```
In [81]: import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
```

```
In [139]: df=pd.read_csv(r"C:\USERS\user\Downloads\C4_framingham - C4_framingham.csv")
```

Out[139]:

	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 [140]:
```

Out[140]: Index(['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
               'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
               'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD'],
              dtype='object')

In [141]: `df=df.head(21)`
`df`

Out[141]:

	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
5	0	43	2.0	0	0.0	0.0	0	1
6	0	63	1.0	0	0.0	0.0	0	0
7	0	45	2.0	1	20.0	0.0	0	0
8	1	52	1.0	0	0.0	0.0	0	1
9	1	43	1.0	1	30.0	0.0	0	1
10	0	50	1.0	0	0.0	0.0	0	0
11	0	43	2.0	0	0.0	0.0	0	0
12	1	46	1.0	1	15.0	0.0	0	1
13	0	41	3.0	0	0.0	1.0	0	1
14	0	39	2.0	1	9.0	0.0	0	0
15	0	38	2.0	1	20.0	0.0	0	1
16	1	48	3.0	1	10.0	0.0	0	1
17	0	46	2.0	1	20.0	0.0	0	0
18	0	38	2.0	1	5.0	0.0	0	0
19	1	41	2.0	0	0.0	0.0	0	0
20	0	42	2.0	1	30.0	0.0	0	0

```
In [142]: a=df[['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMeds',
             'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
             'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD']]
```

Out[142]:

	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
5	0	43	2.0	0	0.0	0.0	0	1
6	0	63	1.0	0	0.0	0.0	0	0
7	0	45	2.0	1	20.0	0.0	0	0
8	1	52	1.0	0	0.0	0.0	0	1
9	1	43	1.0	1	30.0	0.0	0	1
10	0	50	1.0	0	0.0	0.0	0	0
11	0	43	2.0	0	0.0	0.0	0	0
12	1	46	1.0	1	15.0	0.0	0	1
13	0	41	3.0	0	0.0	1.0	0	1
14	0	39	2.0	1	9.0	0.0	0	0
15	0	38	2.0	1	20.0	0.0	0	1
16	1	48	3.0	1	10.0	0.0	0	1
17	0	46	2.0	1	20.0	0.0	0	0
18	0	38	2.0	1	5.0	0.0	0	0
19	1	41	2.0	0	0.0	0.0	0	0
20	0	42	2.0	1	30.0	0.0	0	0

In [143]:

Out[143]:
```
85.0     3
70.0     2
78.0     2
76.0     2
103.0    1
88.0     1
99.0     1
79.0     1
77.0     1
72.0     1
84.0     1
64.0     1
61.0     1
65.0     1
89.0     1
Name: glucose, dtype: int64
```

In [124]:
```python
x=a.drop('glucose',axis=1)
```

```
In [144]: g1={"glucose":{'NaN':1}}
          a=a.replace(g1)
```

```
      male   age   education   currentSmoker   cigsPerDay   BPMeds   prevalentStroke
\
0       1    39       4.0                 0          0.0      0.0                   0
1       0    46       2.0                 0          0.0      0.0                   0
2       1    48       1.0                 1         20.0      0.0                   0
3       0    61       3.0                 1         30.0      0.0                   0
4       0    46       3.0                 1         23.0      0.0                   0
5       0    43       2.0                 0          0.0      0.0                   0
6       0    63       1.0                 0          0.0      0.0                   0
7       0    45       2.0                 1         20.0      0.0                   0
8       1    52       1.0                 0          0.0      0.0                   0
9       1    43       1.0                 1         30.0      0.0                   0
10      0    50       1.0                 0          0.0      0.0                   0
11      0    43       2.0                 0          0.0      0.0                   0
12      1    46       1.0                 1         15.0      0.0                   0
13      0    41       3.0                 0          0.0      1.0                   0
14      0    39       2.0                 1          9.0      0.0                   0
15      0    38       2.0                 1         20.0      0.0                   0
16      1    48       3.0                 1         10.0      0.0                   0
17      0    46       2.0                 1         20.0      0.0                   0
18      0    38       2.0                 1          5.0      0.0                   0
19      1    41       2.0                 0          0.0      0.0                   0
20      0    42       2.0                 1         30.0      0.0                   0

      prevalentHyp   diabetes   totChol   sysBP   diaBP    BMI   heartRate   glucose
\
0                0          0     195.0   106.0    70.0  26.97        80.0      77.0
1                0          0     250.0   121.0    81.0  28.73        95.0      76.0
2                0          0     245.0   127.5    80.0  25.34        75.0      70.0
3                1          0     225.0   150.0    95.0  28.58        65.0     103.0
4                0          0     285.0   130.0    84.0  23.10        85.0      85.0
5                1          0     228.0   180.0   110.0  30.30        77.0      99.0
6                0          0     205.0   138.0    71.0  33.11        60.0      85.0
7                0          0     313.0   100.0    71.0  21.68        79.0      78.0
8                1          0     260.0   141.5    89.0  26.36        76.0      79.0
9                1          0     225.0   162.0   107.0  23.61        93.0      88.0
10               0          0     254.0   133.0    76.0  22.91        75.0      76.0
11               0          0     247.0   131.0    88.0  27.64        72.0      61.0
12               1          0     294.0   142.0    94.0  26.31        98.0      64.0
13               1          0     332.0   124.0    88.0  31.31        65.0      84.0
14               0          0     226.0   114.0    64.0  22.35        85.0       NaN
15               1          0     221.0   140.0    90.0  21.35        95.0      70.0
16               1          0     232.0   138.0    90.0  22.37        64.0      72.0
17               0          0     291.0   112.0    78.0  23.38        80.0      89.0
18               0          0     195.0   122.0    84.5  23.24        75.0      78.0
19               0          0     195.0   139.0    88.0  26.88        85.0      65.0
20               0          0     190.0   108.0    70.5  21.59        72.0      85.0

      TenYearCHD
0              0
1              0
2              0
3              1
```

```
4          0
5          0
6          1
7          0
8          0
9          0
10         0
11         0
12         0
13         0
14         0
15         1
16         0
17         1
18         0
19         0
20         0
```

In [145]:
```python
from sklearn.model_selection import train_test_split
```

In [146]:
```python
from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier()
```

Out[146]: RandomForestClassifier()

In [147]:
```python
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
```

In [148]:
```python
from sklearn.model_selection import GridSearchCV

grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
```

Out[148]:
```
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')
```

In [149]:

Out[149]: 0.5833333333333333

In [150]:

In [151]:
```python
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
```

Out[151]: [Text(2232.0, 1087.2, 'gini = 0.49\nsamples = 6\nvalue = [3, 4]\nclass = No
          ')]

gini = 0.49
samples = 6
value = [3, 4]
class = No

In [ ]: