In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [5]:

```
df=pd.read_csv(r"C:\Users\user\Downloads\C5_health care diabetes.csv")
df
```

Out[5]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunctio
0	6	148	72	35	0	33.6	0.62
1	1	85	66	29	0	26.6	0.35
2	8	183	64	0	0	23.3	0.67
3	1	89	66	23	94	28.1	0.16
4	0	137	40	35	168	43.1	2.28
...	...	...	...	...	...	...	.
763	10	101	76	48	180	32.9	0.17
764	2	122	70	27	0	36.8	0.34
765	5	121	72	23	112	26.2	0.24
766	1	126	60	0	0	30.1	0.34
767	1	93	70	31	0	30.4	0.31

768 rows × 9 columns

In [6]:

```
d=df.fillna(20)
```

In [9]:

```
d['Outcome'].value_counts()
```

Out[9]:

```
0    500
1    268
Name: Outcome, dtype: int64
```

In [10]:

```
x=d.drop('Outcome',axis=1)
y=d['Outcome']
```

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]: paramets = {'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=paramets,cv=2,scoring="acc
         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')
```

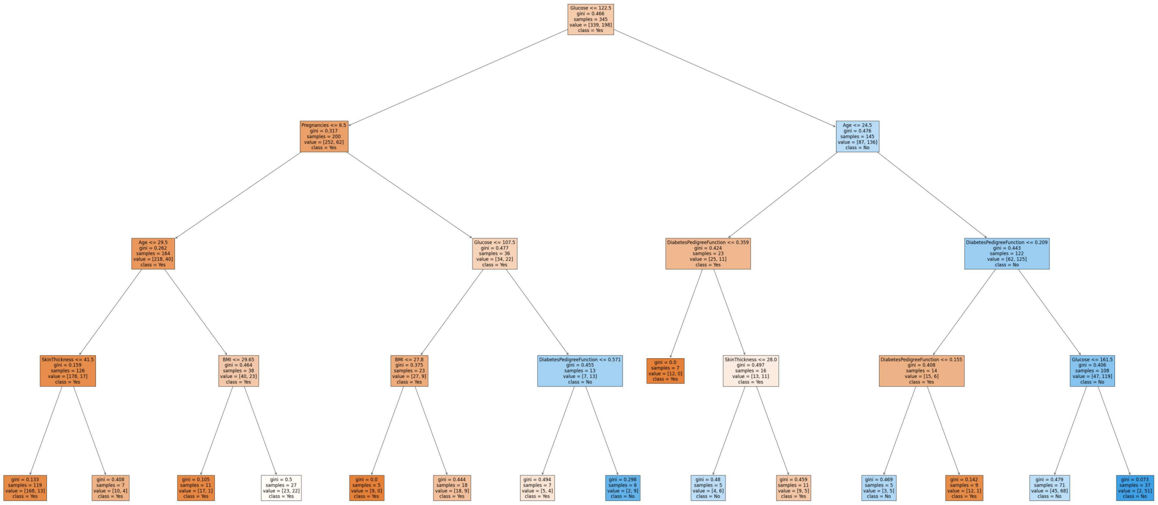
```
In [15]: grid_search.best_score_
```

```
Out[15]: 0.7728180658048049
```

```
In [16]: rfc_best=grid_search.best_estimator_
```

In [17]:

