Assignment 4

Bala Murugan V V

21BIT0725

In [1]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

In [2]:
```
data = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv')
data.head()
```

Out[2]:

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

5 rows × 35 columns

In [3]:
```
data.shape
```

Out[3]:
```
(1470, 35)
```

In [4]:
```
data.describe()
```

Out[4]:

	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000

8 rows × 26 columns

In [5]:
```
data.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
```

In [6]:
```
data.isnull().sum()
```

Out[6]:
```
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
```

In [7]:
```
data.corr()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_8564\2627137660.py:1: FutureWarning: The d
efault 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.
  data.corr()
```

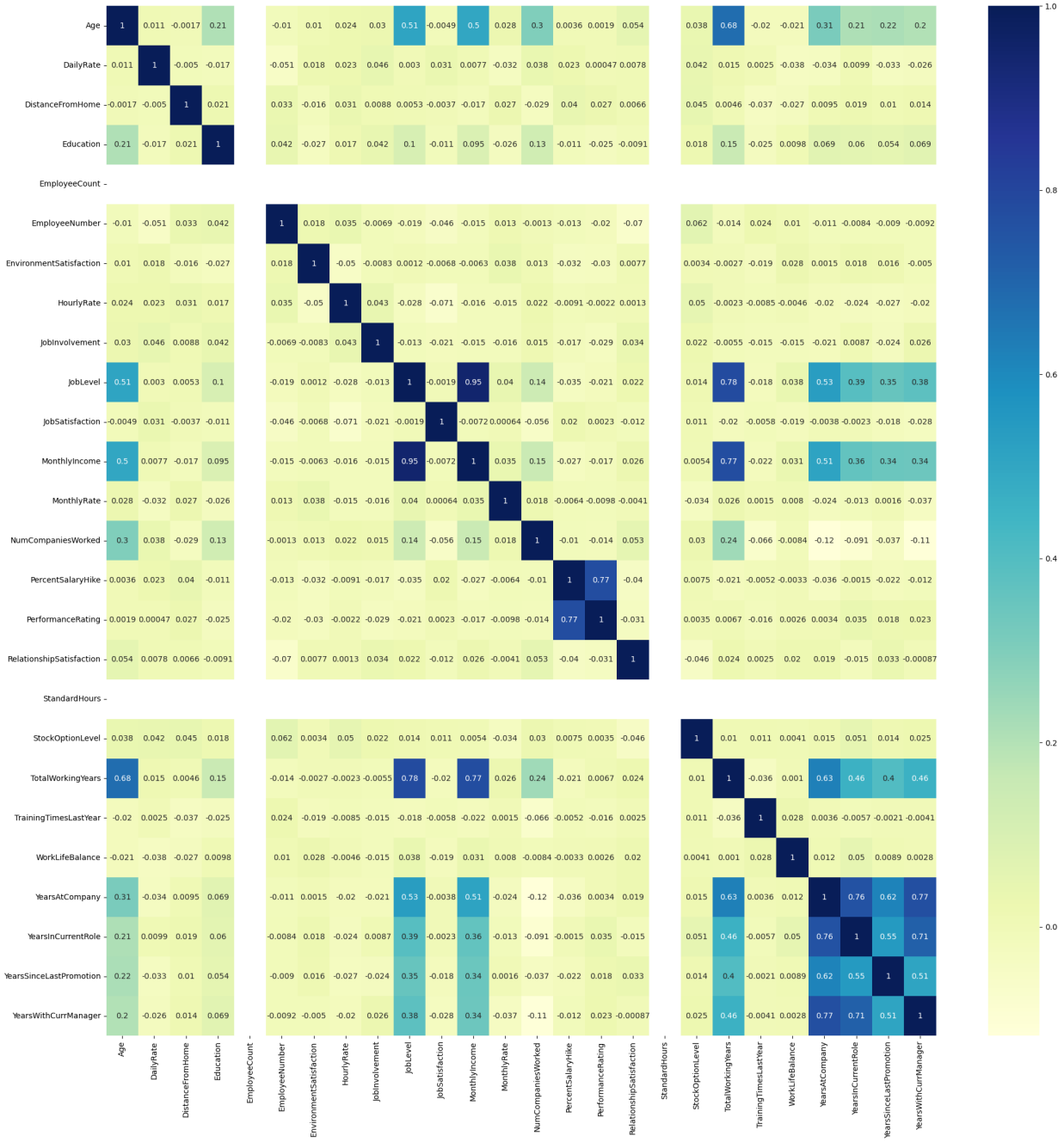
Out[7]:

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

26 rows × 26 columns

In [8]:
```
plt.figure(figsize =(24,24))
sns.heatmap(data.corr(),annot = True,cmap = "YlGnBu")
plt.show()
```

```
C:\Users\Asus\AppData\Local\Temp\ipykernel_8564\2009265951.py:2: FutureWarning: The d
efault 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.
  sns.heatmap(data.corr(),annot = True,cmap = "YlGnBu")
```



In [9]:
```
data  = data.drop(columns = ['EmployeeCount','StandardHours','EmployeeNumber'])
```

In [10]:
```
data.head()
```

Out[10]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	Education
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sci
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sci
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	(
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sci
4	27	No	Travel_Rarely	591	Research & Development	2	1	Me

5 rows × 32 columns

In [11]:
```
attrition_count = pd.DataFrame(data['Attrition'].value_counts())
attrition_count
```

Out[11]:

	Attrition
No	1233
Yes	237

In [12]:
```
plt.pie(attrition_count['Attrition'] , labels = ['No' , 'Yes'] , explode = (0.15,0))
```

Out[12]:
```
([<matplotlib.patches.Wedge at 0x270888b5c10>,
  <matplotlib.patches.Wedge at 0x27088980990>],
 [Text(-1.0930587195656423, 0.606401381579494, 'No'),
  Text(0.961891673217765, -0.5336332157899547, 'Yes')])
```

In [13]:
```
attrition_dummies = pd.get_dummies(data['Attrition'])
data = pd.concat([data, attrition_dummies] , axis = 1)
```

In [14]:
```
data = data.drop(['Attrition','No'],axis = 1)
```

In [15]:
```
data.head()
```

Out[15]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Env
0	41	Travel_Rarely	1102	Sales	1	2	Life Sciences	
1	49	Travel_Frequently	279	Research & Development	8	1	Life Sciences	
2	37	Travel_Rarely	1373	Research & Development	2	2	Other	
3	33	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	
4	27	Travel_Rarely	591	Research & Development	2	1	Medical	

5 rows × 32 columns

We are using for loop for encoding categorical data as there are many categorical data columns in the given dataset.

In [16]:
```
from sklearn.preprocessing import LabelEncoder
l = LabelEncoder()
for columns in data.columns:
    if data[columns].dtype == [np.number,np.float64,np.int64]:
        continue
    else:
        data[columns] = l.fit_transform(data[columns])
```

Spliting dependent and independent variables

In [17]:
```
y = data['Yes']
x = data.drop(['Yes'],axis = 1)
```

In [18]:
```
y.head()
```

Out[18]:
```
0    1
1    0
2    1
3    0
4    0
Name: Yes, dtype: int64
```

In [19]:
```
x.head()
```

Out[19]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Envirc
0	23	2	624	2	0	1	1	
1	31	1	113	1	7	0	1	
2	19	2	805	1	1	1	4	
3	15	1	820	1	2	3	1	
4	9	2	312	1	1	0	3	

5 rows × 31 columns

Feature Scaling

In [20]:
```
from sklearn.preprocessing import MinMaxScaler
ms = MinMaxScaler()
x_Scaled = pd.DataFrame(ms.fit_transform(x),columns=x.columns)
x_Scaled.head()
```

Out[20]:

	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	E
0	0.547619	1.0	0.705085	1.0	0.000000	0.25	0.2	
1	0.738095	0.5	0.127684	0.5	0.250000	0.00	0.2	
2	0.452381	1.0	0.909605	0.5	0.035714	0.25	0.8	
3	0.357143	0.5	0.926554	0.5	0.071429	0.75	0.2	
4	0.214286	1.0	0.352542	0.5	0.035714	0.00	0.6	

5 rows × 31 columns

Train and Test data Split

In [21]:
```
from sklearn.model_selection import train_test_split
```

In [22]:
```
xtrain,xtest,ytrain,ytest = train_test_split(x_Scaled,y,test_size = 0.2,random_state =
```

Logistic Regression

In [23]:
```
from sklearn.linear_model import LogisticRegression

lr = LogisticRegression()
```

In [24]:
```
lr.fit(xtrain,ytrain)
```

Out[24]:
```
LogisticRegression
LogisticRegression()
```

In [25]:
```
lrpred = lr.predict(xtest)
```

In [26]:
```
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report,roc_
```

In [27]:
```
accuracy_score(ytest,lrpred)
```

Out[27]:
```
0.8843537414965986
```

In [28]:
```
confusion_matrix(ytest,lrpred)
```

Out[28]:
```
array([[242,   3],
       [ 31,  18]], dtype=int64)
```

In [29]:
```
print(classification_report(ytest,lrpred))
```

```
              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
```

Decision Tree Classification

In [30]:
```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier()
```

In [31]:
```
dtc.fit(xtrain,ytrain)
```

Out[31]:
```
DecisionTreeClassifier
DecisionTreeClassifier()
```

In [32]:
```
dtcpred = dtc.predict(xtest)
```

In [33]:
```
accuracy_score(ytest,dtcpred)
```

Out[33]:
```
0.7482993197278912
```

In [34]:
```
print(classification_report(ytest,dtcpred))
```

	precision	recall	f1-score	support
0	0.86	0.83	0.85	245
1	0.28	0.33	0.30	49
accuracy			0.75	294
macro avg	0.57	0.58	0.57	294
weighted avg	0.76	0.75	0.76	294

In [35]:
```
from sklearn import tree
plt.figure(figsize=(35,25))
tree.plot_tree(dtc,filled=True)
```

Out[35]:
```
[Text(0.31914702868852457, 0.9722222222222222, 'x[24] <= 0.038\ngini = 0.269\nsamples
= 1176\nvalue = [988, 188]'),
 Text(0.07213114754098361, 0.9166666666666666, 'x[14] <= 0.75\ngini = 0.5\nsamples =
78\nvalue = [39, 39]'),
 Text(0.04262295081967213, 0.8611111111111112, 'x[4] <= 0.554\ngini = 0.426\nsamples
= 39\nvalue = [27, 12]'),
 Text(0.02622950819672131, 0.8055555555555556, 'x[13] <= 0.167\ngini = 0.312\nsamples
= 31\nvalue = [25, 6]'),
 Text(0.013114754098360656, 0.75, 'x[15] <= 0.046\ngini = 0.49\nsamples = 7\nvalue =
[3, 4]'),
 Text(0.006557377049180328, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.019672131147540985, 0.6944444444444444, 'x[14] <= 0.25\ngini = 0.375\nsamples
= 4\nvalue = [3, 1]'),
 Text(0.013114754098360656, 0.6388888888888888, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.02622950819672131, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.03934426229508197, 0.75, 'x[17] <= 0.056\ngini = 0.153\nsamples = 24\nvalue =
[22, 2]'),
 Text(0.03278688524590164, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.04590163934426229, 0.6944444444444444, 'x[7] <= 0.167\ngini = 0.083\nsamples
= 23\nvalue = [22, 1]'),
 Text(0.03934426229508197, 0.6388888888888888, 'x[1] <= 0.75\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(0.03278688524590164, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.04590163934426229, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.05245901639344262, 0.6388888888888888, 'gini = 0.0\nsamples = 21\nvalue = [2
1, 0]'),
 Text(0.05901639344262295, 0.8055555555555556, 'x[20] <= 0.679\ngini = 0.375\nsamples
= 8\nvalue = [2, 6]'),
 Text(0.05245901639344262, 0.75, 'gini = 0.0\nsamples = 6\nvalue = [0, 6]'),
 Text(0.06557377049180328, 0.75, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
 Text(0.10163934426229508, 0.8611111111111112, 'x[9] <= 0.364\ngini = 0.426\nsamples
= 39\nvalue = [12, 27]'),
 Text(0.08524590163934426, 0.8055555555555556, 'x[15] <= 0.231\ngini = 0.133\nsamples
= 14\nvalue = [1, 13]'),
 Text(0.07868852459016394, 0.75, 'gini = 0.0\nsamples = 13\nvalue = [0, 13]'),
 Text(0.09180327868852459, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.1180327868852459, 0.8055555555555556, 'x[19] <= 0.5\ngini = 0.493\nsamples =
25\nvalue = [11, 14]'),
 Text(0.10491803278688525, 0.75, 'x[2] <= 0.108\ngini = 0.484\nsamples = 17\nvalue =
[10, 7]'),
 Text(0.09836065573770492, 0.6944444444444444, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.11147540983606558, 0.6944444444444444, 'x[22] <= 0.167\ngini = 0.408\nsamples
= 14\nvalue = [10, 4]'),
 Text(0.09836065573770492, 0.6388888888888888, 'x[24] <= 0.013\ngini = 0.375\nsamples
= 4\nvalue = [1, 3]'),
 Text(0.09180327868852459, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.10491803278688525, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.12459016393442623, 0.6388888888888888, 'x[12] <= 0.875\ngini = 0.18\nsamples
= 10\nvalue = [9, 1]'),
 Text(0.1180327868852459, 0.5833333333333334, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
```

```
 Text(0.13114754098360656, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.13114754098360656, 0.75, 'x[2] <= 0.162\ngini = 0.219\nsamples = 8\nvalue =
[1, 7]'),
 Text(0.12459016393442623, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.1377049180327869, 0.6944444444444444, 'gini = 0.0\nsamples = 7\nvalue = [0,
7]'),
 Text(0.5661629098360655, 0.9166666666666666, 'x[19] <= 0.5\ngini = 0.235\nsamples =
1098\nvalue = [949, 149]'),
 Text(0.3151127049180328, 0.8611111111111112, 'x[26] <= 0.167\ngini = 0.162\nsamples
= 798\nvalue = [727, 71]'),
 Text(0.18032786885245902, 0.8055555555555556, 'x[2] <= 0.738\ngini = 0.38\nsamples =
47\nvalue = [35, 12]'),
 Text(0.1737704918032787, 0.75, 'x[10] <= 0.5\ngini = 0.463\nsamples = 33\nvalue = [2
1, 12]'),
 Text(0.15081967213114755, 0.6944444444444444, 'x[4] <= 0.446\ngini = 0.42\nsamples =
10\nvalue = [3, 7]'),
 Text(0.14426229508196722, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]'),
 Text(0.15737704918032788, 0.6388888888888888, 'x[5] <= 0.125\ngini = 0.375\nsamples
= 4\nvalue = [3, 1]'),
 Text(0.15081967213114755, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.16393442622950818, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.19672131147540983, 0.6944444444444444, 'x[25] <= 0.583\ngini = 0.34\nsamples
= 23\nvalue = [18, 5]'),
 Text(0.18360655737704917, 0.6388888888888888, 'x[29] <= 0.333\ngini = 0.117\nsamples
= 16\nvalue = [15, 1]'),
 Text(0.17704918032786884, 0.5833333333333334, 'gini = 0.0\nsamples = 15\nvalue = [1
5, 0]'),
 Text(0.1901639344262295, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.2098360655737705, 0.6388888888888888, 'x[30] <= 0.529\ngini = 0.49\nsamples =
7\nvalue = [3, 4]'),
 Text(0.20327868852459016, 0.5833333333333334, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(0.21639344262295082, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.18688524590163935, 0.75, 'gini = 0.0\nsamples = 14\nvalue = [14, 0]'),
 Text(0.44989754098360657, 0.8055555555555556, 'x[27] <= 0.986\ngini = 0.145\nsamples
= 751\nvalue = [692, 59]'),
 Text(0.4433401639344262, 0.75, 'x[27] <= 0.125\ngini = 0.143\nsamples = 750\nvalue =
[692, 58]'),
 Text(0.3081967213114754, 0.6944444444444444, 'x[7] <= 0.167\ngini = 0.218\nsamples =
257\nvalue = [225, 32]'),
 Text(0.2573770491803279, 0.6388888888888888, 'x[30] <= 0.147\ngini = 0.355\nsamples
= 65\nvalue = [50, 15]'),
 Text(0.22950819672131148, 0.5833333333333334, 'x[30] <= 0.029\ngini = 0.303\nsamples
= 59\nvalue = [48, 11]'),
 Text(0.20655737704918034, 0.5277777777777778, 'x[10] <= 0.5\ngini = 0.463\nsamples =
22\nvalue = [14, 8]'),
 Text(0.19344262295081968, 0.4722222222222222, 'x[9] <= 0.179\ngini = 0.198\nsamples
= 9\nvalue = [8, 1]'),
 Text(0.18688524590163935, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.2, 0.4166666666666667, 'gini = 0.0\nsamples = 8\nvalue = [8, 0]'),
 Text(0.21967213114754097, 0.4722222222222222, 'x[9] <= 0.4\ngini = 0.497\nsamples =
13\nvalue = [6, 7]'),
```

```
 Text(0.21311475409836064, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(0.2262295081967213, 0.4166666666666667, 'x[4] <= 0.286\ngini = 0.346\nsamples =
9\nvalue = [2, 7]'),
 Text(0.21967213114754097, 0.3611111111111111, 'x[5] <= 0.5\ngini = 0.444\nsamples =
3\nvalue = [2, 1]'),
 Text(0.21311475409836064, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.2262295081967213, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.23278688524590163, 0.3611111111111111, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]'),
 Text(0.25245901639344265, 0.5277777777777778, 'x[13] <= 0.167\ngini = 0.149\nsamples
= 37\nvalue = [34, 3]'),
 Text(0.2459016393442623, 0.4722222222222222, 'x[26] <= 0.5\ngini = 0.5\nsamples = 6
\nvalue = [3, 3]'),
 Text(0.23934426229508196, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.25245901639344265, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.25901639344262295, 0.4722222222222222, 'gini = 0.0\nsamples = 31\nvalue = [3
1, 0]'),
 Text(0.28524590163934427, 0.5833333333333334, 'x[8] <= 0.5\ngini = 0.444\nsamples =
6\nvalue = [2, 4]'),
 Text(0.2786885245901639, 0.5277777777777778, 'x[1] <= 0.75\ngini = 0.444\nsamples =
3\nvalue = [2, 1]'),
 Text(0.2721311475409836, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.28524590163934427, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.29180327868852457, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.3590163934426229, 0.6388888888888888, 'x[0] <= 0.321\ngini = 0.161\nsamples =
192\nvalue = [175, 17]'),
 Text(0.3114754098360656, 0.5833333333333334, 'x[6] <= 0.1\ngini = 0.294\nsamples = 6
7\nvalue = [55, 12]'),
 Text(0.30491803278688523, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.3180327868852459, 0.5277777777777778, 'x[26] <= 0.5\ngini = 0.26\nsamples = 6
5\nvalue = [55, 10]'),
 Text(0.3016393442622951, 0.4722222222222222, 'x[9] <= 0.679\ngini = 0.469\nsamples =
16\nvalue = [10, 6]'),
 Text(0.29508196721311475, 0.4166666666666667, 'x[6] <= 0.4\ngini = 0.444\nsamples =
9\nvalue = [3, 6]'),
 Text(0.28852459016393445, 0.3611111111111111, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.3016393442622951, 0.3611111111111111, 'x[20] <= 0.464\ngini = 0.245\nsamples
= 7\nvalue = [1, 6]'),
 Text(0.29508196721311475, 0.3055555555555556, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]'),
 Text(0.3081967213114754, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.3081967213114754, 0.4166666666666667, 'gini = 0.0\nsamples = 7\nvalue = [7,
0]'),
 Text(0.3344262295081967, 0.4722222222222222, 'x[2] <= 0.04\ngini = 0.15\nsamples = 4
9\nvalue = [45, 4]'),
 Text(0.32786885245901637, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.34098360655737703, 0.4166666666666667, 'x[2] <= 0.942\ngini = 0.117\nsamples
= 48\nvalue = [45, 3]'),
```

```
 Text(0.3344262295081967, 0.3611111111111111, 'x[5] <= 0.875\ngini = 0.081\nsamples =
