In [2]:
```python
# In Machine Learning, Entropy is a measure to calculate the impurity of the grou
#Entropy is used in tree algorithms such as Decision tree to decide where to spli
#Entropy helps to check the homogeneity of the data.

#In our slides 14, we split our data by company.
#Can you please choose a company and compute its entropy?
#Please interpret your result.

import pandas as pd
import numpy as np

df = pd.read_csv('OneDrive\Desktop\salaries.csv')
#dfOrig.head()

#selecting the company
newdf = df[df.company=='google']
#newdf.head()

#prepping for analysis
inputs = df.drop('salary_more_then_100k', axis = 1)
target = df['salary_more_then_100k']

from sklearn.preprocessing import LabelEncoder
le_company = LabelEncoder()
le_job = LabelEncoder()
le_degree = LabelEncoder()

inputs['company_n'] = le_company.fit_transform(inputs['company'])
inputs['job_n'] = le_company.fit_transform(inputs['job'])
inputs['degree_n'] = le_company.fit_transform(inputs['degree'])
inputs_n = inputs.drop(['company','job','degree'],axis='columns')

inputs_n.head()
```

Out[2]:

	company_n	job_n	degree_n
0	2	2	0
1	2	2	1
2	2	0	0
3	2	0	1
4	2	1	0

from sklearn import tree model = tree.DecisionTreeClassifier() model.fit(inputs_n, target)
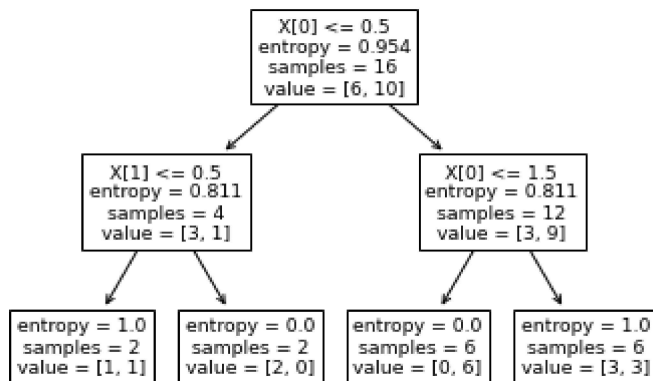model.score(inputs_n, target)

In [7]:
```python
from sklearn import tree
model=tree.DecisionTreeClassifier(criterion = 'entropy', max_depth = 2)
model.fit(inputs_n,target)
model.score(inputs_n,target)
```

Out[7]: 0.75

In [8]:
```python
tree.plot_tree(model)
```

Out[8]: [Text(0.5, 0.8333333333333334, 'X[0] <= 0.5\nentropy = 0.954\nsamples = 16\nval
         ue = [6, 10]'),
          Text(0.25, 0.5, 'X[1] <= 0.5\nentropy = 0.811\nsamples = 4\nvalue = [3, 1]'),
          Text(0.125, 0.16666666666666666, 'entropy = 1.0\nsamples = 2\nvalue = [1,
         1]'),
          Text(0.375, 0.16666666666666666, 'entropy = 0.0\nsamples = 2\nvalue = [2,
         0]'),
          Text(0.75, 0.5, 'X[0] <= 1.5\nentropy = 0.811\nsamples = 12\nvalue = [3, 9]'),
          Text(0.625, 0.16666666666666666, 'entropy = 0.0\nsamples = 6\nvalue = [0,
         6]'),
          Text(0.875, 0.16666666666666666, 'entropy = 1.0\nsamples = 6\nvalue = [3,
         3]')]

In [ ]: