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

In [2]:
```
```

In [3]:
```
df = pd.read_csv("C5_health care diabetes.csv")
```

Out[3]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunctio
0	6	148	72	35	0	33.6	0.6
1	1	85	66	29	0	26.6	0.3
2	8	183	64	0	0	23.3	0.6
3	1	89	66	23	94	28.1	0.1
4	0	137	40	35	168	43.1	2.2
...	...	...	...	...	...	...	
763	10	101	76	48	180	32.9	0.1
764	2	122	70	27	0	36.8	0.3
765	5	121	72	23	112	26.2	0.2
766	1	126	60	0	0	30.1	0.3
767	1	93	70	31	0	30.4	0.3

768 rows × 9 columns

In [4]:
```
```

```
<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 [5]:

Out[5]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

In [6]:
```
f_m=df[['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age']]
```

In [7]:

Out[7]:
```
(768, 8)
```

In [8]:

Out[8]:
```
(768,)
```

In [9]:

In [10]:

In [11]:
```
logr=LogisticRegression()
```

Out[11]:
```
LogisticRegression()
```

In [12]:

In [13]:
```
prediction=logr.predict(observation)
```

Out[13]:
```
array([1], dtype=int64)
```

In [14]:

Out[14]:
```
array([0, 1], dtype=int64)
```

In [15]:

Out[15]:
```
0.00029236948687560993
```

In [16]:

Out[16]:
```
0.9997076305131244
```

RANDOM VARIABLE

In [17]:

Out[17]:
```
0    500
1    268
Name: Outcome, dtype: int64
```

In [18]:
```
x=df[['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age']]
```

In [19]:
```
#g1={"Verified":{'True':1,'False':2}}
#df=df.replace(g1)
```

In [20]:
```
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 [21]:
```
rfc=RandomForestClassifier()
```

Out[21]:
```
RandomForestClassifier()
```

In [22]:
```
parameters={'max_depth':[1,2,3,4,5],
            'min_samples_leaf':[5,10,15,20,25],
```

In [23]:
```
from sklearn.model_selection import GridSearchCV
grid_search =GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="ac
```

Out[23]:
```
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 [24]:

Out[24]:
```
0.7728527437163624
```

In [25]:

In [26]:
```
from sklearn.tree import plot_tree
plt.figure(figsize=(80,40))
plot_tree(rfc_best.estimators_[5],feature_names=x.columns,class_names=['Yes','
```

Out[26]:
```
[Text(2232.0, 1956.96, 'BMI <= 29.85\ngini = 0.456\nsamples = 332\nvalue = [3
48, 189]\nclass = Yes'),
 Text(892.8, 1522.0800000000002, 'SkinThickness <= 8.5\ngini = 0.24\nsamples
= 125\nvalue = [179, 29]\nclass = Yes'),
 Text(446.4, 1087.2, 'gini = 0.35\nsamples = 40\nvalue = [48, 14]\nclass = Ye
s'),
 Text(1339.1999999999998, 1087.2, 'Age <= 28.5\ngini = 0.184\nsamples = 85\nv
alue = [131, 15]\nclass = Yes'),
 Text(892.8, 652.3200000000002, 'gini = 0.0\nsamples = 55\nvalue = [102, 0]\n
class = Yes'),
 Text(1785.6, 652.3200000000002, 'gini = 0.449\nsamples = 30\nvalue = [29, 1
5]\nclass = Yes'),
 Text(3571.2, 1522.0800000000002, 'BMI <= 40.8\ngini = 0.5\nsamples = 207\nva
lue = [169, 160]\nclass = Yes'),
 Text(3124.7999999999997, 1087.2, 'BloodPressure <= 84.5\ngini = 0.494\nsampl
es = 172\nvalue = [153, 123]\nclass = Yes'),
 Text(2678.3999999999996, 652.3200000000002, 'Insulin <= 9.0\ngini = 0.481\ns
amples = 147\nvalue = [143, 96]\nclass = Yes'),
 Text(2232.0, 217.44000000000005, 'gini = 0.496\nsamples = 70\nvalue = [52, 6
2]\nclass = No'),
 Text(3124.7999999999997, 217.44000000000005, 'gini = 0.396\nsamples = 77\nva
lue = [91, 34]\nclass = Yes'),
 Text(3571.2, 652.3200000000002, 'gini = 0.394\nsamples = 25\nvalue = [10, 2
7]\nclass = No'),
 Text(4017.6, 1087.2, 'gini = 0.422\nsamples = 35\nvalue = [16, 37]\nclass =
No')]
```