```
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[17]: [Text(2271.8571428571427, 1956.96, 'Glucose <= 122.5\ngini = 0.466\nsamples =
         345\nvalue = [339, 198]\nclass = Yes'),
          Text(1275.4285714285713, 1522.0800000000002, 'Pregnancies <= 6.5\ngini = 0.3
         17\nsamples = 200\nvalue = [252, 62]\nclass = Yes'),
          Text(637.7142857142857, 1087.2, 'Age <= 29.5\ngini = 0.262\nsamples = 164\nv
         alue = [218, 40]\nclass = Yes'),
          Text(318.85714285714283, 652.3200000000002, 'SkinThickness <= 41.5\ngini =
         0.159\nsamples = 126\nvalue = [178, 17]\nclass = Yes'),
          Text(159.42857142857142, 217.44000000000005, 'gini = 0.133\nsamples = 119\nv
         alue = [168, 13]\nclass = Yes'),
          Text(478.2857142857142, 217.44000000000005, 'gini = 0.408\nsamples = 7\nvalu
         e = [10, 4]\nclass = Yes'),
          Text(956.5714285714284, 652.3200000000002, 'BMI <= 29.65\ngini = 0.464\nsamp
         les = 38\nvalue = [40, 23]\nclass = Yes'),
          Text(797.1428571428571, 217.44000000000005, 'gini = 0.105\nsamples = 11\nval
         ue = [17, 1]\nclass = Yes'),
          Text(1116.0, 217.44000000000005, 'gini = 0.5\nsamples = 27\nvalue = [23, 22]
         \nclass = Yes'),
          Text(1913.1428571428569, 1087.2, 'Glucose <= 107.5\ngini = 0.477\nsamples =
         36\nvalue = [34, 22]\nclass = Yes'),
          Text(1594.2857142857142, 652.3200000000002, 'BMI <= 27.8\ngini = 0.375\nsamp
         les = 23\nvalue = [27, 9]\nclass = Yes'),
          Text(1434.8571428571427, 217.44000000000005, 'gini = 0.0\nsamples = 5\nvalue
         = [9, 0]\nclass = Yes'),
          Text(1753.7142857142856, 217.44000000000005, 'gini = 0.444\nsamples = 18\nva
         lue = [18, 9]\nclass = Yes'),
          Text(2232.0, 652.3200000000002, 'DiabetesPedigreeFunction <= 0.571\ngini =
         0.455\nsamples = 13\nvalue = [7, 13]\nclass = No'),
          Text(2072.5714285714284, 217.44000000000005, 'gini = 0.494\nsamples = 7\nval
         ue = [5, 4]\nclass = Yes'),
          Text(2391.428571428571, 217.44000000000005, 'gini = 0.298\nsamples = 6\nvalu
         e = [2, 9]\nclass = No'),
          Text(3268.285714285714, 1522.0800000000002, 'Age <= 24.5\ngini = 0.476\nsamp
         les = 145\nvalue = [87, 136]\nclass = No'),
          Text(2710.285714285714, 1087.2, 'DiabetesPedigreeFunction <= 0.359\ngini =
         0.424\nsamples = 23\nvalue = [25, 11]\nclass = Yes'),
          Text(2550.8571428571427, 652.3200000000002, 'gini = 0.0\nsamples = 7\nvalue
         = [12, 0]\nclass = Yes'),
          Text(2869.7142857142853, 652.3200000000002, 'SkinThickness <= 28.0\ngini =
         0.497\nsamples = 16\nvalue = [13, 11]\nclass = Yes'),
          Text(2710.285714285714, 217.44000000000005, 'gini = 0.48\nsamples = 5\nvalue
         = [4, 6]\nclass = No'),
          Text(3029.142857142857, 217.44000000000005, 'gini = 0.459\nsamples = 11\nval
         ue = [9, 5]\nclass = Yes'),
          Text(3826.2857142857138, 1087.2, 'DiabetesPedigreeFunction <= 0.209\ngini =
         0.443\nsamples = 122\nvalue = [62, 125]\nclass = No'),
          Text(3507.428571428571, 652.3200000000002, 'DiabetesPedigreeFunction <= 0.15
         5\ngini = 0.408\nsamples = 14\nvalue = [15, 6]\nclass = Yes'),
          Text(3347.9999999999995, 217.44000000000005, 'gini = 0.469\nsamples = 5\nval
         ue = [3, 5]\nclass = No'),
          Text(3666.8571428571427, 217.44000000000005, 'gini = 0.142\nsamples = 9\nval
         ue = [12, 1]\nclass = Yes'),
          Text(4145.142857142857, 652.3200000000002, 'Glucose <= 161.5\ngini = 0.406\n
         samples = 108\nvalue = [47, 119]\nclass = No'),
          Text(3985.7142857142853, 217.44000000000005, 'gini = 0.479\nsamples = 71\nva
         lue = [45, 68]\nclass = No'),
 Text(4304.571428571428, 217.44000000000005, 'gini = 0.073\nsamples = 37\nval
ue = [2, 51]\nclass = No')]
```


Glucose <= 122.5 gini = 0.466 samples = 345 value = [339, 198] class = Yes
Pregnancies <= 6.5 gini = 0.317 samples = 200 value = [252, 62] class = Yes
Age <= 24.5 gini = 0.476 samples = 145 value = [87, 136] class = No
Age <= 29.5 gini = 0.262 samples = 164 value = [218, 40] class = Yes
Glucose <= 107.5 gini = 0.477 samples = 36 value = [34, 22] class = Yes
DiabetesPedigreeFunction <= 0.359 gini = 0.424 samples = 23 value = [25, 11] class = Yes
DiabetesPedigreeFunction <= 0.209 gini = 0.443 samples = 122 value = [62, 125] class = No
SkinThickness <= 41.5 gini = 0.159 samples = 126 value = [178, 17] class = Yes
BMI <= 29.65 gini = 0.464 samples = 38 value = [40, 23] class = Yes
BMI <= 27.8 gini = 0.375 samples = 23 value = [27, 9] class = Yes
DiabetesPedigreeFunction <= 0.571 gini = 0.455 samples = 13 value = [7, 13] class = No
gini = 0.0 samples = 7 value = [12, 0] class = Yes
SkinThickness <= 28.0 gini = 0.497 samples = 16 value = [13, 11] class = Yes
DiabetesPedigreeFunction <= 0.155 gini = 0.408 samples = 14 value = [15, 6] class = Yes
Glucose <= 161.5 gini = 0.406 samples = 108 value = [47, 119] class = No
gini = 0.133 samples = 119 value = [168, 13] class = Yes
gini = 0.408 samples = 7 value = [10, 4] class = Yes
gini = 0.105 samples = 11 value = [17, 1] class = Yes
gini = 0.5 samples = 27 value = [23, 22] class = Yes
gini = 0.0 samples = 5 value = [9, 0] class = Yes
gini = 0.444 samples = 18 value = [18, 9] class = Yes
gini = 0.494 samples = 7 value = [5, 4] class = Yes
gini = 0.298 samples = 6 value = [2, 9] class = No
gini = 0.48 samples = 5 value = [4, 6] class = No
gini = 0.459 samples = 11 value = [9, 5] class = Yes
gini = 0.469 samples = 5 value = [3, 5] class = No
gini = 0.142 samples = 9 value = [12, 1] class = Yes
gini = 0.479 samples = 71 value = [45, 68] class = No
gini = 0.073 samples = 37 value = [2, 51] class = No

In [ ]: