**Logistic Regression**

**Import packages:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sea

**Import data :**

data=pd.read\_csv('general\_data.csv')

#check nul

data.isnull().sum

da=data.dropna()

# convert catagorical to numaricl

from sklearn import preprocessing

label\_encoder=preprocessing.LabelEncoder()

da["Attrition"]=label\_encoder.fit\_transform(da["Attrition"])

da["Department"]=label\_encoder.fit\_transform(da["Department"])

da["BusinessTravel"]=label\_encoder.fit\_transform(da["BusinessTravel"])

da["EducationField"]=label\_encoder.fit\_transform(da["EducationField"])

da["Gender"]=label\_encoder.fit\_transform(da["Gender"])

da["JobRole"]=label\_encoder.fit\_transform(da["JobRole"])

da["MaritalStatus"]=label\_encoder.fit\_transform(da["MaritalStatus"])

da["EducationField"]=label\_encoder.fit\_transform(da["EducationField"])

del da['Over18']

**Split x and y:**

y=da.iloc[:,1].values#convert array also

x=da[['Age','BusinessTravel', 'Department', 'DistanceFromHome',

'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',

'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',

'NumCompaniesWorked', 'PercentSalaryHike', 'StandardHours',

'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',

'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager']].values

#split the data set

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

xtrain,xtest,ytrain,ytest=train\_test\_split(x, y,test\_size=0.2,random\_state=2)

**Fitting the logistic regression:**

from sklearn.linear\_model import LogisticRegression

cls=LogisticRegression(random\_state=0)

cls.fit(xtrain,ytrain)

#predict the test model

ypre=cls.predict(xtest)

**Making confusion matrix:**

from sklearn.metrics import confusion\_matrix

cm=confusion\_matrix(ytest,ypre)

