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

```
In [2]: df=pd.read_csv(r"C4_framingham.csv")
        df
```

Out[2]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp
0	1	39	4.0	0	0.0	0.0	0	0
1	0	46	2.0	0	0.0	0.0	0	0
2	1	48	1.0	1	20.0	0.0	0	0
3	0	61	3.0	1	30.0	0.0	0	1
4	0	46	3.0	1	23.0	0.0	0	0
...	...	...	...	...	...	...	...	...
4233	1	50	1.0	1	1.0	0.0	0	1
4234	1	51	3.0	1	43.0	0.0	0	0
4235	0	48	2.0	1	20.0	NaN	0	0
4236	0	44	1.0	1	15.0	0.0	0	0
4237	0	52	2.0	0	0.0	0.0	0	0

4238 rows × 16 columns

```
In [3]:  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4238 entries, 0 to 4237
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   male             4238 non-null   int64
 1   age              4238 non-null   int64
 2   education        4133 non-null   float64
 3   currentSmoker    4238 non-null   int64
 4   cigsPerDay       4209 non-null   float64
 5   BPMeds           4185 non-null   float64
 6   prevalentStroke  4238 non-null   int64
 7   prevalentHyp     4238 non-null   int64
 8   diabetes         4238 non-null   int64
 9   totChol          4188 non-null   float64
 10  sysBP            4238 non-null   float64
 11  diaBP            4238 non-null   float64
 12  BMI              4219 non-null   float64
 13  heartRate        4237 non-null   float64
 14  glucose          3850 non-null   float64
 15  TenYearCHD       4238 non-null   int64
dtypes: float64(9), int64(7)
memory usage: 529.9 KB
```

```
In [4]:  df=df.dropna()
```

```
In [5]:  df.isnull().sum()
```

```
Out[5]:  male               0
         age                0
         education          0
         currentSmoker      0
         cigsPerDay         0
         BPMeds             0
         prevalentStroke    0
         prevalentHyp       0
         diabetes           0
         totChol            0
         sysBP              0
         diaBP              0
         BMI                0
         heartRate          0
         glucose            0
         TenYearCHD         0
         dtype: int64
```

```
In [6]: df.describe()
```

Out[6]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevaler
count	3656.000000	3656.000000	3656.000000	3656.000000	3656.000000	3656.000000	3656
mean	0.443654	49.557440	1.979759	0.489059	9.022155	0.030361	0
std	0.496883	8.561133	1.022657	0.499949	11.918869	0.171602	0
min	0.000000	32.000000	1.000000	0.000000	0.000000	0.000000	0
25%	0.000000	42.000000	1.000000	0.000000	0.000000	0.000000	0
50%	0.000000	49.000000	2.000000	0.000000	0.000000	0.000000	0
75%	1.000000	56.000000	3.000000	1.000000	20.000000	0.000000	0
max	1.000000	70.000000	4.000000	1.000000	70.000000	1.000000	1

```
In [7]: df.columns
```

Out[7]: Index(['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
       'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
       'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD'],
       dtype='object')

```
In [8]: df["TenYearCHD"].value_counts()
```

Out[8]: 0   3099
        1   557
        Name: TenYearCHD, dtype: int64

```
In [9]: df1=df[['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
          'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
          'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD']]
```

```
In [10]: x=df1.drop("TenYearCHD",axis=1)
         y=df1["TenYearCHD"]
```

```
In [11]: from sklearn.model_selection import train_test_split
         x_train,x_test,y_train,y_test=train_test_split(x,y,train_size=0.70)
```

```
In [12]: from sklearn.ensemble import RandomForestClassifier
         rfc=RandomForestClassifier()
         rfc.fit(x_train,y_train)
```

