In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("drug200.csv")
df.head()
```

Out[1]:

	Age	Sex	BP	Cholesterol	Na_to_K	Drug
0	23	F	HIGH	HIGH	25.355	drugY
1	47	M	LOW	HIGH	13.093	drugC
2	47	M	LOW	HIGH	10.114	drugC
3	28	F	NORMAL	HIGH	7.798	drugX
4	61	F	LOW	HIGH	18.043	drugY

In [2]:
```
df.describe()
```

Out[2]:

	Age	Na_to_K
count	200.000000	200.000000
mean	44.315000	16.084485
std	16.544315	7.223956
min	15.000000	6.269000
25%	31.000000	10.445500
50%	45.000000	13.936500
75%	58.000000	19.380000
max	74.000000	38.247000

In [3]:
```
x = df[["Age","Sex","BP","Cholesterol","Na_to_K"]].values
```

In [4]:
```
x[0:5]
```

Out[4]:
```
array([[23, 'F', 'HIGH', 'HIGH', 25.355],
       [47, 'M', 'LOW', 'HIGH', 13.093],
       [47, 'M', 'LOW', 'HIGH', 10.113999999999999],
       [28, 'F', 'NORMAL', 'HIGH', 7.797999999999999],
       [61, 'F', 'LOW', 'HIGH', 18.043]], dtype=object)
```

In [5]:
```
from sklearn.preprocessing import LabelEncoder
label = LabelEncoder()
```

In [6]:
```
label.fit(["F","M"])
x[:,1] = label.transform(x[:,1])
```

In [7]:
```
label.fit(["HIGH","NORMAL","LOW"])
x[:,2] = label.transform(x[:,2])
```

In [8]:
```
label.fit(["HIGH","NORMAL"])
x[:,3] = label.transform(x[:,3])
```

In [9]:
```
x[0:5]
```

Out[9]:
```
array([[23, 0, 0, 0, 25.355],
       [47, 1, 1, 0, 13.093],
       [47, 1, 1, 0, 10.113999999999999],
       [28, 0, 2, 0, 7.797999999999999],
       [61, 0, 1, 0, 18.043]], dtype=object)
```

In [10]:
```
y = df["Drug"]
y.head()
```

Out[10]:
```
0    drugY
1    drugC
2    drugC
3    drugX
4    drugY
Name: Drug, dtype: object
```

In [11]:
```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.25, random_state = 3)
```

In [12]:
```
print(x_train.shape)
print(x_test.shape)
```

```
(150, 5)
(50, 5)
```

In [18]:
```
from sklearn.tree import DecisionTreeClassifier
rissh = DecisionTreeClassifier(criterion="entropy",max_depth=4)
```

In [28]:
```
rissh.fit(x_train,y_train)
```

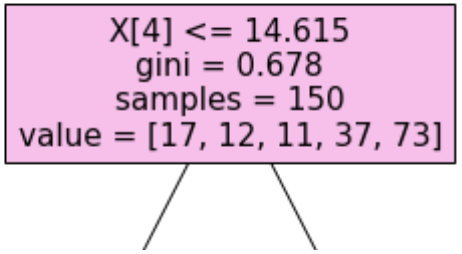
Out[28]:
```
DecisionTreeClassifier()
```

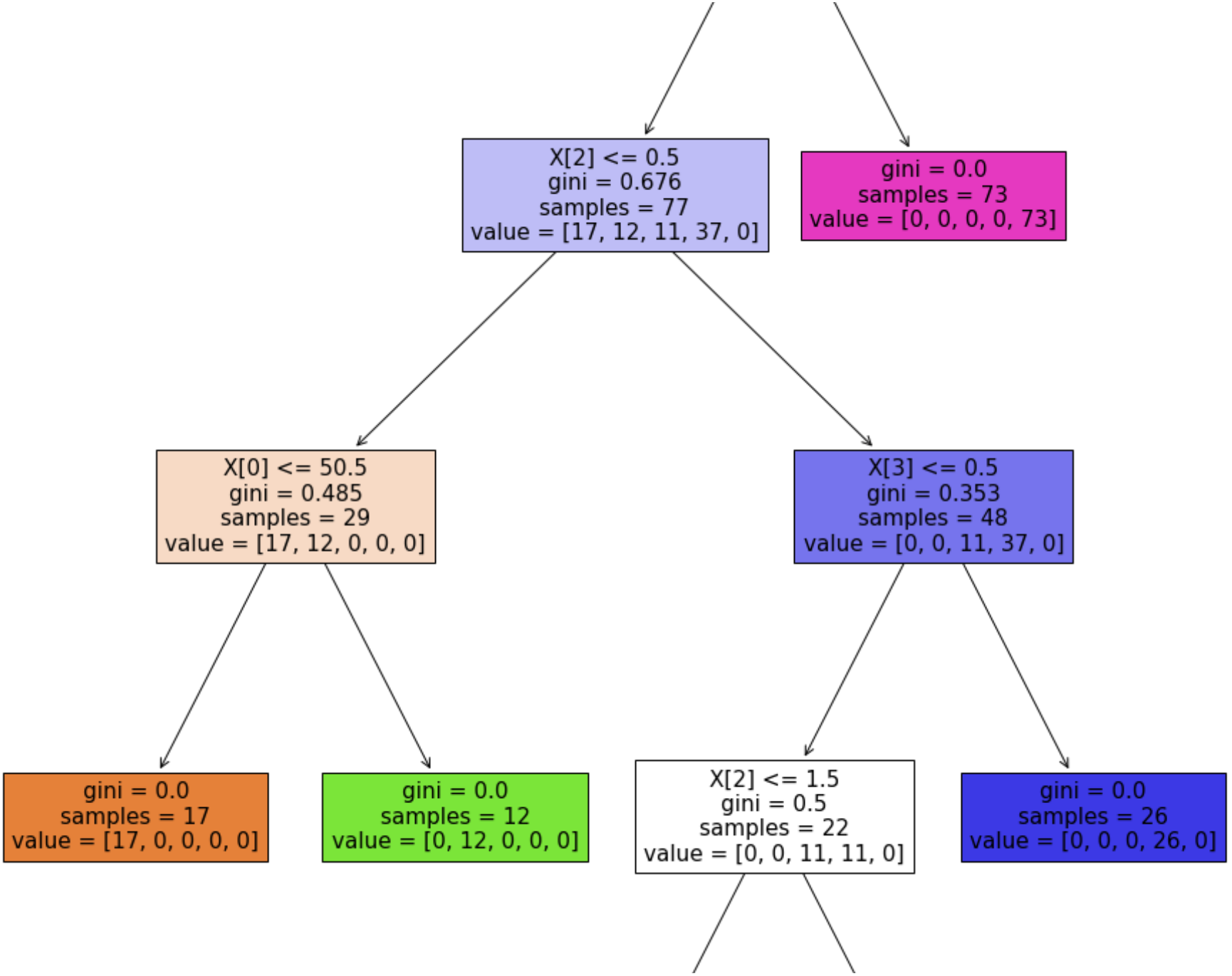
In [29]:
```
y_pred = rissh.predict(x_test)
y_pred[0:5]
```

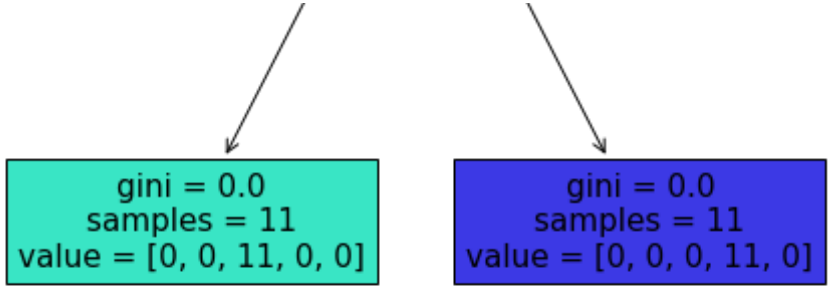
Out[29]:
```
array(['drugY', 'drugX', 'drugX', 'drugX', 'drugX'], dtype=object)
```

In [33]:
```
plt.figure(figsize=(16,20))
tree.plot_tree(rissh,filled = True)
```

Out[33]:
```
[Text(558.0, 978.48, 'X[4] <= 14.615\ngini = 0.678\nsamples = 150\nvalue = [17, 12, 11, 37, 73]'),
 Text(446.4, 761.0400000000001, 'X[2] <= 0.5\ngini = 0.676\nsamples = 77\nvalue = [17, 12, 11, 37, 0]'),
 Text(223.2, 543.6, 'X[0] <= 50.5\ngini = 0.485\nsamples = 29\nvalue = [17, 12, 0, 0, 0]'),
 Text(111.6, 326.1600000000001, 'gini = 0.0\nsamples = 17\nvalue = [17, 0, 0, 0, 0]'),
 Text(334.79999999999995, 326.1600000000001, 'gini = 0.0\nsamples = 12\nvalue = [0, 12, 0, 0, 0]'),
 Text(669.5999999999999, 543.6, 'X[3] <= 0.5\ngini = 0.353\nsamples = 48\nvalue = [0, 0, 11, 37, 0]'),
 Text(558.0, 326.1600000000001, 'X[2] <= 1.5\ngini = 0.5\nsamples = 22\nvalue = [0, 0, 11, 11, 0]'),
 Text(446.4, 108.72000000000003, 'gini = 0.0\nsamples = 11\nvalue = [0, 0, 11, 0, 0]'),
 Text(669.5999999999999, 108.72000000000003, 'gini = 0.0\nsamples = 11\nvalue = [0, 0, 0, 11, 0]'),
 Text(781.1999999999999, 326.1600000000001, 'gini = 0.0\nsamples = 26\nvalue = [0, 0, 0, 26, 0]'),
 Text(669.5999999999999, 761.0400000000001, 'gini = 0.0\nsamples = 73\nvalue = [0, 0, 0, 0, 73]')]
```







In [21]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)*100
```

Out[21]: 98.0