Author - AYUSH CHHOKER

DATA SCIENCE AND BUSINESS ANALYTICS INTERN¶ ¶

Task 6 : Prediction using Decision Tree

In [6]:

```
# Importing libraries in Python
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import sklearn.datasets as datasets
```

In [7]:

```
# Loading the iris dataset
iris=datasets.load_iris()
```

In [8]:

```
# Forming the iris dataframe
X = pd.DataFrame(iris.data, columns=iris.feature_names)
print(X.head(5))

y=iris.target
print(y)
```

```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (c
m)
0                5.1               3.5                1.4               0.
2
1                4.9               3.0                1.4               0.
2
2                4.7               3.2                1.3               0.
2
3                4.6               3.1                1.5               0.
2
4                5.0               3.6                1.4               0.
2
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2]
```

In [9]:

```
X.isnull().sum()
```

Out[9]:

```
sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
dtype: int64
```

In [10]:

```
Y = iris.target
Y
```

Out[10]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

split dataset into train and test sets

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_stat
e = 0)
```

Training the Decision Tree Classification model on the Training set

In [12]:

```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
classifier.fit(X_train, y_train)
```

Out[12]:

```
DecisionTreeClassifier(criterion='entropy', random_state=0)
```

Predicting the test set data

In [13]:

```
classifier.predict(X_test)
```

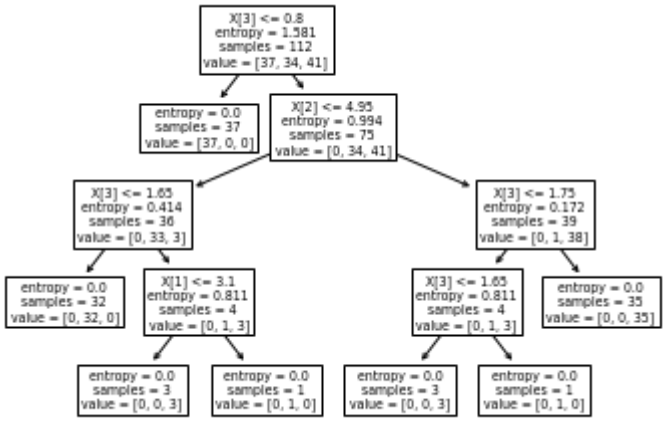
Out[13]:

```
array([2, 1, 0, 2, 0, 2, 0, 1, 1, 1, 2, 1, 1, 1, 1, 0, 1, 1, 0, 0, 2, 1,
       0, 0, 2, 0, 0, 1, 1, 0, 2, 1, 0, 2, 2, 1, 0, 2])
```

Visualization Using Scikit-learn library

In [14]:

```
from sklearn import tree
tree.plot_tree(classifier);
```



In [15]:

```
fn=['sepal length (cm)','sepal width (cm)','petal length (cm)','petal width (cm)']
cn=['setosa', 'versicolor', 'virginica']
fig, axes = plt.subplots(nrows = 1,ncols = 1,figsize = (4,4), dpi=300)
tree.plot_tree(classifier,
               feature_names = fn,
               class_names=cn,
               filled = True);
fig.savefig('imagename.png')
```


petal width (cm) <= 0.8
entropy = 1.581
samples = 112
value = [37, 34, 41]
class = virginica
entropy = 0.0
samples = 37
value = [37, 0, 0]
class = setosa
petal length (cm) <= 4.95
entropy = 0.994
samples = 75
value = [0, 34, 41]
class = virginica
petal width (cm) <= 1.65
entropy = 0.414
samples = 36
value = [0, 33, 3]
class = versicolor
petal width (cm) <= 1.75
entropy = 0.172
samples = 39
value = [0, 1, 38]
class = virginica
entropy = 0.0
samples = 32
value = [0, 32, 0]
class = versicolor
sepal width (cm) <= 3.1
entropy = 0.811
samples = 4
value = [0, 1, 3]
class = virginica
petal width (cm) <= 1.65
entropy = 0.811
samples = 4
value = [0, 1, 3]
class = virginica
entropy = 0.0
samples = 35
value = [0, 0, 35]
class = virginica
entropy = 0.0
samples = 3
value = [0, 0, 3]
class = virginica
entropy = 0.0
samples = 1
value = [0, 1, 0]
class = versicolor
entropy = 0.0
samples = 3
value = [0, 0, 3]
class = virginica
entropy = 0.0
samples = 1
value = [0, 1, 0]
class = versicolor