```python
import numpy as np
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
```

```python
my_data = pd.read_csv("drug200.csv", delimiter=",")
my_data[0:5]
```

	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

```python
my_data.shape
```

```
(200, 6)
```

```python
X = my_data[['Age', 'Sex', 'BP', 'Cholesterol', 'Na_to_K']].values
X[0:5]
```

```
array([[23, 'F', 'HIGH', 'HIGH', 25.355],
       [47, 'M', 'LOW', 'HIGH', 13.093],
       [47, 'M', 'LOW', 'HIGH', 10.114],
       [28, 'F', 'NORMAL', 'HIGH', 7.798],
       [61, 'F', 'LOW', 'HIGH', 18.043]], dtype=object)
```

```python
from sklearn import preprocessing
le_sex = preprocessing.LabelEncoder()
le_sex.fit(['F','M'])
X[:,1] = le_sex.transform(X[:,1])


le_BP = preprocessing.LabelEncoder()
le_BP.fit([ 'LOW', 'NORMAL', 'HIGH'])
X[:,2] = le_BP.transform(X[:,2])


le_Chol = preprocessing.LabelEncoder()
le_Chol.fit([ 'NORMAL', 'HIGH'])
X[:,3] = le_Chol.transform(X[:,3])

X[0:5]
```

```
array([[23, 0, 0, 0, 25.355],
       [47, 1, 1, 0, 13.093],
       [47, 1, 1, 0, 10.114],
       [28, 0, 2, 0, 7.798],
       [61, 0, 1, 0, 18.043]], dtype=object)
```

```python
y = my_data["Drug"]
y[0:5]
```

```
0    drugY
1    drugC
2    drugC
3    drugX
4    drugY
Name: Drug, dtype: object
```

```python
from sklearn.model_selection import train_test_split
X_trainset, X_testset, y_trainset, y_testset = train_test_split(X, y, test_size=0.3, random_state=3)
```

```python
print(X_trainset.shape)
print(y_trainset.shape)
```

```
(140, 5)
(140,)
```

```python
print(X_testset.shape)
print(y_testset.shape)
```

```
(60, 5)
(60,)
```

```python
drugTree = DecisionTreeClassifier(criterion="entropy", max_depth = 4)
drugTree
```

```
▼              DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy', max_depth=4)
```

```python
drugTree.fit(X_trainset,y_trainset)
```

```
Out[ ]:  ▼                    DecisionTreeClassifier
         DecisionTreeClassifier(criterion='entropy', max_depth=4)
```

```
In [ ]: predTree = drugTree.predict(X_testset)
```

```
In [ ]: print (predTree [0:5])
        print (y_testset [0:5])

        ['drugY' 'drugX' 'drugX' 'drugX' 'drugX']
        40      drugY
        51      drugX
        139     drugX
        197     drugX
        170     drugX
        Name: Drug, dtype: object
```

```
In [ ]: from sklearn import metrics
        import matplotlib.pyplot as plt
        print("DecisionTrees's Accuracy: ", metrics.accuracy_score(y_testset, predTree))

        DecisionTrees's Accuracy:  0.9833333333333333
```

```
In [ ]: from io import StringIO
        import pydotplus
        import matplotlib.image as mpimg
        from sklearn import tree
        %matplotlib inline
```

```
In [ ]: dot_data = StringIO()
        filename = "drugtree.png"
        featureNames = my_data.columns[0:5]
        targetNames = my_data["Drug"].unique().tolist()
        out=tree.export_graphviz(drugTree,feature_names=featureNames, out_file=dot_data, class_names= np.unique(y_trainset), filled=Tr
        graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
        graph.write_png(filename)
        img = mpimg.imread(filename)
        plt.figure(figsize=(100, 200))
        plt.imshow(img,interpolation='nearest')
```

```
Out[ ]: <matplotlib.image.AxesImage at 0x7bc19f8e0a30>
```