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('fram.csv')
df
```

age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBl
39	4.0	0	0.0	0.0	0	0	0	195.0	106.
46	2.0	0	0.0	0.0	0	0	0	250.0	121.
48	1.0	1	20.0	0.0	0	0	0	245.0	127.
61	3.0	1	30.0	0.0	0	1	0	225.0	150.
46	3.0	1	23.0	0.0	0	0	0	285.0	130.
...	...	...	...	...	...	...	...	...	.
50	1.0	1	1.0	0.0	0	1	0	313.0	179.
51	3.0	1	43.0	0.0	0	0	0	207.0	126.
48	2.0	1	20.0	NaN	0	0	0	248.0	131.
44	1.0	1	15.0	0.0	0	0	0	210.0	126.
52	2.0	0	0.0	0.0	0	0	0	269.0	133.

In [3]:

```
df.columns
```

Out[3]:

```
Index(['male', 'age', 'education', 'currentSmoker', 'cigsPerDay', 'BPMed
s',
       'prevalentStroke', 'prevalentHyp', 'diabetes', 'totChol', 'sysBP',
       'diaBP', 'BMI', 'heartRate', 'glucose', 'TenYearCHD'],
      dtype='object')
```

In [4]:

```
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
```

```
df['glucose'].value_counts()
```

In [6]:

```
x=df[['male', 'age','currentSmoker','prevalentStroke', 'prevalentHyp', 'diabetes']]
y=df['TenYearCHD']
```

d={"Verified":{'True':1,'False ':2}} df=df.replace(df) print(df)

In [7]:

```
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 [8]:

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

Out[8]:

```
RandomForestClassifier()
```

Depth of Tree

In [9]:

```
parameters={"max_depth":[1,2,3,4,5],"min_samples_leaf":[5,23,45,76,78],'n_estimators':[10
```

Cross Validate

In [10]:

```
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[10]:

```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2, 3, 4, 5],
                         'min_samples_leaf': [5, 23, 45, 76, 78],
                         'n_estimators': [10, 23, 45, 65, 7]},
             scoring='accuracy')
```

Score

In [11]:

```
grid_search.best_score_
```

Out[11]:

```
0.8544458475709404
```

In [12]:

```
rfc_best=grid_search.best_estimator_
```

In [13]:

```
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'],filled
```

Out[13]:

```
[Text(2039.5862068965519, 1993.2, 'male <= 0.5\ngini = 0.231\nsamples = 80
4\nvalue = [1102, 169]\nclass = Yes'),
 Text(692.6896551724138, 1630.8000000000002, 'currentSmoker <= 0.5\ngini =
0.179\nsamples = 463\nvalue = [662, 73]\nclass = Yes'),
 Text(538.7586206896552, 1268.4, 'gini = 0.196\nsamples = 266\nvalue = [37
9, 47]\nclass = Yes'),
 Text(846.6206896551724, 1268.4, 'age <= 55.5\ngini = 0.154\nsamples = 197
\nvalue = [283, 26]\nclass = Yes'),
 Text(307.86206896551727, 906.0, 'age <= 38.5\ngini = 0.13\nsamples = 167
\nvalue = [252, 19]\nclass = Yes'),
 Text(153.93103448275863, 543.5999999999999, 'gini = 0.0\nsamples = 27\nva
lue = [45, 0]\nclass = Yes'),
 Text(461.79310344827593, 543.5999999999999, 'prevalentHyp <= 0.5\ngini =
0.154\nsamples = 140\nvalue = [207, 19]\nclass = Yes'),
 Text(307.86206896551727, 181.19999999999982, 'gini = 0.145\nsamples = 122
\nvalue = [187, 16]\nclass = Yes'),
 Text(1385.3793103448277, 906.0, 'age <= 59.5\ngini = 0.301\nsamples = 30
\nvalue = [31, 7]\nclass = Yes'),
 Text(1077.5172413793105, 543.5999999999999, 'prevalentHyp <= 0.5\ngini =
0.401\nsamples = 14\nvalue = [12, 5]\nclass = Yes'),
 Text(923.5862068965519, 181.19999999999982, 'gini = 0.298\nsamples = 8\nv
alue = [9, 2]\nclass = Yes'),
 Text(1231.448275862069, 181.19999999999982, 'gini = 0.49\nsamples = 6\nva
lue = [4, 3]\nclass = Yes'),
 Text(1693.2413793103449, 543.5999999999999, 'prevalentHyp <= 0.5\ngini =
0.18\nsamples = 16\nvalue = [18, 2]\nclass = Yes'),
 Text(1539.3103448275863, 181.19999999999982, 'gini = 0.278\nsamples = 5\n
value = [5, 1]\nclass = Yes'),
 Text(1847.1724137931037, 181.19999999999982, 'gini = 0.133\nsamples = 11
```