Workshop - Decision Trees

This workshop deals with understanding the working of decision trees.

Author: Xaltius Pte. Ltd.

Copyright Xaltius Pte. Ltd., 2018

Not for distribution.

```
# Importing libraries in Python
import sklearn.datasets as datasets
import pandas as pd

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

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

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

```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)
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 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]
```

Now let us define the Decision Tree Algorithm

```
# Defining the decision tree algorithm
from sklearn.tree import DecisionTreeClassifier
dtree=DecisionTreeClassifier()
dtree.fit(df,y)

print('Decision Tree Classifer Created')
```
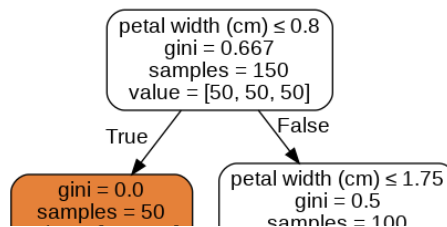
```
    Decision Tree Classifer Created
```

Let us visualize the Decision Tree to understand it better.

```python
# Install required libraries
!pip install pydotplus
!apt-get install graphviz -y
```

```python
# Import necessary libraries for graph viz
from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
import pydotplus

# Visualize the graph
dot_data = StringIO()
export_graphviz(dtree, out_file=dot_data, feature_names=iris.feature_names,
                filled=True, rounded=True,
                special_characters=True)
graph = pydotplus.graph_from_dot_data(dot_data.getvalue())
Image(graph.create_png())
```

petal width (cm) ≤ 0.8
gini = 0.667
samples = 150
value = [50, 50, 50]

True   False

gini = 0.0
samples = 50

petal width (cm) ≤ 1.75
gini = 0.5
samples = 100

You can now feed any new/test data to this classifer and it would be able to predict the right class accordingly.

samples = 54
value = [0, 49, 5]

samples = 46
value = [0, 1, 45]

petal width (cm) ≤ 1.65
gini = 0.041
samples = 48
value = [0, 47, 1]

petal width (cm) ≤ 1.55
gini = 0.444
samples = 6
value = [0, 2, 4]

sepal width (cm) ≤ 3.1
gini = 0.444
samples = 3
value = [0, 1, 2]

gini = 0.0
samples = 43
value = [0, 0, 43]

gini = 0.0
samples = 47
value = [0, 47, 0]

gini = 0.0
samples = 1
value = [0, 0, 1]

gini = 0.0
samples = 3
value = [0, 0, 3]

sepal length (cm) ≤ 6.95
gini = 0.444
samples = 3
value = [0, 2, 1]

gini = 0.0
samples = 2
value = [0, 0, 2]

gini = 0.0
samples = 1
value = [0, 1, 0]

gini = 0.0
samples = 2
value = [0, 2, 0]

gini = 0.0
samples = 1
value = [0, 0, 1]