47\nvalue = [45, 2]'),
 Text(0.32131147540983607, 0.3055555555555556, 'x[10] <= 0.167\ngini = 0.043\nsamples
= 45\nvalue = [44, 1]'),
 Text(0.31475409836065577, 0.25, 'x[13] <= 0.5\ngini = 0.444\nsamples = 3\nvalue =
[2, 1]'),
 Text(0.3081967213114754, 0.19444444444444445, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.32131147540983607, 0.19444444444444445, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.32786885245901637, 0.25, 'gini = 0.0\nsamples = 42\nvalue = [42, 0]'),
 Text(0.3475409836065574, 0.3055555555555556, 'x[20] <= 0.393\ngini = 0.5\nsamples =
2\nvalue = [1, 1]'),
 Text(0.34098360655737703, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.3540983606557377, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.3475409836065574, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.4065573770491803, 0.5833333333333334, 'x[6] <= 0.9\ngini = 0.077\nsamples = 1
25\nvalue = [120, 5]'),
 Text(0.39344262295081966, 0.5277777777777778, 'x[0] <= 0.393\ngini = 0.05\nsamples =
118\nvalue = [115, 3]'),
 Text(0.38688524590163936, 0.4722222222222222, 'x[2] <= 0.96\ngini = 0.185\nsamples =
29\nvalue = [26, 3]'),
 Text(0.380327868852459, 0.4166666666666667, 'x[10] <= 0.167\ngini = 0.133\nsamples =
28\nvalue = [26, 2]'),
 Text(0.36721311475409835, 0.3611111111111111, 'x[16] <= 0.682\ngini = 0.5\nsamples =
2\nvalue = [1, 1]'),
 Text(0.36065573770491804, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.3737704918032787, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.39344262295081966, 0.3611111111111111, 'x[30] <= 0.147\ngini = 0.074\nsamples
= 26\nvalue = [25, 1]'),
 Text(0.38688524590163936, 0.3055555555555556, 'gini = 0.0\nsamples = 24\nvalue = [2
4, 0]'),
 Text(0.4, 0.3055555555555556, 'x[17] <= 0.111\ngini = 0.5\nsamples = 2\nvalue = [1,
1]'),
 Text(0.39344262295081966, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.4065573770491803, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.39344262295081966, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.4, 0.4722222222222222, 'gini = 0.0\nsamples = 89\nvalue = [89, 0]'),
 Text(0.419672131147541, 0.5277777777777778, 'x[2] <= 0.593\ngini = 0.408\nsamples =
7\nvalue = [5, 2]'),
 Text(0.4131147540983607, 0.4722222222222222, 'x[13] <= 0.333\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.4065573770491803, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.419672131147541, 0.4166666666666667, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.4262295081967213, 0.4722222222222222, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(0.5784836065573771, 0.6944444444444444, 'x[27] <= 0.847\ngini = 0.1\nsamples =
493\nvalue = [467, 26]'),
 Text(0.5438524590163935, 0.6388888888888888, 'x[13] <= 0.5\ngini = 0.094\nsamples =
486\nvalue = [462, 24]'),
 Text(0.49426229508196723, 0.5833333333333334, 'x[12] <= 0.938\ngini = 0.154\nsamples
= 191\nvalue = [175, 16]'),
 Text(0.48770491803278687, 0.5277777777777778, 'x[16] <= 0.488\ngini = 0.145\nsamples
= 190\nvalue = [175, 15]'),
```

```
 Text(0.4672131147540984, 0.4722222222222222, 'x[16] <= 0.478\ngini = 0.221\nsamples
= 95\nvalue = [83, 12]'),
 Text(0.460655737704918, 0.4166666666666667, 'x[30] <= 0.794\ngini = 0.207\nsamples =
94\nvalue = [83, 11]'),
 Text(0.4540983606557377, 0.3611111111111111, 'x[5] <= 0.375\ngini = 0.192\nsamples =
93\nvalue = [83, 10]'),
 Text(0.4262295081967213, 0.3055555555555556, 'x[6] <= 0.9\ngini = 0.363\nsamples = 2
1\nvalue = [16, 5]'),
 Text(0.419672131147541, 0.25, 'x[15] <= 0.751\ngini = 0.266\nsamples = 19\nvalue =
[16, 3]'),
 Text(0.4065573770491803, 0.19444444444444445, 'x[4] <= 0.982\ngini = 0.117\nsamples
= 16\nvalue = [15, 1]'),
 Text(0.4, 0.1388888888888889, 'gini = 0.0\nsamples = 14\nvalue = [14, 0]'),
 Text(0.4131147540983607, 0.1388888888888889, 'x[14] <= 0.5\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(0.4065573770491803, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.419672131147541, 0.08333333333333333, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.43278688524590164, 0.19444444444444445, 'x[22] <= 0.833\ngini = 0.444\nsample
s = 3\nvalue = [1, 2]'),
 Text(0.4262295081967213, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.43934426229508194, 0.1388888888888889, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.43278688524590164, 0.25, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(0.4819672131147541, 0.3055555555555556, 'x[28] <= 0.139\ngini = 0.129\nsamples
= 72\nvalue = [67, 5]'),
 Text(0.46557377049180326, 0.25, 'x[15] <= 0.565\ngini = 0.444\nsamples = 6\nvalue =
[4, 2]'),
 Text(0.45901639344262296, 0.19444444444444445, 'x[2] <= 0.625\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.4524590163934426, 0.1388888888888889, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.46557377049180326, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.4721311475409836, 0.19444444444444445, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.49836065573770494, 0.25, 'x[9] <= 0.993\ngini = 0.087\nsamples = 66\nvalue =
[63, 3]'),
 Text(0.4852459016393443, 0.19444444444444445, 'x[25] <= 0.583\ngini = 0.061\nsamples
= 64\nvalue = [62, 2]'),
 Text(0.4786885245901639, 0.1388888888888889, 'gini = 0.0\nsamples = 51\nvalue = [51,
0]'),
 Text(0.4918032786885246, 0.1388888888888889, 'x[3] <= 0.75\ngini = 0.26\nsamples = 1
3\nvalue = [11, 2]'),
 Text(0.4852459016393443, 0.08333333333333333, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
 Text(0.49836065573770494, 0.08333333333333333, 'x[24] <= 0.308\ngini = 0.5\nsamples
= 4\nvalue = [2, 2]'),
 Text(0.4918032786885246, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.5049180327868853, 0.027777777777777776, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.5114754098360655, 0.19444444444444445, 'x[28] <= 0.5\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(0.5049180327868853, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.5180327868852459, 0.1388888888888889, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
```

```
 Text(0.4672131147540984, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.4737704918032787, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.5081967213114754, 0.4722222222222222, 'x[17] <= 0.5\ngini = 0.061\nsamples =
95\nvalue = [92, 3]'),
 Text(0.5016393442622951, 0.4166666666666667, 'gini = 0.0\nsamples = 76\nvalue = [76,
0]'),
 Text(0.5147540983606558, 0.4166666666666667, 'x[30] <= 0.088\ngini = 0.266\nsamples
= 19\nvalue = [16, 3]'),
 Text(0.5016393442622951, 0.3611111111111111, 'x[22] <= 0.833\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.49508196721311476, 0.3055555555555556, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.5081967213114754, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.5278688524590164, 0.3611111111111111, 'x[15] <= 0.239\ngini = 0.117\nsamples
= 16\nvalue = [15, 1]'),
 Text(0.521311475409836, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.5344262295081967, 0.3055555555555556, 'gini = 0.0\nsamples = 15\nvalue = [15,
0]'),
 Text(0.5008196721311475, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.5934426229508196, 0.5833333333333334, 'x[20] <= 0.036\ngini = 0.053\nsamples
= 295\nvalue = [287, 8]'),
 Text(0.5704918032786885, 0.5277777777777778, 'x[29] <= 0.7\ngini = 0.159\nsamples =
46\nvalue = [42, 4]'),
 Text(0.5639344262295082, 0.4722222222222222, 'x[10] <= 0.167\ngini = 0.124\nsamples
= 45\nvalue = [42, 3]'),
 Text(0.5475409836065573, 0.4166666666666667, 'x[24] <= 0.308\ngini = 0.5\nsamples =
2\nvalue = [1, 1]'),
 Text(0.5409836065573771, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.5540983606557377, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.580327868852459, 0.4166666666666667, 'x[24] <= 0.705\ngini = 0.089\nsamples =
43\nvalue = [41, 2]'),
 Text(0.5672131147540984, 0.3611111111111111, 'x[12] <= 0.062\ngini = 0.048\nsamples
= 41\nvalue = [40, 1]'),
 Text(0.5606557377049181, 0.3055555555555556, 'x[0] <= 0.595\ngini = 0.375\nsamples =
4\nvalue = [3, 1]'),
 Text(0.5540983606557377, 0.25, 'gini = 0.0\nsamples = 3\nvalue = [3, 0]'),
 Text(0.5672131147540984, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.5737704918032787, 0.3055555555555556, 'gini = 0.0\nsamples = 37\nvalue = [37,
0]'),
 Text(0.5934426229508196, 0.3611111111111111, 'x[27] <= 0.236\ngini = 0.5\nsamples =
2\nvalue = [1, 1]'),
 Text(0.5868852459016394, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.6, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(0.5770491803278689, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.6163934426229508, 0.5277777777777778, 'x[15] <= 0.037\ngini = 0.032\nsamples
= 249\nvalue = [245, 4]'),
 Text(0.6, 0.4722222222222222, 'x[15] <= 0.034\ngini = 0.32\nsamples = 5\nvalue = [4,
1]'),
 Text(0.5934426229508196, 0.4166666666666667, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(0.6065573770491803, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
```

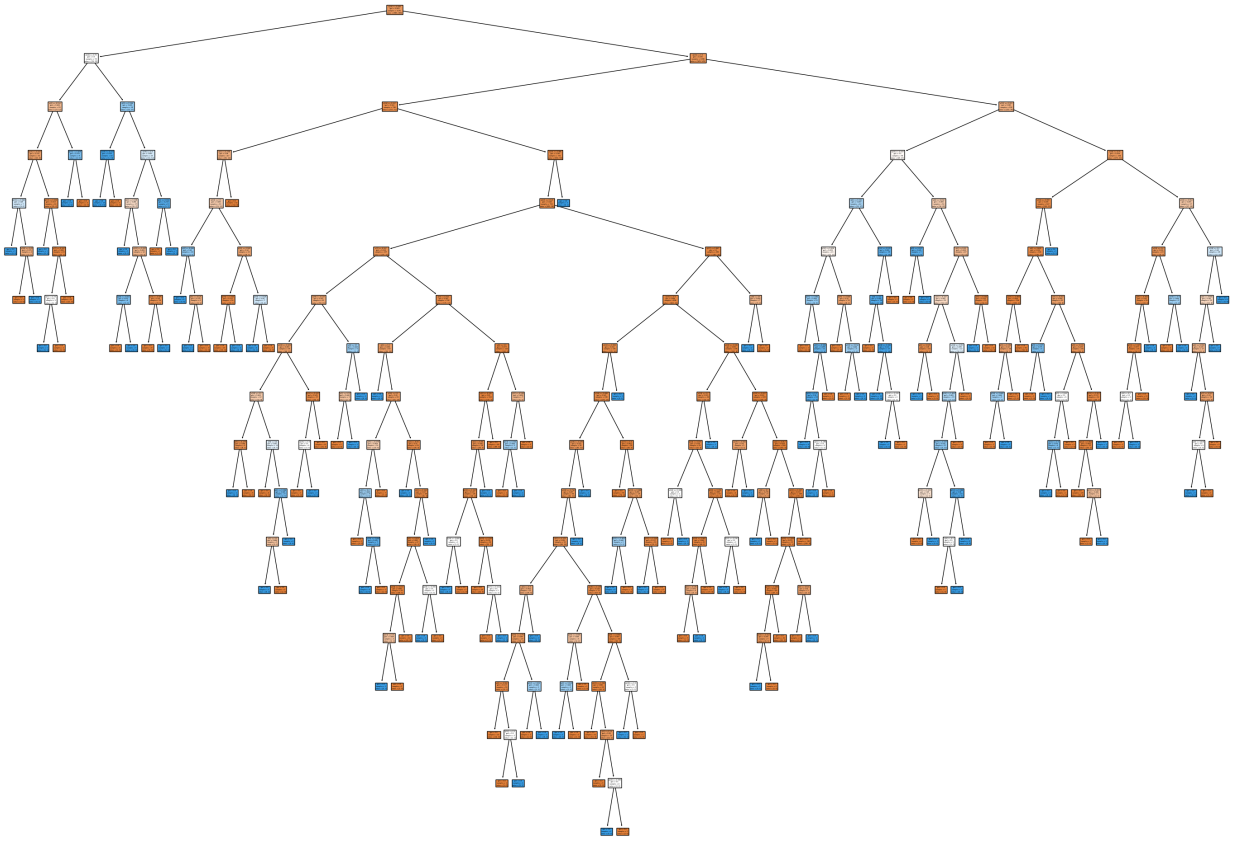
```
1]'),
 Text(0.6327868852459017, 0.4722222222222222, 'x[2] <= 0.017\ngini = 0.024\nsamples =
244\nvalue = [241, 3]'),
 Text(0.6196721311475409, 0.4166666666666667, 'x[8] <= 0.5\ngini = 0.278\nsamples = 6
\nvalue = [5, 1]'),
 Text(0.6131147540983607, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.6262295081967213, 0.3611111111111111, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
 Text(0.6459016393442623, 0.4166666666666667, 'x[22] <= 0.167\ngini = 0.017\nsamples
= 238\nvalue = [236, 2]'),
 Text(0.639344262295082, 0.3611111111111111, 'x[26] <= 0.833\ngini = 0.073\nsamples =
53\nvalue = [51, 2]'),
 Text(0.6262295081967213, 0.3055555555555556, 'x[30] <= 0.088\ngini = 0.041\nsamples
= 48\nvalue = [47, 1]'),
 Text(0.6196721311475409, 0.25, 'x[12] <= 0.312\ngini = 0.245\nsamples = 7\nvalue =
[6, 1]'),
 Text(0.6131147540983607, 0.19444444444444445, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.6262295081967213, 0.19444444444444445, 'gini = 0.0\nsamples = 6\nvalue = [6,
0]'),
 Text(0.6327868852459017, 0.25, 'gini = 0.0\nsamples = 41\nvalue = [41, 0]'),
 Text(0.6524590163934426, 0.3055555555555556, 'x[0] <= 0.631\ngini = 0.32\nsamples =
5\nvalue = [4, 1]'),
 Text(0.6459016393442623, 0.25, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
 Text(0.659016393442623, 0.25, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.6524590163934426, 0.3611111111111111, 'gini = 0.0\nsamples = 185\nvalue = [18
5, 0]'),
 Text(0.6131147540983607, 0.6388888888888888, 'x[8] <= 0.5\ngini = 0.408\nsamples = 7
\nvalue = [5, 2]'),
 Text(0.6065573770491803, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.6196721311475409, 0.5833333333333334, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
 Text(0.45645491803278687, 0.75, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(0.8172131147540984, 0.8611111111111112, 'x[15] <= 0.351\ngini = 0.385\nsamples
= 300\nvalue = [222, 78]'),
 Text(0.728688524590164, 0.8055555555555556, 'x[23] <= 0.167\ngini = 0.5\nsamples = 9
6\nvalue = [49, 47]'),
 Text(0.6950819672131148, 0.75, 'x[4] <= 0.161\ngini = 0.459\nsamples = 42\nvalue =
[15, 27]'),
 Text(0.6721311475409836, 0.6944444444444444, 'x[16] <= 0.418\ngini = 0.499\nsamples
= 23\nvalue = [12, 11]'),
 Text(0.659016393442623, 0.6388888888888888, 'x[15] <= 0.062\ngini = 0.426\nsamples =
13\nvalue = [4, 9]'),
 Text(0.6524590163934426, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.6655737704918033, 0.5833333333333334, 'x[9] <= 0.993\ngini = 0.298\nsamples =
11\nvalue = [2, 9]'),
 Text(0.659016393442623, 0.5277777777777778, 'x[21] <= 0.5\ngini = 0.18\nsamples = 10
\nvalue = [1, 9]'),
 Text(0.6524590163934426, 0.4722222222222222, 'gini = 0.0\nsamples = 8\nvalue = [0,
8]'),
 Text(0.6655737704918033, 0.4722222222222222, 'x[2] <= 0.298\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(0.659016393442623, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.6721311475409836, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.6721311475409836, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1,
```

```
0]'),
 Text(0.6852459016393443, 0.6388888888888888, 'x[25] <= 0.583\ngini = 0.32\nsamples =
10\nvalue = [8, 2]'),
 Text(0.6786885245901639, 0.5833333333333334, 'gini = 0.0\nsamples = 7\nvalue = [7,
0]'),
 Text(0.6918032786885245, 0.5833333333333334, 'x[15] <= 0.154\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.6852459016393443, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.6983606557377049, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.7180327868852459, 0.6944444444444444, 'x[11] <= 0.125\ngini = 0.266\nsamples
= 19\nvalue = [3, 16]'),
 Text(0.7114754098360656, 0.6388888888888888, 'x[9] <= 0.2\ngini = 0.198\nsamples = 1
8\nvalue = [2, 16]'),
 Text(0.7049180327868853, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.7180327868852459, 0.5833333333333334, 'x[28] <= 0.306\ngini = 0.111\nsamples
= 17\nvalue = [1, 16]'),
 Text(0.7114754098360656, 0.5277777777777778, 'gini = 0.0\nsamples = 15\nvalue = [0,
15]'),
 Text(0.7245901639344262, 0.5277777777777778, 'x[15] <= 0.102\ngini = 0.5\nsamples =
2\nvalue = [1, 1]'),
 Text(0.7180327868852459, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.7311475409836066, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.7245901639344262, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.7622950819672131, 0.75, 'x[0] <= 0.202\ngini = 0.466\nsamples = 54\nvalue =
[34, 20]'),
 Text(0.7442622950819672, 0.6944444444444444, 'x[0] <= 0.107\ngini = 0.245\nsamples =
7\nvalue = [1, 6]'),
 Text(0.7377049180327869, 0.6388888888888888, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.7508196721311475, 0.6388888888888888, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]'),
 Text(0.780327868852459, 0.6944444444444444, 'x[2] <= 0.62\ngini = 0.418\nsamples = 4
7\nvalue = [33, 14]'),
 Text(0.7639344262295082, 0.6388888888888888, 'x[2] <= 0.147\ngini = 0.482\nsamples =
32\nvalue = [19, 13]'),
 Text(0.7508196721311475, 0.5833333333333334, 'x[2] <= 0.025\ngini = 0.18\nsamples =
10\nvalue = [9, 1]'),
 Text(0.7442622950819672, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.7573770491803279, 0.5277777777777778, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
 Text(0.7770491803278688, 0.5833333333333334, 'x[16] <= 0.89\ngini = 0.496\nsamples =
22\nvalue = [10, 12]'),
 Text(0.7704918032786885, 0.5277777777777778, 'x[26] <= 0.833\ngini = 0.465\nsamples
= 19\nvalue = [7, 12]'),
 Text(0.7639344262295082, 0.4722222222222222, 'x[17] <= 0.167\ngini = 0.415\nsamples
= 17\nvalue = [5, 12]'),
 Text(0.7508196721311475, 0.4166666666666667, 'x[20] <= 0.321\ngini = 0.49\nsamples =
7\nvalue = [4, 3]'),
 Text(0.7442622950819672, 0.3611111111111111, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(0.7573770491803279, 0.3611111111111111, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.7770491803278688, 0.4166666666666667, 'x[12] <= 0.188\ngini = 0.18\nsamples =
```

```
10\nvalue = [1, 9]'),
 Text(0.7704918032786885, 0.3611111111111111, 'x[8] <= 0.5\ngini = 0.5\nsamples = 2\n
value = [1, 1]'),
 Text(0.7639344262295082, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.7770491803278688, 0.3055555555555556, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.7836065573770492, 0.3611111111111111, 'gini = 0.0\nsamples = 8\nvalue = [0,
8]'),
 Text(0.7770491803278688, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.7836065573770492, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.7967213114754098, 0.6388888888888888, 'x[9] <= 0.064\ngini = 0.124\nsamples =
15\nvalue = [14, 1]'),
 Text(0.7901639344262295, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.8032786885245902, 0.5833333333333334, 'gini = 0.0\nsamples = 14\nvalue = [14,
0]'),
 Text(0.9057377049180327, 0.8055555555555556, 'x[14] <= 0.75\ngini = 0.258\nsamples =
204\nvalue = [173, 31]'),
 Text(0.8475409836065574, 0.75, 'x[15] <= 0.997\ngini = 0.138\nsamples = 147\nvalue =
[136, 11]'),
 Text(0.840983606557377, 0.6944444444444444, 'x[4] <= 0.482\ngini = 0.128\nsamples =
146\nvalue = [136, 10]'),
 Text(0.8229508196721311, 0.6388888888888888, 'x[27] <= 0.069\ngini = 0.038\nsamples
= 104\nvalue = [102, 2]'),
 Text(0.8163934426229508, 0.5833333333333334, 'x[9] <= 0.193\ngini = 0.32\nsamples =
10\nvalue = [8, 2]'),
 Text(0.8098360655737705, 0.5277777777777778, 'x[16] <= 0.469\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(0.8032786885245902, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.8163934426229508, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.8229508196721311, 0.5277777777777778, 'gini = 0.0\nsamples = 7\nvalue = [7,
0]'),
 Text(0.8295081967213115, 0.5833333333333334, 'gini = 0.0\nsamples = 94\nvalue = [94,
0]'),
 Text(0.8590163934426229, 0.6388888888888888, 'x[7] <= 0.167\ngini = 0.308\nsamples =
42\nvalue = [34, 8]'),
 Text(0.8426229508196721, 0.5833333333333334, 'x[9] <= 0.307\ngini = 0.375\nsamples =
4\nvalue = [1, 3]'),
 Text(0.8360655737704918, 0.5277777777777778, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.8491803278688524, 0.5277777777777778, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.8754098360655738, 0.5833333333333334, 'x[0] <= 0.393\ngini = 0.229\nsamples =
38\nvalue = [33, 5]'),
 Text(0.8622950819672132, 0.5277777777777778, 'x[9] <= 0.643\ngini = 0.5\nsamples = 6
\nvalue = [3, 3]'),
 Text(0.8557377049180328, 0.4722222222222222, 'x[21] <= 0.5\ngini = 0.375\nsamples =
4\nvalue = [1, 3]'),
 Text(0.8491803278688524, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.8622950819672132, 0.4166666666666667, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.8688524590163934, 0.4722222222222222, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.8885245901639345, 0.5277777777777778, 'x[25] <= 0.917\ngini = 0.117\nsamples
```

```
= 32\nvalue = [30, 2]'),
 Text(0.8819672131147541, 0.4722222222222222, 'x[12] <= 0.812\ngini = 0.062\nsamples
= 31\nvalue = [30, 1]'),
 Text(0.8754098360655738, 0.4166666666666667, 'gini = 0.0\nsamples = 28\nvalue = [28,
0]'),
 Text(0.8885245901639345, 0.4166666666666667, 'x[25] <= 0.417\ngini = 0.444\nsamples
= 3\nvalue = [2, 1]'),
 Text(0.8819672131147541, 0.3611111111111111, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(0.8950819672131147, 0.3611111111111111, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.8950819672131147, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.8540983606557377, 0.6944444444444444, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.9639344262295082, 0.75, 'x[12] <= 0.812\ngini = 0.456\nsamples = 57\nvalue =
[37, 20]'),
 Text(0.940983606557377, 0.6944444444444444, 'x[29] <= 0.4\ngini = 0.238\nsamples = 2
9\nvalue = [25, 4]'),
 Text(0.9278688524590164, 0.6388888888888888, 'x[9] <= 0.964\ngini = 0.142\nsamples =
26\nvalue = [24, 2]'),
 Text(0.921311475409836, 0.5833333333333334, 'x[10] <= 0.167\ngini = 0.077\nsamples =
25\nvalue = [24, 1]'),
 Text(0.9147540983606557, 0.5277777777777778, 'x[8] <= 0.5\ngini = 0.5\nsamples = 2\n
value = [1, 1]'),
 Text(0.9081967213114754, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.921311475409836, 0.4722222222222222, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.9278688524590164, 0.5277777777777778, 'gini = 0.0\nsamples = 23\nvalue = [23,
0]'),
 Text(0.9344262295081968, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(0.9540983606557377, 0.6388888888888888, 'x[10] <= 0.5\ngini = 0.444\nsamples =
3\nvalue = [1, 2]'),
 Text(0.9475409836065574, 0.5833333333333334, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(0.9606557377049181, 0.5833333333333334, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.9868852459016394, 0.6944444444444444, 'x[29] <= 0.1\ngini = 0.49\nsamples = 2
8\nvalue = [12, 16]'),
 Text(0.980327868852459, 0.6388888888888888, 'x[4] <= 0.804\ngini = 0.48\nsamples = 2
0\nvalue = [12, 8]'),
 Text(0.9737704918032787, 0.5833333333333334, 'x[4] <= 0.018\ngini = 0.415\nsamples =
17\nvalue = [12, 5]'),
 Text(0.9672131147540983, 0.5277777777777778, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(0.980327868852459, 0.5277777777777778, 'x[7] <= 0.167\ngini = 0.32\nsamples = 1
5\nvalue = [12, 3]'),
 Text(0.9737704918032787, 0.4722222222222222, 'x[0] <= 0.405\ngini = 0.5\nsamples = 6
\nvalue = [3, 3]'),
 Text(0.9672131147540983, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(0.980327868852459, 0.4166666666666667, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(0.9868852459016394, 0.4722222222222222, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
 Text(0.9868852459016394, 0.5833333333333334, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
```

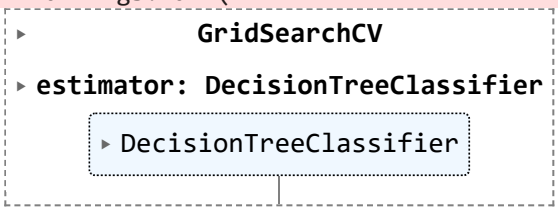
```
 Text(0.9934426229508196, 0.6388888888888888, 'gini = 0.0\nsamples = 8\nvalue = [0,
8]')]
```



In [36]:
```
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=dtc,param_grid=parameter,cv=5,scoring="accuracy")
grid_search.fit(xtrain,ytrain)
```

```
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\model_selection\_validation.py:42
5: 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:\ProgramData\anaconda3\Lib\site-packages\sklearn\model_selection\_validatio
n.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py", line 1144, in wr
apper
    estimator._validate_params()
  File "C:\ProgramData\anaconda3\Lib\site-packages\sklearn\base.py", line 637, in _va
lidate_params
    validate_parameter_constraints(
  File "C:\ProgramData\anaconda3\Lib\site-packages\sklearn\utils\_param_validation.p
y", line 95, in validate_parameter_constraints
    raise InvalidParameterError(
sklearn.utils._param_validation.InvalidParameterError: The 'max_features' parameter o
f DecisionTreeClassifier must be an int in the range [1, inf), a float in the range
(0.0, 1.0], a str among {'log2', 'sqrt'} or None. Got 'auto' instead.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\ProgramData\anaconda3\Lib\site-packages\sklearn\model_selection\_search.py:976: Us
erWarning: One or more of the test scores are non-finite: [       nan        nan 0.84
013704 0.84013704 0.84013704 0.84013704
        nan        nan 0.83674721 0.83928597 0.83928597 0.84013704
        nan        nan 0.84778219 0.84524702 0.84097367 0.8409881
        nan        nan 0.8341832  0.82229354 0.84438514 0.83248467
        nan        nan 0.84269744 0.83758745 0.8180238  0.84097367
        nan        nan 0.84013704 0.84013704 0.84013704 0.83843491
        nan        nan 0.84184638 0.84013704 0.83759106 0.84013704
        nan        nan 0.85287414 0.83673639 0.83332492 0.8409881
        nan        nan 0.83078255 0.83842409 0.83844212 0.83417959
        nan        nan 0.84013343 0.82993509 0.83758384 0.84268662]
  warnings.warn(
```

