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"C5_health care diabetes.csv")
df
```

Out[2]:

	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 [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

In [4]:
```
df=df.dropna()
```

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
Pregnancies                 0
Glucose                     0
BloodPressure               0
SkinThickness               0
Insulin                     0
BMI                         0
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64
```

In [6]:
```
df.describe()
```

Out[6]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	Diabetes
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

In [8]:
```
df['Outcome'].value_counts()
```

Out[8]:
```
0    500
1    268
Name: Outcome, dtype: int64
```

In [9]:
```
df1=df[['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome']]
```

In [10]:
```
x=df1.drop('Outcome',axis=1)
y=df1['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]:
```
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]:
```
0.7765840869999445
```

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(2368.4, 1993.2, 'Pregnancies <= 4.5\ngini = 0.448\nsamples = 352\nvalue
= [355, 182]\nclass = Yes'),
 Text(1339.2, 1630.8000000000002, 'Age <= 30.5\ngini = 0.35\nsamples = 216\nv
alue = [263, 77]\nclass = Yes'),
 Text(793.6, 1268.4, 'Insulin <= 186.5\ngini = 0.27\nsamples = 172\nvalue =
[229, 44]\nclass = Yes'),
 Text(396.8, 906.0, 'Glucose <= 130.5\ngini = 0.24\nsamples = 150\nvalue = [2
04, 33]\nclass = Yes'),
 Text(198.4, 543.5999999999999, 'Pregnancies <= 2.5\ngini = 0.07\nsamples = 1
22\nvalue = [185, 7]\nclass = Yes'),
 Text(99.2, 181.19999999999982, 'gini = 0.041\nsamples = 90\nvalue = [139, 3]
\nclass = Yes'),
 Text(297.6, 181.19999999999982, 'gini = 0.147\nsamples = 32\nvalue = [46, 4]
\nclass = Yes'),
 Text(595.2, 543.5999999999999, 'Pregnancies <= 0.5\ngini = 0.488\nsamples =
28\nvalue = [19, 26]\nclass = No'),
 Text(496.0, 181.19999999999982, 'gini = 0.34\nsamples = 15\nvalue = [5, 18]
\nclass = No'),
 Text(694.4, 181.19999999999982, 'gini = 0.463\nsamples = 13\nvalue = [14, 8]
\nclass = Yes'),
 Text(1190.4, 906.0, 'DiabetesPedigreeFunction <= 0.488\ngini = 0.424\nsample
s = 22\nvalue = [25, 11]\nclass = Yes'),
 Text(992.0, 543.5999999999999, 'Insulin <= 277.0\ngini = 0.1\nsamples = 11\n
value = [18, 1]\nclass = Yes'),
 Text(892.8000000000001, 181.19999999999982, 'gini = 0.0\nsamples = 6\nvalue
= [13, 0]\nclass = Yes'),
 Text(1091.2, 181.19999999999982, 'gini = 0.278\nsamples = 5\nvalue = [5, 1]
\nclass = Yes'),
 Text(1388.8, 543.5999999999999, 'Age <= 23.5\ngini = 0.484\nsamples = 11\nva
lue = [7, 10]\nclass = No'),
 Text(1289.6000000000001, 181.19999999999982, 'gini = 0.346\nsamples = 5\nval
ue = [7, 2]\nclass = Yes'),
 Text(1488.0, 181.19999999999982, 'gini = 0.0\nsamples = 6\nvalue = [0, 8]\nc
lass = No'),
 Text(1884.8, 1268.4, 'Glucose <= 114.5\ngini = 0.5\nsamples = 44\nvalue = [3
4, 33]\nclass = Yes'),
 Text(1686.4, 906.0, 'Age <= 38.0\ngini = 0.124\nsamples = 11\nvalue = [14,
1]\nclass = Yes'),
 Text(1587.2, 543.5999999999999, 'gini = 0.0\nsamples = 6\nvalue = [9, 0]\ncl
ass = Yes'),
 Text(1785.6000000000001, 543.5999999999999, 'gini = 0.278\nsamples = 5\nvalu
e = [5, 1]\nclass = Yes'),
 Text(2083.2000000000003, 906.0, 'Age <= 33.5\ngini = 0.473\nsamples = 33\nva
lue = [20, 32]\nclass = No'),
 Text(1984.0, 543.5999999999999, 'gini = 0.0\nsamples = 7\nvalue = [0, 9]\ncl
ass = No'),
 Text(2182.4, 543.5999999999999, 'Glucose <= 154.5\ngini = 0.498\nsamples = 2
6\nvalue = [20, 23]\nclass = No'),
 Text(2083.2000000000003, 181.19999999999982, 'gini = 0.473\nsamples = 17\nva
lue = [16, 10]\nclass = Yes'),
 Text(2281.6, 181.19999999999982, 'gini = 0.36\nsamples = 9\nvalue = [4, 13]
\nclass = No'),
 Text(3397.6, 1630.8000000000002, 'Glucose <= 128.5\ngini = 0.498\nsamples =
136\nvalue = [92, 105]\nclass = No'),
 Text(2876.8, 1268.4, 'SkinThickness <= 37.5\ngini = 0.427\nsamples = 72\nval
ue = [74, 33]\nclass = Yes'),
 Text(2678.4, 906.0, 'BMI <= 39.05\ngini = 0.447\nsamples = 59\nvalue = [61,
```

```
31]\nclass = Yes'),
 Text(2579.2000000000003, 543.5999999999999, 'Insulin <= 24.5\ngini = 0.415\n
samples = 54\nvalue = [60, 25]\nclass = Yes'),
 Text(2480.0, 181.19999999999982, 'gini = 0.438\nsamples = 43\nvalue = [46, 2
2]\nclass = Yes'),
 Text(2678.4, 181.19999999999982, 'gini = 0.291\nsamples = 11\nvalue = [14,
3]\nclass = Yes'),
 Text(2777.6, 543.5999999999999, 'gini = 0.245\nsamples = 5\nvalue = [1, 6]\n
class = No'),
 Text(3075.2000000000003, 906.0, 'DiabetesPedigreeFunction <= 0.321\ngini =
0.231\nsamples = 13\nvalue = [13, 2]\nclass = Yes'),
 Text(2976.0, 543.5999999999999, 'gini = 0.0\nsamples = 8\nvalue = [10, 0]\nc
lass = Yes'),
 Text(3174.4, 543.5999999999999, 'gini = 0.48\nsamples = 5\nvalue = [3, 2]\nc
lass = Yes'),
 Text(3918.4, 1268.4, 'BloodPressure <= 75.5\ngini = 0.32\nsamples = 64\nvalu
e = [18, 72]\nclass = No'),
 Text(3571.2000000000003, 906.0, 'Glucose <= 166.5\ngini = 0.402\nsamples = 3
1\nvalue = [12, 31]\nclass = No'),
 Text(3372.8, 543.5999999999999, 'Insulin <= 132.5\ngini = 0.493\nsamples = 1
9\nvalue = [11, 14]\nclass = No'),
 Text(3273.6, 181.19999999999982, 'gini = 0.415\nsamples = 13\nvalue = [5, 1
2]\nclass = No'),
 Text(3472.0, 181.19999999999982, 'gini = 0.375\nsamples = 6\nvalue = [6, 2]
\nclass = Yes'),
 Text(3769.6, 543.5999999999999, 'BMI <= 33.25\ngini = 0.105\nsamples = 12\nv
alue = [1, 17]\nclass = No'),
 Text(3670.4, 181.19999999999982, 'gini = 0.0\nsamples = 7\nvalue = [0, 12]\n
class = No'),
 Text(3868.8, 181.19999999999982, 'gini = 0.278\nsamples = 5\nvalue = [1, 5]
\nclass = No'),
 Text(4265.6, 906.0, 'Insulin <= 157.5\ngini = 0.223\nsamples = 33\nvalue =
[6, 41]\nclass = No'),
 Text(4166.400000000001, 543.5999999999999, 'Age <= 47.5\ngini = 0.284\nsampl
es = 24\nvalue = [6, 29]\nclass = No'),
 Text(4067.2000000000003, 181.19999999999982, 'gini = 0.351\nsamples = 17\nva
lue = [5, 17]\nclass = No'),
 Text(4265.6, 181.19999999999982, 'gini = 0.142\nsamples = 7\nvalue = [1, 12]
\nclass = No'),
 Text(4364.8, 543.5999999999999, 'gini = 0.0\nsamples = 9\nvalue = [0, 12]\nc
lass = No')]
```