In [1]:
```
# Load libraries
import pandas as pd
from sklearn.tree import DecisionTreeClassifier # Import Decision Tree Classifier
from sklearn.model_selection import train_test_split # Import train_test_split funct
from sklearn import metrics #Import scikit-learn metrics module for accuracy calcula
```

In [5]:
```
import matplotlib.pyplot as plt
```

In [3]:
```
# Load dataset
pima = pd.read_csv("C:/Users/91920/Python/diabetes.csv",)
```

In [4]:
```
pima.head()
```

Out[4]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	6	148	72	35	0	33.6	0.627	50
1	1	85	66	29	0	26.6	0.351	31
2	8	183	64	0	0	23.3	0.672	32
3	1	89	66	23	94	28.1	0.167	21
4	0	137	40	35	168	43.1	2.288	33

In [6]:
```
print("dimension of diabetes data: {}".format(pima.shape))
```

```
dimension of diabetes data: (768, 9)
```

In [5]:
```
print(pima.groupby('Outcome').size())
```

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

In [8]:
```
X_train, X_test, y_train, y_test = train_test_split(pima.loc[:, pima.columns != 'Out
```

In [9]:
```
print(y_train.value_counts())
print(y_test.value_counts())
```

```
0    375
1    201
Name: Outcome, dtype: int64
0    125
1     67
Name: Outcome, dtype: int64
```

In [10]:
```
feature_name=list(X_train.columns)
class_name = list(y_train.unique())
feature_name
```

Out[10]:
```
['Pregnancies',
```

```
'Glucose',
'BloodPressure',
'SkinThickness',
'Insulin',
'BMI',
'DiabetesPedigreeFunction',
'Age']
```

In [11]:
```
class_name
```

Out[11]:
```
[0, 1]
```

In [13]:
```
# Create Decision Tree classifer object
clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(X_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(X_test)
```

In [14]:
```
# Model Accuracy, how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
```

```
Accuracy: 0.734375
```

In [18]:
```
X_test=[]
```

Out[18]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	A
635	13	104	72	0	0	31.2	0.465	
698	4	127	88	11	155	34.5	0.598	
637	2	94	76	18	66	31.6	0.649	
402	5	136	84	41	88	35.0	0.286	
425	4	184	78	39	277	37.0	0.264	
...	...	...	...	...	...	...	...	...
62	5	44	62	0	0	25.0	0.587	
477	7	114	76	17	110	23.8	0.466	
311	0	106	70	37	148	39.4	0.605	
116	5	124	74	0	0	34.0	0.220	
541	3	128	72	25	190	32.4	0.549	

192 rows × 8 columns

In [ ]:
```
```

In [17]:
```
from sklearn import tree
plt.figure(figsize=(70,30))
tree.plot_tree(clf,filled=True)
```

Out[17]:
```
[Text(2174.509359137056, 1572.5571428571427, 'X[1] <= 132.5\ngini = 0.454\nsamples = 576\nvalue = [375, 201]'),
 Text(1150.6094543147208, 1456.0714285714284, 'X[7] <= 30.5\ngini = 0.325\nsamples = 397\nvalue = [316, 81]'),
 Text(801.7709390862943, 1339.5857142857142, 'X[5] <= 45.4\ngini = 0.192\nsamples = 241\nvalue = [215, 26]'),
 Text(641.9124365482234, 1223.1, 'X[0] <= 7.5\ngini = 0.163\nsamples = 235\nvalue = [214, 21]'),
 Text(441.15989847715736, 1106.6142857142856, 'X[6] <= 0.509\ngini = 0.144\nsamples = 231\nvalue = [213, 18]'),
 Text(158.61928934010152, 990.1285714285714, 'X[2] <= 19.0\ngini = 0.049\nsamples = 158\nvalue = [154, 4]'),
 Text(79.30964467005076, 873.6428571428571, 'X[1] <= 116.5\ngini = 0.278\nsamples = 6\nvalue = [5, 1]'),
 Text(39.65482233502538, 757.1571428571428, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
 Text(118.96446700507613, 757.1571428571428, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(237.92893401015226, 873.6428571428571, 'X[4] <= 37.0\ngini = 0.039\nsamples = 152\nvalue = [149, 3]'),
 Text(198.2741116751269, 757.1571428571428, 'X[4] <= 34.0\ngini = 0.083\nsamples = 69\nvalue = [66, 3]'),
 Text(118.96446700507613, 640.6714285714286, 'X[1] <= 111.5\ngini = 0.058\nsamples = 67\nvalue = [65, 2]'),
 Text(79.30964467005076, 524.1857142857143, 'gini = 0.0\nsamples = 45\nvalue = [45, 0]'),
 Text(158.61928934010152, 524.1857142857143, 'X[7] <= 24.5\ngini = 0.165\nsamples = 22\nvalue = [20, 2]'),
 Text(118.96446700507613, 407.70000000000005, 'gini = 0.0\nsamples = 13\nvalue = [13, 0]'),
 Text(198.2741116751269, 407.70000000000005, 'X[1] <= 113.0\ngini = 0.346\nsamples = 9\nvalue = [7, 2]'),
 Text(158.61928934010152, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(237.92893401015226, 291.2142857142858, 'X[2] <= 65.0\ngini = 0.219\nsamples = 8\nvalue = [7, 1]'),
 Text(198.2741116751269, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(277.5837563451777, 174.7285714285715, 'gini = 0.0\nsamples = 7\nvalue = [7, 0]'),
 Text(277.5837563451777, 640.6714285714286, 'X[5] <= 32.6\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(237.92893401015226, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(317.23857868020303, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(277.5837563451777, 757.1571428571428, 'gini = 0.0\nsamples = 83\nvalue = [83, 0]'),
 Text(723.7005076142132, 990.1285714285714, 'X[5] <= 32.1\ngini = 0.31\nsamples = 73\nvalue = [59, 14]'),
 Text(555.1675126903554, 873.6428571428571, 'X[1] <= 127.5\ngini = 0.097\nsamples = 39\nvalue = [37, 2]'),
 Text(475.8578680203045, 757.1571428571428, 'X[5] <= 23.55\ngini = 0.054\nsamples = 36\nvalue = [35, 1]'),
 Text(436.20304568527916, 640.6714285714286, 'X[0] <= 2.5\ngini = 0.32\nsamples = 5\nvalue = [4, 1]'),
 Text(396.5482233502538, 524.1857142857143, 'gini = 0.0\nsamples = 4\nvalue = [4, 0]'),
 Text(475.8578680203045, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(515.5126903553299, 640.6714285714286, 'gini = 0.0\nsamples = 31\nvalue = [31, 0]'),
 Text(634.4771573604061, 757.1571428571428, 'X[1] <= 128.5\ngini = 0.444\nsamples = 3\nvalue = [2, 1]'),
 Text(594.8223350253807, 640.6714285714286, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(674.1319796954315, 640.6714285714286, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
```

```
 Text(892.233502538071, 873.6428571428571, 'X[1] <= 92.0\ngini = 0.457\nsamples = 34
\nvalue = [22, 12]'),
 Text(793.0964467005076, 757.1571428571428, 'X[6] <= 1.272\ngini = 0.198\nsamples =
9\nvalue = [8, 1]'),
 Text(753.4416243654822, 640.6714285714286, 'gini = 0.0\nsamples = 8\nvalue = [8,
0]'),
 Text(832.7512690355329, 640.6714285714286, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(991.3705583756345, 757.1571428571428, 'X[2] <= 68.0\ngini = 0.493\nsamples = 2
5\nvalue = [14, 11]'),
 Text(912.0609137055837, 640.6714285714286, 'X[5] <= 34.25\ngini = 0.48\nsamples = 1
5\nvalue = [6, 9]'),
 Text(872.4060913705583, 524.1857142857143, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(951.715736040609, 524.1857142857143, 'X[4] <= 114.5\ngini = 0.496\nsamples = 1
1\nvalue = [6, 5]'),
 Text(872.4060913705583, 407.70000000000005, 'X[3] <= 31.0\ngini = 0.32\nsamples = 5
\nvalue = [4, 1]'),
 Text(832.7512690355329, 291.2142857142858, 'X[1] <= 113.0\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(793.0964467005076, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(872.4060913705583, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(912.0609137055837, 291.2142857142858, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(1031.0253807106599, 407.70000000000005, 'X[7] <= 25.5\ngini = 0.444\nsamples =
6\nvalue = [2, 4]'),
 Text(991.3705583756345, 291.2142857142858, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(1070.6802030456852, 291.2142857142858, 'X[0] <= 1.5\ngini = 0.444\nsamples = 3
\nvalue = [2, 1]'),
 Text(1031.0253807106599, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1110.3350253807107, 174.7285714285715, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(1070.6802030456852, 640.6714285714286, 'X[3] <= 13.0\ngini = 0.32\nsamples = 1
0\nvalue = [8, 2]'),
 Text(1031.0253807106599, 524.1857142857143, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(1110.3350253807107, 524.1857142857143, 'gini = 0.0\nsamples = 8\nvalue = [8,
0]'),
 Text(842.6649746192893, 1106.6142857142856, 'X[5] <= 32.15\ngini = 0.375\nsamples =
4\nvalue = [1, 3]'),
 Text(803.0101522842639, 990.1285714285714, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(882.3197969543147, 990.1285714285714, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(961.6294416243654, 1223.1, 'X[6] <= 0.844\ngini = 0.278\nsamples = 6\nvalue =
[1, 5]'),
 Text(921.97461928934, 1106.6142857142856, 'gini = 0.0\nsamples = 5\nvalue = [0,
5]'),
 Text(1001.2842639593908, 1106.6142857142856, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1499.447969543147, 1339.5857142857142, 'X[5] <= 26.9\ngini = 0.457\nsamples =
156\nvalue = [101, 55]'),
 Text(1259.0406091370558, 1223.1, 'X[5] <= 9.8\ngini = 0.064\nsamples = 30\nvalue =
[29, 1]'),
 Text(1219.3857868020305, 1106.6142857142856, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1298.6954314720813, 1106.6142857142856, 'gini = 0.0\nsamples = 29\nvalue = [2
9, 0]'),
 Text(1739.8553299492385, 1223.1, 'X[1] <= 107.5\ngini = 0.49\nsamples = 126\nvalue
= [72, 54]'),
 Text(1378.0050761421319, 1106.6142857142856, 'X[6] <= 0.636\ngini = 0.414\nsamples
= 65\nvalue = [46, 19]'),
 Text(1189.6446700507613, 990.1285714285714, 'X[7] <= 31.5\ngini = 0.31\nsamples = 4
7\nvalue = [38, 9]'),
 Text(1110.3350253807107, 873.6428571428571, 'X[2] <= 32.0\ngini = 0.444\nsamples =
```

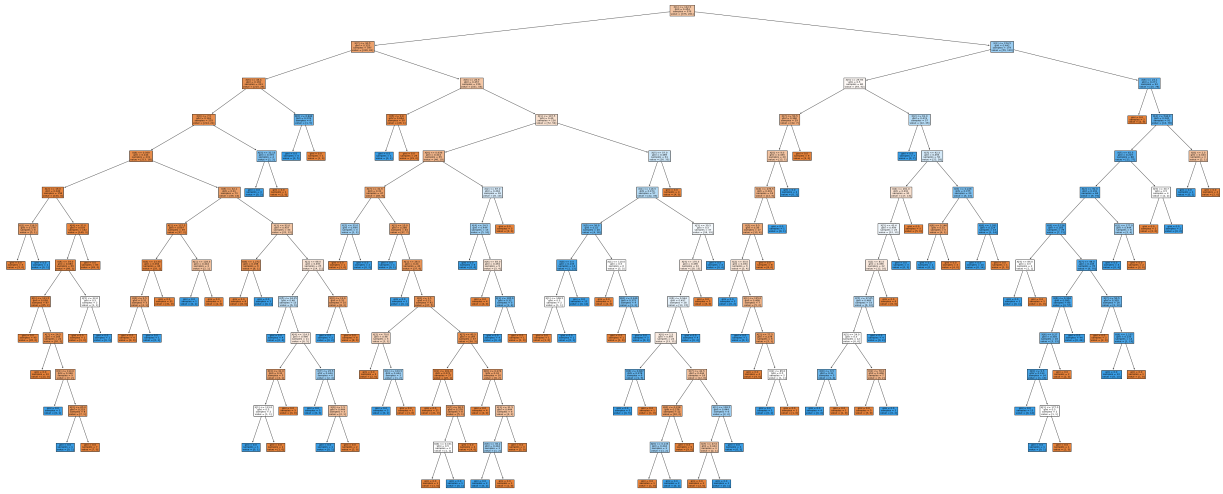
```
3\nvalue = [1, 2]'),
 Text(1070.6802030456852, 757.1571428571428, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1149.989847715736, 757.1571428571428, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(1268.9543147208121, 873.6428571428571, 'X[2] <= 15.0\ngini = 0.268\nsamples =
44\nvalue = [37, 7]'),
 Text(1229.2994923857868, 757.1571428571428, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1308.6091370558374, 757.1571428571428, 'X[1] <= 28.5\ngini = 0.24\nsamples = 4
3\nvalue = [37, 6]'),
 Text(1268.9543147208121, 640.6714285714286, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1348.263959390863, 640.6714285714286, 'X[0] <= 1.5\ngini = 0.21\nsamples = 42
\nvalue = [37, 5]'),
 Text(1209.472081218274, 524.1857142857143, 'X[2] <= 70.0\ngini = 0.48\nsamples = 5
\nvalue = [3, 2]'),
 Text(1169.8172588832488, 407.70000000000005, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(1249.1269035532994, 407.70000000000005, 'X[1] <= 103.5\ngini = 0.444\nsamples
= 3\nvalue = [1, 2]'),
 Text(1209.472081218274, 291.2142857142858, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(1288.781725888325, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1487.0558375634516, 524.1857142857143, 'X[7] <= 45.5\ngini = 0.149\nsamples =
37\nvalue = [34, 3]'),
 Text(1407.746192893401, 407.70000000000005, 'X[1] <= 101.0\ngini = 0.071\nsamples =
27\nvalue = [26, 1]'),
 Text(1368.0913705583755, 291.2142857142858, 'gini = 0.0\nsamples = 21\nvalue = [21,
0]'),
 Text(1447.4010152284263, 291.2142857142858, 'X[7] <= 36.5\ngini = 0.278\nsamples =
6\nvalue = [5, 1]'),
 Text(1407.746192893401, 174.7285714285715, 'X[6] <= 0.151\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(1368.0913705583755, 58.24285714285725, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1447.4010152284263, 58.24285714285725, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1487.0558375634516, 174.7285714285715, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(1566.3654822335025, 407.70000000000005, 'X[6] <= 0.274\ngini = 0.32\nsamples =
10\nvalue = [8, 2]'),
 Text(1526.7106598984772, 291.2142857142858, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(1606.0203045685278, 291.2142857142858, 'X[3] <= 31.5\ngini = 0.444\nsamples =
6\nvalue = [4, 2]'),
 Text(1566.3654822335025, 174.7285714285715, 'X[5] <= 36.2\ngini = 0.444\nsamples =
3\nvalue = [1, 2]'),
 Text(1526.7106598984772, 58.24285714285725, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(1606.0203045685278, 58.24285714285725, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1645.6751269035533, 174.7285714285715, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(1566.3654822335025, 990.1285714285714, 'X[2] <= 87.0\ngini = 0.494\nsamples =
18\nvalue = [8, 10]'),
 Text(1526.7106598984772, 873.6428571428571, 'X[4] <= 24.5\ngini = 0.444\nsamples =
15\nvalue = [5, 10]'),
 Text(1487.0558375634516, 757.1571428571428, 'gini = 0.0\nsamples = 6\nvalue = [0,
6]'),
 Text(1566.3654822335025, 757.1571428571428, 'X[4] <= 85.0\ngini = 0.494\nsamples =
9\nvalue = [5, 4]'),
 Text(1526.7106598984772, 640.6714285714286, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(1606.0203045685278, 640.6714285714286, 'X[1] <= 105.0\ngini = 0.32\nsamples =
5\nvalue = [1, 4]'),
 Text(1566.3654822335025, 524.1857142857143, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
```

```
 Text(1645.6751269035533, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1606.0203045685278, 873.6428571428571, 'gini = 0.0\nsamples = 3\nvalue = [3,
0]'),
 Text(2101.7055837563453, 1106.6142857142856, 'X[7] <= 57.0\ngini = 0.489\nsamples =
61\nvalue = [26, 35]'),
 Text(2062.0507614213197, 990.1285714285714, 'X[1] <= 116.5\ngini = 0.474\nsamples =
57\nvalue = [22, 35]'),
 Text(1883.6040609137056, 873.6428571428571, 'X[5] <= 36.0\ngini = 0.32\nsamples = 2
0\nvalue = [4, 16]'),
 Text(1804.2944162436547, 757.1571428571428, 'X[0] <= 0.5\ngini = 0.133\nsamples = 1
4\nvalue = [1, 13]'),
 Text(1764.6395939086294, 640.6714285714286, 'X[1] <= 108.5\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(1724.984771573604, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(1804.2944162436547, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(1843.94923857868, 640.6714285714286, 'gini = 0.0\nsamples = 12\nvalue = [0, 1
2]'),
 Text(1962.9137055837564, 757.1571428571428, 'X[1] <= 110.0\ngini = 0.5\nsamples = 6
\nvalue = [3, 3]'),
 Text(1923.2588832487309, 640.6714285714286, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(2002.5685279187817, 640.6714285714286, 'X[6] <= 0.249\ngini = 0.375\nsamples =
4\nvalue = [1, 3]'),
 Text(1962.9137055837564, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(2042.223350253807, 524.1857142857143, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(2240.4974619289337, 873.6428571428571, 'X[5] <= 39.5\ngini = 0.5\nsamples = 37
\nvalue = [18, 19]'),
 Text(2200.8426395939086, 757.1571428571428, 'X[4] <= 170.5\ngini = 0.496\nsamples =
33\nvalue = [18, 15]'),
 Text(2161.187817258883, 640.6714285714286, 'X[6] <= 0.944\ngini = 0.497\nsamples =
28\nvalue = [13, 15]'),
 Text(2121.5329949238576, 524.1857142857143, 'X[0] <= 1.5\ngini = 0.497\nsamples = 2
4\nvalue = [13, 11]'),
 Text(2022.3959390862944, 407.70000000000005, 'X[6] <= 0.667\ngini = 0.278\nsamples
= 6\nvalue = [1, 5]'),
 Text(1982.741116751269, 291.2142857142858, 'gini = 0.0\nsamples = 5\nvalue = [0,
5]'),
 Text(2062.0507614213197, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(2220.6700507614214, 407.70000000000005, 'X[7] <= 42.5\ngini = 0.444\nsamples =
18\nvalue = [12, 6]'),
 Text(2141.3604060913704, 291.2142857142858, 'X[6] <= 0.228\ngini = 0.278\nsamples =
12\nvalue = [10, 2]'),
 Text(2101.7055837563453, 174.7285714285715, 'X[6] <= 0.135\ngini = 0.444\nsamples =
3\nvalue = [1, 2]'),
 Text(2062.0507614213197, 58.24285714285725, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(2141.3604060913704, 58.24285714285725, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(2181.015228426396, 174.7285714285715, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
 Text(2299.979695431472, 291.2142857142858, 'X[1] <= 128.5\ngini = 0.444\nsamples =
6\nvalue = [2, 4]'),
 Text(2260.3248730964465, 174.7285714285715, 'X[7] <= 47.0\ngini = 0.444\nsamples =
3\nvalue = [2, 1]'),
 Text(2220.6700507614214, 58.24285714285725, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(2299.979695431472, 58.24285714285725, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(2339.6345177664975, 174.7285714285715, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(2200.8426395939086, 524.1857142857143, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(2240.4974619289337, 640.6714285714286, 'gini = 0.0\nsamples = 5\nvalue = [5,
```

```
0]'),
 Text(2280.1522842639592, 757.1571428571428, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(2141.3604060913704, 990.1285714285714, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(3198.409263959391, 1456.0714285714284, 'X[1] <= 154.5\ngini = 0.442\nsamples =
179\nvalue = [59, 120]'),
 Text(2726.2690355329946, 1339.5857142857142, 'X[5] <= 29.95\ngini = 0.5\nsamples =
86\nvalue = [44, 42]'),
 Text(2518.0812182741115, 1223.1, 'X[7] <= 50.5\ngini = 0.366\nsamples = 29\nvalue =
