import libraries

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.svm import SVC
from sklearn.metrics import confusion_matrix,accuracy_score
```

```
<frozen importlib._bootstrap>:219: RuntimeWarning: numpy.ufunc size changed, may indicate binary incomp
atibility. Expected 192 from C header, got 216 from PyObject
<frozen importlib._bootstrap>:219: RuntimeWarning: numpy.ufunc size changed, may indicate binary incomp
atibility. Expected 192 from C header, got 216 from PyObject
```

In [3]:

```
dataset = pd.read_csv("social.csv")
dataset.head()
```

Out[3]:

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

```
In [8]: dataset['Gender'].value_counts()
```

```
Out[8]: Female    204
        Male      196
        Name: Gender, dtype: int64
```

```
In [4]: dataset.shape
```

```
Out[4]: (400, 5)
```

```
In [6]: dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   User ID          400 non-null    int64 
 1   Gender           400 non-null    object
 2   Age              400 non-null    int64 
 3   EstimatedSalary  400 non-null    int64 
 4   Purchased        400 non-null    int64 
dtypes: int64(4), object(1)
memory usage: 15.8+ KB
```

In [7]:

```
dataset.describe()
```

Out[7]:

	User ID	Age	EstimatedSalary	Purchased
count	4.000000e+02	400.000000	400.000000	400.000000
mean	1.569154e+07	37.655000	69742.500000	0.357500
std	7.165832e+04	10.482877	34096.960282	0.479864
min	1.556669e+07	18.000000	15000.000000	0.000000
25%	1.562676e+07	29.750000	43000.000000	0.000000
50%	1.569434e+07	37.000000	70000.000000	0.000000
75%	1.575036e+07	46.000000	88000.000000	1.000000
max	1.581524e+07	60.000000	150000.000000	1.000000

In [9]:

```
# seperate out data in feaures, target

X = dataset.iloc[:,[2,3]].values
y = dataset.iloc[:,[4]].values
```

In [10]:

```
# features set shape
X.shape
```

Out[10]:

```
(400, 2)
```

In [11]:

```
# target shape
y.shape
```

Out[11]:

```
(400, 1)
```

In [14]:

```
# split the data in training and testing set

X_train, X_test, y_train,y_test = train_test_split(X,y, test_size = 0.25, random_state = 42)
```

In [15]:

```
print("X_train shape : " , X_train.shape)
print("X_test shape : " , X_test.shape)
print("y_train shape : " , y_train.shape)
print("y_test shape : " , y_test.shape)
```

```
X_train shape :  (300, 2)
X_test shape :  (100, 2)
y_train shape :  (300, 1)
y_test shape :  (100, 1)
```

In [16]:

```
# model
clf = SVC(kernel = 'rbf',random_state = 0)
clf.fit(X_train,y_train)
```

```
/Users/kunalshriwas/opt/anaconda3/lib/python3.8/site-packages/sklearn/utils/validation.py:73: DataConve
rsionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y t
o (n_samples, ), for example using ravel().
  return f(**kwargs)
```

Out[16]:

```
SVC(random_state=0)
```

In [17]:

```
# prediction

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

Out[17]:

```
array([0, 0, 0, 0, 0, 0, 1, 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, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1,
       0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 1])
```

In [19]:

```
# accuracy

accuracy = accuracy_score(y_test,y_pred)
print(accuracy)
```

```
0.75
```

In [20]:

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

```
[[60  3]
 [22 15]]
```

In [22]:

```
# True negative = 60
# False positive = 3
# True positive = 15
# False negative = 22
```

In [23]:

```
# Scaling features and then modeling
```

In [24]:

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train_scaled = sc.fit_transform(X_train)
X_test_scaled = sc.fit_transform(X_test)
```

In [25]:

```
# model
clf1 = SVC(kernel = 'rbf',random_state = 0)
clf1.fit(X_train_scaled,y_train)
```

```
/Users/kunalshriwas/opt/anaconda3/lib/python3.8/site-packages/sklearn/utils/validation.py:73: DataConve
rsionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y t
o (n_samples, ), for example using ravel().
  return f(**kwargs)
```

Out[25]:

```
SVC(random_state=0)
```

In [26]:

```
# prediction

y_pred1 = clf1.predict(X_test_scaled)
y_pred1
```

Out[26]:

```
array([1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0,
       1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1,
       0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1,
       1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1,
       0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1])
```

In [27]:

```
# accuracy after scaling

accuracy1 = accuracy_score(y_test,y_pred1)
print(accuracy1)
```

```
0.95
```

In [28]:

```
cm1 = confusion_matrix(y_test,y_pred1)
print(cm1)
```

```
[[59  4]
 [ 1 36]]
```

In [ ]:

```
```