import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

dataset=pd.read\_csv(‘C:\\Users\\STUDENT\\Desktop\\diabetes.csv’)

dataset.head()

dataset.shape

dataset.isnull().values.any()

dataset.info()

dataset.describe()

dataset.isnull().sum()

sns.countplot(x = ‘Outcome’,data = dataset)

sns.pairplot(data = dataset, hue = ‘Outcome’)

plt.show()

sns.heatmap(dataset.corr(), annot = True)

plt.show()

dataset\_new = dataset

dataset\_new[[“Glucose”, “BloodPressure”, “SkinThickness”, “Insulin”, “BMI”]] = dataset\_new[[“Glucose”, “BloodPressure”, “SkinThickness”, “Insulin”, “BMI”]].replace(0, np.NaN)

dataset\_new.isnull().sum()

dataset\_new[“Glucose”].fillna(dataset\_new[“Glucose”].mean(), inplace = True)

dataset\_new[“BloodPressure”].fillna(dataset\_new[“BloodPressure”].mean(), inplace = True)

dataset\_new[“SkinThickness”].fillna(dataset\_new[“SkinThickness”].mean(), inplace = True)

dataset\_new[“Insulin”].fillna(dataset\_new[“Insulin”].mean(), inplace = True)

dataset\_new[“BMI”].fillna(dataset\_new[“BMI”].mean(), inplace = True)

dataset\_new.isnull().sum()

y = dataset\_new[‘Outcome’]

X = dataset\_new.drop(‘Outcome’, axis=1)

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

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, y, test\_size = 0.20, random\_state = 42, stratify = dataset\_new[‘Outcome’] )

from sklearn.linear\_model import LogisticRegression

Model = LogisticRegression()

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

Y\_predict = model.predict(X\_test)

Y\_predict

From sklearn.metrics import confusion\_matrix

Cm = confusion\_matrix(Y\_test, y\_predict)

Cm

Sns.heatmap(pd.DataFrame(cm), annot=True)

From sklearn.metrics import accuracy\_score

accuracy =accuracy\_score(Y\_test, y\_predict)

accuracy

y\_predict = model.predict([[1,148,72,35,79.799,33.6,0.627,50]])

print(y\_predict)

if y\_predict==1:

print(“Diabetic”)

else:

print(“Non Diabetic”)