Out[36]:



In [37]:
```
grid_search.best_params_
```

Out[37]:
```
{'criterion': 'entropy',
 'max_depth': 3,
 'max_features': 'sqrt',
 'splitter': 'best'}
```

In [38]:
```
dtc_cv = DecisionTreeClassifier(criterion = 'entropy',
 max_depth = 5,
```

```
 max_features = 'sqrt',
 splitter = 'best')
```

In [39]:
```
dtc_cv.fit(xtrain,ytrain)
```

Out[39]:
```
DecisionTreeClassifier(criterion='entropy', max_depth=5, max_features='sqr
t')
```

In [40]:
```
dtc_cv_pred = dtc_cv.predict(xtest)
```

In [41]:
```
print(classification_report(ytest,dtc_cv_pred))
```

```
              precision    recall  f1-score   support

           0       0.86      0.96      0.91       245
           1       0.50      0.18      0.27        49

    accuracy                           0.83       294
   macro avg       0.68      0.57      0.59       294
weighted avg       0.80      0.83      0.80       294
```

Random Forest Classifier

In [42]:
```
from sklearn.ensemble import RandomForestClassifier

rfc = RandomForestClassifier()
```

In [43]:
```
forest_params = [{'max_depth': list(range(10, 15)), 'max_features': [2, 4, 6, 8, 10]}]
```

In [44]:
```
rfc_cv = GridSearchCV(rfc, forest_params, cv=5)
```

In [45]:
```
rfc_cv.fit(xtrain, ytrain)
```

Out[45]:
```
GridSearchCV
estimator: RandomForestClassifier
RandomForestClassifier
```

In [46]:
```
rfc_pred = rfc_cv.predict(xtest)
```

In [47]:
```
accuracy_score(ytest,rfc_pred)
```

Out[47]:
```
0.8469387755102041
```

In [48]:
```
print(classification_report(ytest,rfc_pred))
```

	precision	recall	f1-score	support
0	0.85	0.99	0.92	245
1	0.75	0.12	0.21	49
accuracy			0.85	294
macro avg	0.80	0.56	0.56	294
weighted avg	0.83	0.85	0.80	294