import numpy as np

import pandas as pd

import os

for dirname, \_, filenames in os.walk('/kaggle/input'):

    for filename in filenames:

        print(os.path.join(dirname, filename))

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import confusion\_matrix

data= pd.read\_csv(r"/content/Churn\_Modelling.csv")

data.head(8)

data.info

df = data.copy()

gender= pd.get\_dummies(df['Gender'],drop\_first=True)

geo=pd.get\_dummies(df['Geography'],drop\_first=True)

df= pd.concat((df,gender,geo),axis=1)

cat\_features = ['Spain','Germany','Male','NumOfProducts','HasCrCard','IsActiveMember']

num\_features = ['Tenure','Balance','EstimatedSalary','Age','CreditScore']

for i in cat\_features:

    ax=sns.countplot(x=df[i], data=df);

    plt.title(i)

    plt.show()

for i in num\_features:

    plt.hist(df[i],   edgecolor='black',linewidth=2)

    plt.title(i)

    plt.show()

df.isnull().sum()

df = df.drop(['Gender','Geography'],axis=1)

Remove = ['RowNumber', 'CustomerId', 'Surname','EstimatedSalary','HasCrCard','CreditScore','Germany']

df = df.drop(Remove,axis=1)

X\_train,X\_test,Y\_train,Y\_test= train\_test\_split(X,Y,test\_size=0.3,random\_state=1234,stratify=Y)

from sklearn.ensemble import RandomForestClassifier

model1 = RandomForestClassifier(random\_state=42)

model = LogisticRegression()

model1.fit(X\_train,Y\_train)

Y\_predict = model1.predict(X\_test)

cm = confusion\_matrix(Y\_test,Y\_predict)

cm

model1.score(X\_test,Y\_test)