Machine Learning Lab (PMCA507P)

> **Reg No:** 23MCA1030
>
> **Name :** Vinayak Kumar Singh

Exercise 5 : Decision Tree Classifier (Without splitting the data)

Collab url : https://colab.research.google.com/drive/1JtuP4rL6rUxooGvayhYVBa7k9prskq5L?usp=sharing

```
import numpy as np
import pandas as pd
from sklearn import metrics
```

```
df=pd.read_csv("/content/tennis.csv")
value=['Outlook','Temprature','Humidity','Wind']
df
```

	outlook	temp	humidity	windy	play
0	sunny	hot	high	weak	no
1	sunny	hot	high	strong	no
2	overcast	hot	high	weak	yes
3	rainy	mild	high	weak	yes
4	rainy	cool	normal	weak	yes
5	rainy	cool	normal	strong	no
6	overcast	cool	normal	strong	yes
7	sunny	mild	high	weak	no
8	sunny	cool	normal	weak	yes
9	rainy	mild	normal	weak	yes
10	sunny	mild	normal	strong	yes
11	overcast	mild	high	strong	yes
12	overcast	hot	normal	weak	yes
13	rainy	mild	high	strong	no

```
len(df)
```

```
14
```

```
df.shape
```

```
(14, 5)
```

```
df.head()
```

	outlook	temp	humidity	windy	play
0	sunny	hot	high	weak	no
1	sunny	hot	high	strong	no
2	overcast	hot	high	weak	yes
3	rainy	mild	high	weak	yes
4	rainy	cool	normal	weak	yes

```
df.tail()
```

	outlook	temp	humidity	windy	play
9	rainy	mild	normal	weak	yes
10	sunny	mild	normal	strong	yes
11	overcast	mild	high	strong	yes
12	overcast	hot	normal	weak	yes
13	rainy	mild	high	strong	no

```
df.describe()
```

	outlook	temp	humidity	windy	play
count	14	14	14	14	14
unique	3	3	2	2	2
top	sunny	mild	high	weak	yes
freq	5	6	7	8	9

```
from sklearn import preprocessing
string_to_int= preprocessing.LabelEncoder()
df=df.apply(string_to_int.fit_transform)
df
```

	outlook	temp	humidity	windy	play
0	2	1	0	1	0
1	2	1	0	0	0
2	0	1	0	1	1
3	1	2	0	1	1
4	1	0	1	1	1
5	1	0	1	0	0
6	0	0	1	0	1
7	2	2	0	1	0
8	2	0	1	1	1
9	1	2	1	1	1
10	2	2	1	0	1
11	0	2	0	0	1
12	0	1	1	1	1
13	1	2	0	0	0

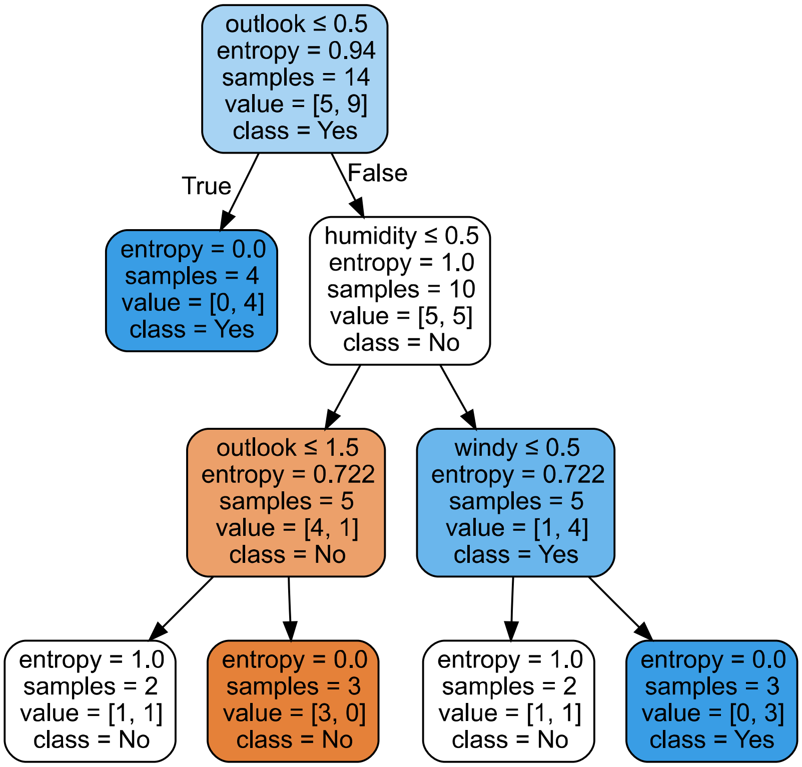
```
feature_cols = ['outlook','temp','humidity','windy']
X = df[feature_cols ]
y = df.play
```

```
from sklearn.tree import export_graphviz
import graphviz

from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier(criterion="entropy", max_depth=3)
clf = clf.fit(X, y)

dot_data = export_graphviz(clf, out_file=None, feature_names=feature_cols, class_names=["No", "Yes"], filled=True, rounded=True, special_ch

graph = graphviz.Source(dot_data)
graph
```


outlook ≤ 0.5
entropy = 0.94
samples = 14
value = [5, 9]
class = Yes
True
False
entropy = 0.0
samples = 4
value = [0, 4]
class = Yes
humidity ≤ 0.5
entropy = 1.0
samples = 10
value = [5, 5]
class = No
outlook ≤ 1.5
entropy = 0.722
samples = 5
value = [4, 1]
class = No
windy ≤ 0.5
entropy = 0.722
samples = 5
value = [1, 4]
class = Yes
entropy = 1.0
samples = 2
value = [1, 1]
class = No
entropy = 0.0
samples = 3
value = [3, 0]
class = No
entropy = 1.0
samples = 2
value = [1, 1]
class = No
entropy = 0.0
samples = 3
value = [0, 3]
class = Yes