import numpy as np

import pandas as pd

from sklearn.preprocessing import StandardScaler

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

from sklearn import svm

from sklearn.metrics import accuracy\_score

#uploading data

diabetes\_dataset = pd.read\_csv(r"C:\Users\HP\Desktop\shreya programs\diabetes.csv")

#printing firt 5 rows

diabetes\_dataset.head()

#number of rows and columns in dataset

diabetes\_dataset.shape

#getting the statistical measures of the data

diabetes\_dataset.describe()

diabetes\_dataset['Outcome'].value\_counts()

diabetes\_dataset.groupby('Outcome').mean()

#separating data and labels

X = diabetes\_dataset.drop(columns = 'Outcome', axis=1)

Y = diabetes\_dataset['Outcome']

print(X)

print(Y)

scaler = StandardScaler()

scaler.fit(X)

standardized\_data = scaler.transform(X)

print(standardized\_data)

X = standardized\_data

Y = diabetes\_dataset['Outcome']

print(X)

print(Y)

#Train test split

X\_train, X\_test,Y\_train, Y\_test = train\_test\_split(X, Y, test\_size = 0.2, stratify=Y, random\_state=2)

print(X.shape, X\_train.shape, X\_test.shape)

#Training the model

classifier = svm.SVC(kernel = 'linear')

#training the support vector machine classifier

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

#accuracy score on the training data

X\_train\_prediction = classifier.predict(X\_train)

training\_data\_accuracy = accuracy\_score(X\_train\_prediction, Y\_train)

print('Accuracy score of the training data: ', training\_data\_accuracy)

#accuracy score on the test data

X\_test\_prediction = classifier.predict(X\_test)

test\_data\_accuracy = accuracy\_score(X\_test\_prediction, Y\_test)

print('Accuracy score of the test data : ', test\_data\_accuracy)

#Making a predictive system

input\_data = (4,110,92,0,0,37.6,0.191,30)

#changing the input\_data to numpy array

input\_data\_as\_numpy\_array = np.asarray(input\_data)

#reshape the array as we are predicting for one instance

input\_data\_reshaped = input\_data\_as\_numpy\_array.reshape(1,-1)

#standardize the input data

std\_data = scaler.transform(input\_data\_reshaped)

print(std\_data)

prediction = classifier.predict(std\_data)

print(prediction)

if (prediction[0] == 0):

print('The person is non diabetic')

else:

print('The person is diabetic')