DECISION TREE CLASSIFIER MODEL ¶

In [32]:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
import matplotlib.pyplot as plt
%matplotlib inline
```

In [33]:

```
df= pd.read_csv('/Users/anjali98/lc.csv')
df.head()
df.describe()
```

Out[33]:

	Age	Smokes	Alcohol	Result
count	107.000000	107.000000	107.000000	107.000000
mean	43.635514	16.046729	3.280374	0.485981
std	15.229931	6.724193	2.188281	0.502155
min	18.000000	0.000000	0.000000	0.000000
25%	32.000000	12.000000	2.000000	0.000000
50%	42.000000	20.000000	3.000000	0.000000
75%	56.000000	20.000000	4.500000	1.000000
max	77.000000	34.000000	8.000000	1.000000

In [34]:

```
y= df['Result']
X= df.drop(['Result'], axis=1)
```

In [35]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.4, random_st
```

In [36]:

```
clf_entropy=DecisionTreeClassifier(criterion='entropy',random_state=100,max_depth=3,
clf_entropy.fit(X_train,y_train)
```

Out[36]:

```
DecisionTreeClassifier(criterion='entropy', max_depth=3, min_samples_l
eaf=5,
                       random_state=100)
```

In [37]:

```
y_pred=clf_entropy.predict(X_test)
```

FINAL ACCURACY SCORE

In [38]:

```
ac = (accuracy_score(y_test,y_pred)*100)
print('Accuarcy score : ', ac)
```

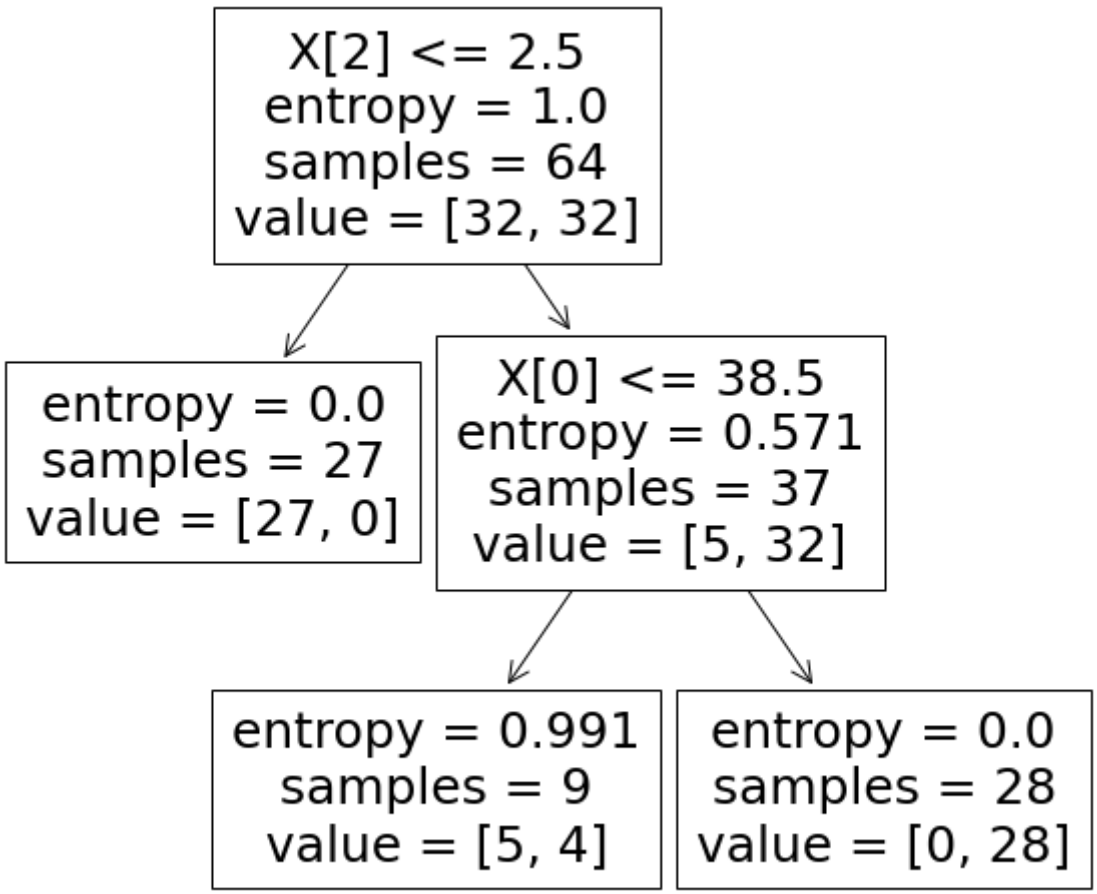
```
Accuarcy score :  97.67441860465115
```

In [39]:

```
plt.figure(figsize=(10,9))
from sklearn import tree
tree.plot_tree(clf_entropy.fit(X_train, y_train))
```

Out[39]:

```
[Text(223.2, 407.7, 'X[2] <= 2.5\nentropy = 1.0\nsamples = 64\nvalue =
[32, 32]'),
 Text(111.6, 244.62, 'entropy = 0.0\nsamples = 27\nvalue = [27, 0]'),
 Text(334.79999999999995, 244.62, 'X[0] <= 38.5\nentropy = 0.571\nsamp
les = 37\nvalue = [5, 32]'),
 Text(223.2, 81.53999999999996, 'entropy = 0.991\nsamples = 9\nvalue =
[5, 4]'),
 Text(446.4, 81.53999999999996, 'entropy = 0.0\nsamples = 28\nvalue =
[0, 28]')]
```



In [ ]: