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

In [2]:
```python
da=pd.read_csv(r"C:\Users\user\Downloads\framingham.csv")
da
```

Out[2]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
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]:
```python
df=da.fillna(value=0)
df
```

Out[3]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
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	0.0	0	0	
4236	0	44	1.0	1	15.0	0.0	0	0	

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabe
4237	0	52	2.0	0	0.0	0.0	0	0	

4238 rows × 16 columns

In [4]:
```python
df['TenYearCHD'].value_counts()
```

Out[4]:
```
0    3594
1     644
Name: TenYearCHD, dtype: int64
```

In [5]:
```python
df['TenYearCHD'].value_counts()
```

Out[5]:
```
0    3594
1     644
Name: TenYearCHD, dtype: int64
```

In [6]:
```python
x=df.drop('TenYearCHD',axis=1)
y=df['TenYearCHD']
```

In [7]:
```python
g1={"TenYearCHD":{'1':3,'0':4}}
df=df.replace(g1)
print(df)
```

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

      prevalentStroke  prevalentHyp  diabetes  totChol  sysBP  diaBP   BMI  \
0                   0             0         0    195.0  106.0   70.0  26.97
1                   0             0         0    250.0  121.0   81.0  28.73
2                   0             0         0    245.0  127.5   80.0  25.34
3                   0             1         0    225.0  150.0   95.0  28.58
4                   0             0         0    285.0  130.0   84.0  23.10
...               ...           ...       ...      ...    ...    ...    ...
4233                0             1         0    313.0  179.0   92.0  25.97
4234                0             0         0    207.0  126.5   80.0  19.71
4235                0             0         0    248.0  131.0   72.0  22.00
4236                0             0         0    210.0  126.5   87.0  19.16
4237                0             0         0    269.0  133.5   83.0  21.47

      heartRate  glucose  TenYearCHD
0          80.0     77.0           0
1          95.0     76.0           0
2          75.0     70.0           0
3          65.0    103.0           1
4          85.0     85.0           0
```

```
...         ...       ...              ...
4233        66.0      86.0              1
4234        65.0      68.0              0
4235        84.0      86.0              0
4236        86.0       0.0              0
4237        80.0     107.0              0

[4238 rows x 16 columns]
```

In [8]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.70)
```

In [9]:
```python
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[9]: RandomForestClassifier()

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

In [11]:
```python
from sklearn.model_selection import GridSearchCV
grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

Out[11]: 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 [12]:
```python
grid_search.best_score_
```

Out[12]: 0.845789382459268

In [13]:
```python
rfc_best=grid_search.best_estimator_
```

In [14]:
```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','No'],fill
```

Out[14]: [Text(2265.818181818182, 1993.2, 'age <= 58.5\ngini = 0.26\nsamples = 808\nvalue = [107
6, 195]\nclass = Yes'),
 Text(1285.0909090909092, 1630.8000000000002, 'diaBP <= 97.25\ngini = 0.203\nsamples = 6
59\nvalue = [913, 118]\nclass = Yes'),
 Text(676.3636363636364, 1268.4, 'diaBP <= 70.75\ngini = 0.183\nsamples = 582\nvalue =
[821, 93]\nclass = Yes'),
 Text(270.54545454545456, 906.0, 'BMI <= 19.015\ngini = 0.089\nsamples = 110\nvalue = [1
64, 8]\nclass = Yes'),
 Text(135.27272727272728, 543.5999999999999, 'gini = 0.375\nsamples = 6\nvalue = [6, 2]

```
\nclass = Yes'),
 Text(405.81818181818187, 543.5999999999999, 'cigsPerDay <= 13.5\ngini = 0.07\nsamples =
104\nvalue = [158, 6]\nclass = Yes'),
 Text(270.54545454545456, 181.19999999999982, 'gini = 0.035\nsamples = 70\nvalue = [109,
2]\nclass = Yes'),
 Text(541.0909090909091, 181.19999999999982, 'gini = 0.14\nsamples = 34\nvalue = [49, 4]
\nclass = Yes'),
 Text(1082.1818181818182, 906.0, 'totChol <= 348.0\ngini = 0.203\nsamples = 472\nvalue =
[657, 85]\nclass = Yes'),
 Text(946.909090909091, 543.5999999999999, 'cigsPerDay <= 27.5\ngini = 0.191\nsamples =
464\nvalue = [651, 78]\nclass = Yes'),
 Text(811.6363636363637, 181.19999999999982, 'gini = 0.168\nsamples = 415\nvalue = [597,
61]\nclass = Yes'),
 Text(1082.1818181818182, 181.19999999999982, 'gini = 0.364\nsamples = 49\nvalue = [54,
17]\nclass = Yes'),
 Text(1217.4545454545455, 543.5999999999999, 'gini = 0.497\nsamples = 8\nvalue = [6, 7]
\nclass = No'),
 Text(1893.818181818182, 1268.4, 'sysBP <= 176.5\ngini = 0.336\nsamples = 77\nvalue = [9
2, 25]\nclass = Yes'),
 Text(1758.5454545454547, 906.0, 'male <= 0.5\ngini = 0.264\nsamples = 62\nvalue = [81,
15]\nclass = Yes'),
 Text(1488.0, 543.5999999999999, 'heartRate <= 83.5\ngini = 0.091\nsamples = 26\nvalue =
[40, 2]\nclass = Yes'),
 Text(1352.7272727272727, 181.19999999999982, 'gini = 0.0\nsamples = 14\nvalue = [24, 0]
\nclass = Yes'),
 Text(1623.2727272727275, 181.19999999999982, 'gini = 0.198\nsamples = 12\nvalue = [16,
2]\nclass = Yes'),
 Text(2029.0909090909092, 543.5999999999999, 'BMI <= 27.595\ngini = 0.366\nsamples = 36
\nvalue = [41, 13]\nclass = Yes'),
 Text(1893.818181818182, 181.19999999999982, 'gini = 0.239\nsamples = 22\nvalue = [31,
5]\nclass = Yes'),
 Text(2164.3636363636365, 181.19999999999982, 'gini = 0.494\nsamples = 14\nvalue = [10,
8]\nclass = Yes'),
 Text(2029.0909090909092, 906.0, 'gini = 0.499\nsamples = 15\nvalue = [11, 10]\nclass =
Yes'),
 Text(3246.545454545455, 1630.8000000000002, 'BMI <= 20.585\ngini = 0.436\nsamples = 149
\nvalue = [163, 77]\nclass = Yes'),
 Text(3111.2727272727275, 1268.4, 'gini = 0.188\nsamples = 9\nvalue = [2, 17]\nclass = N
o'),
 Text(3381.818181818182, 1268.4, 'sysBP <= 166.0\ngini = 0.396\nsamples = 140\nvalue =
[161, 60]\nclass = Yes'),
 Text(2840.727272727273, 906.0, 'education <= 3.5\ngini = 0.332\nsamples = 111\nvalue =
[143, 38]\nclass = Yes'),
 Text(2570.1818181818185, 543.5999999999999, 'cigsPerDay <= 9.5\ngini = 0.285\nsamples =
98\nvalue = [135, 28]\nclass = Yes'),
 Text(2434.909090909091, 181.19999999999982, 'gini = 0.233\nsamples = 79\nvalue = [116,
18]\nclass = Yes'),
 Text(2705.4545454545455, 181.19999999999982, 'gini = 0.452\nsamples = 19\nvalue = [19,
10]\nclass = Yes'),
 Text(3111.2727272727275, 543.5999999999999, 'heartRate <= 73.0\ngini = 0.494\nsamples =
13\nvalue = [8, 10]\nclass = No'),
 Text(2976.0, 181.19999999999982, 'gini = 0.32\nsamples = 7\nvalue = [2, 8]\nclass = N
o'),
 Text(3246.545454545455, 181.19999999999982, 'gini = 0.375\nsamples = 6\nvalue = [6, 2]
\nclass = Yes'),
 Text(3922.909090909091, 906.0, 'diaBP <= 102.5\ngini = 0.495\nsamples = 29\nvalue = [1
8, 22]\nclass = No'),
 Text(3652.3636363636365, 543.5999999999999, 'cigsPerDay <= 8.0\ngini = 0.444\nsamples =
18\nvalue = [16, 8]\nclass = Yes'),
 Text(3517.0909090909095, 181.19999999999982, 'gini = 0.469\nsamples = 12\nvalue = [10,
6]\nclass = Yes'),
 Text(3787.636363636364, 181.19999999999982, 'gini = 0.375\nsamples = 6\nvalue = [6, 2]
\nclass = Yes'),
 Text(4193.454545454546, 543.5999999999999, 'cigsPerDay <= 0.5\ngini = 0.219\nsamples =
11\nvalue = [2, 14]\nclass = No'),
```

```
  Text(4058.1818181818185, 181.19999999999982, 'gini = 0.0\nsamples = 5\nvalue = [0, 7]\n
class = No'),
  Text(4328.727272727273, 181.19999999999982, 'gini = 0.346\nsamples = 6\nvalue = [2, 7]
\nclass = No')]
```