

import libraries

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

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

df.head()

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Emplo
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	

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.shape

(1470, 35)

df.Attrition.value_counts()

```
No      1233
Yes     237
Name: Attrition, dtype: int64

df.corr()

<ipython-input-17-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to 'ignore'. To silence this warning, you can pass numeric_only=False.
df.corr()
```

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction
Age	1.000000	0.010661	-0.001686	0.208034	NaN	-0.010145	0.01
DailyRate	0.010661	1.000000	-0.004985	-0.016806	NaN	-0.050990	0.01
DistanceFromHome	-0.001686	-0.004985	1.000000	0.021042	NaN	0.032916	-0.01
Education	0.208034	-0.016806	0.021042	1.000000	NaN	0.042070	-0.02
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeNumber	-0.010145	-0.050990	0.032916	0.042070	NaN	1.000000	0.01
EnvironmentSatisfaction	0.010146	0.018355	-0.016075	-0.027128	NaN	0.017621	1.00
HourlyRate	0.024287	0.023381	0.031131	0.016775	NaN	0.035179	-0.04
JobInvolvement	0.029820	0.046135	0.008783	0.042438	NaN	-0.006888	-0.00
JobLevel	0.509604	0.002966	0.005303	0.101589	NaN	-0.018519	0.00
JobSatisfaction	-0.004892	0.030571	-0.003669	-0.011296	NaN	-0.046247	-0.00
MonthlyIncome	0.497855	0.007707	-0.017014	0.094961	NaN	-0.014829	-0.00
MonthlyRate	0.028051	-0.032182	0.027473	-0.026084	NaN	0.012648	0.00
NumCompaniesWorked	0.299635	0.038153	-0.029251	0.126317	NaN	-0.001251	0.01
PercentSalaryHike	0.003634	0.022704	0.040235	-0.011111	NaN	-0.012944	-0.00
PerformanceRating	0.001904	0.000473	0.027110	-0.024539	NaN	-0.020359	-0.00
RelationshipSatisfaction	0.053535	0.007846	0.006557	-0.009118	NaN	-0.069861	0.00
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	0.037510	0.042143	0.044872	0.018422	NaN	0.062227	0.00
TotalWorkingYears	0.680381	0.014515	0.004628	0.148280	NaN	-0.014365	-0.00
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	-0.025100	NaN	0.023603	-0.01
WorkLifeBalance	-0.021490	-0.037848	-0.026556	0.009819	NaN	0.010309	0.00
YearsAtCompany	0.311309	-0.034055	0.009508	0.069114	NaN	-0.011240	0.00
YearsInCurrentRole	0.212901	0.009932	0.018845	0.060236	NaN	-0.008416	0.01
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	0.054254	NaN	-0.009019	0.01
YearsWithCurrManager	0.202089	-0.026363	0.014406	0.069065	NaN	-0.009197	-0.00

26 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

https://colab.research.google.com/drive/1wtvIbM7oMnk_EpLhMgPFwMa8QkpLnynxT#scrollTo=ieoTZ5RbsFvd&printMode=true

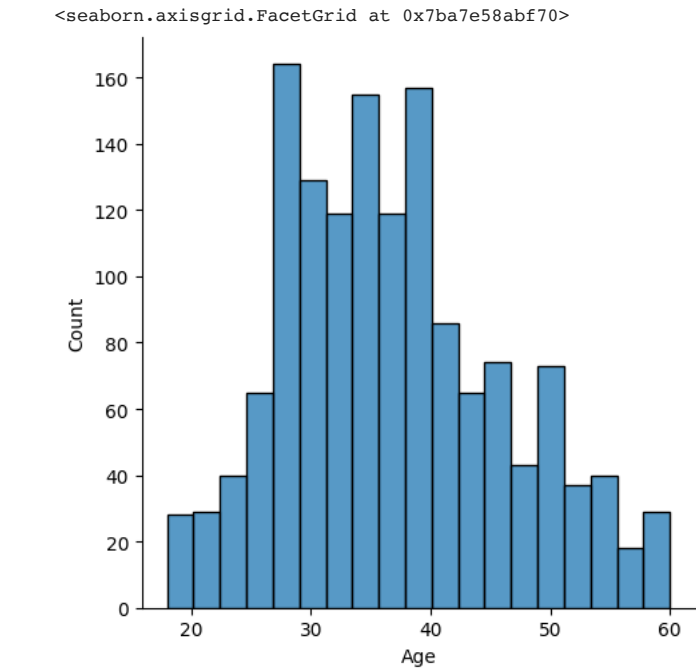
2/21

```
RelationshipSatisfaction    False
StandardHours               False
StockOptionLevel           False
TotalWorkingYears          False
TrainingTimesLastYear       False
WorkLifeBalance            False
YearsAtCompany              False
YearsInCurrentRole          False
YearsSinceLastPromotion     False
YearsWithCurrManager        False
dtype: bool
```

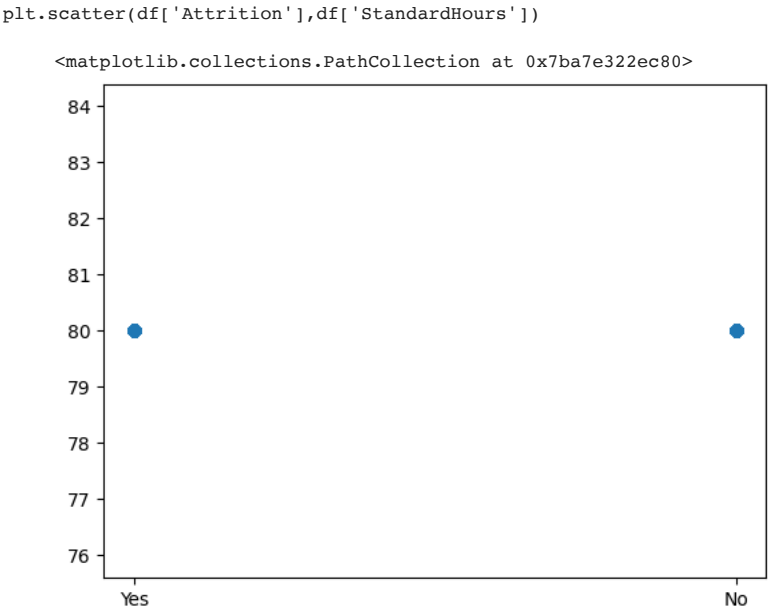
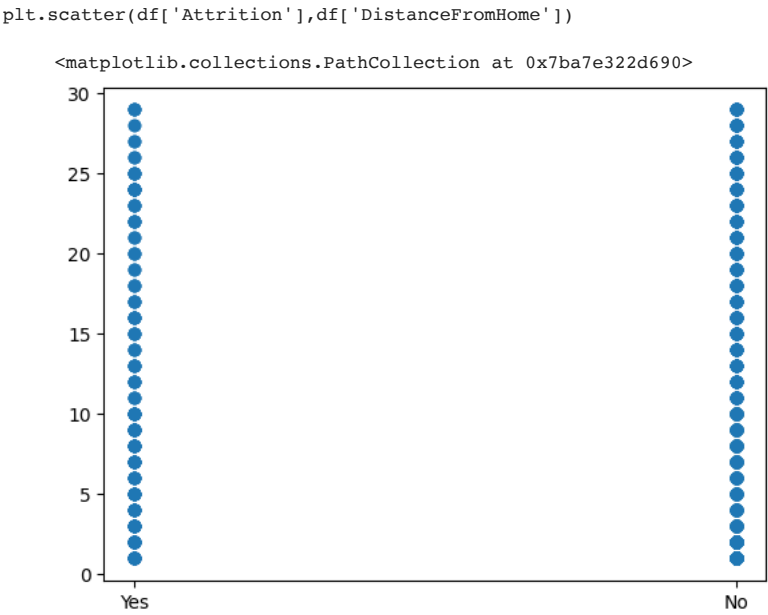
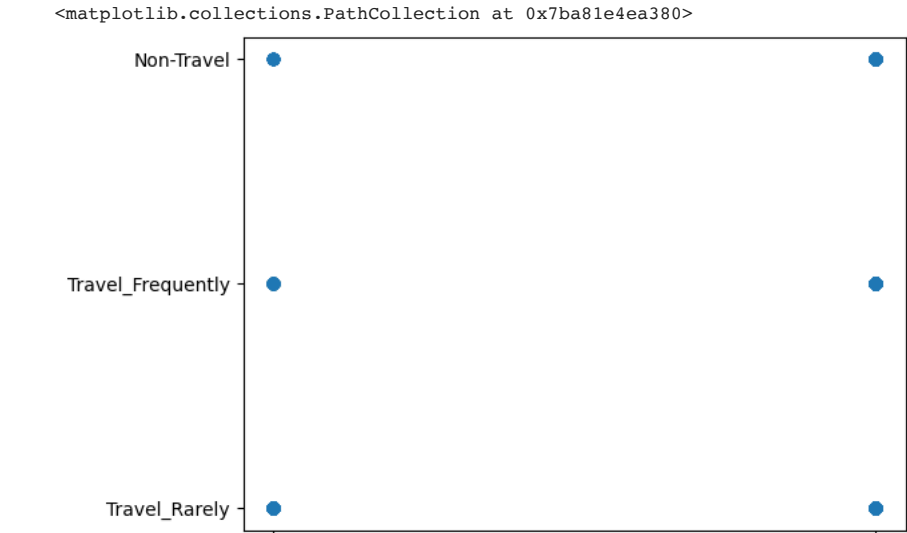
```
df.isnull().sum()
```

```
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
```

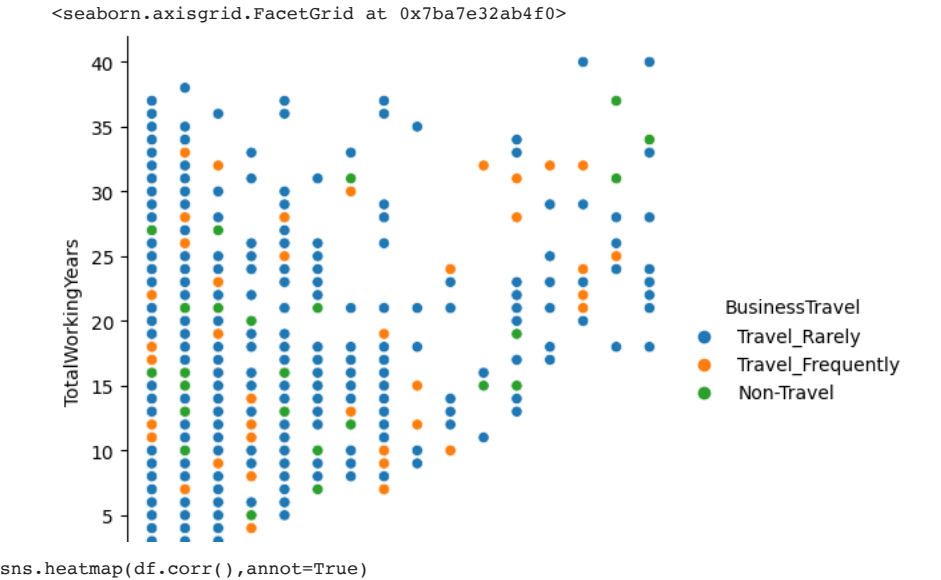
```
sns.displot(df["Age"])
```



```
plt.scatter(df["Attrition"],df["BusinessTravel"])
```



```
sns.relplot(x="YearsSinceLastPromotion",y="TotalWorkingYears",data=df,hue="BusinessTravel")
```

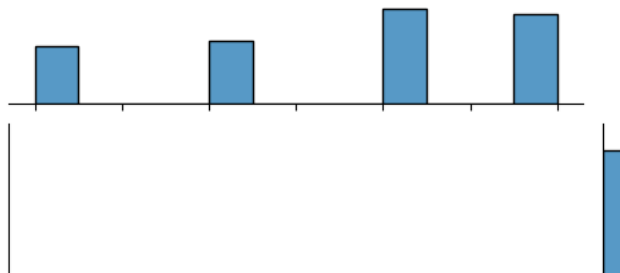


```
<ipython-input-25-8df7bcac526d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a
sns.heatmap(df.corr(),annot=True)
<Axes: >
```



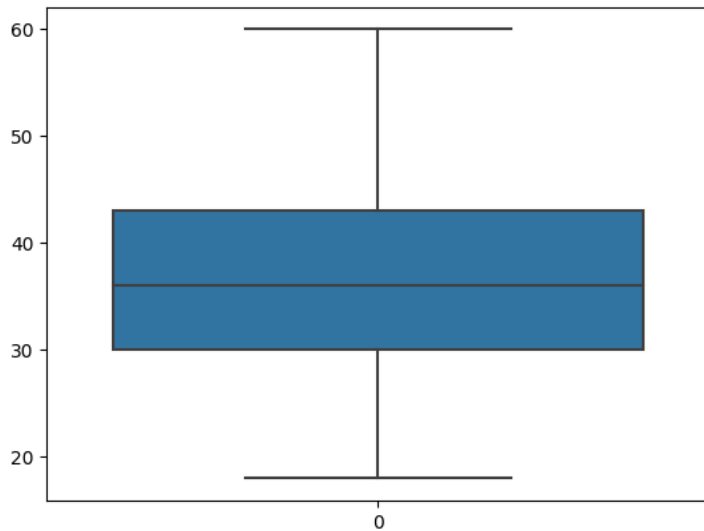
```
sns.jointplot(x="RelationshipSatisfaction",y="Attrition",data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7ba7e137dcf0>
```



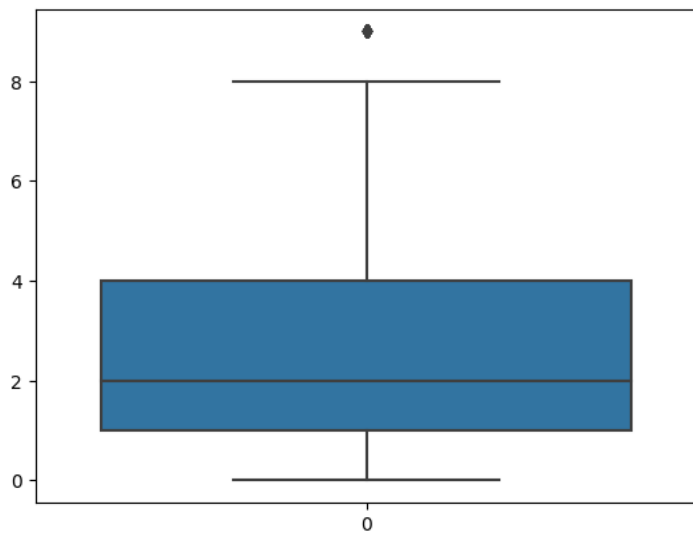
```
sns.boxplot(df.Age)
```

```
<Axes: >
```



```
sns.boxplot(df.NumCompaniesWorked)
```

```
<Axes: >
```



```
q1=df.NumCompaniesWorked.quantile(0.25)
q3=df.NumCompaniesWorked.quantile(0.75)
```

```
print(q1)
print(q3)
```

```
1.0
4.0
```

```
IQR=q3-q1
```

```
IQR
```

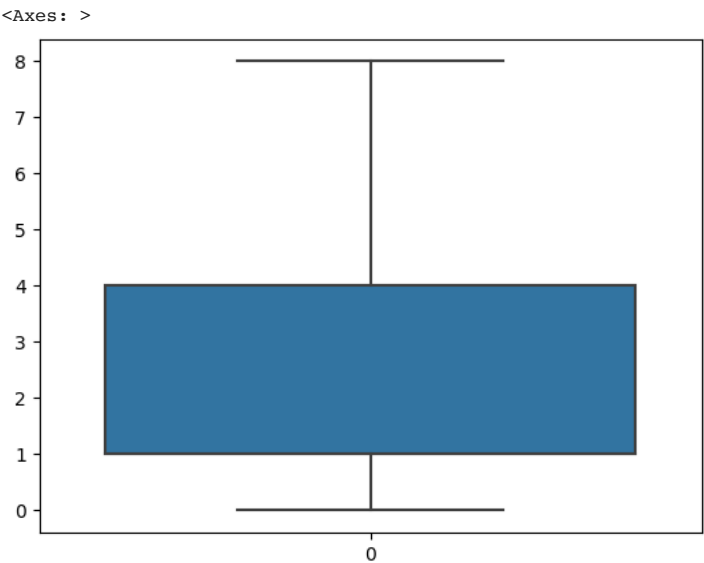
```
3.0
```

```
upper_limit=q3+1.5*IQR
upper_limit
```

8.5

```
df=df[df.NumCompaniesWorked<upper_limit]
```

```
sns.boxplot(df.NumCompaniesWorked)
```



```
#dependent variable
y=df.Attrition
```

```
y.head()
```

```
0    Yes
1    No
2    Yes
3    No
5    No
Name: Attrition, dtype: object
```

```
#independent variable
x=df.drop(["Attrition"],axis=1)
```

```
x.head()
```

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1
1	49	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2
2	37	Travel_Rarely	1373	Research & Development	2	2	Other	1	4
3	33	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5
5	32	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8

5 rows x 34 columns

```
x.shape
```

(1418, 34)

```
y.shape
```

(1418,)

```
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Edu
0	41	Yes	Travel_Rarely	1102	Sales		1
1	49	No	Travel_Frequently	279	Research & Development		8
2	37	Yes	Travel_Rarely	1373	Research & Development		2
3	33	No	Travel_Frequently	1392	Research & Development		3
5	32	No	Travel_Frequently	1005	Research & Development		2

5 rows x 35 columns

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
x["BusinessTravel"]=le.fit_transform(x["BusinessTravel"])
```

```
x["BusinessTravel"]
```

0	2
1	1
2	2
3	1
5	1
..	
1465	1
1466	2
1467	2
1468	1
1469	2

Name: BusinessTravel, Length: 1418, dtype: int64

```
x.head()
```

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Edu
0	41	2	1102	Sales		1	2
1	49	1	279	Research & Development		8	1
2	37	2	1373	Research & Development		2	2
3	33	1	1392	Research & Development		3	4
5	32	1	1005	Research & Development		2	2

5 rows x 34 columns

```
x["Department"]=le.fit_transform(x["Department"])
```

```
x["Department"]
```

0	2
1	1
2	1
3	1
5	1
..	
1465	1
1466	1
1467	1
1468	2
1469	1

Name: Department, Length: 1418, dtype: int64

```
x.head()
```


	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	2	1102	2	1	2	Life Sciences	1	1
1	49	1	279	1	8	1	Life Sciences	1	2
2	37	2	1373	1	2	2	Other	1	4

x["EducationField"]=le.fit_transform(x["EducationField"])

x["EducationField"]

```
0      1
1      1
2      4
3      1
5      1
..
1465   3
1466   3
1467   1
1468   3
1469   3
Name: EducationField, Length: 1418, dtype: int64
```

x.head()

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	2	1102	2	1	2	1	1	1
1	49	1	279	1	8	1	1	1	2
2	37	2	1373	1	2	2	4	1	4
3	33	1	1392	1	3	4	1	1	5
5	32	1	1005	1	2	2	1	1	8