[22, 7]'),
 Text(2478.426395939086, 1106.6142857142856, 'X[0] <= 9.0\ngini = 0.455\nsamples = 2
0\nvalue = [13, 7]'),
 Text(2438.771573604061, 990.1285714285714, 'X[4] <= 276.5\ngini = 0.401\nsamples =
18\nvalue = [13, 5]'),
 Text(2399.1167512690354, 873.6428571428571, 'X[3] <= 9.5\ngini = 0.36\nsamples = 17
\nvalue = [13, 4]'),
 Text(2359.46192893401, 757.1571428571428, 'X[2] <= 55.0\ngini = 0.494\nsamples = 9
\nvalue = [5, 4]'),
 Text(2319.807106598985, 640.6714285714286, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(2399.1167512690354, 640.6714285714286, 'X[1] <= 135.0\ngini = 0.408\nsamples =
7\nvalue = [5, 2]'),
 Text(2359.46192893401, 524.1857142857143, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(2438.771573604061, 524.1857142857143, 'X[2] <= 77.5\ngini = 0.278\nsamples = 6
\nvalue = [5, 1]'),
 Text(2399.1167512690354, 407.70000000000005, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(2478.426395939086, 407.70000000000005, 'X[2] <= 85.0\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(2438.771573604061, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(2518.0812182741115, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(2438.771573604061, 757.1571428571428, 'gini = 0.0\nsamples = 8\nvalue = [8,
0]'),
 Text(2478.426395939086, 873.6428571428571, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(2518.0812182741115, 990.1285714285714, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(2557.736040609137, 1106.6142857142856, 'gini = 0.0\nsamples = 9\nvalue = [9,
0]'),
 Text(2934.456852791878, 1223.1, 'X[2] <= 61.0\ngini = 0.474\nsamples = 57\nvalue =
[22, 35]'),
 Text(2894.8020304568527, 1106.6142857142856, 'gini = 0.0\nsamples = 7\nvalue = [0,
7]'),
 Text(2974.1116751269033, 1106.6142857142856, 'X[7] <= 42.5\ngini = 0.493\nsamples =
50\nvalue = [22, 28]'),
 Text(2874.97461928934, 990.1285714285714, 'X[4] <= 261.0\ngini = 0.491\nsamples = 3
0\nvalue = [17, 13]'),
 Text(2835.319796954315, 873.6428571428571, 'X[5] <= 43.4\ngini = 0.499\nsamples = 2
5\nvalue = [12, 13]'),
 Text(2795.6649746192893, 757.1571428571428, 'X[2] <= 81.0\ngini = 0.496\nsamples =
22\nvalue = [12, 10]'),
 Text(2756.0101522842638, 640.6714285714286, 'X[5] <= 37.65\ngini = 0.469\nsamples =
16\nvalue = [6, 10]'),
 Text(2716.3553299492387, 524.1857142857143, 'X[3] <= 24.0\ngini = 0.5\nsamples = 12
\nvalue = [6, 6]'),
 Text(2637.0456852791876, 407.70000000000005, 'X[0] <= 9.5\ngini = 0.32\nsamples = 5
\nvalue = [1, 4]'),
 Text(2597.3908629441626, 291.2142857142858, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(2676.700507614213, 291.2142857142858, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(2795.6649746192893, 407.70000000000005, 'X[2] <= 76.0\ngini = 0.408\nsamples =
7\nvalue = [5, 2]'),
 Text(2756.0101522842638, 291.2142857142858, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
```

```
 Text(2835.319796954315, 291.2142857142858, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(2795.6649746192893, 524.1857142857143, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(2835.319796954315, 640.6714285714286, 'gini = 0.0\nsamples = 6\nvalue = [6,
0]'),
 Text(2874.97461928934, 757.1571428571428, 'gini = 0.0\nsamples = 3\nvalue = [0,
3]'),
 Text(2914.6294416243654, 873.6428571428571, 'gini = 0.0\nsamples = 5\nvalue = [5,
0]'),
 Text(3073.248730964467, 990.1285714285714, 'X[6] <= 0.226\ngini = 0.375\nsamples =
20\nvalue = [5, 15]'),
 Text(2993.939086294416, 873.6428571428571, 'X[6] <= 0.146\ngini = 0.32\nsamples = 5
\nvalue = [4, 1]'),
 Text(2954.284263959391, 757.1571428571428, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(3033.5939086294416, 757.1571428571428, 'gini = 0.0\nsamples = 4\nvalue = [4,
0]'),
 Text(3152.5583756345177, 873.6428571428571, 'X[6] <= 1.391\ngini = 0.124\nsamples =
15\nvalue = [1, 14]'),
 Text(3112.903553299492, 757.1571428571428, 'gini = 0.0\nsamples = 14\nvalue = [0, 1
4]'),
 Text(3192.213197969543, 757.1571428571428, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(3670.549492385787, 1339.5857142857142, 'X[5] <= 23.1\ngini = 0.271\nsamples =
93\nvalue = [15, 78]'),
 Text(3630.8946700507613, 1223.1, 'gini = 0.0\nsamples = 2\nvalue = [2, 0]'),
 Text(3710.204314720812, 1223.1, 'X[4] <= 544.0\ngini = 0.245\nsamples = 91\nvalue =
[13, 78]'),
 Text(3593.718274111675, 1106.6142857142856, 'X[7] <= 62.5\ngini = 0.219\nsamples =
88\nvalue = [11, 77]'),
 Text(3479.710659898477, 990.1285714285714, 'X[2] <= 92.0\ngini = 0.191\nsamples = 8
4\nvalue = [9, 75]'),
 Text(3370.659898477157, 873.6428571428571, 'X[6] <= 0.126\ngini = 0.163\nsamples =
78\nvalue = [7, 71]'),
 Text(3271.522842639594, 757.1571428571428, 'X[4] <= 84.0\ngini = 0.5\nsamples = 2\n
value = [1, 1]'),
 Text(3231.8680203045683, 640.6714285714286, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(3311.1776649746193, 640.6714285714286, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(3469.7969543147206, 757.1571428571428, 'X[7] <= 48.0\ngini = 0.145\nsamples =
76\nvalue = [6, 70]'),
 Text(3390.48730964467, 640.6714285714286, 'X[6] <= 0.304\ngini = 0.095\nsamples = 6
0\nvalue = [3, 57]'),
 Text(3350.8324873096444, 524.1857142857143, 'X[6] <= 0.27\ngini = 0.305\nsamples =
16\nvalue = [3, 13]'),
 Text(3311.1776649746193, 407.70000000000005, 'X[0] <= 9.5\ngini = 0.133\nsamples =
14\nvalue = [1, 13]'),
 Text(3271.522842639594, 291.2142857142858, 'gini = 0.0\nsamples = 12\nvalue = [0, 1
2]'),
 Text(3350.8324873096444, 291.2142857142858, 'X[1] <= 177.0\ngini = 0.5\nsamples = 2
\nvalue = [1, 1]'),
 Text(3311.1776649746193, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(3390.48730964467, 174.7285714285715, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(3390.48730964467, 407.70000000000005, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(3430.1421319796955, 524.1857142857143, 'gini = 0.0\nsamples = 44\nvalue = [0,
44]'),
 Text(3549.1065989847716, 640.6714285714286, 'X[7] <= 50.5\ngini = 0.305\nsamples =
16\nvalue = [3, 13]'),
 Text(3509.451776649746, 524.1857142857143, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(3588.7614213197967, 524.1857142857143, 'X[6] <= 1.157\ngini = 0.133\nsamples =
14\nvalue = [1, 13]'),
 Text(3549.1065989847716, 407.70000000000005, 'gini = 0.0\nsamples = 13\nvalue = [0,
13]'),
```

```
 Text(3628.416243654822, 407.70000000000005, 'gini = 0.0\nsamples = 1\nvalue = [1,
0]'),
 Text(3588.7614213197967, 873.6428571428571, 'X[1] <= 177.0\ngini = 0.444\nsamples =
6\nvalue = [2, 4]'),
 Text(3549.1065989847716, 757.1571428571428, 'gini = 0.0\nsamples = 4\nvalue = [0,
4]'),
 Text(3628.416243654822, 757.1571428571428, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(3707.725888324873, 990.1285714285714, 'X[5] <= 30.7\ngini = 0.5\nsamples = 4\n
value = [2, 2]'),
 Text(3668.0710659898477, 873.6428571428571, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]'),
 Text(3747.3807106598983, 873.6428571428571, 'gini = 0.0\nsamples = 2\nvalue = [0,
2]'),
 Text(3826.690355329949, 1106.6142857142856, 'X[0] <= 2.5\ngini = 0.444\nsamples = 3
\nvalue = [2, 1]'),
 Text(3787.035532994924, 990.1285714285714, 'gini = 0.0\nsamples = 1\nvalue = [0,
1]'),
 Text(3866.3451776649745, 990.1285714285714, 'gini = 0.0\nsamples = 2\nvalue = [2,
0]')]
```



In [ ]:

```
!conda install graphviz

!pip install pydotplus
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done
```

In [19]:

```
## https://graphviz.org/download/
## https://stackoverflow.com/questions/28312534/graphvizs-executables-are-not-found-
```

export_graphviz function converts decision tree classifier into dot file and pydotplus convert this dot file to png or displayable form on Jupyter.

In [11]:

```
from six import StringIO
```

In [24]:

```
import six
import sys
sys.modules['sklearn.externals.six'] = six
```

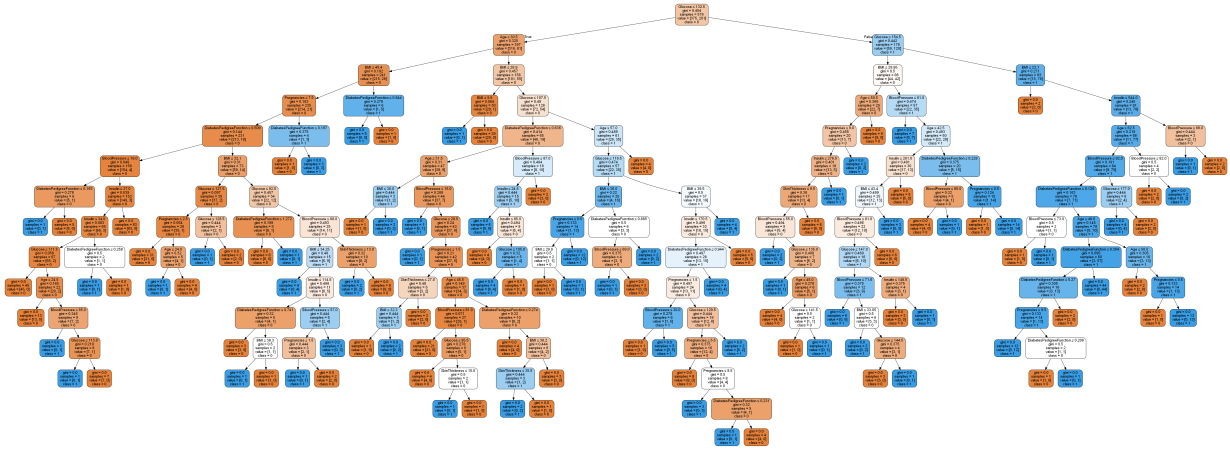
In [25]:

```
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
```

In [26]:
```
import graphviz
```

In [27]:
```
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True,feature_names = feature_name,class_names=['0
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('diabetes.png')
Image(graph.create_png())
```

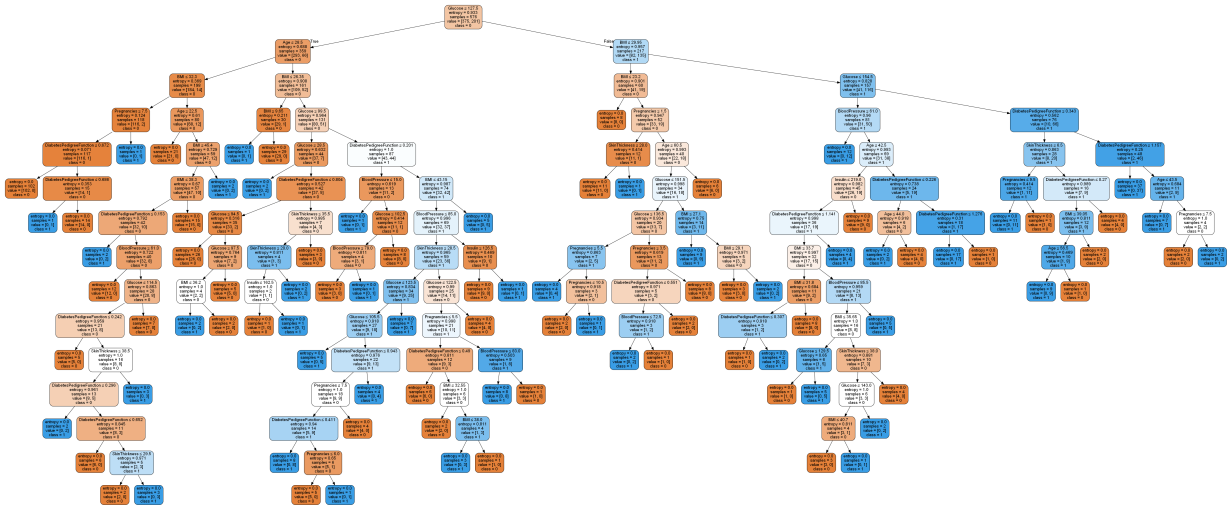
Out[27]:



Create Decision Tree classifer object clf = DecisionTreeClassifier(criterion="entropy") # Train Decision Tree Classifer clf = clf.fit(X_train,y_train) #Predict the response for test dataset y_pred = clf.predict(X_test) # Model Accuracy, how often is the classifier correct? print("Accuracy:",metrics.accuracy_score(y_test, y_pred)) print('Testing Set Evaluation F1-Score=>',f1_score(y_test,y_pred))

In [17]:
```
from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus
dot_data = StringIO()
export_graphviz(clf, out_file=dot_data,
                filled=True, rounded=True,
                special_characters=True, feature_names = feature_name,class_names=['
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
graph.write_png('diabetes.png')
Image(graph.create_png())
```

Out[17]:



In [ ]:
```
```

In [ ]:

https://www.analyticsvidhya.com/blog/2020/03/beginners-guide-random-forest-hyperparameter-tuning/?utm_source=blog&utm_medium=decision-tree-vs-random-forest-algorithm

In [18]:

```
# Building  Random Forest Classifier
from sklearn.metrics import f1_score
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(criterion = 'entropy', random_state = 42)
rfc.fit(X_train, y_train)

# Evaluating on Training set
rfc_pred_train = rfc.predict(X_train)
print('Training Set Evaluation F1-Score=>',f1_score(y_train,rfc_pred_train))
```

```
Training Set Evaluation F1-Score=> 1.0
```

In [19]:

```
# Evaluating on Test set
rfc_pred_test = rfc.predict(X_test)
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
print('Testing Set Evaluation F1-Score=>',f1_score(y_test,rfc_pred_test))
```

```
Accuracy: 0.7135416666666666
Testing Set Evaluation F1-Score=> 0.5689655172413793
```

In [ ]: