Chapter 4: Demo

```
### Stolen car: Red, SUV, Domestic => stolen Yes or No???
from google.colab import drive
drive.mount("/content/gdrive", force_remount=True)
%cd '/content/gdrive/My Drive/LDS6_MachineLearning/practice_2020/Chapter6_Decision_Tree/'
```

```
     Mounted at /content/gdrive
     /content/gdrive/My Drive/LDS6_MachineLearning/practice_2020/Chapter6_Decision_Tree
```

```
from sklearn import datasets
from IPython.display import Image
from sklearn import tree
import pydotplus
import pandas as pd
import numpy as np
```

```
X = np.array([["Red", "Sports", "Domestic"],
              ["Red", "Sports", "Domestic"],
              ["Red", "Sports", "Domestic"],
              ["Yellow", "Sports", "Domestic"],
              ["Yellow", "Sports", "Imported"],
              ["Yellow", "SUV", "Imported"],
              ["Yellow", "SUV", "Imported"],
              ["Yellow", "SUV", "Domestic"],
              ["Red", "SUV", "Imported"],
              ["Red", "Sports", "Imported"]])
Y = np.array(["Yes", "No", "Yes", "No", "Yes", "No", "Yes", "No", "No", "Yes"])
```

```
X = pd.DataFrame(X, columns = ['color', 'type', 'origin'])
X_now = pd.get_dummies(X)
X_now
```

	color_Red	color_Yellow	type_SUV	type_Sports	origin_Domestic	origin_Imported
0	1	0	0	1	1	0
1	1	0	0	1	1	0
2	1	0	0	1	1	0

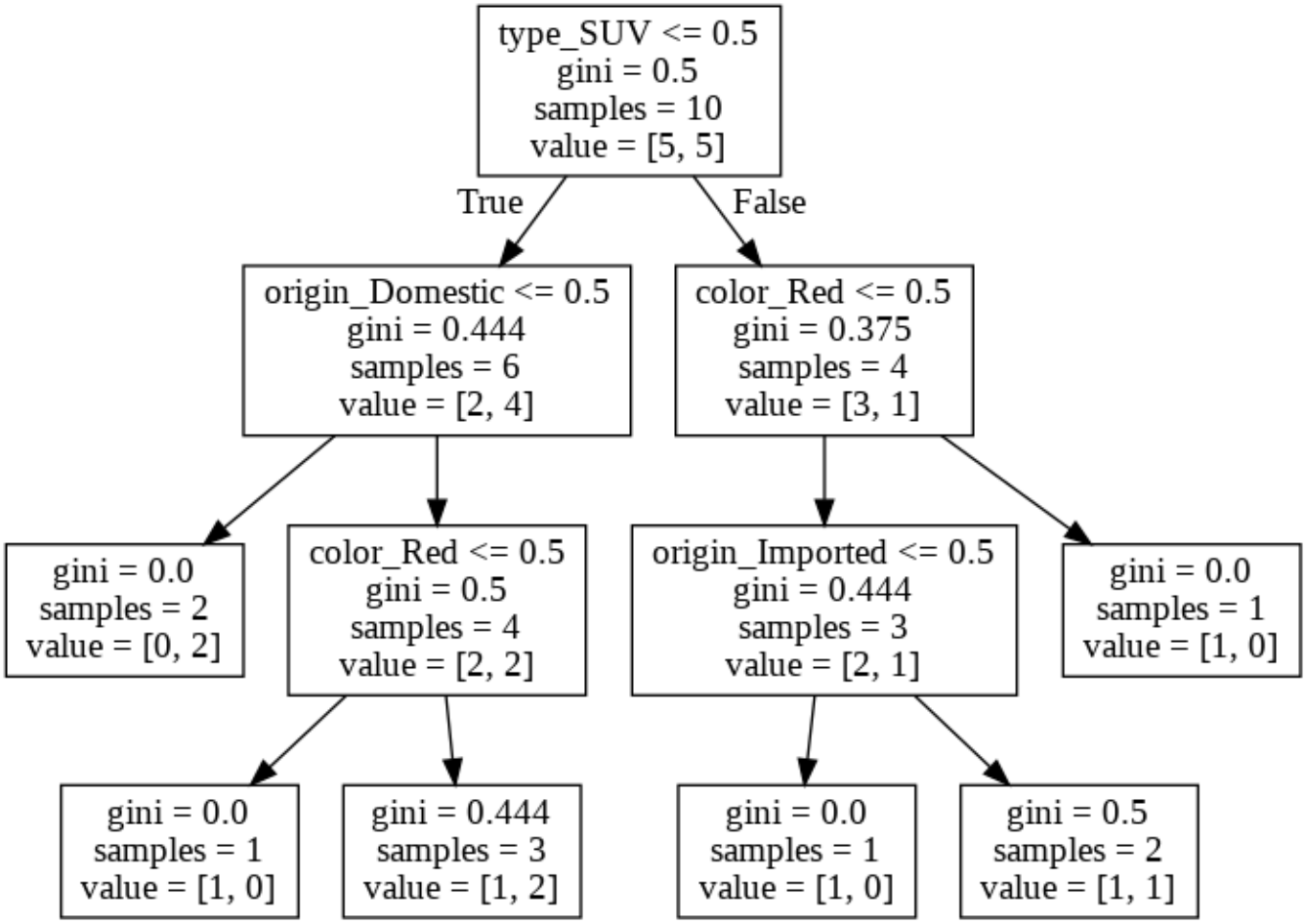
```
Y = pd.DataFrame(Y, columns = ['stolen'])
Y_now = pd.get_dummies(Y, drop_first=True)
Y_now
```

	stolen_Yes
0	1
1	0
2	1
3	0
4	1
5	0
6	1
7	0
8	0
9	1

```
from sklearn.tree import DecisionTreeClassifier


clf = DecisionTreeClassifier()
model = clf.fit(X_now, Y_now)


X_new = [[1,0,1,0,1,0]] #"Red", "SUV", "Domestic"


#Predict Output
predicted= model.predict(X_new) # Red, SUV, Domestic
predicted
```

```
array([0], dtype=uint8)
```

```
dot_data = tree.export_graphviz(model, out_file=None,
                                feature_names=X_now.columns
                                )
graph = pydotplus.graph_from_dot_data(dot_data)
Image(graph.create_png())
```


type_SUV <= 0.5
gini = 0.5
samples = 10
value = [5, 5]
True
False
origin_Domestic <= 0.5
gini = 0.444
samples = 6
value = [2, 4]
color_Red <= 0.5
gini = 0.375
samples = 4
value = [3, 1]
gini = 0.0
samples = 2
value = [0, 2]
color_Red <= 0.5
gini = 0.5
samples = 4
value = [2, 2]
origin_Imported <= 0.5
gini = 0.444
samples = 3
value = [2, 1]
gini = 0.0
samples = 1
value = [1, 0]
gini = 0.0
samples = 1
value = [1, 0]
gini = 0.444
samples = 3
value = [1, 2]
gini = 0.0
samples = 1
value = [1, 0]
gini = 0.5
samples = 2
value = [1, 1]