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

Out[2]:

_	battery_power	blue	clock_speed	dual_sim	fc	four_g	int_memory	m_dep	mobile_wt	n_cores	 px_height	px_width	ram	sc_h	sc_w
	842	0	2.2	0	1	0	7	0.6	188	2	 20	756	2549	9	7
	1 1021	1	0.5	1	0	1	53	0.7	136	3	 905	1988	2631	17	3
	2 563	1	0.5	1	2	1	41	0.9	145	5	 1263	1716	2603	11	2
	3 615	1	2.5	0	0	0	10	8.0	131	6	 1216	1786	2769	16	8
	4 1821	1	1.2	0	13	1	44	0.6	141	2	 1208	1212	1411	8	2
9	5 794	1	0.5	1	0	1	2	8.0	106	6	 1222	1890	668	13	4
9	6 1965	1	2.6	1	0	0	39	0.2	187	4	 915	1965	2032	11	10
9	7 1911	0	0.9	1	1	1	36	0.7	108	8	 868	1632	3057	9	1
9	8 1512	0	0.9	0	4	1	46	0.1	145	5	 336	670	869	18	10
9	9 510	1	2.0	1	5	1	45	0.9	168	6	 483	754	3919	19	4

00 rows × 21 columns

localhost:8888/notebooks/mobile_price.ipynb

```
In [3]: df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 2000 entries, 0 to 1999
        Data columns (total 21 columns):
             Column
                             Non-Null Count Dtype
             battery power 2000 non-null
                                             int64
             blue
                             2000 non-null
                                             int64
             clock speed
                             2000 non-null
                                             float64
             dual sim
                             2000 non-null
                                             int64
                             2000 non-null
             fc
                                             int64
             four g
                             2000 non-null
                                             int64
                             2000 non-null
             int memory
                                             int64
                             2000 non-null
             m dep
                                             float64
             mobile wt
                             2000 non-null
                                             int64
             n cores
                             2000 non-null
                                             int64
                             2000 non-null
         10
             рс
                                             int64
                             2000 non-null
             px height
                                             int64
         12 px width
                             2000 non-null
                                             int64
                             2000 non-null
         13
             ram
                                             int64
                             2000 non-null
             sc h
                                             int64
         15 sc w
                             2000 non-null
                                             int64
                             2000 non-null
         16 talk time
                                             int64
                             2000 non-null
         17 three g
                                             int64
         18 touch screen
                             2000 non-null
                                             int64
         19 wifi
                             2000 non-null
                                             int64
         20 price range
                             2000 non-null
                                             int64
        dtypes: float64(2), int64(19)
        memory usage: 328.3 KB
In [4]: x=df.drop('blue',axis=1)
        y=df['blue']
```

```
In [5]: t={"three_g":{"Yes":1,"No":0}}
    df=df.replace(t)
    print(df)
```

	battery_pow		clock	_speed	dual_sim		_g in	t_memo	ry \	
0		342 0		2.2	0		0		7	
1		21 1		0.5	1		1		53	
2	5	663 1		0.5	1	2	1		41	
3	6	515 1		2.5	0	0	0		10	
4	18	321 1		1.2	0	13	1		44	
1005		704 1					1	•		
1995		794 1		0.5	1		1		2	
1996		965 1		2.6	1		0		39	
1997		911 0		0.9	1		1		36	
1998		512 0		0.9	0		1		46	
1999	5	510 1		2.0	1	5	1		45	
	m_dep mobi	lle_wt n_	cores	р	x_height	px_width	ram	sc_h	SC_W	\
0	0.6	188	2		20	756	2549	9	7	
1	0.7	136	3		905	1988	2631	17	3	
2	0.9	145	5		1263	1716	2603	11	2	
3	0.8	131	6		1216	1786	2769	16	8	
4	0.6	141	2		1208	1212	1411	8	2	
	• • •				• • •					
1995	0.8	106	6		1222	1890	668	13	4	
1996	0.2	187	4		915	1965	2032	11	10	
1997	0.7	108	8		868	1632	3057	9	1	
1998	0.1	145	5		336	670	869	18	10	
1999	0.9	168	6	•••	483	754	3919	19	4	
•			touch_	screen		ice_range				
0	19	0		0	1	1				
1	7	1		1	0	2				
2	9	1		1	0	2				
3	11	1		0	0	2				
4	15	1		1	0	1				
1005		• • •		• • •	•••	• • • •				
1995	19	1		1	0	0				
1996	16	1		1	1	2				
1997	5	1		1	0	3				
1998	19	1		1	1	0				
1999	2	1		1	1	3				

[2000 rows x 21 columns]

```
In [6]: from sklearn.model selection import train test split
         x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.7,random_state=42)
         x_train.shape,x_test.shape
 Out[6]: ((1400, 20), (600, 20))
 In [7]: from sklearn.ensemble import RandomForestClassifier
         rfc=RandomForestClassifier()
         rfc.fit(x train,y train)
 Out[7]:
          ▼ RandomForestClassifier
          RandomForestClassifier()
 In [8]: rf=RandomForestClassifier()
 In [9]: params={'max depth':[2,3,5,10,20],'min samples leaf':[5,10,20,50,100,200],'n estimators':[10,25,30,50,100,200]}
In [10]: from sklearn.model selection import GridSearchCV
         grid search=GridSearchCV(estimator=rf,param grid=params,cv=2,scoring="accuracy")
         grid search.fit(x train,y train)
Out[10]:
                       GridSearchCV
           ▶ estimator: RandomForestClassifier
                ▶ RandomForestClassifier
        grid search.best score
In [11]:
Out[11]: 0.5364285714285715
```

```
In [12]: rf best=grid search.best estimator
           print(rf best)
           RandomForestClassifier(max depth=20, min samples leaf=100, n estimators=30)
In [13]: from sklearn.tree import plot tree
           plt.figure(figsize=(80,40))
           plot tree(rf best.estimators [5],feature names=x.columns,class names=["Yes","No"],filled=True);
                                                                                                          px height <= 1221.5
                                                                                                              gini = 0.498
                                                                                                            samples = 882
                                                                                                           value = [743, 657]
                                                                                                              class = Yes
                                                                                           n cores <= 7.5
                                                                                                                                gini = 0.47
                                                                                            gini = 0.499
                                                                                                                              samples = 107
                                                                                           samples = 775
                                                                                                                             value = [109, 66]
                                                                                          value = [634, 591]
                                                                                                                               class = Yes
                                                                                             class = Yes
                                                                          ram <= 1877.0
                                                                                                             gini = 0.485
                                                                           gini = 0.498
                                                                                                            samples = 101
                                                                          samples = 674
                                                                                                            value = [62, 88]
                                                                        value = [572, 503]
                                                                                                               class = No
                                                                            class = Yes
                                     mobile wt <= 158.5
                                                                                                               fc <= 4.5
                                        qini = 0.486
                                                                                                               aini = 0.5
                                       samples = 280
                                                                                                            samples = 394
                                     value = [263, 187]
                                                                                                           value = [309, 316]
                                         class = Yes
                                                                                                               class = No
                                                                                         clock speed <= 1.65
                       gini = 0.478
                                                          gini = 0.495
                                                                                                                               gini = 0.496
                                                                                            qini = 0.497
                      samples = 173
                                                        samples = 107
                                                                                                                              samples = 150
                                                                                           samples = 244
                    value = [169, 110]
                                                        value = [94, 77]
                                                                                                                            value = [132, 111]
                                                                                         value = [177, 205]
                        class = Yes
                                                          class = Yes
                                                                                                                               class = Yes
                                                                                             class = No
                                                                           gini = 0.499
                                                                                                              gini = 0.47
                                                                          samples = 136
                                                                                                            samples = 108
                                                                        value = [118, 108]
                                                                                                            value = [59, 97]
                                                                           class = Yes
                                                                                                               class = No
```

```
In [14]: from sklearn.tree import plot tree
           plt.figure(figsize=(80,40))
           plot tree(rf best.estimators [7],feature names=x.columns,class names=["Yes","No"],filled=True);
                                 mobile wt \leq 96.5
                                    aini = 0.499
                                   samples = 889
                                 value = [677, 723]
                                     class = No
                                                int memory \leq 11.5
                     gini = 0.472
                                                    qini = 0.5
                    samples = 111
                                                  samples = 778
                   value = [65, 105]
                                                value = [612, 618]
                      class = No
                                                    class = No
                                                                 ram <= 1017.5
                                     gini = 0.49
                                                                    qini = 0.5
                                   samples = 144
                                                                 samples = 634
                                  value = [98, 131]
                                                                value = [514, 487]
                                     class = No
                                                                   class = Yes
                                                                               talk time \leq 10.5
                                                   aini = 0.479
                                                                                   aini = 0.5
                                                  samples = 124
                                                                                samples = 510
                                                 value = [124, 82]
                                                                               value = [390, 405]
                                                    class = Yes
                                                                                  class = No
                                                   sc h <= 13.5
                                                                                                                 fc <= 2.5
                                                   qini = 0.494
                                                                                                                aini = 0.499
                                                  samples = 244
                                                                                                               samples = 266
                                                 value = [160, 198]
                                                                                                             value = [230, 207]
                                                    class = No
                                                                                                                class = Yes
                                     gini = 0.5
                                                                  gini = 0.469
                                                                                                  gini = 0.5
                                                                                                                               gini = 0.494
                                   samples = 141
                                                                 samples = 103
                                                                                               samples = 115
                                                                                                                              samples = 151
                                 value = [104, 105]
                                                                 value = [56, 93]
                                                                                               value = [93, 97]
                                                                                                                            value = [137, 110]
                                     class = No
                                                                                                  class = No
                                                                                                                                class = Yes
                                                                   class = No
In [15]: rf best.feature importances
Out[15]: array([0.11505021, 0.0733561, 0.01468286, 0.07053421, 0.01729881,
                   0.11712097, 0.07411121, 0.07676443, 0.05037934, 0.03578947,
                   0.07619111, 0.05351856, 0.08817483, 0.01732954, 0.03832682,
                   0.03229431, 0.01137491, 0.0128924 , 0.00719067, 0.01761924])
```

```
In [16]: imp_df=pd.DataFrame({"varname":x_train.columns,"imp":rf_best.feature_importances_})
imp_df.sort_values(by="imp",ascending=False)
```

Out[16]:

	varname	imp
5	int_memory	0.117121
0	battery_power	0.115050
12	ram	0.088175
7	mobile_wt	0.076764
10	px_height	0.076191
6	m_dep	0.074111
1	clock_speed	0.073356
3	fc	0.070534
11	px_width	0.053519
8	n_cores	0.050379
14	sc_w	0.038327
9	рс	0.035789
15	talk_time	0.032294
19	price_range	0.017619
13	sc_h	0.017330
4	four_g	0.017299
2	dual_sim	0.014683
17	touch_screen	0.012892
16	three_g	0.011375
18	wifi	0.007191