

DECISION TREE

In [1]:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
import matplotlib.pyplot as plt
%matplotlib inline
```

In [57]:

```
df=pd.read_csv('C:/Users/USER/Desktop/petrol_consumption.csv')

df.head()
df.describe()
```

Out[57]:

	Petrol_tax	Average_income	Paved_Highways	Population_Driver_licence(%)	Petrol_Consumption
count	48.000000	48.000000	48.000000	48.000000	48
mean	7.668333	4241.833333	5565.416667	0.570333	576
std	0.950770	573.623768	3491.507166	0.055470	111
min	5.000000	3063.000000	431.000000	0.451000	344
25%	7.000000	3739.000000	3110.250000	0.529750	509
50%	7.500000	4298.000000	4735.500000	0.564500	568
75%	8.125000	4578.750000	7156.000000	0.595250	632
max	10.000000	5342.000000	17782.000000	0.724000	968

In [58]:

```
y= df['target']
X= df.drop(['target'], axis=1)
```

In [59]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.6, random_state =1)
```

In [60]:

```
clf_entropy=DecisionTreeClassifier(criterion='entropy',random_state=100,max_depth=2,min_sam  
clf_entropy.fit(X_train,y_train)
```

Out[60]:

```
DecisionTreeClassifier(criterion='entropy', max_depth=2, min_samples_leaf=5,  
                      random_state=100)
```

In [61]:

```
y_pred=clf_entropy.predict(X_test)
```

In [62]:

```
ac = (accuracy_score(y_test,y_pred)*100)  
print('Accuarcy score : ', ac)
```

Accuarcy score : 62.06896551724138