

KDDM Lab4

Name : Kunal Sanjay Patil

PRN : 20190802025

In [1]:

```
import numpy as np
import pandas as pd
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
```

In [2]:

```
df=pd.read_csv('Car1.csv')
```

In [3]:

```
df.head()
```

Out[3]:

	Manufacturer	Sales_in_thousands	__year_resale_value	Price_in_thousands	Engine_size	Ho
0	Acura	16.919	16.360	21.50	1.8	
1	Acura	39.384	19.875	28.40	3.2	
2	Acura	14.114	18.225	25.50	3.2	
3	Acura	8.588	29.725	42.00	3.5	
4	Audi	20.397	22.255	23.99	1.8	

In [4]:

```
df.shape
```

Out[4]:

```
(157, 7)
```

In [5]:

```
df.isna().sum()
```

Out[5]:

```
Manufacturer      0
Sales_in_thousands  0
__year_resale_value  0
Price_in_thousands  0
Engine_size       0
Horsepower        2
Wheelbase         0
dtype: int64
```

In [6]:

```
X = df.iloc[:,1:].values
y = df.iloc[:,0].values
```

In [7]:

```
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer = imputer.fit(X[:,1:])
X[:,1:] = imputer.transform(X[:,1:])
```

In [8]:

```
labelencoder_y = LabelEncoder()
y = labelencoder_y.fit_transform(y)
print(y)
```

```
[ 0  0  0  0  1  1  1  2  2  2  3  3  3  3  4  4  4  4  4  5  5  5  5  5
  5  5  5  5  6  6  6  6  6  6  6  7  7  7  7  7  7  7  7  7  7  8  8
  8  8  8  8  8  8  8  8  8  9  9  9  9  9 10 10 10 11 12 13 13 13 14 14
 14 14 14 14 15 15 15 18 18 18 18 18 18 18 17 17 17 17 17 17 16 16 16 16
 16 16 16 16 16 19 19 19 19 19 19 19 19 20 20 20 20 20 20 21 21 21 21 22 22
 22 22 22 22 23 23 23 24 24 25 25 25 25 25 26 26 27 27 27 27 27 27 27 27
 27 28 28 28 28 28 28 29 29 29 29 29 29]
```

In [9]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

In [10]:

```
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.transform(X_test)
```

In []:

