Experiment No- 2.4

Student Name: Yash Kumar **UID: 20BCS9256**
Branch: CSE **Section/Group: 616-B**
Semester: 5 **Date of Performance: 10-10-22**
Subject Name: Machine Learning **Subject Code: 20CSP -317**

1. **Aim/Overview of the practical:** Implementing Decision Tree on any dataset and analyse the accuracy.

2. **Task to be done/ Which logistics used:** Analysing accuracy by implementing the Decision Tree on any dataset.

3. **Steps of experiment/Code:**

    1. **Importing libraries such as panda and reading the dataset "data.csv".**

```
import pandas
```

```
df= pandas.read_csv("data.csv")
print(df)
```

```
    Age  Experience  Rank Nationality   Go
0    36          10     9          UK   NO
1    42          12     4         USA   NO
2    23           4     6           N   NO
3    52           4     4         USA   NO
4    43          21     8         USA  YES
5    44          14     5          UK   NO
6    66           3     7           N  YES
7    35          14     9          UK  YES
8    52          13     7           N  YES
9    35           5     9           N  YES
10   24           3     5         USA   NO
11   18           3     7          UK  YES
12   45           9     9          UK  YES
```

2. Mapping categorical values to numerical value:

```
d = {'UK': 0, 'USA':1, 'N':2}
df['Nationality'] = df['Nationality'].map(d)
d= {'YES':1 , 'NO':0}
df['Go'] = df['Go'].map(d)
print(df)
```

```
    Age  Experience  Rank  Nationality  Go
0    36          10     9            0   0
1    42          12     4            1   0
2    23           4     6            2   0
3    52           4     4            1   0
4    43          21     8            1   1
5    44          14     5            0   0
6    66           3     7            2   1
7    35          14     9            0   1
8    52          13     7            2   1
9    35           5     9            2   1
10   24           3     5            1   0
11   18           3     7            0   1
12   45           9     9            0   1
```

3. Assigning one feature to variable y and other features to variable x.

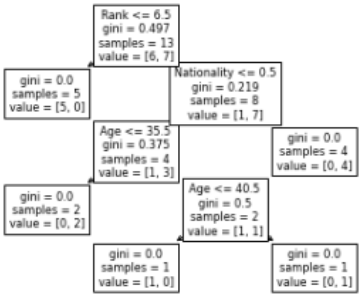
```
features =['Age','Experience', 'Rank','Nationality']

x= df[features]
y= df['Go']

print(x)
print(y)
```

```
    Age  Experience  Rank  Nationality
0    36          10     9            0
1    42          12     4            1
2    23           4     6            2
3    52           4     4            1
4    43          21     8            1
5    44          14     5            0
6    66           3     7            2
7    35          14     9            0
8    52          13     7            2
9    35           5     9            2
10   24           3     5            1
11   18           3     7            0
12   45           9     9            0
0    0
1    0
2    0
3    0
4    1
5    0
6    1
7    1
8    1
9    1
10   0
11   1
12   1
Name: Go, dtype: int64
```

4. Implementing Decision Tree Classifier:

```
import pandas
import numpy as np
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt

df= pandas.read_csv("data.csv")
d = {'UK': 0, 'USA':1, 'N':2}
df['Nationality'] = df['Nationality'].map(d)
d= {'YES':1 , 'NO':0}
df['Go'] = df['Go'].map(d)

features =['Age','Experience', 'Rank','Nationality']

x= df[features]
y= df['Go']

dtree = DecisionTreeClassifier()
dtree = dtree.fit(x,y)

tree.plot_tree(dtree, feature_names= features)
```

```
[Text(133.92000000000002, 195.696, 'Rank <= 6.5\ngini = 0.497\nsamples = 13\nvalue = [6, 7]'),
 Text(66.96000000000001, 152.208, 'gini = 0.0\nsamples = 5\nvalue = [5, 0]'),
 Text(200.88000000000002, 152.208, 'Nationality <= 0.5\ngini = 0.219\nsamples = 8\nvalue = [1, 7]'),
 Text(133.92000000000002, 108.72, 'Age <= 35.5\ngini = 0.375\nsamples = 4\nvalue = [1, 3]'),
 Text(66.96000000000001, 65.232, 'gini = 0.0\nsamples = 2\nvalue = [0, 2]'),
 Text(200.88000000000002, 65.232, 'Age <= 40.5\ngini = 0.5\nsamples = 2\nvalue = [1, 1]'),
 Text(133.92000000000002, 21.744, 'gini = 0.0\nsamples = 1\nvalue = [1, 0]'),
 Text(267.84000000000003, 21.744, 'gini = 0.0\nsamples = 1\nvalue = [0, 1]'),
 Text(267.84000000000003, 108.72, 'gini = 0.0\nsamples = 4\nvalue = [0, 4]')]
```



5. **Predicting the values.**

```
: print(dtree.predict([[40,10,7,1]]))
[1]

: print(dtree.predict([[40,10,6,1]]))
[0]
```

Learning Outcomes (What I have learnt):

1. I have learnt about implementing Decision Tree classifier on any dataset.

2. I have learnt about assigning few features to one variable and rest to other.

3. I have learnt about various libraries which are supported by python such as sklearn, pandas, matplotlib.

4. I have learnt about the various functions provided by various libraries.

5. I have understood the experiment very well.

Evaluation Grid:

	Parameters	Marks Obtained	Maximum Marks
1.	Student Performance (Conduct of experiment) objectives/Outcomes.		12
2.	Viva Voce		10
3.	Submission of Work Sheet (Record)		8
	Total		30