

In [12]:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
```

In [13]:

```
data=pd.read_csv('diabetes.csv')
data.head()
```

Out[13]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
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

In [14]:

```
not_zero=['Glucose','BloodPressure','SkinThickness','Insulin','BMI']
for column in not_zero:
    mean=int(data[column].mean())
    data[column]=data[column].replace(0,mean)
```

In [15]:

```
x=data.iloc[:, :8]
y=data.iloc[:, 8]
```

In [16]:

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=13)
```

In [17]:

```
dtc=DecisionTreeClassifier()
dtc.fit(x_train,y_train)
```

Out[17]:

```
DecisionTreeClassifier()
```

In [18]:

```
y_predict=dtc.predict(x_test)
```

In [19]:



```
accuracy_score(y_test,y_predict)
```

Out[19]:

```
0.6753246753246753
```

In [20]:



```
confusion_matrix(y_test,y_predict)
```

Out[20]:

```
array([[104,  40],
       [ 35,  52]], dtype=int64)
```

In [22]:



```
'''KNeighborsClassifier's accuracy_score = 0.7056277056277056
and confusion_matrix = array([[120,  24],
                             [ 44,  43]], dtype=int64)'''
```

Out[22]:

```
"KNeighborsClassifier's accuracy_score = 0.7056277056277056\nand confusion_matrix = array([[120,  24],\n                             [ 44,  43]], dtype=int64)"
```

In []:

