

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

```
In [2]: df=pd.read_csv(r"C:\Users\user\Downloads\C4_framingham.csv")
df
```

```
Out[2]:
```

	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaB
1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.
0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.
1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.
0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.
0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.
...	...	...	...	...	...	...	...	...	...	...	...
1	50	1.0	1	1.0	0.0	0	1	0	313.0	179.0	92.
1	51	3.0	1	43.0	0.0	0	0	0	207.0	126.5	80.
0	48	2.0	1	20.0	NaN	0	0	0	248.0	131.0	72.
0	44	1.0	1	15.0	0.0	0	0	0	210.0	126.5	87.
0	52	2.0	0	0.0	0.0	0	0	0	269.0	133.5	83.

× 16 columns

```
In [9]: d=df.fillna(20)
```

```
In [10]: d['TenYearCHD'].value_counts()
```

```
Out[10]: 0    3594
1      644
Name: TenYearCHD, dtype: int64
```

```
In [11]: x=d.drop('TenYearCHD',axis=1)
y=d['TenYearCHD']
```

```
In [12]: 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 [13]: from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
```

```
Out[13]: RandomForestClassifier()
```

```
In [14]: params = {'max_depth':[1,2,3,4,5],
                  'min_samples_leaf':[5,10,15,20,25],
                  'n_estimators':[10,20,30,40,50]}
```

```
In [15]: from sklearn.model_selection import GridSearchCV
grid_search= GridSearchCV(estimator = rfc,param_grid=params,cv=2,scoring="accuracy")
grid_search.fit(x_train,y_train)
```

```
Out[15]: 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')
```

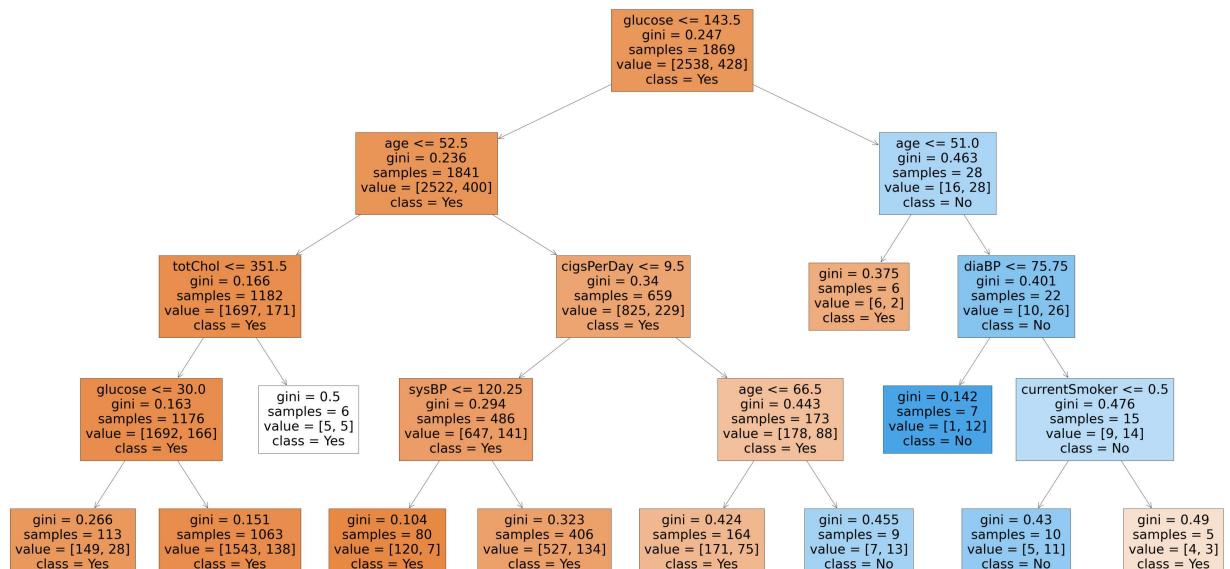
```
In [16]: grid_search.best_score_
```

```
Out[16]: 0.8479433580579905
```

```
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','No'],filled=True)
```

```
Out[18]: [Text(2441.25, 1956.96, 'glucose <= 143.5\ngini = 0.247\nsamples = 1869\nvalue = [2538, 428]\n\nclass = Yes'),
Text(1534.5, 1522.0800000000002, 'age <= 52.5\ngini = 0.236\nsamples = 1841\nvalue = [2522, 400]\n\nclass = Yes'),
Text(837.0, 1087.2, 'totChol <= 351.5\ngini = 0.166\nsamples = 1182\nvalue = [1697, 171]\n\nclass = Yes'),
Text(558.0, 652.3200000000002, 'glucose <= 30.0\ngini = 0.163\nsamples = 1176\nvalue = [1692, 166]\n\nclass = Yes'),
Text(279.0, 217.44000000000005, 'gini = 0.266\nsamples = 113\nvalue = [149, 28]\n\nclass = Yes'),
Text(837.0, 217.44000000000005, 'gini = 0.151\nsamples = 1063\nvalue = [1543, 138]\n\nclass = Yes'),
Text(1116.0, 652.3200000000002, 'gini = 0.5\nsamples = 6\nvalue = [5, 5]\n\nclass = Yes'),
Text(2232.0, 1087.2, 'cigsPerDay <= 9.5\ngini = 0.34\nsamples = 659\nvalue = [825, 229]\n\nclass = Yes'),
Text(1674.0, 652.3200000000002, 'sysBP <= 120.25\ngini = 0.294\nsamples = 486\nvalue = [647, 141]\n\nclass = Yes'),
Text(1395.0, 217.44000000000005, 'gini = 0.104\nsamples = 80\nvalue = [120, 7]\n\nclass = Yes'),
Text(1953.0, 217.44000000000005, 'gini = 0.323\nsamples = 406\nvalue = [527, 134]\n\nclass = Yes'),
Text(2790.0, 652.3200000000002, 'age <= 66.5\ngini = 0.443\nsamples = 173\nvalue = [178, 88]\n\nclass = Yes'),
Text(2511.0, 217.44000000000005, 'gini = 0.424\nsamples = 164\nvalue = [171, 75]\n\nclass = Yes'),
Text(3069.0, 217.44000000000005, 'gini = 0.455\nsamples = 9\nvalue = [7, 13]\n\nclass = No'),
Text(3348.0, 1522.0800000000002, 'age <= 51.0\ngini = 0.463\nsamples = 28\nvalue = [16, 28]\n\nclass = No'),
Text(3069.0, 1087.2, 'gini = 0.375\nsamples = 6\nvalue = [6, 2]\n\nclass = Yes'),
Text(3627.0, 1087.2, 'diaBP <= 75.75\ngini = 0.401\nsamples = 22\nvalue = [10, 26]\n\nclass = No'),
Text(3348.0, 652.3200000000002, 'gini = 0.142\nsamples = 7\nvalue = [1, 12]\n\nclass = No'),
Text(3906.0, 652.3200000000002, 'currentSmoker <= 0.5\ngini = 0.476\nsamples = 15\nvalue = [9, 14]\n\nclass = No'),
Text(3627.0, 217.44000000000005, 'gini = 0.43\nsamples = 10\nvalue = [5, 11]\n\nclass = No'),
Text(4185.0, 217.44000000000005, 'gini = 0.49\nsamples = 5\nvalue = [4, 3]\n\nclass = Yes')]
```



In [ ]: