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
Text(0.4858428446005268, 0.7941176470588235, 'x[27] <= 0.975\ngini =
0.145\nsamples = 751\nvalue = [692, 59]'),
Text(0.478819139596137, 0.7352941176470589, 'x[30] <= 0.321\ngini =
0.143\nsamples = 750\nvalue = [692, 58]'),
Text(0.36479367866549606, 0.6764705882352942, 'x[9] <= 0.167\ngini =
0.218\nsamples = 257\nvalue = [225, 32]'),
Text(0.32133450395083407, 0.6176470588235294, 'x[33] <= 0.147\ngini =
0.355\nsamples = 65\nvalue = [50, 15]'),
```

```
Text(0.29850746268656714, 0.5588235294117647, 'x[33] <= 0.029\ngini = 0.303\nsamples = 59\nvalue = [48, 11]'),
Text(0.2739244951712028, 0.5, 'x[12] <= 0.5\ngini = 0.463\nsamples = 22\nvalue = [14, 8]'),
Text(0.2598770851624232, 0.4411764705882353, 'x[11] <= 0.179\ngini = 0.198\nsamples = 9\nvalue = [8, 1]'),
Text(0.2528533801580334, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.266900790166813, 0.38235294117647056, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.28797190517998245, 0.4411764705882353, 'x[11] <= 0.4\ngini = 0.497\nsamples = 13\nvalue = [6, 7]'),
Text(0.28094820017559263, 0.38235294117647056, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.29499561018437226, 0.38235294117647056, 'x[4] <= 0.286\ngini = 0.346\nsamples = 9\nvalue = [2, 7]'),
Text(0.28797190517998245, 0.3235294117647059, 'x[19] <= 0.944\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.28094820017559263, 0.2647058823529412, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.29499561018437226, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.30201931518876207, 0.3235294117647059, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.3230904302019315, 0.5, 'x[15] <= 0.167\ngini = 0.149\nsamples = 37\nvalue = [34, 3]'),
Text(0.3160667251975417, 0.4411764705882353, 'x[29] <= 0.5\ngini = 0.5\nsamples = 6\nvalue = [3, 3]'),
Text(0.3090430201931519, 0.38235294117647056, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.3230904302019315, 0.38235294117647056, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.3301141352063213, 0.4411764705882353, 'gini = 0.0\nsamples = 31\nvalue = [31, 0]'),
Text(0.34416154521510095, 0.5588235294117647, 'x[8] <= 0.065\ngini = 0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.33713784021071114, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.35118525021949076, 0.5, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.40825285338015804, 0.6176470588235294, 'x[0] <= 0.321\ngini = 0.161\nsamples = 192\nvalue = [175, 17]'),
Text(0.3722563652326602, 0.5588235294117647, 'x[6] <= 0.1\ngini = 0.294\nsamples = 67\nvalue = [55, 12]'),
Text(0.3652326602282704, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.37928007023705007, 0.5, 'x[29] <= 0.5\ngini = 0.26\nsamples = 65\nvalue = [55, 10]'),
Text(0.3617208077260755, 0.4411764705882353, 'x[11] <= 0.679\ngini =
```

```
0.469\nsamples = 16\nvalue = [10, 6]'),  
  Text(0.3546971027216857, 0.38235294117647056, 'x[4] <= 0.018\ngini =  
0.444\nsamples = 9\nvalue = [3, 6]'),  
  Text(0.3476733977172959, 0.3235294117647059, 'gini = 0.0\nsamples =  
2\nvalue = [2, 0]'),  
  Text(0.3617208077260755, 0.3235294117647059, 'x[1] <= 0.25\ngini =  
0.245\nsamples = 7\nvalue = [1, 6]'),  
  Text(0.3546971027216857, 0.2647058823529412, 'gini = 0.0\nsamples =  
1\nvalue = [1, 0]'),  
  Text(0.3687445127304653, 0.2647058823529412, 'gini = 0.0\nsamples =  
6\nvalue = [0, 6]'),  
  Text(0.3687445127304653, 0.38235294117647056, 'gini = 0.0\nsamples =  
7\nvalue = [7, 0]'),  
  Text(0.3968393327480246, 0.4411764705882353, 'x[2] <= 0.037\ngini =  
0.15\nsamples = 49\nvalue = [45, 4]'),  
  Text(0.38981562774363476, 0.38235294117647056, 'gini = 0.0\nsamples =  
1\nvalue = [0, 1]'),  
  Text(0.4038630377524144, 0.38235294117647056, 'x[2] <= 0.938\ngini =  
0.117\nsamples = 48\nvalue = [45, 3]'),  
  Text(0.3968393327480246, 0.3235294117647059, 'x[5] <= 0.875\ngini =  
0.081\nsamples = 47\nvalue = [45, 2]'),  
  Text(0.38279192273924495, 0.2647058823529412, 'x[12] <= 0.167\ngini =  
0.043\nsamples = 45\nvalue = [44, 1]'),  
  Text(0.37576821773485514, 0.20588235294117646, 'x[14] <= 0.625\ngini  
= 0.444\nsamples = 3\nvalue = [2, 1]'),  
  Text(0.3687445127304653, 0.14705882352941177, 'gini = 0.0\nsamples =  
2\nvalue = [2, 0]'),  
  Text(0.38279192273924495, 0.14705882352941177, 'gini = 0.0\nsamples =  
1\nvalue = [0, 1]'),  
  Text(0.38981562774363476, 0.20588235294117646, 'gini = 0.0\nsamples =  
42\nvalue = [42, 0]'),  
  Text(0.4108867427568042, 0.2647058823529412, 'x[27] <= 0.125\ngini =  
0.5\nsamples = 2\nvalue = [1, 1]'),  
  Text(0.4038630377524144, 0.20588235294117646, 'gini = 0.0\nsamples =  
1\nvalue = [1, 0]'),  
  Text(0.417910447761194, 0.20588235294117646, 'gini = 0.0\nsamples =  
1\nvalue = [0, 1]'),  
  Text(0.4108867427568042, 0.3235294117647059, 'gini = 0.0\nsamples =  
1\nvalue = [0, 1]'),  
  Text(0.4442493415276558, 0.5588235294117647, 'x[8] <= 0.022\ngini =  
0.077\nsamples = 125\nvalue = [120, 5]'),  
  Text(0.42493415276558383, 0.5, 'x[2] <= 0.578\ngini = 0.5\nsamples =  
4\nvalue = [2, 2]'),  
  Text(0.417910447761194, 0.4411764705882353, 'gini = 0.0\nsamples = 2\  
value = [0, 2]'),  
  Text(0.43195785776997364, 0.4411764705882353, 'gini = 0.0\nsamples =  
2\nvalue = [2, 0]'),  
  Text(0.4635645302897278, 0.5, 'x[18] <= 0.968\ngini = 0.048\nsamples  
= 121\nvalue = [118, 3]'),
```



```
Text(0.44600526777875327, 0.4411764705882353, 'x[2] <= 0.98\ngini = 0.033\nsamples = 118\nvalue = [116, 2]'),
Text(0.43195785776997364, 0.38235294117647056, 'x[14] <= 0.938\ngini = 0.017\nsamples = 114\nvalue = [113, 1]'),
Text(0.42493415276558383, 0.3235294117647059, 'gini = 0.0\nsamples = 107\nvalue = [107, 0]'),
Text(0.43898156277436345, 0.3235294117647059, 'x[16] <= 0.25\ngini = 0.245\nsamples = 7\nvalue = [6, 1]'),
Text(0.43195785776997364, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.44600526777875327, 0.2647058823529412, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),
Text(0.46005267778753295, 0.38235294117647056, 'x[19] <= 0.833\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.45302897278314314, 0.3235294117647059, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.46707638279192276, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4811237928007024, 0.4411764705882353, 'x[6] <= 0.9\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.4741000877963126, 0.38235294117647056, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.4881474978050922, 0.38235294117647056, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5928446005267779, 0.6764705882352942, 'x[31] <= 0.964\ngini = 0.1\nsamples = 493\nvalue = [467, 26]'),
Text(0.5649692712906058, 0.6176470588235294, 'x[15] <= 0.167\ngini = 0.097\nsamples = 491\nvalue = [466, 25]'),
Text(0.5302897278314311, 0.5588235294117647, 'x[14] <= 0.938\ngini = 0.185\nsamples = 97\nvalue = [87, 10]'),
Text(0.5232660228270413, 0.5, 'x[19] <= 0.944\ngini = 0.17\nsamples = 96\nvalue = [87, 9]'),
Text(0.5092186128182616, 0.4411764705882353, 'x[0] <= 0.44\ngini = 0.142\nsamples = 91\nvalue = [84, 7]'),
Text(0.5021949078138718, 0.38235294117647056, 'x[22] <= 0.75\ngini = 0.241\nsamples = 50\nvalue = [43, 7]'),
Text(0.495171202809482, 0.3235294117647059, 'x[33] <= 0.794\ngini = 0.215\nsamples = 49\nvalue = [43, 6]'),
Text(0.4881474978050922, 0.2647058823529412, 'x[19] <= 0.389\ngini = 0.187\nsamples = 48\nvalue = [43, 5]'),
Text(0.46707638279192276, 0.20588235294117646, 'x[17] <= 0.399\ngini = 0.097\nsamples = 39\nvalue = [37, 2]'),
Text(0.46005267778753295, 0.14705882352941177, 'gini = 0.0\nsamples = 32\nvalue = [32, 0]'),
Text(0.4741000877963126, 0.14705882352941177, 'x[12] <= 0.5\ngini = 0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.46707638279192276, 0.08823529411764706, 'x[10] <= 0.5\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.46005267778753295, 0.029411764705882353, 'gini = 0.0\nsamples
```

```
= 1\nvalue = [1, 0]'),
  Text(0.4741000877963126, 0.029411764705882353, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
  Text(0.4811237928007024, 0.08823529411764706, 'gini = 0.0\nsamples =
4\nvalue = [4, 0]'),
  Text(0.5092186128182616, 0.20588235294117646, 'x[2] <= 0.262\ngini =
0.444\nsamples = 9\nvalue = [6, 3]'),
  Text(0.5021949078138718, 0.14705882352941177, 'x[17] <= 0.075\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
  Text(0.495171202809482, 0.08823529411764706, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
  Text(0.5092186128182616, 0.08823529411764706, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
  Text(0.5162423178226514, 0.14705882352941177, 'gini = 0.0\nsamples =
5\nvalue = [5, 0]'),
  Text(0.5021949078138718, 0.2647058823529412, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.5092186128182616, 0.3235294117647059, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.5162423178226514, 0.38235294117647056, 'gini = 0.0\nsamples =
41\nvalue = [41, 0]'),
  Text(0.5373134328358209, 0.4411764705882353, 'x[17] <= 0.531\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
  Text(0.5302897278314311, 0.38235294117647056, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
  Text(0.5443371378402108, 0.38235294117647056, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
  Text(0.5373134328358209, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
  Text(0.5996488147497805, 0.5588235294117647, 'x[22] <= 0.036\ngini =
0.073\nsamples = 394\nvalue = [379, 15]'),
  Text(0.57243195785777, 0.5, 'x[32] <= 0.7\ngini = 0.198\nsamples =
63\nvalue = [56, 7]'),
  Text(0.5654082528533801, 0.4411764705882353, 'x[28] <= 0.417\ngini =
0.175\nsamples = 62\nvalue = [56, 6]'),
  Text(0.5583845478489904, 0.38235294117647056, 'x[11] <= 0.071\ngini =
0.278\nsamples = 36\nvalue = [30, 6]'),
  Text(0.5513608428446005, 0.3235294117647059, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.5654082528533801, 0.3235294117647059, 'x[15] <= 0.5\ngini =
0.245\nsamples = 35\nvalue = [30, 5]'),
  Text(0.5513608428446005, 0.2647058823529412, 'x[27] <= 0.275\ngini =
0.469\nsamples = 8\nvalue = [5, 3]'),
  Text(0.5443371378402108, 0.20588235294117646, 'x[16] <= 0.25\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
  Text(0.5373134328358209, 0.14705882352941177, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
  Text(0.5513608428446005, 0.14705882352941177, 'gini = 0.0\nsamples =
3\nvalue = [0, 3]'),
```

```
Text(0.5583845478489904, 0.20588235294117646, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.5794556628621598, 0.2647058823529412, 'x[0] <= 0.595\ngini = 0.137\nsamples = 27\nvalue = [25, 2]'),
Text(0.57243195785777, 0.20588235294117646, 'gini = 0.0\nsamples = 21\nvalue = [21, 0]'),
Text(0.5864793678665496, 0.20588235294117646, 'x[9] <= 0.333\ngini = 0.444\nsamples = 6\nvalue = [4, 2]'),
Text(0.5794556628621598, 0.14705882352941177, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.5935030728709394, 0.14705882352941177, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.57243195785777, 0.38235294117647056, 'gini = 0.0\nsamples = 26\nvalue = [26, 0]'),
Text(0.5794556628621598, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6268656716417911, 0.5, 'x[2] <= 0.03\ngini = 0.047\nsamples = 331\nvalue = [323, 8]'),
Text(0.5935030728709394, 0.4411764705882353, 'x[17] <= 0.346\ngini = 0.298\nsamples = 11\nvalue = [9, 2]'),
Text(0.5864793678665496, 0.38235294117647056, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.6005267778753293, 0.38235294117647056, 'x[14] <= 0.625\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.5935030728709394, 0.3235294117647059, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.607550482879719, 0.3235294117647059, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.6602282704126426, 0.4411764705882353, 'x[0] <= 0.655\ngini = 0.037\nsamples = 320\nvalue = [314, 6]'),
Text(0.6356453028972783, 0.38235294117647056, 'x[17] <= 0.056\ngini = 0.015\nsamples = 265\nvalue = [263, 2]'),
Text(0.6215978928884986, 0.3235294117647059, 'x[28] <= 0.25\ngini = 0.32\nsamples = 5\nvalue = [4, 1]'),
Text(0.6145741878841089, 0.2647058823529412, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6286215978928885, 0.2647058823529412, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.6496927129060579, 0.3235294117647059, 'x[33] <= 0.088\ngini = 0.008\nsamples = 260\nvalue = [259, 1]'),
Text(0.6426690079016681, 0.2647058823529412, 'x[14] <= 0.312\ngini = 0.083\nsamples = 23\nvalue = [22, 1]'),
Text(0.6356453028972783, 0.20588235294117646, 'x[24] <= 0.167\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.6286215978928885, 0.14705882352941177, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6426690079016681, 0.14705882352941177, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.6496927129060579, 0.20588235294117646, 'gini = 0.0\nsamples =
```

```
19\nvalue = [19, 0]'),
Text(0.6567164179104478, 0.2647058823529412, 'gini = 0.0\nsamples =
237\nvalue = [237, 0]'),
Text(0.684811237928007, 0.38235294117647056, 'x[17] <= 0.49\ngini =
0.135\nsamples = 55\nvalue = [51, 4]'),
Text(0.6777875329236172, 0.3235294117647059, 'x[17] <= 0.32\ngini =
0.346\nsamples = 18\nvalue = [14, 4]'),
Text(0.6707638279192274, 0.2647058823529412, 'gini = 0.0\nsamples =
13\nvalue = [13, 0]'),
Text(0.684811237928007, 0.2647058823529412, 'x[16] <= 0.25\ngini =
0.32\nsamples = 5\nvalue = [1, 4]'),
Text(0.6777875329236172, 0.20588235294117646, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
Text(0.6918349429323969, 0.20588235294117646, 'gini = 0.0\nsamples =
4\nvalue = [0, 4]'),
Text(0.6918349429323969, 0.3235294117647059, 'gini = 0.0\nsamples =
37\nvalue = [37, 0]'),
Text(0.62071992976295, 0.6176470588235294, 'x[32] <= 0.6\ngini = 0.5\
nsamples = 2\nvalue = [1, 1]'),
Text(0.6136962247585601, 0.5588235294117647, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.6277436347673397, 0.5588235294117647, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
Text(0.4928665496049166, 0.7352941176470589, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.8147497805092186, 0.8529411764705882, 'x[17] <= 0.157\ngini =
0.385\nsamples = 300\nvalue = [222, 78]'),
Text(0.723441615452151, 0.7941176470588235, 'x[26] <= 0.167\ngini =
0.5\nsamples = 96\nvalue = [49, 47]'),
Text(0.6874451273046532, 0.7352941176470589, 'x[4] <= 0.161\ngini =
0.459\nsamples = 42\nvalue = [15, 27]'),
Text(0.6628621597892889, 0.6764705882352942, 'x[8] <= 0.415\ngini =
0.499\nsamples = 23\nvalue = [12, 11]'),
Text(0.6488147497805092, 0.6176470588235294, 'x[18] <= 0.561\ngini =
0.355\nsamples = 13\nvalue = [3, 10]'),
Text(0.6417910447761194, 0.5588235294117647, 'gini = 0.0\nsamples =
8\nvalue = [0, 8]'),
Text(0.655838454784899, 0.5588235294117647, 'x[9] <= 0.333\ngini =
0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.6488147497805092, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
Text(0.6628621597892889, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
Text(0.6769095697980685, 0.6176470588235294, 'x[3] <= 0.75\ngini =
0.18\nsamples = 10\nvalue = [9, 1]'),
Text(0.6698858647936786, 0.5588235294117647, 'gini = 0.0\nsamples =
8\nvalue = [8, 0]'),
Text(0.6839332748024582, 0.5588235294117647, 'x[5] <= 0.625\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
```

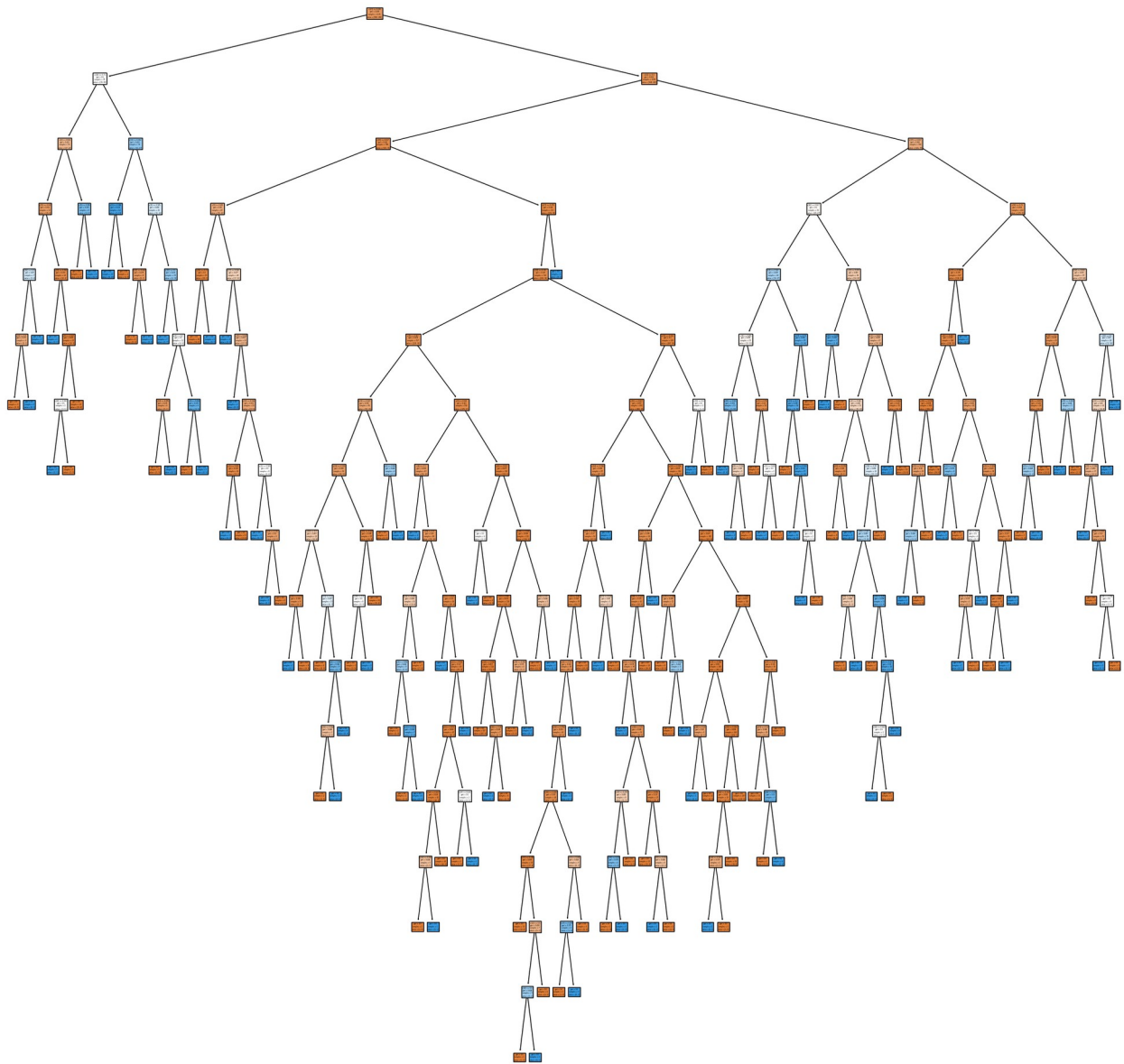
```
Text(0.6769095697980685, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6909569798068481, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7120280948200176, 0.6764705882352942, 'x[27] <= 0.35\ngini = 0.266\nsamples = 19\nvalue = [3, 16]'),
Text(0.7050043898156277, 0.6176470588235294, 'x[11] <= 0.2\ngini = 0.198\nsamples = 18\nvalue = [2, 16]'),
Text(0.6979806848112379, 0.5588235294117647, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7120280948200176, 0.5588235294117647, 'x[31] <= 0.393\ngini = 0.111\nsamples = 17\nvalue = [1, 16]'),
Text(0.7050043898156277, 0.5, 'gini = 0.0\nsamples = 15\nvalue = [0, 15]'),
Text(0.7190517998244074, 0.5, 'x[0] <= 0.44\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.7120280948200176, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7260755048287972, 0.4411764705882353, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7190517998244074, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7594381035996488, 0.7352941176470589, 'x[0] <= 0.202\ngini = 0.466\nsamples = 54\nvalue = [34, 20]'),
Text(0.7401229148375769, 0.6764705882352942, 'x[12] <= 0.833\ngini = 0.245\nsamples = 7\nvalue = [1, 6]'),
Text(0.733099209833187, 0.6176470588235294, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.7471466198419666, 0.6176470588235294, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7787532923617209, 0.6764705882352942, 'x[2] <= 0.622\ngini = 0.418\nsamples = 47\nvalue = [33, 14]'),
Text(0.7611940298507462, 0.6176470588235294, 'x[2] <= 0.145\ngini = 0.482\nsamples = 32\nvalue = [19, 13]'),
Text(0.7471466198419666, 0.5588235294117647, 'x[30] <= 0.679\ngini = 0.18\nsamples = 10\nvalue = [9, 1]'),
Text(0.7401229148375769, 0.5, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),
Text(0.7541703248463565, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7752414398595259, 0.5588235294117647, 'x[18] <= 0.87\ngini = 0.496\nsamples = 22\nvalue = [10, 12]'),
Text(0.7682177348551361, 0.5, 'x[8] <= 0.41\ngini = 0.465\nsamples = 19\nvalue = [7, 12]'),
Text(0.7541703248463565, 0.4411764705882353, 'x[18] <= 0.715\ngini = 0.469\nsamples = 8\nvalue = [5, 3]'),
Text(0.7471466198419666, 0.38235294117647056, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.7611940298507462, 0.38235294117647056, 'gini = 0.0\nsamples =
```

```
3\nvalue = [0, 3]'),
Text(0.7822651448639157, 0.4411764705882353, 'x[0] <= 0.25\ngini =
0.298\nsamples = 11\nvalue = [2, 9]'),
Text(0.7752414398595259, 0.38235294117647056, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
Text(0.7892888498683055, 0.38235294117647056, 'x[3] <= 0.25\ngini =
0.18\nsamples = 10\nvalue = [1, 9]'),
Text(0.7822651448639157, 0.3235294117647059, 'x[18] <= 0.063\ngini =
0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.7752414398595259, 0.2647058823529412, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.7892888498683055, 0.2647058823529412, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
Text(0.7963125548726954, 0.3235294117647059, 'gini = 0.0\nsamples =
8\nvalue = [0, 8]'),
Text(0.7822651448639157, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
Text(0.7963125548726954, 0.6176470588235294, 'x[11] <= 0.064\ngini =
0.124\nsamples = 15\nvalue = [14, 1]'),
Text(0.7892888498683055, 0.5588235294117647, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
Text(0.8033362598770851, 0.5588235294117647, 'gini = 0.0\nsamples =
14\nvalue = [14, 0]'),
Text(0.9060579455662863, 0.7941176470588235, 'x[16] <= 0.75\ngini =
0.258\nsamples = 204\nvalue = [173, 31]'),
Text(0.8507462686567164, 0.7352941176470589, 'x[17] <= 0.992\ngini =
0.138\nsamples = 147\nvalue = [136, 11]'),
Text(0.8437225636523266, 0.6764705882352942, 'x[4] <= 0.482\ngini =
0.128\nsamples = 146\nvalue = [136, 10]'),
Text(0.8244073748902546, 0.6176470588235294, 'x[30] <= 0.179\ngini =
0.038\nsamples = 104\nvalue = [102, 2]'),
Text(0.8173836698858647, 0.5588235294117647, 'x[11] <= 0.193\ngini =
0.32\nsamples = 10\nvalue = [8, 2]'),
Text(0.810359964881475, 0.5, 'x[17] <= 0.599\ngini = 0.444\nsamples =
3\nvalue = [1, 2]'),
Text(0.8033362598770851, 0.4411764705882353, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
Text(0.8173836698858647, 0.4411764705882353, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
Text(0.8244073748902546, 0.5, 'gini = 0.0\nsamples = 7\nvalue = [7,
0]'),
Text(0.8314310798946444, 0.5588235294117647, 'gini = 0.0\nsamples =
94\nvalue = [94, 0]'),
Text(0.8630377524143986, 0.6176470588235294, 'x[9] <= 0.167\ngini =
0.308\nsamples = 42\nvalue = [34, 8]'),
Text(0.845478489903424, 0.5588235294117647, 'x[2] <= 0.736\ngini =
0.375\nsamples = 4\nvalue = [1, 3]'),
Text(0.8384547848990342, 0.5, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
Text(0.8525021949078139, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1,
```

```
0]'),
  Text(0.8805970149253731, 0.5588235294117647, 'x[0] <= 0.393\ngini =
0.229\nsamples = 38\nvalue = [33, 5]'),
  Text(0.8665496049165935, 0.5, 'x[2] <= 0.51\ngini = 0.5\nsamples = 6\
nvalue = [3, 3]'),
  Text(0.8595258999122037, 0.4411764705882353, 'x[12] <= 0.333\ngini =
0.375\nsamples = 4\nvalue = [3, 1]'),
  Text(0.8525021949078139, 0.38235294117647056, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.8665496049165935, 0.38235294117647056, 'gini = 0.0\nsamples =
3\nvalue = [3, 0]'),
  Text(0.8735733099209834, 0.4411764705882353, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
  Text(0.8946444249341527, 0.5, 'x[28] <= 0.917\ngini = 0.117\nsamples
= 32\nvalue = [30, 2]'),
  Text(0.887620719929763, 0.4411764705882353, 'x[8] <= 0.992\ngini =
0.062\nsamples = 31\nvalue = [30, 1]'),
  Text(0.8805970149253731, 0.38235294117647056, 'gini = 0.0\nsamples =
30\nvalue = [30, 0]'),
  Text(0.8946444249341527, 0.38235294117647056, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.9016681299385426, 0.4411764705882353, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.8577699736611062, 0.6764705882352942, 'gini = 0.0\nsamples =
1\nvalue = [0, 1]'),
  Text(0.961369622475856, 0.7352941176470589, 'x[14] <= 0.812\ngini =
0.456\nsamples = 57\nvalue = [37, 20]'),
  Text(0.9367866549604916, 0.6764705882352942, 'x[32] <= 0.4\ngini =
0.238\nsamples = 29\nvalue = [25, 4]'),
  Text(0.922739244951712, 0.6176470588235294, 'x[8] <= 0.071\ngini =
0.142\nsamples = 26\nvalue = [24, 2]'),
  Text(0.9157155399473222, 0.5588235294117647, 'x[0] <= 0.321\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
  Text(0.9086918349429324, 0.5, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
  Text(0.922739244951712, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
  Text(0.9297629499561019, 0.5588235294117647, 'gini = 0.0\nsamples =
23\nvalue = [23, 0]'),
  Text(0.9508340649692713, 0.6176470588235294, 'x[29] <= 0.833\ngini =
0.444\nsamples = 3\nvalue = [1, 2]'),
  Text(0.9438103599648815, 0.5588235294117647, 'gini = 0.0\nsamples =
2\nvalue = [0, 2]'),
  Text(0.9578577699736611, 0.5588235294117647, 'gini = 0.0\nsamples =
1\nvalue = [1, 0]'),
  Text(0.9859525899912204, 0.6764705882352942, 'x[32] <= 0.1\ngini =
0.49\nsamples = 28\nvalue = [12, 16]'),
  Text(0.9789288849868305, 0.6176470588235294, 'x[12] <= 0.833\ngini =
0.48\nsamples = 20\nvalue = [12, 8]'),
```

```
Text(0.9719051799824407, 0.5588235294117647, 'x[30] <= 0.036\ngini = 0.415\nsamples = 17\nvalue = [12, 5]'),  
Text(0.9648814749780509, 0.5, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),  
Text(0.9789288849868305, 0.5, 'x[18] <= 0.505\ngini = 0.32\nsamples = 15\nvalue = [12, 3]'),  
Text(0.9719051799824407, 0.4411764705882353, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),  
Text(0.9859525899912204, 0.4411764705882353, 'x[18] <= 0.706\ngini = 0.5\nsamples = 6\nvalue = [3, 3]'),  
Text(0.9789288849868305, 0.38235294117647056, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),  
Text(0.9929762949956101, 0.38235294117647056, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),  
Text(0.9859525899912204, 0.5588235294117647, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),  
Text(0.9929762949956101, 0.6176470588235294, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]')]
```





```

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']
}

grid_search =
GridSearchCV(estimator=dt,param_grid=parameter,cv=5,scoring='accuracy'
)
grid_search.fit(x_train,y_train)

```

```
C:\Users\tejos\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 "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\
model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\base.py",
line 1144, in wrapper
    estimator._validate_params()
  File "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\base.py",
line 637, in _validate_params
    validate_parameter_constraints(
  File "C:\Users\tejos\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
{'sqrt', 'log2'} or None. Got 'auto' instead.
```

```
warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\tejos\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	0.83928597	0.84013704	0.82993148	0.83758384
nan	nan	0.83420483	0.84013704	0.83417598	0.84183916
nan	nan	0.83671114	0.84014064	0.83332853	0.84099171
nan	nan	0.83078255	0.83248467	0.82737108	0.83163722
nan	nan	0.84013704	0.84013704	0.84013704	0.84013704
nan	nan	0.84013704	0.84013704	0.83928597	0.83758384
nan	nan	0.84012982	0.84013704	0.83673999	0.84183556
nan	nan	0.84183195	0.84098449	0.83077533	0.8409881
nan	nan	0.83164082	0.83929318	0.82993509	0.84862964]

```
warnings.warn(
```

```
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')

grid_search.best_params_

{'criterion': 'entropy',
 'max_depth': 5,
 'max_features': 'log2',
 'splitter': 'random'}

dt_cv =
DecisionTreeClassifier(criterion='entropy',max_depth=3,max_features='s
qrt')
dt_cv.fit(x_train,y_train)

DecisionTreeClassifier(criterion='entropy', max_depth=3,
max_features='sqrt')

pred = dt_cv.predict(x_test)

print(classification_report(y_test,pred))

```

	precision	recall	f1-score	support
0	0.84	1.00	0.91	245
1	1.00	0.04	0.08	49
accuracy			0.84	294
macro avg	0.92	0.52	0.50	294
weighted avg	0.87	0.84	0.77	294

## Random Forest

```

from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()

forest_params = [{'max_depth':list(range(10,15)), 'max_features':
['auto', 'sqrt', 'log2']}]

rf_cv = GridSearchCV(rf,param_grid = forest_params,cv =
10,scoring='accuracy')
rf_cv.fit(x_train,y_train)

C:\Users\tejos\anaconda\Lib\site-packages\sklearn\model_selection\
_validation.py:425: FitFailedWarning:
50 fits failed out of a total of 150.
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 "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\  
model_selection\_validation.py", line 732, in _fit_and_score  
    estimator.fit(X_train, y_train, **fit_params)  
  File "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\base.py",  
line 1144, in wrapper  
    estimator._validate_params()  
  File "C:\Users\tejos\anaconda\Lib\site-packages\sklearn\base.py",  
line 637, in _validate_params  
    validate_parameter_constraints(  
  File "C:\Users\tejos\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  
{ 'sqrt', 'log2' } or None. Got 'auto' instead.  
  
warnings.warn(some_fits_failed_message, FitFailedWarning)  
C:\Users\tejos\anaconda\Lib\site-packages\sklearn\model_selection\  
_search.py:976: UserWarning: One or more of the test scores are non-  
finite: [          nan 0.8605389  0.85797479          nan 0.85967695  
0.8605389  
          nan 0.86224105 0.85543966          nan 0.85543966 0.85714907  
          nan 0.85799652 0.85799652]  
warnings.warn(  
  
GridSearchCV(cv=10, estimator=RandomForestClassifier(),  
             param_grid=[{'max_depth': [10, 11, 12, 13, 14],  
                           'max_features': ['auto', 'sqrt', 'log2']}],  
             scoring='accuracy')  
  
pred = rf_cv.predict(x_test)  
accuracy_score(y_test, pred)  
0.8503401360544217  
  
print(classification_report(y_test, pred))
```

	precision	recall	f1-score	support
0	0.86	0.99	0.92	245
1	0.73	0.16	0.27	49
accuracy			0.85	294
macro avg	0.79	0.58	0.59	294

weighted avg	0.83	0.85	0.81	294
--------------	------	------	------	-----

```
rf_cv.best_params_
```

```
{'max_depth': 12, 'max_features': 'sqrt'}
```

```
rf.fit(x_train,y_train)
```

```
RandomForestClassifier()
```

```
probabo = rf.predict_proba(x_test)[: ,1]
```

```
probabo
```

```
array([0.07, 0.03, 0.14, 0.17, 0.69, 0.18, 0.29, 0.08, 0.07, 0.17,
0.06,
      0.16, 0.08, 0.37, 0.05, 0.02, 0.16, 0.15, 0.12, 0.24, 0.56,
0.05,
      0.01, 0.06, 0.29, 0.27, 0.03, 0.06, 0.64, 0.02, 0.07, 0.12,
0.26,
      0.1 , 0.09, 0.04, 0.18, 0.11, 0.11, 0.08, 0.14, 0.08, 0.11, 0.1
,
      0.04, 0.46, 0.33, 0.02, 0.68, 0.44, 0.16, 0.45, 0.25, 0.15, 0.4
,
      0.16, 0.06, 0.06, 0.08, 0.34, 0.08, 0.16, 0.08, 0.14, 0.25,
0.12,
      0.22, 0.16, 0.05, 0.17, 0.15, 0.35, 0.17, 0.14, 0.1 , 0.12, 0.1
,
      0.03, 0.32, 0.06, 0.07, 0.07, 0.15, 0.12, 0.14, 0.03, 0.04, 0.2
,
      0.06, 0.1 , 0.59, 0.1 , 0.16, 0.22, 0.08, 0.07, 0.1 , 0.34,
0.11,
      0.11, 0.22, 0.2 , 0.28, 0.08, 0.07, 0.05, 0.13, 0.43, 0.24,
0.18,
      0.2 , 0.07, 0.04, 0.05, 0.11, 0.14, 0.06, 0.14, 0.07, 0.01,
0.15,
      0.02, 0.01, 0.63, 0.26, 0.13, 0.03, 0.08, 0.1 , 0.09, 0.02,
0.31,
      0.37, 0.21, 0.23, 0.16, 0.36, 0.18, 0.1 , 0.1 , 0.11, 0.12, 0.1
,
      0.1 , 0.17, 0.11, 0.08, 0.06, 0.14, 0.09, 0.07, 0.23, 0.08,
0.29,
      0.16, 0.17, 0.07, 0.17, 0.17, 0.04, 0.33, 0.03, 0.3 , 0.62,
0.05,
      0.22, 0.15, 0.06, 0.1 , 0.09, 0.07, 0.15, 0.11, 0.2 , 0.11,
0.48,
      0.07, 0.24, 0.21, 0.13, 0.02, 0. , 0.02, 0.52, 0.03, 0.08,
0.22,
      0.08, 0.2 , 0.23, 0.29, 0.63, 0.19, 0.06, 0.14, 0.06, 0.1 ,
0.04,
      0.39, 0.14, 0.27, 0.16, 0.28, 0.08, 0.2 , 0.16, 0.46, 0.06,
```

```

0.07,
    0.01, 0.1 , 0.14, 0.15, 0.2 , 0.05, 0.05, 0.06, 0.15, 0.14,
0.46,
    0.13, 0.41, 0.35, 0.16, 0.16, 0.06, 0.06, 0.31, 0.53, 0.13,
0.06,
    0.26, 0.04, 0.1 , 0.03, 0.25, 0.28, 0.04, 0.16, 0.11, 0.21,
0.12,
    0.1 , 0.08, 0.19, 0.1 , 0.08, 0.25, 0.09, 0.13, 0.16, 0.18,
0.39,
    0.05, 0.23, 0.15, 0.03, 0.57, 0.24, 0.29, 0.07, 0.08, 0.16,
0.09,
    0.16, 0.23, 0.04, 0.19, 0.16, 0.04, 0.1 , 0.01, 0.06, 0.09,
0.03,
    0.09, 0.1 , 0.46, 0.13, 0.34, 0.26, 0.12, 0.18, 0.15, 0.15,
0.24,
    0.03, 0.13, 0.03, 0.08, 0.16, 0.27, 0.07, 0.12])
fpr,tpr,threshsholds = roc_curve(y_test,probabo)
plt.plot(fpr,tpr)
[<matplotlib.lines.Line2D at 0x1a1c5d78d50>]

```