5 rows x 34 columns

non_numeric_columns = x.select_dtypes(exclude=['number']).columns

print(non_numeric_columns)

```
Index(['Gender', 'JobRole', 'MaritalStatus', 'Over18', 'OverTime'], dtype='object')
```

x["Gender"]=le.fit_transform(x["Gender"])

x["Gender"]

```
0      0
1      1
2      1
3      0
5      1
..
1465   1
1466   1
1467   1
1468   1
1469   1
Name: Gender, Length: 1418, dtype: int64
```

x.head()

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	2	1102	2	1	2	1	1	1
1	49	1	279	1	8	1	1	1	2
2	37	2	1373	1	2	2	4	1	4
3	33	1	1392	1	3	4	1	1	5
5	32	1	1005	1	2	2	1	1	8

5 rows x 34 columns

```
x["JobRole"]=le.fit_transform(x["JobRole"])
```

```
x["JobRole"]
```

```
0      7
1      6
2      2
3      6
5      2
..
1465    2
1466    0
1467    4
1468    7
1469    2
Name: JobRole, Length: 1418, dtype: int64
```

```
x["MaritalStatus"]=le.fit_transform(x["MaritalStatus"])
```

```
x["MaritalStatus"]
```

```
0      2
1      1
2      2
3      1
5      2
..
1465    1
1466    1
1467    1
1468    1
1469    1
Name: MaritalStatus, Length: 1418, dtype: int64
```

```
x["Over18"]=le.fit_transform(x["Over18"])
```

```
x["Over18"]
```

```
0      0
1      0
2      0
3      0
5      0
..
1465    0
1466    0
1467    0
1468    0
1469    0
Name: Over18, Length: 1418, dtype: int64
```

```
x["OverTime"]=le.fit_transform(x["OverTime"])
```

```
x["OverTime"]
```

```
0      1
1      0
2      1
3      1
5      0
..
1465    0
1466    0
1467    1
1468    0
1469    0
Name: OverTime, Length: 1418, dtype: int64
```

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

```
x_scaled
```


28/09/2023, 23:01Assignment4.ipynb - Colaboratory

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Emj
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
5	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	
...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	

1418 rows x 35 columns

▾ Evaluation of classification model

```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_auc_score,roc_curve
```

```
accuracy_score(y_test,pred)

0.8697183098591549
```

```
confusion_matrix(y_test,pred)

array([[237,  3],
       [ 34, 10]])
```

```
pd.crosstab(y_test,pred)
```

col_0	No	Yes
Attrition		
No	237	3
Yes	34	10

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

	precision	recall	f1-score	support
No	0.87	0.99	0.93	240
Yes	0.77	0.23	0.35	44
accuracy			0.87	284
macro avg	0.82	0.61	0.64	284
weighted avg	0.86	0.87	0.84	284

```
#ROC_AUC Curve
```

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

array([0.13657895, 0.03742004, 0.08053736, 0.08659374, 0.023358 ,
       0.10563069, 0.13815154, 0.00229225, 0.06771379, 0.12744425,
       0.08172802, 0.05965762, 0.0638561 , 0.12855128, 0.2275486 ,
       0.08936636, 0.06240484, 0.09603478, 0.21199145, 0.05717384,
       0.01180209, 0.00367791, 0.07898725, 0.02473968, 0.11962886,
       0.12904799, 0.0184306 , 0.0365714 , 0.02049336, 0.10008116,
       0.16143025, 0.03099261, 0.05571065, 0.04469354, 0.21600549,
```

```

0.42230677, 0.2197372, 0.5227653, 0.18101958, 0.10182865,
0.03088844, 0.18054679, 0.08248226, 0.01733578, 0.19733818,
0.06725397, 0.01197982, 0.01366601, 0.02702768, 0.18659878,
0.04323244, 0.00445696, 0.05192806, 0.1866853, 0.1632088,
0.27853238, 0.07437663, 0.09816652, 0.00573849, 0.00449716,
0.0059488, 0.03111943, 0.00839901, 0.00669404, 0.04253402,
0.18695255, 0.19941885, 0.03278527, 0.00238087, 0.01663221,
0.58136087, 0.1578733, 0.21711936, 0.03898385, 0.04521495,
0.03220001, 0.06616953, 0.19809653, 0.10991992, 0.22934288,
0.05904098, 0.02037218, 0.66970453, 0.26829173, 0.08216447,
0.04010601, 0.11590138, 0.27057603, 0.22694055, 0.20450222,
0.56793147, 0.22053355, 0.36393157, 0.01755166, 0.01233427,
0.01492107, 0.2081514, 0.12205625, 0.40315397, 0.04856193,
0.07330096, 0.25379683, 0.14516211, 0.28647266, 0.02781388,
0.18391223, 0.26396952, 0.01946723, 0.28598072, 0.04347479,
0.15563751, 0.13357455, 0.00963796, 0.02116195, 0.07528362,
0.05922541, 0.11977388, 0.00903596, 0.36455439, 0.05168354,
0.20310448, 0.01231492, 0.05158269, 0.57453501, 0.07656055,
0.03508536, 0.30385493, 0.0309728, 0.42983322, 0.02371366,
0.05130702, 0.02103465, 0.04602763, 0.01905589, 0.32734204,
0.19614051, 0.06294798, 0.0186783, 0.00440507, 0.12521514,
0.35937712, 0.01824423, 0.03851794, 0.36623505, 0.0761209,
0.26592758, 0.03553327, 0.02772604, 0.0193432, 0.28332535,
0.31642215, 0.02571374, 0.12136821, 0.32580669, 0.13472202,
0.06624905, 0.08617629, 0.03661786, 0.01839348, 0.15357873,
0.39926896, 0.71257736, 0.89315923, 0.00546009, 0.00246771,
0.02778452, 0.05857899, 0.36399558, 0.01646451, 0.14794275,
0.47711028, 0.0384135, 0.01739, 0.04238425, 0.20976761,
0.54481958, 0.02510394, 0.01863455, 0.24136931, 0.06312414,
0.03643677, 0.00616726, 0.1100783, 0.15064248, 0.07821613,
0.10409581, 0.20971698, 0.13795456, 0.28657845, 0.02226441,
0.23272876, 0.23596972, 0.16844684, 0.00414635, 0.03126561,
0.44815074, 0.01643598, 0.10900941, 0.01603778, 0.0333788,
0.27797218, 0.14158042, 0.05577601, 0.09399929, 0.24091949,
0.09998247, 0.01242131, 0.02205424, 0.1890573, 0.06235382,
0.09115454, 0.00728886, 0.19906759, 0.1575069, 0.20840636,
0.13738917, 0.05410298, 0.18636277, 0.08545779, 0.2373784,
0.04893286, 0.28718093, 0.07707427, 0.25024676, 0.11690009,
0.05663235, 0.06336832, 0.1402614, 0.09635028, 0.5603858,
0.07966128, 0.18409077, 0.00949154, 0.04702311, 0.16756119,
0.03001824, 0.51420487, 0.00555785, 0.09370631, 0.01171392,
0.12695966, 0.03659918, 0.3821563, 0.13188418, 0.17530265,
0.19609367, 0.10202889, 0.74603311, 0.05622724, 0.15448205,
0.17941515, 0.07061508, 0.07724554, 0.11220407, 0.19871038,
0.08215216, 0.00188234, 0.15323164, 0.06851284, 0.02069078,
0.71737346, 0.17804198, 0.15215912, 0.00469619, 0.23093543,
0.03742954, 0.06874542, 0.45373149, 0.6448183, 0.09910567,
0.3574995, 0.02215789, 0.00967421, 0.07067802, 0.35407627,
0.31550123, 0.01930184, 0.08248221, 0.07689043, 0.01921869,
0.13324521, 0.08754501, 0.22298726, 0.42007529])

```

DecisionTreeClassifier

```

from sklearn.tree import DecisionTreeClassifier
dtc=DecisionTreeClassifier()

```

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

```

▼ DecisionTreeClassifier
DecisionTreeClassifier()

```

```
pred=dtc.predict(x_test)
```

```
pred
```

```

array(['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', 'Yes', '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', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No',
       'No', 'No', 'Yes', 'No', 'No', 'Yes', 'No', 'Yes', 'No', 'No',
       'No', 'No', 'Yes', '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',
       'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No',
       'No', 'Yes', 'No', 'Yes', 'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes',
       'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'No', 'No',
       'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No',

```

```
'No', 'Yes', 'No', 'No', 'No', 'Yes', 'Yes', 'No', 'Yes', 'No',
'No', 'No', 'No', 'No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'No',
'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'Yes', 'Yes', 'No',
'Yes', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No', 'No', 'No',
'Yes', 'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'No', 'No',
'Yes', 'No', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No',
'Yes', 'No', 'No', 'No', 'No', 'No', 'Yes', 'No', 'No', 'No',
'No', 'No', 'No', 'No', 'No', 'No', 'No', dtype=object)
```

y_test

```
451    No
639    No
832    No
1287   No
1277   No
..
521    No
550    No
1113   No
335    No
917    No
Name: Attrition, Length: 284, dtype: object
```

df

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	Em
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	
5	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	
...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	

1418 rows x 35 columns

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

probability

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

Hyper parameter tuning

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

```
[Text(0.45738636363636365, 0.9722222222222222, 'x[21] <= 0.5\ngini = 0.268\nsamples = 1134\nvalue = [953, 181]'),
Text(0.15522727272727272, 0.9166666666666666, 'x[17] <= 0.048\ngini = 0.185\nsamples = 806\nvalue = [723, 83]'),
Text(0.014545454545454545, 0.8611111111111112, 'x[2] <= 0.169\ngini = 0.5\nsamples = 18\nvalue = [9, 9]'),
Text(0.007272727272727273, 0.8055555555555556, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.02181818181818182, 0.8055555555555556, 'x[15] <= 0.167\ngini = 0.459\nsamples = 14\nvalue = [9, 5]'),
Text(0.014545454545454545, 0.75, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.02909090909090909, 0.75, 'x[0] <= 0.262\ngini = 0.298\nsamples = 11\nvalue = [9, 2]'),
Text(0.02181818181818182, 0.6944444444444444, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.03636363636363636, 0.6944444444444444, 'x[2] <= 0.62\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.02909090909090909, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.04363636363636364, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.2959090909090909, 0.8611111111111112, 'x[26] <= 0.167\ngini = 0.17\nsamples = 788\nvalue = [714, 74]'),
Text(0.14772727272727273, 0.8055555555555556, 'x[29] <= 0.167\ngini = 0.249\nsamples = 337\nvalue = [288, 49]'),
Text(0.08, 0.75, 'x[4] <= 0.25\ngini = 0.498\nsamples = 17\nvalue = [9, 8]'),
Text(0.06545454545454546, 0.6944444444444444, 'x[32] <= 0.333\ngini = 0.32\nsamples = 10\nvalue = [8, 2]'),
Text(0.05818181818181818, 0.6388888888888888, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.07272727272727272, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.09454545454545454, 0.6944444444444444, 'x[4] <= 0.696\ngini = 0.245\nsamples = 7\nvalue = [1, 6]'),
Text(0.08727272727272728, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
Text(0.10181818181818182, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.21545454545454545, 0.75, 'x[6] <= 0.9\ngini = 0.223\nsamples = 320\nvalue = [279, 41]'),
Text(0.15818181818181817, 0.6944444444444444, 'x[15] <= 0.167\ngini = 0.185\nsamples = 290\nvalue = [260, 30]'),
Text(0.11636363636363636, 0.6388888888888888, 'x[30] <= 0.163\ngini = 0.346\nsamples = 54\nvalue = [42, 12]'),
Text(0.09818181818181818, 0.5833333333333334, 'x[19] <= 0.188\ngini = 0.493\nsamples = 25\nvalue = [14, 11]'),
Text(0.08363636363636363, 0.5277777777777778, 'x[5] <= 0.75\ngini = 0.165\nsamples = 11\nvalue = [10, 1]'),
Text(0.07636363636363637, 0.4722222222222222, 'gini = 0.0\nsamples = 10\nvalue = [10, 0]'),
Text(0.09090909090909091, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.11272727272727273, 0.5277777777777778, 'x[0] <= 0.452\ngini = 0.408\nsamples = 14\nvalue = [4, 10]'),
Text(0.10545454545454545, 0.4722222222222222, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]'),
Text(0.12, 0.4722222222222222, 'x[30] <= 0.063\ngini = 0.444\nsamples = 6\nvalue = [4, 2]'),
Text(0.11272727272727273, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.12727272727272726, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.13454545454545455, 0.5833333333333334, 'x[17] <= 0.095\ngini = 0.067\nsamples = 29\nvalue = [28, 1]'),
Text(0.12727272727272726, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.14181818181818182, 0.5277777777777778, 'gini = 0.0\nsamples = 28\nvalue = [28, 0]'),
Text(0.2, 0.6388888888888888, 'x[12] <= 0.167\ngini = 0.141\nsamples = 236\nvalue = [218, 18]'),
Text(0.16363636363636364, 0.5833333333333334, 'x[2] <= 0.217\ngini = 0.444\nsamples = 12\nvalue = [8, 4]'),
Text(0.15636363636363637, 0.5277777777777778, 'gini = 0.0\nsamples = 6\nvalue = [6, 0]'),
Text(0.1709090909090909, 0.5277777777777778, 'x[24] <= 0.167\ngini = 0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.16363636363636364, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.17818181818181818, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.23636363636363636, 0.5833333333333334, 'x[4] <= 0.982\ngini = 0.117\nsamples = 224\nvalue = [210, 14]'),
Text(0.21454545454545454, 0.5277777777777778, 'x[2] <= 0.939\ngini = 0.104\nsamples = 219\nvalue = [207, 12]'),
Text(0.19272727272727272, 0.4722222222222222, 'x[2] <= 0.044\ngini = 0.089\nsamples = 214\nvalue = [204, 10]'),
Text(0.1709090909090909, 0.4166666666666667, 'x[18] <= 0.905\ngini = 0.305\nsamples = 16\nvalue = [13, 3]'),
Text(0.16363636363636364, 0.3611111111111111, 'x[32] <= 0.4\ngini = 0.133\nsamples = 14\nvalue = [13, 1]'),
Text(0.15636363636363637, 0.3055555555555556, 'gini = 0.0\nsamples = 13\nvalue = [13, 0]'),
Text(0.1709090909090909, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.17818181818181818, 0.3611111111111111, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.21454545454545454, 0.4166666666666667, 'x[6] <= 0.1\ngini = 0.068\nsamples = 198\nvalue = [191, 7]'),
Text(0.19272727272727272, 0.3611111111111111, 'x[12] <= 0.5\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.18545454545454546, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.2, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.23636363636363636, 0.3611111111111111, 'x[33] <= 0.088\ngini = 0.06\nsamples = 195\nvalue = [189, 6]'),
Text(0.21454545454545454, 0.3055555555555556, 'x[4] <= 0.643\ngini = 0.145\nsamples = 51\nvalue = [47, 4]'),
Text(0.2, 0.25, 'x[4] <= 0.054\ngini = 0.083\nsamples = 46\nvalue = [44, 2]'),
Text(0.19272727272727272, 0.19444444444444445, 'x[19] <= 0.812\ngini = 0.278\nsamples = 12\nvalue = [10, 2]'),
Text(0.18545454545454546, 0.13888888888888889, 'x[0] <= 0.274\ngini = 0.165\nsamples = 11\nvalue = [10, 1]'),
Text(0.17818181818181818, 0.08333333333333333, 'x[19] <= 0.438\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.1709090909090909, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.18545454545454546, 0.027777777777777776, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.19272727272727272, 0.08333333333333333, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
Text(0.2, 0.13888888888888889, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.20727272727272728, 0.19444444444444445, 'gini = 0.0\nsamples = 34\nvalue = [34, 0]'),
Text(0.2290909090909091, 0.25, 'x[14] <= 0.375\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.22181818181818183, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.23636363636363636, 0.19444444444444445, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.25818181818181818, 0.3055555555555556, 'x[0] <= 0.655\ngini = 0.027\nsamples = 144\nvalue = [142, 2]'),
Text(0.2509090909090909, 0.25, 'gini = 0.0\nsamples = 115\nvalue = [115, 0]'),
Text(0.26545454545454544, 0.25, 'x[24] <= 0.167\ngini = 0.128\nsamples = 29\nvalue = [27, 2]'),
Text(0.2509090909090909, 0.19444444444444445, 'x[27] <= 0.587\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.24363636363636362, 0.13888888888888889, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.25818181818181818, 0.13888888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.28, 0.19444444444444445, 'x[0] <= 0.69\ngini = 0.071\nsamples = 27\nvalue = [26, 1]'),
Text(0.27272727272727272, 0.13888888888888889, 'x[9] <= 0.5\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
Text(0.26545454545454544, 0.08333333333333333, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.28, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.28727272727272727, 0.13888888888888889, 'gini = 0.0\nsamples = 24\nvalue = [24, 0]'),
Text(0.23636363636363636, 0.4722222222222222, 'x[6] <= 0.7\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.2290909090909091, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.24363636363636362, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.25818181818181818, 0.5277777777777778, 'x[1] <= 0.75\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.2509090909090909, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.26545454545454544, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.27272727272727272, 0.6944444444444444, 'x[0] <= 0.274\ngini = 0.464\nsamples = 30\nvalue = [19, 11]'),
Text(0.25818181818181818, 0.6388888888888888, 'x[22] <= 0.036\ngini = 0.32\nsamples = 10\nvalue = [2, 8]'),
Text(0.2509090909090909, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.26545454545454544, 0.5833333333333334, 'gini = 0.0\nsamples = 8\nvalue = [0, 8]'),
Text(0.28727272727272727, 0.6388888888888888, 'x[11] <= 0.321\ngini = 0.255\nsamples = 20\nvalue = [17, 3]'),
Text(0.28, 0.5833333333333334, 'x[11] <= 0.25\ngini = 0.48\nsamples = 5\nvalue = [2, 3]'),
```



```

Text(0.2727272727272727, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.2872727272727273, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.2945454545454545, 0.5833333333333334, 'gini = 0.0\nsamples = 15\nvalue = [15, 0]'),
Text(0.4440909090909091, 0.8055555555555556, 'x[30] <= 0.787\ngini = 0.105\nsamples = 451\nvalue = [426, 25]'),
Text(0.3990909090909091, 0.75, 'x[17] <= 0.072\ngini = 0.094\nsamples = 447\nvalue = [425, 22]'),
Text(0.3309090909090909, 0.6944444444444444, 'x[29] <= 0.5\ngini = 0.249\nsamples = 55\nvalue = [47, 8]'),
Text(0.3163636363636364, 0.6388888888888889, 'x[11] <= 0.129\ngini = 0.475\nsamples = 18\nvalue = [11, 7]'),
Text(0.3090909090909091, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.3236363636363637, 0.5833333333333334, 'x[11] <= 0.507\ngini = 0.391\nsamples = 15\nvalue = [11, 4]'),
Text(0.3163636363636364, 0.5277777777777778, 'gini = 0.0\nsamples = 9\nvalue = [9, 0]'),
Text(0.3309090909090909, 0.5277777777777778, 'x[8] <= 0.383\ngini = 0.444\nsamples = 6\nvalue = [2, 4]'),
Text(0.3236363636363637, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.3381818181818182, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.3454545454545455, 0.6388888888888889, 'x[8] <= 0.022\ngini = 0.053\nsamples = 37\nvalue = [36, 1]'),
Text(0.3381818181818182, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.3527272727272727, 0.5833333333333334, 'gini = 0.0\nsamples = 36\nvalue = [36, 0]'),
Text(0.4672727272727273, 0.6944444444444444, 'x[2] <= 0.997\ngini = 0.069\nsamples = 392\nvalue = [378, 14]'),
Text(0.4290909090909091, 0.6388888888888889, 'x[22] <= 0.036\ngini = 0.064\nsamples = 390\nvalue = [377, 13]'),
Text(0.3745454545454546, 0.5833333333333334, 'x[17] <= 0.352\ngini = 0.183\nsamples = 59\nvalue = [53, 6]'),
Text(0.36, 0.5277777777777778, 'x[4] <= 0.071\ngini = 0.053\nsamples = 37\nvalue = [36, 1]'),
Text(0.3527272727272727, 0.4722222222222222, 'x[9] <= 0.833\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
Text(0.3454545454545455, 0.4166666666666667, 'x[27] <= 0.137\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.3381818181818182, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.3527272727272727, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.36, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.3672727272727273, 0.4722222222222222, 'gini = 0.0\nsamples = 31\nvalue = [31, 0]'),
Text(0.3890909090909091, 0.5277777777777778, 'x[2] <= 0.486\ngini = 0.351\nsamples = 22\nvalue = [17, 5]'),
Text(0.3818181818181818, 0.4722222222222222, 'x[33] <= 0.353\ngini = 0.494\nsamples = 9\nvalue = [4, 5]'),
Text(0.3745454545454546, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]'),
Text(0.3890909090909091, 0.4166666666666667, 'x[13] <= 0.375\ngini = 0.32\nsamples = 5\nvalue = [4, 1]'),
Text(0.3818181818181818, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.3963636363636364, 0.3611111111111111, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.3963636363636364, 0.4722222222222222, 'gini = 0.0\nsamples = 13\nvalue = [13, 0]'),
Text(0.4836363636363637, 0.5833333333333334, 'x[8] <= 0.986\ngini = 0.041\nsamples = 331\nvalue = [324, 7]'),
Text(0.4618181818181818, 0.5277777777777778, 'x[26] <= 0.833\ngini = 0.036\nsamples = 327\nvalue = [321, 6]'),
Text(0.44, 0.4722222222222222, 'x[2] <= 0.117\ngini = 0.02\nsamples = 294\nvalue = [291, 3]'),
Text(0.4181818181818182, 0.4166666666666667, 'x[8] <= 0.868\ngini = 0.114\nsamples = 33\nvalue = [31, 2]'),
Text(0.4109090909090909, 0.3611111111111111, 'gini = 0.0\nsamples = 30\nvalue = [30, 0]'),
Text(0.4254545454545455, 0.3611111111111111, 'x[22] <= 0.357\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4181818181818182, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.4327272727272727, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.4618181818181818, 0.4166666666666667, 'x[0] <= 0.179\ngini = 0.008\nsamples = 261\nvalue = [260, 1]'),
Text(0.4545454545454546, 0.3611111111111111, 'x[19] <= 0.438\ngini = 0.153\nsamples = 12\nvalue = [11, 1]'),
Text(0.4472727272727273, 0.3055555555555556, 'gini = 0.0\nsamples = 10\nvalue = [10, 0]'),
Text(0.4618181818181818, 0.3055555555555556, 'x[6] <= 0.4\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.4545454545454546, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.4690909090909091, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4690909090909091, 0.3611111111111111, 'gini = 0.0\nsamples = 249\nvalue = [249, 0]'),
Text(0.4836363636363637, 0.4722222222222222, 'x[2] <= 0.117\ngini = 0.165\nsamples = 33\nvalue = [30, 3]'),
Text(0.4763636363636364, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4909090909090909, 0.4166666666666667, 'x[24] <= 0.167\ngini = 0.117\nsamples = 32\nvalue = [30, 2]'),
Text(0.4836363636363637, 0.3611111111111111, 'x[4] <= 0.268\ngini = 0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.4763636363636364, 0.3055555555555556, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
Text(0.4909090909090909, 0.3055555555555556, 'x[4] <= 0.589\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.4836363636363637, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.4981818181818182, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.4981818181818182, 0.3611111111111111, 'gini = 0.0\nsamples = 25\nvalue = [25, 0]'),
Text(0.5054545454545455, 0.5277777777777778, 'x[3] <= 0.75\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.4981818181818182, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.5127272727272727, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.5054545454545455, 0.6388888888888889, 'x[11] <= 0.521\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.4981818181818182, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5127272727272727, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.4890909090909091, 0.75, 'x[11] <= 0.95\ngini = 0.375\nsamples = 4\nvalue = [1, 3]'),
Text(0.4818181818181818, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue = [0, 3]'),
Text(0.4963636363636364, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7595454545454546, 0.9166666666666667, 'x[17] <= 0.13\ngini = 0.419\nsamples = 328\nvalue = [230, 98]'),
Text(0.6318181818181818, 0.8611111111111112, 'x[26] <= 0.167\ngini = 0.498\nsamples = 107\nvalue = [50, 57]'),
Text(0.5818181818181818, 0.8055555555555556, 'x[4] <= 0.161\ngini = 0.427\nsamples = 55\nvalue = [17, 38]'),
Text(0.5490909090909091, 0.75, 'x[8] <= 0.4\ngini = 0.499\nsamples = 27\nvalue = [14, 13]'),
Text(0.5272727272727273, 0.6944444444444444, 'x[18] <= 0.561\ngini = 0.337\nsamples = 14\nvalue = [3, 11]'),
Text(0.52, 0.6388888888888889, 'gini = 0.0\nsamples = 9\nvalue = [0, 9]'),
Text(0.5345454545454546, 0.6388888888888889, 'x[28] <= 0.583\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.5272727272727273, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.5418181818181819, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.5709090909090909, 0.6944444444444444, 'x[2] <= 0.33\ngini = 0.26\nsamples = 13\nvalue = [11, 2]'),
Text(0.5636363636363637, 0.6388888888888889, 'x[15] <= 0.333\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.5563636363636364, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5709090909090909, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.5781818181818182, 0.6388888888888889, 'gini = 0.0\nsamples = 10\nvalue = [10, 0]'),
Text(0.6145454545454546, 0.75, 'x[4] <= 0.875\ngini = 0.191\nsamples = 28\nvalue = [3, 25]'),
Text(0.6072727272727273, 0.6944444444444444, 'x[12] <= 0.833\ngini = 0.137\nsamples = 27\nvalue = [2, 25]'),
Text(0.5927272727272728, 0.6388888888888889, 'x[19] <= 0.062\ngini = 0.077\nsamples = 25\nvalue = [1, 24]'),
Text(0.5854545454545455, 0.5833333333333334, 'x[23] <= 0.5\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.5781818181818182, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.5927272727272728, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6, 0.5833333333333334, 'gini = 0.0\nsamples = 23\nvalue = [0, 23]'),
Text(0.6218181818181818, 0.6388888888888889, 'x[6] <= 0.4\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.6145454545454546, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.6290909090909091, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.6218181818181818, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),

```

```

Text(0.6818181818181818, 0.8055555555555556, 'x[33] <= 0.029\ngini = 0.464\nsamples = 52\nvalue = [33, 19]'),
Text(0.6581818181818182, 0.75, 'x[2] <= 0.717\ngini = 0.457\nsamples = 17\nvalue = [6, 11]'),
Text(0.6436363636363637, 0.6944444444444444, 'x[5] <= 0.125\ngini = 0.18\nsamples = 10\nvalue = [1, 9]'),
Text(0.6363636363636364, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.6509090909090909, 0.6388888888888888, 'gini = 0.0\nsamples = 9\nvalue = [0, 9]'),
Text(0.6727272727272727, 0.6944444444444444, 'x[30] <= 0.013\ngini = 0.408\nsamples = 7\nvalue = [5, 2]'),
Text(0.6654545454545454, 0.6388888888888888, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.68, 0.6388888888888888, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
Text(0.7054545454545454, 0.75, 'x[17] <= 0.057\ngini = 0.353\nsamples = 35\nvalue = [27, 8]'),
Text(0.6981818181818182, 0.6944444444444444, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.7127272727272728, 0.6944444444444444, 'x[32] <= 0.433\ngini = 0.298\nsamples = 33\nvalue = [27, 6]'),
Text(0.6945454545454546, 0.6388888888888888, 'x[0] <= 0.202\ngini = 0.231\nsamples = 30\nvalue = [26, 4]'),
Text(0.68, 0.5833333333333334, 'x[17] <= 0.079\ngini = 0.5\nsamples = 4\nvalue = [2, 2]'),
Text(0.6727272727272727, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.6872727272727273, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
Text(0.7090909090909091, 0.5833333333333334, 'x[19] <= 0.562\ngini = 0.142\nsamples = 26\nvalue = [24, 2]'),
Text(0.7018181818181818, 0.5277777777777778, 'gini = 0.0\nsamples = 21\nvalue = [21, 0]'),
Text(0.7163636363636363, 0.5277777777777778, 'x[18] <= 0.772\ngini = 0.48\nsamples = 5\nvalue = [3, 2]'),
Text(0.7090909090909091, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.7236363636363636, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.7309090909090909, 0.6388888888888888, 'x[18] <= 0.312\ngini = 0.444\nsamples = 3\nvalue = [1, 2]'),
Text(0.7236363636363636, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
Text(0.7381818181818182, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
Text(0.8872727272727273, 0.8611111111111112, 'x[16] <= 0.75\ngini = 0.302\nsamples = 221\nvalue = [180, 41]'),
Text(0.8363636363636363, 0.8055555555555556, 'x[19] <= 0.562\ngini = 0.192\nsamples = 158\nvalue = [141, 17]'),
Text(0.7963636363636364, 0.75, 'x[0] <= 0.119\ngini = 0.136\nsamples = 136\nvalue = [126, 10]'),
Text(0.7890909090909091, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.8036363636363636, 0.6944444444444444, 'x[4] <= 0.589\ngini = 0.124\nsamples = 135\nvalue = [126, 9]'),
Text(0.7672727272727273, 0.6388888888888888, 'x[12] <= 0.167\ngini = 0.039\nsamples = 101\nvalue = [99, 2]'),
Text(0.7527272727272727, 0.5833333333333334, 'x[8] <= 0.141\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.7454545454545455, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.76, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.7818181818181819, 0.5833333333333334, 'x[28] <= 0.917\ngini = 0.02\nsamples = 97\nvalue = [96, 1]'),
Text(0.7745454545454545, 0.5277777777777778, 'gini = 0.0\nsamples = 93\nvalue = [93, 0]'),
Text(0.7890909090909091, 0.5277777777777778, 'x[18] <= 0.303\ngini = 0.375\nsamples = 4\nvalue = [3, 1]'),
Text(0.7818181818181819, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.7963636363636364, 0.4722222222222222, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
Text(0.84, 0.6388888888888888, 'x[14] <= 0.812\ngini = 0.327\nsamples = 34\nvalue = [27, 7]'),
Text(0.8254545454545454, 0.5833333333333334, 'x[11] <= 0.95\ngini = 0.137\nsamples = 27\nvalue = [25, 2]'),
Text(0.8181818181818182, 0.5277777777777778, 'x[4] <= 0.982\ngini = 0.074\nsamples = 26\nvalue = [25, 1]'),
Text(0.8109090909090909, 0.4722222222222222, 'gini = 0.0\nsamples = 24\nvalue = [24, 0]'),
Text(0.8254545454545454, 0.4722222222222222, 'x[1] <= 0.5\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
Text(0.8181818181818182, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
Text(0.8272727272727273, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]')

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']
}

Text(0.8836363636363637, 0.5833333333333334, 'x[31] <= 0.083\ngini = 0.32\nsamples = 3\nvalue = [1, 4]'),

grid_search=GridSearchCV(estimator=dtc,param_grid=parameter,cv=5,scoring="accuracy")
Text(0.8981818181818182, 0.5833333333333334, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),

grid_search.fit(x_train,y_train)

```

19/21

[illegible]

```
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269: FutureWarning: `max_features='auto'` has been depre
warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269: FutureWarning: `max_features='auto'` has been depre
grid_search.best_params_

{'criterion': 'gini',
 'max_depth': 3,
 'max_features': 'auto',
 'splitter': 'best'}

/usr/local/lib/python3.10/dist-packages/sklearn/tree/_classes.py:269: FutureWarning: `max_features='auto'` has been depre
dtc_cv=DecisionTreeClassifier(criterion= 'entropy',
max_depth=3,
max_features='sqrt',
splitter='best')
dtc_cv.fit(x_train,y_train)
```

▼

DecisionTreeClassifier

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

```
! estimator: DecisionTreeClassifier
pred=dtc_cv.predict(x_test)
!
print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
No	0.85	1.00	0.92	240
Yes	0.00	0.00	0.00	44
accuracy			0.85	284
macro avg	0.42	0.50	0.46	284
weighted avg	0.71	0.85	0.77	284

```
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.10/dist-packages/sklearn/metrics/_classification.py:1344: UndefinedMetricWarning: Precision and F-
_warn_prf(average, modifier, msg_start, len(result))
```

► RANDOM FOREST

[] ↪ 8 cells hidden