Out[12]: RandomForestClassifier()

```
In [13]: parameters={'max_depth':[1,2,3,4,5],
                     'min_samples_leaf':[5,10,15,20,25],
                     'n_estimators':[10,20,30,40,50]}
```

```python
In [14]: from sklearn.model_selection import GridSearchCV
         grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accu
         grid_search.fit(x_train,y_train)
```

```
Out[14]: GridSearchCV(cv=2, estimator=RandomForestClassifier(),
                      param_grid={'max_depth': [1, 2, 3, 4, 5],
                                  'min_samples_leaf': [5, 10, 15, 20, 25],
                                  'n_estimators': [10, 20, 30, 40, 50]},
                      scoring='accuracy')
```

```python
In [15]: grid_search.best_score_
```

```
Out[15]: 0.8495507354378421
```

```python
In [16]: parameters={'max_depth':[1,2,3,4,5],
                     'min_samples_leaf':[5,10,15,20,25],
                     'n_estimators':[10,20,30,40,50]}
```

```python
In [17]: rfc_best=grid_search.best_estimator_
```

```python
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','N
```

Out[18]: [Text(2557.5, 1993.2, 'prevalentHyp <= 0.5\ngini = 0.261\nsamples = 1626\nval
ue = [2164, 395]\nclass = Yes'),
 Text(1488.0, 1630.8000000000002, 'education <= 1.5\ngini = 0.189\nsamples =
1111\nvalue = [1580, 187]\nclass = Yes'),
 Text(744.0, 1268.4, 'age <= 46.5\ngini = 0.265\nsamples = 412\nvalue = [564,
105]\nclass = Yes'),
 Text(372.0, 906.0, 'male <= 0.5\ngini = 0.081\nsamples = 154\nvalue = [248,
11]\nclass = Yes'),
 Text(186.0, 543.5999999999999, 'glucose <= 65.5\ngini = 0.096\nsamples = 94
\nvalue = [151, 8]\nclass = Yes'),
 Text(93.0, 181.19999999999982, 'gini = 0.355\nsamples = 9\nvalue = [10, 3]\n
class = Yes'),
 Text(279.0, 181.19999999999982, 'gini = 0.066\nsamples = 85\nvalue = [141,
5]\nclass = Yes'),
 Text(558.0, 543.5999999999999, 'sysBP <= 131.0\ngini = 0.058\nsamples = 60\n
value = [97, 3]\nclass = Yes'),
 Text(465.0, 181.19999999999982, 'gini = 0.0\nsamples = 48\nvalue = [81, 0]\n
class = Yes'),
 Text(651.0, 181.19999999999982, 'gini = 0.266\nsamples = 12\nvalue = [16, 3]
\nclass = Yes'),
 Text(1116.0, 906.0, 'diaBP <= 87.75\ngini = 0.353\nsamples = 258\nvalue = [3
16, 94]\nclass = Yes'),
 Text(930.0, 543.5999999999999, 'age <= 65.5\ngini = 0.308\nsamples = 225\nva
lue = [289, 68]\nclass = Yes'),
 Text(837.0, 181.19999999999982, 'gini = 0.284\nsamples = 214\nvalue = [280,
58]\nclass = Yes'),
 Text(1023.0, 181.19999999999982, 'gini = 0.499\nsamples = 11\nvalue = [9, 1
0]\nclass = No'),
 Text(1302.0, 543.5999999999999, 'totChol <= 272.0\ngini = 0.5\nsamples = 33
\nvalue = [27, 26]\nclass = Yes'),
 Text(1209.0, 181.19999999999982, 'gini = 0.48\nsamples = 22\nvalue = [16, 2
4]\nclass = No'),
 Text(1395.0, 181.19999999999982, 'gini = 0.26\nsamples = 11\nvalue = [11, 2]
\nclass = Yes'),
 Text(2232.0, 1268.4, 'male <= 0.5\ngini = 0.138\nsamples = 699\nvalue = [101
6, 82]\nclass = Yes'),
 Text(1860.0, 906.0, 'diaBP <= 77.75\ngini = 0.1\nsamples = 394\nvalue = [59
3, 33]\nclass = Yes'),
 Text(1674.0, 543.5999999999999, 'age <= 51.5\ngini = 0.052\nsamples = 215\nv
alue = [327, 9]\nclass = Yes'),
 Text(1581.0, 181.19999999999982, 'gini = 0.022\nsamples = 175\nvalue = [270,
3]\nclass = Yes'),
 Text(1767.0, 181.19999999999982, 'gini = 0.172\nsamples = 40\nvalue = [57,
6]\nclass = Yes'),
 Text(2046.0, 543.5999999999999, 'BMI <= 23.43\ngini = 0.152\nsamples = 179\n
value = [266, 24]\nclass = Yes'),
 Text(1953.0, 181.19999999999982, 'gini = 0.25\nsamples = 74\nvalue = [99, 1
7]\nclass = Yes'),
 Text(2139.0, 181.19999999999982, 'gini = 0.077\nsamples = 105\nvalue = [167,
7]\nclass = Yes'),
 Text(2604.0, 906.0, 'currentSmoker <= 0.5\ngini = 0.186\nsamples = 305\nvalu
e = [423, 49]\nclass = Yes'),
 Text(2418.0, 543.5999999999999, 'glucose <= 73.5\ngini = 0.143\nsamples = 11
9\nvalue = [166, 14]\nclass = Yes'),
 Text(2325.0, 181.19999999999982, 'gini = 0.039\nsamples = 36\nvalue = [49,
1]\nclass = Yes'),
 Text(2511.0, 181.19999999999982, 'gini = 0.18\nsamples = 83\nvalue = [117, 1

```
3]\nclass = Yes'),
 Text(2790.0, 543.5999999999999, 'diaBP <= 92.5\ngini = 0.211\nsamples = 186
\nvalue = [257, 35]\nclass = Yes'),
 Text(2697.0, 181.19999999999982, 'gini = 0.201\nsamples = 180\nvalue = [251,
32]\nclass = Yes'),
 Text(2883.0, 181.19999999999982, 'gini = 0.444\nsamples = 6\nvalue = [6, 3]
\nclass = Yes'),
 Text(3627.0, 1630.8000000000002, 'BMI <= 19.67\ngini = 0.387\nsamples = 515
\nvalue = [584, 208]\nclass = Yes'),
 Text(3534.0, 1268.4, 'gini = 0.426\nsamples = 7\nvalue = [4, 9]\nclass = N
o'),
 Text(3720.0, 1268.4, 'sysBP <= 157.75\ngini = 0.38\nsamples = 508\nvalue =
[580, 199]\nclass = Yes'),
 Text(3348.0, 906.0, 'age <= 46.5\ngini = 0.301\nsamples = 312\nvalue = [375,
85]\nclass = Yes'),
 Text(3162.0, 543.5999999999999, 'glucose <= 71.0\ngini = 0.122\nsamples = 89
\nvalue = [115, 8]\nclass = Yes'),
 Text(3069.0, 181.19999999999982, 'gini = 0.298\nsamples = 25\nvalue = [27,
6]\nclass = Yes'),
 Text(3255.0, 181.19999999999982, 'gini = 0.043\nsamples = 64\nvalue = [88,
2]\nclass = Yes'),
 Text(3534.0, 543.5999999999999, 'age <= 64.5\ngini = 0.353\nsamples = 223\nv
alue = [260, 77]\nclass = Yes'),
 Text(3441.0, 181.19999999999982, 'gini = 0.333\nsamples = 204\nvalue = [243,
65]\nclass = Yes'),
 Text(3627.0, 181.19999999999982, 'gini = 0.485\nsamples = 19\nvalue = [17, 1
2]\nclass = Yes'),
 Text(4092.0, 906.0, 'male <= 0.5\ngini = 0.459\nsamples = 196\nvalue = [205,
114]\nclass = Yes'),
 Text(3906.0, 543.5999999999999, 'diaBP <= 81.0\ngini = 0.409\nsamples = 132
\nvalue = [157, 63]\nclass = Yes'),
 Text(3813.0, 181.19999999999982, 'gini = 0.219\nsamples = 5\nvalue = [1, 7]
\nclass = No'),
 Text(3999.0, 181.19999999999982, 'gini = 0.389\nsamples = 127\nvalue = [156,
56]\nclass = Yes'),
 Text(4278.0, 543.5999999999999, 'age <= 49.0\ngini = 0.5\nsamples = 64\nvalu
e = [48, 51]\nclass = No'),
 Text(4185.0, 181.19999999999982, 'gini = 0.328\nsamples = 19\nvalue = [23,
6]\nclass = Yes'),
 Text(4371.0, 181.19999999999982, 'gini = 0.459\nsamples = 45\nvalue = [25, 4
5]\nclass = No')]
```

In [ ]: