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]}
```

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')
```

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

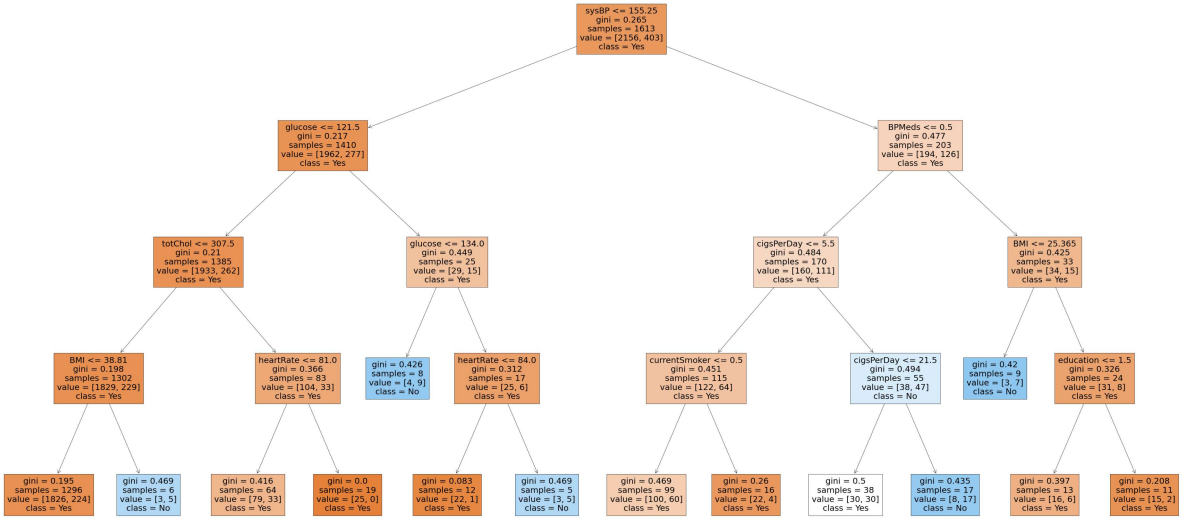
Out[15]:
```
0.8452517225371383
```

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]}
```

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

In [18]:
```
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(2325.0, 1956.96, 'sysBP <= 155.25\ngini = 0.265\nsamples = 1613\nvalue
= [2156, 403]\nclass = Yes'),
 Text(1209.0, 1522.0800000000002, 'glucose <= 121.5\ngini = 0.217\nsamples =
1410\nvalue = [1962, 277]\nclass = Yes'),
 Text(744.0, 1087.2, 'totChol <= 307.5\ngini = 0.21\nsamples = 1385\nvalue =
[1933, 262]\nclass = Yes'),
 Text(372.0, 652.3200000000002, 'BMI <= 38.81\ngini = 0.198\nsamples = 1302\n
value = [1829, 229]\nclass = Yes'),
 Text(186.0, 217.44000000000005, 'gini = 0.195\nsamples = 1296\nvalue = [182
6, 224]\nclass = Yes'),
 Text(558.0, 217.44000000000005, 'gini = 0.469\nsamples = 6\nvalue = [3, 5]\n
class = No'),
 Text(1116.0, 652.3200000000002, 'heartRate <= 81.0\ngini = 0.366\nsamples =
83\nvalue = [104, 33]\nclass = Yes'),
 Text(930.0, 217.44000000000005, 'gini = 0.416\nsamples = 64\nvalue = [79, 3
3]\nclass = Yes'),
 Text(1302.0, 217.44000000000005, 'gini = 0.0\nsamples = 19\nvalue = [25, 0]
\nclass = Yes'),
 Text(1674.0, 1087.2, 'glucose <= 134.0\ngini = 0.449\nsamples = 25\nvalue =
[29, 15]\nclass = Yes'),
 Text(1488.0, 652.3200000000002, 'gini = 0.426\nsamples = 8\nvalue = [4, 9]\n
class = No'),
 Text(1860.0, 652.3200000000002, 'heartRate <= 84.0\ngini = 0.312\nsamples =
17\nvalue = [25, 6]\nclass = Yes'),
 Text(1674.0, 217.44000000000005, 'gini = 0.083\nsamples = 12\nvalue = [22,
1]\nclass = Yes'),
 Text(2046.0, 217.44000000000005, 'gini = 0.469\nsamples = 5\nvalue = [3, 5]
\nclass = No'),
 Text(3441.0, 1522.0800000000002, 'BPMeds <= 0.5\ngini = 0.477\nsamples = 203
\nvalue = [194, 126]\nclass = Yes'),
 Text(2976.0, 1087.2, 'cigsPerDay <= 5.5\ngini = 0.484\nsamples = 170\nvalue
= [160, 111]\nclass = Yes'),
 Text(2604.0, 652.3200000000002, 'currentSmoker <= 0.5\ngini = 0.451\nsamples
= 115\nvalue = [122, 64]\nclass = Yes'),
 Text(2418.0, 217.44000000000005, 'gini = 0.469\nsamples = 99\nvalue = [100,
60]\nclass = Yes'),
 Text(2790.0, 217.44000000000005, 'gini = 0.26\nsamples = 16\nvalue = [22, 4]
\nclass = Yes'),
 Text(3348.0, 652.3200000000002, 'cigsPerDay <= 21.5\ngini = 0.494\nsamples =
55\nvalue = [38, 47]\nclass = No'),
 Text(3162.0, 217.44000000000005, 'gini = 0.5\nsamples = 38\nvalue = [30, 30]
\nclass = Yes'),
 Text(3534.0, 217.44000000000005, 'gini = 0.435\nsamples = 17\nvalue = [8, 1
7]\nclass = No'),
 Text(3906.0, 1087.2, 'BMI <= 25.365\ngini = 0.425\nsamples = 33\nvalue = [3
4, 15]\nclass = Yes'),
 Text(3720.0, 652.3200000000002, 'gini = 0.42\nsamples = 9\nvalue = [3, 7]\nc
lass = No'),
 Text(4092.0, 652.3200000000002, 'education <= 1.5\ngini = 0.326\nsamples = 2
4\nvalue = [31, 8]\nclass = Yes'),
 Text(3906.0, 217.44000000000005, 'gini = 0.397\nsamples = 13\nvalue = [16,
6]\nclass = Yes'),
 Text(4278.0, 217.44000000000005, 'gini = 0.208\nsamples = 11\nvalue = [15,
2]\nclass = Yes')]
```


sysBP <= 155.25
gini = 0.265
samples = 1613
value = [2156, 403]
class = Yes
glucose <= 121.5
gini = 0.217
samples = 1410
value = [1962, 277]
class = Yes
BPMeds <= 0.5
gini = 0.477
samples = 203
value = [194, 126]
class = Yes
totChol <= 307.5
gini = 0.21
samples = 1385
value = [1933, 262]
class = Yes
glucose <= 134.0
gini = 0.449
samples = 25
value = [29, 15]
class = Yes
cigsPerDay <= 5.5
gini = 0.484
samples = 170
value = [160, 111]
class = Yes
BMI <= 25.365
gini = 0.425
samples = 33
value = [34, 15]
class = Yes
BMI <= 38.81
gini = 0.198
samples = 1302
value = [1829, 229]
class = Yes
heartRate <= 81.0
gini = 0.366
samples = 83
value = [104, 33]
class = Yes
gini = 0.426
samples = 8
value = [4, 9]
class = No
heartRate <= 84.0
gini = 0.312
samples = 17
value = [25, 6]
class = Yes
currentSmoker <= 0.5
gini = 0.451
samples = 115
value = [122, 64]
class = Yes
cigsPerDay <= 21.5
gini = 0.494
samples = 55
value = [38, 47]
class = No
gini = 0.42
samples = 9
value = [3, 7]
class = No
education <= 1.5
gini = 0.326
samples = 24
value = [31, 8]
class = Yes
gini = 0.195
samples = 1296
value = [1826, 224]
class = Yes
gini = 0.469
samples = 6
value = [3, 5]
class = No
gini = 0.416
samples = 64
value = [79, 33]
class = Yes
gini = 0.0
samples = 19
value = [25, 0]
class = Yes
gini = 0.083
samples = 12
value = [22, 1]
class = Yes
gini = 0.469
samples = 5
value = [3, 5]
class = No
gini = 0.469
samples = 99
value = [100, 60]
class = Yes
gini = 0.26
samples = 16
value = [22, 4]
class = Yes
gini = 0.5
samples = 38
value = [30, 30]
class = Yes
gini = 0.435
samples = 17
value = [8, 17]
class = No
gini = 0.397
samples = 13
value = [16, 6]
class = Yes
gini = 0.208
samples = 11
value = [15, 2]
class = Yes