In [33]:
```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,confusion_matrix
from sklearn.preprocessing import LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```
cm_data = pd.read_csv('Company_Data.csv')
cm_data
```

Out[3]:

	Sales	CompPrice	Income	Advertising	Population	Price	ShelveLoc	Age	Education	Urbar
0	9.50	138	73	11	276	120	Bad	42	17	Yes
1	11.22	111	48	16	260	83	Good	65	10	Yes
2	10.06	113	35	10	269	80	Medium	59	12	Yes
3	7.40	117	100	4	466	97	Medium	55	14	Yes
4	4.15	141	64	3	340	128	Bad	38	13	Yes
...	...	...	...	...	...	...	...	...	...	..
395	12.57	138	108	17	203	128	Good	33	14	Yes
396	6.14	139	23	3	37	120	Medium	55	11	No
397	7.41	162	26	12	368	159	Medium	40	18	Yes
398	5.94	100	79	7	284	95	Bad	50	12	Yes
399	9.71	134	37	0	27	120	Good	49	16	Yes

400 rows × 11 columns

In [7]:
```
cm_data.loc[cm_data['Sales'] >=8,'sales'] = 'High'
cm_data.loc[cm_data['Sales'] <8 ,'sales'] = 'Low'
```

In [8]:
```
cm_data.drop('Sales',axis = 1,inplace = True)
```

In [10]:
```
cm_data.shape
```

Out[10]:
```
(400, 11)
```

In [11]:
```
cm_data.isna().sum()
```

Out[11]:
```
CompPrice      0
Income         0
Advertising    0
Population     0
Price          0
ShelveLoc      0
Age            0
Education      0
Urban          0
US             0
sales          0
dtype: int64
```

In [13]:
```
cm_data.dtypes
```

Out[13]:
```
CompPrice       int64
Income          int64
Advertising     int64
Population      int64
Price           int64
ShelveLoc      object
Age             int64
Education       int64
Urban          object
US             object
sales          object
dtype: object
```

In [14]:
```
le = LabelEncoder()
cm_data['ShelveLoc'] = le.fit_transform(cm_data['ShelveLoc'])
cm_data['ShelveLoc'].unique()
```

Out[14]:
```
array([0, 1, 2])
```

In [15]:
```
cm_data['Urban'] = le.fit_transform(cm_data['Urban'])
cm_data['Urban'].unique()
```

Out[15]:
```
array([1, 0])
```

In [16]:
```
cm_data['US'] = le.fit_transform(cm_data['US'])
cm_data['US'].unique()
```

Out[16]:
```
array([1, 0])
```

In [17]:
```
cm_data.head()
```

Out[17]:

	CompPrice	Income	Advertising	Population	Price	ShelveLoc	Age	Education	Urban	US	sa
0	138	73	11	276	120	0	42	17	1	1	H
1	111	48	16	260	83	1	65	10	1	1	H
2	113	35	10	269	80	2	59	12	1	1	H
3	117	100	4	466	97	2	55	14	1	1	L
4	141	64	3	340	128	0	38	13	1	0	L

In [18]:
```
x = cm_data.drop('sales',axis=1)
y = cm_data[['sales']]
```

In [19]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.20,random_state=
```

In [20]:
```
x_train.shape,y_train.shape
```

Out[20]:
```
((320, 10), (320, 1))
```

In [21]:
```
x_test.shape,y_test.shape
```

Out[21]:
```
((80, 10), (80, 1))
```

In [24]:
```
rf = RandomForestClassifier()
```

In [25]:
```
rf.fit(x_train,y_train)
```

Out[25]:
```
RandomForestClassifier()
```

In [26]:
```
y_pred_train = rf.predict(x_train)
```

In [27]:
```
print('Accuracy score :',accuracy_score(y_train,y_pred_train))
```

```
Accuracy score : 1.0
```

In [28]:
```
print('Confusion Metric :\n',confusion_matrix(y_train,y_pred_train))
```

```
Confusion Metric :
 [[133   0]
 [  0 187]]
```

In [29]:
```
y_pred_test = rf.predict(x_test)
```

In [30]:
```
print('Accuracy score :', accuracy_score(y_test,y_pred_test))
```

```
Accuracy score : 0.75
```

In [32]:
```
print('Confusion Metric :\n',confusion_matrix(y_test,y_pred_test))
```

```
Confusion Metric :
 [[16 15]
 [ 5 44]]
```

In [35]:
```
gb = GradientBoostingClassifier()
gb.fit(x_train,y_train)
```

Out[35]:
```
GradientBoostingClassifier()
```

In [36]:
```
y_pred = gb.predict(x_test)
```

In [37]:
```
print('Accuracy score :', accuracy_score(y_test,y_pred))
```

```
Accuracy score : 0.775
```

In [38]:
```
print('Confusion Metric :\n',confusion_matrix(y_test,y_pred))
```

```
Confusion Metric :
 [[20 11]
 [ 7 42]]
```

In [ ]:
```

```