

In [1]:

```
import numpy as nd
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
dataset = pd.read_csv("datasets_1846_3197_Social_Network_Ads.csv")
```

In [3]:

```
dataset
```

Out[3]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	10000	0
1	15810944	Male	35	50000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0
...
395	15691863	Female	46	41000	1
396	15706071	Male	51	23000	1
397	15654296	Female	50	20000	1
398	15755018	Male	36	33000	0
399	15594041	Female	49	36000	1

400 rows × 5 columns

In [4]:

```
x = dataset.iloc[:,[2,3]].values
```

In [5]:

x

Out[5]:

```
array([[ 19, 10000],
       [ 35, 50000],
       [ 26, 43000],
       [ 27, 57000],
       [ 19, 76000],
       [ 27, 58000],
       [ 27, 84000],
       [ 32, 130000],
       [ 25, 33000],
       [ 35, 65000],
       [ 26, 80000],
       [ 26, 52000],
       [ 20, 86000],
       [ 32, 18000],
       [ 18, 82000],
       [ 29, 80000],
       [ 47, 25000],
       [ 45, 26000])
```

In [6]:

```
y = dataset.iloc[:,4].values
```

In [7]:

y

Out[7]:

```
array([0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1,
       0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0,
       1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1,
       1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1,
       1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1,
       0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0,
       1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1,
       0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1,
       1, 1, 0, 1], dtype=int64)
```

In [8]:

```
from sklearn.model_selection import train_test_split
```

In [9]:

```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.10)
```

In [10]:

```
x_train.shape
```

Out[10]:

```
(360, 2)
```

In [11]:

```
y_train.shape
```

Out[11]:

```
(360,)
```

In [12]:

```
from sklearn.preprocessing import StandardScaler
```

In [13]:

```
sc = StandardScaler()
```

In [14]:

```
x_train = sc.fit_transform(x_train)
```

In [15]:

```
x_test = sc.transform(x_test)
```

In [16]:

```
from sklearn.neighbors import KNeighborsClassifier
```

In [17]:

```
classifier = KNeighborsClassifier(n_neighbors=5,metric='minkowski')
```

In [18]:

```
classifier.fit(x_train,y_train)
```

Out[18]:

```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',  
                    metric_params=None, n_jobs=None, n_neighbors=5, p=2,  
                    weights='uniform')
```

In [19]:

```
y_pred = classifier.predict(x_test)
```

In [20]:

```
y_pred
```

Out[20]:

```
array([0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 1, 1,  
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1], dtype=int64)
```

In [21]:

```
y_test
```

Out[21]:

```
array([0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1,  
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1], dtype=int64)
```

In [22]:

```
from sklearn.metrics import confusion_matrix
```

In [23]:

```
cm = confusion_matrix(y_test,y_pred)
```

In [24]:

```
cm
```

Out[24]:

```
array([[22,  4],  
       [ 4, 10]], dtype=int64)
```

In [25]:

```
(22+10)/(22+10+4+4)
```

Out[25]:

```
0.8
```

In []:

