In [81]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [152]:
```
df=pd.read_csv(r"C:\USERS\user\Downloads\C5_health care diabetes - C5_health c
```

Out[152]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunctic
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 [153]:
```

```

Out[153]:
```
Index(['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'],
      dtype='object')
```

In [154]:
```
df=df.head(21)
```

Out[154]:

	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
5	5	116	74	0	0	25.6	0.20
6	3	78	50	32	88	31.0	0.24
7	10	115	0	0	0	35.3	0.13
8	2	197	70	45	543	30.5	0.15
9	8	125	96	0	0	0.0	0.23
10	4	110	92	0	0	37.6	0.19
11	10	168	74	0	0	38.0	0.53
12	10	139	80	0	0	27.1	1.44
13	1	189	60	23	846	30.1	0.39
14	5	166	72	19	175	25.8	0.58
15	7	100	0	0	0	30.0	0.48
16	0	118	84	47	230	45.8	0.55
17	7	107	74	0	0	29.6	0.25
18	1	103	30	38	83	43.3	0.18
19	1	115	70	30	96	34.6	0.52
20	3	126	88	41	235	39.3	0.70

In [156]:
```
a=df[['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin',
       'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome']]
```

Out[156]:

	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
5	5	116	74	0	0	25.6	0.20
6	3	78	50	32	88	31.0	0.24
7	10	115	0	0	0	35.3	0.13
8	2	197	70	45	543	30.5	0.15
9	8	125	96	0	0	0.0	0.23
10	4	110	92	0	0	37.6	0.19
11	10	168	74	0	0	38.0	0.53
12	10	139	80	0	0	27.1	1.44
13	1	189	60	23	846	30.1	0.39
14	5	166	72	19	175	25.8	0.58
15	7	100	0	0	0	30.0	0.48
16	0	118	84	47	230	45.8	0.55
17	7	107	74	0	0	29.6	0.25
18	1	103	30	38	83	43.3	0.18
19	1	115	70	30	96	34.6	0.52
20	3	126	88	41	235	39.3	0.70

In [157]:

```
Out[157]: 32    3
          31    3
          33    2
          30    2
          34    1
          54    1
          50    1
          51    1
          27    1
          21    1
          53    1
          57    1
          26    1
          59    1
          29    1
          Name: Age, dtype: int64
```

In [159]:

```
x=a.drop('Age',axis=1)
```

In [160]:
```
g1={"Age":{'29':1}}
a=a.replace(g1)
```

```
    Pregnancies  Glucose  BloodPressure  SkinThickness  Insulin   BMI  \
0             6      148             72             35        0  33.6
1             1       85             66             29        0  26.6
2             8      183             64              0        0  23.3
3             1       89             66             23       94  28.1
4             0      137             40             35      168  43.1
5             5      116             74              0        0  25.6
6             3       78             50             32       88  31.0
7            10      115              0              0        0  35.3
8             2      197             70             45      543  30.5
9             8      125             96              0        0   0.0
10            4      110             92              0        0  37.6
11           10      168             74              0        0  38.0
12           10      139             80              0        0  27.1
13            1      189             60             23      846  30.1
14            5      166             72             19      175  25.8
15            7      100              0              0        0  30.0
16            0      118             84             47      230  45.8
17            7      107             74              0        0  29.6
18            1      103             30             38       83  43.3
19            1      115             70             30       96  34.6
20            3      126             88             41      235  39.3

    DiabetesPedigreeFunction  Age  Outcome
0                      0.627   50        1
1                      0.351   31        0
2                      0.672   32        1
3                      0.167   21        0
4                      2.288   33        1
5                      0.201   30        0
6                      0.248   26        1
7                      0.134   29        0
8                      0.158   53        1
9                      0.232   54        1
10                     0.191   30        0
11                     0.537   34        1
12                     1.441   57        0
13                     0.398   59        1
14                     0.587   51        1
15                     0.484   32        1
16                     0.551   31        1
17                     0.254   31        1
18                     0.183   33        0
19                     0.529   32        1
20                     0.704   27        0
```

In [161]:
```
from sklearn.model_selection import train_test_split
```

In [162]:
```
from sklearn.ensemble import RandomForestClassifier

rfc=RandomForestClassifier()
```

Out[162]:
```
RandomForestClassifier()
```

In [179]:
```
parameters={'max_depth':[1,2],
            'min_samples_leaf':[5,10],
```

In [180]:
```
from sklearn.model_selection import GridSearchCV

grid_search=GridSearchCV(estimator=rfc,param_grid=parameters,cv=2,scoring="acc
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\model_selection\_split.py:
666: UserWarning: The least populated class in y has only 1 members, which is
less than n_splits=2.
  warnings.warn(("The least populated class in y has only %d"
```

Out[180]:
```
GridSearchCV(cv=2, estimator=RandomForestClassifier(),
             param_grid={'max_depth': [1, 2], 'min_samples_leaf': [5, 10],
                         'n_estimators': [10, 20]},
             scoring='accuracy')
```

In [181]:

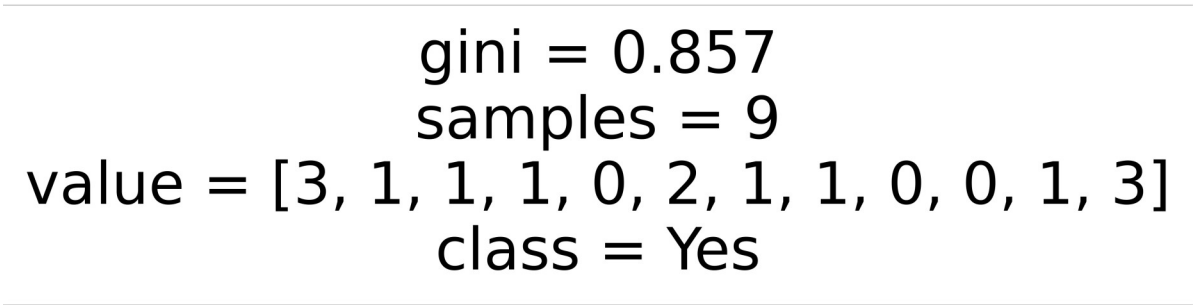
Out[181]:
```
0.14285714285714285
```

In [182]:

In [183]:
```
from sklearn.tree import plot_tree

plt.figure(figsize=(80,40))
```

Out[183]:
```
[Text(2232.0, 1087.2, 'gini = 0.857\nsamples = 9\nvalue = [3, 1, 1, 1, 0, 2,
1, 1, 0, 0, 1, 3]\nclass = Yes')]
```



In [ ]: