Aim:

To write the Implementation of Logistic Regression

Procedure:

To avoid the outliers, I have used logistic regression to overcome this problem, first I have used isnull function to find all the null values in the data set later I have replaced those values with the mean of each column using fillna function. Then I have plotted a scatter plot where it shows a variation of SibSp and Age column later I have used sklearn to import the x,y training and testing data set.after giving an 0.2 for testing 0.8 is taken for training the data set.after which from sklearn we get the logistic regression model where y\_test and y\_pred is used and an accuracy score is predicted using sklearn again.

Source code:

import pandas as pd

import numpy as np

df=pd.read\_csv("C:/Users/vknsr/Downloads/Titanic.csv")

df.head()

df.isnull().sum()

df.fillna(df["Age"].mean(),inplace=True)

df.isnull().sum()

import matplotlib.pyplot as plt

plt.scatter(df.SibSp,df.Age,color='r',marker="\*")

plt.xlabel("SibSp")

plt.ylabel("Age")

plt.title("Relationship b/w Age and SibSp")

df.describe()

from sklearn.model\_selection

import train\_test\_split

x=df.Age.values.reshape(-1,1)

y=df.SibSp.values.reshape(-1,1)

SEED=42

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=SEED)

from sklearn.linear\_model import LogisticRegression

reg=LogisticRegression()

reg.fit(x\_train,y\_train)

y\_pred=reg.predict(x\_test)

y\_pred

from sklearn.metrics import

accuracy\_score accuracy\_score(y\_test,y\_pred)

Output: 0.6871508379888268

Result: The above result is verified and proved