HEART DISEASE PREDICTION



***Introduction:***

**Heart disease** describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you’re born with (congenital heart defects), among others.

The term “heart disease” is often used interchangeably with the term “cardiovascular disease”. Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart’s muscle, valves or rhythm, also are considered forms of heart disease.

Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions.

This makes heart disease a major concern to be dealt with. But it is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.

Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.

# Why is disease prediction important?

Disease prediction has the potential to benefit stakeholders such as the government and health insurance companies. It can identify patients at risk of disease or health conditions.

# Objectives:

* The main objective of this project is to develop a heart prediction system, the system can discover and extract hidden knowledge associated with diseases from heart dataset
* This system aims to exploit machine learning techniques on medical data set to assist in the prediction of the heart disease.
* Reduce the cost of medical tests.
* To help avoid human biases.

# Problem statement:

Heart disease prediction using machine learning algorithm.

# INPUT:

We are taking dataset as an input. The dataset used in this project is the Cleveland Heart Disease dataset taken from the UCI repository.

The dataset consists of 303 individuals data. There are 12 columns in the dataset, which are described below. We are giving following parameters as an input,

1. ***Age***: displays the age of the individual.
2. ***Sex***: displays the gender of the individual using the following format :

1 = male

* 1. = female

1. ***Chest-pain type***: displays the type of chest-pain experienced by the individual using the following format :
   1. = typical angina
   2. = atypical angina
   3. = non — anginal pain
   4. = asymptotic
2. ***Resting Blood Pressure***: displays the resting blood pressure value of an individual in mmHg (unit)
3. ***Serum Cholestrol***: displays the serum cholesterol in mg/dl (unit)
4. ***Fasting Blood Sugar***: compares the fasting blood sugar value of an individual with 120mg/dl.

If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)

1. ***Resting ECG*** : displays resting electrocardiographic results
   1. = normal
   2. = having ST-T wave abnormality
   3. = left ventricular hyperthrophy
2. ***Max heart rate achieved*** : displays the max heart rate achieved by an individual.
3. ***Exercise induced angina*** :

1 = yes

* 1. = no

1. ***Peak exercise ST segment*** :
   1. = upsloping
   2. = flat
   3. = downsloping
2. ***Thal*** : displays the thalassemia :
   1. = normal
   2. = fixed defect
   3. = reversible defect
3. ***Target***: Displays whether the individual is suffering from heart disease or not :
   1. = no disease

1= disease

Why these parameters:

In the actual dataset, we had 76 features but for our study, we chose only the above 14 because :

1. Age: Age is the most important risk factor in developing cardiovascular or heart diseases, with approximately a tripling of risk with each decade of life. Coronary fatty streaks can begin to form in adolescence. It is estimated that 82 percent of people who die of coronary heart disease are 65 and older. Simultaneously, the risk of stroke doubles every decade after age 55.
2. Sex: Men are at greater risk of heart disease than pre-menopausal women. Once past menopause, it has been argued that a woman’s risk is similar to a man’s although more recent data from the WHO disputes this. If a female has diabetes, she is more likely to develop heart disease than a male with diabetes.
3. Angina (Chest Pain): Angina is chest pain or discomfort caused when your heart muscle doesn’t get enough oxygen-rich blood. It may feel like pressure or squeezing in your chest. The discomfort also can occur in your shoulders, arms, neck, jaw, or back. Angina pain may even feel like indigestion.
4. Resting Blood Pressure: Over time, high blood pressure can damage arteries that feed your heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases your risk even more. Normal blood pressure is between 90/60 mmHg and 120/80mmHg, Low blood pressure is 90/60mmHg and High blood pressure is 140/90mmHg.
5. Serum Cholesterol: A high level of low-density lipoprotein (LDL) cholesterol (the “bad” cholesterol) is most likely to narrow arteries. A high level of triglycerides, a type of blood fat related to your diet, also ups your risk of a heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the “good” cholesterol) lowers your risk of a heart attack.
6. Fasting Blood Sugar: Not producing enough of a hormone secreted by your pancreas (insulin) or not responding to insulin properly causes your body’s blood sugar levels to rise, increasing your risk of a heart attack. Normal fasting blood sugar level is 99mm/Dl.
7. Resting ECG: Resting ECG(electrocariography) is a test that can detect abnormalities including arrtythmias, evidence of coronary heart disease, left ventricular hypertrophy and bundle branch blocks. It measures the electrical activity of the heart.
8. Max heart rate achieved: The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%. A normal resting heart rate for adults ranges from 60 to 100 beats per minute.
9. Exercise induced angina: The pain or discomfort associated with angina usually feels tight, gripping or squeezing, and can vary from mild to severe. Angina is usually felt in the center of your chest but may spread to either or both of your shoulders, or your back, neck, jaw or arm. It can even be felt in your hand.

# OUTPUT:

*Whether the individual is suffering from heart disease or not.* *Here, target = 1 implies that the person is suffering from heart disease and target = 0 implies the person is not suffering.*

# Approach:

In this project I am using the Logistic Regression. **Logistic Regression:**

Logistic Regression is a “Supervised machine learning” algorithm that can be used to model the probability of a certain class or event. It is used when the data is linearly separable and the outcome is binary or dichotomous in nature.That means Logistic regression is usually used for Binary classification problems.**Binary Classification** refers to predictingthe output variable that is discrete in **two** classes.

A few examples of Binary classification are Yes/No, Pass/Fail, Win/Lose, Cancerous/Non-cancerous, etc

STEP-1: Importing the libraries

**Code:** *import numpy as np import pandas as pd*

*from sklearn.model\_selection import train\_test\_split from sklearn.linear\_model import LogisticRegression from sklearn.metrics import accuracy\_score*

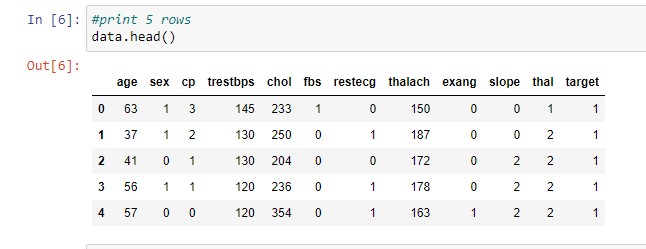
STEP-2: Data collection and preprocessing

**Code:**

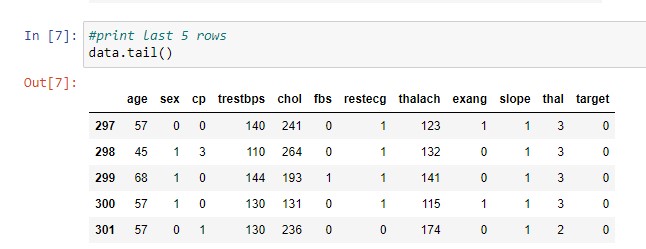
*#loading the csv data to a pandas dataframe data=pd.read\_csv('heart\_disease\_data.csv')*

*#print first 5 rows data.head()*

**Output:**



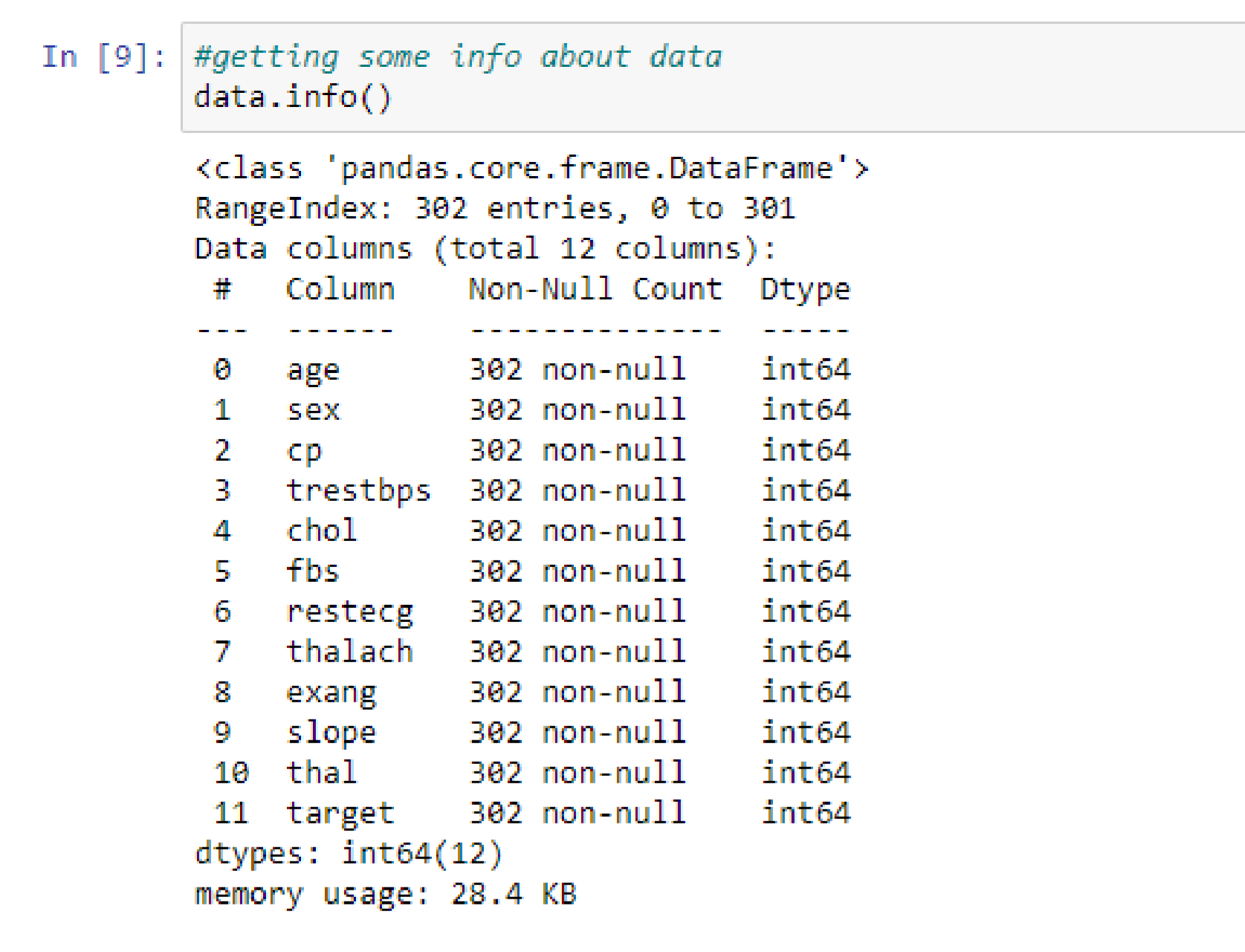
*#print last 5 rows data.tail()*



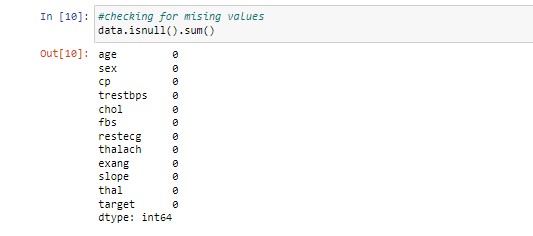
*#no. of rows and colunms in a dataset data.shape*



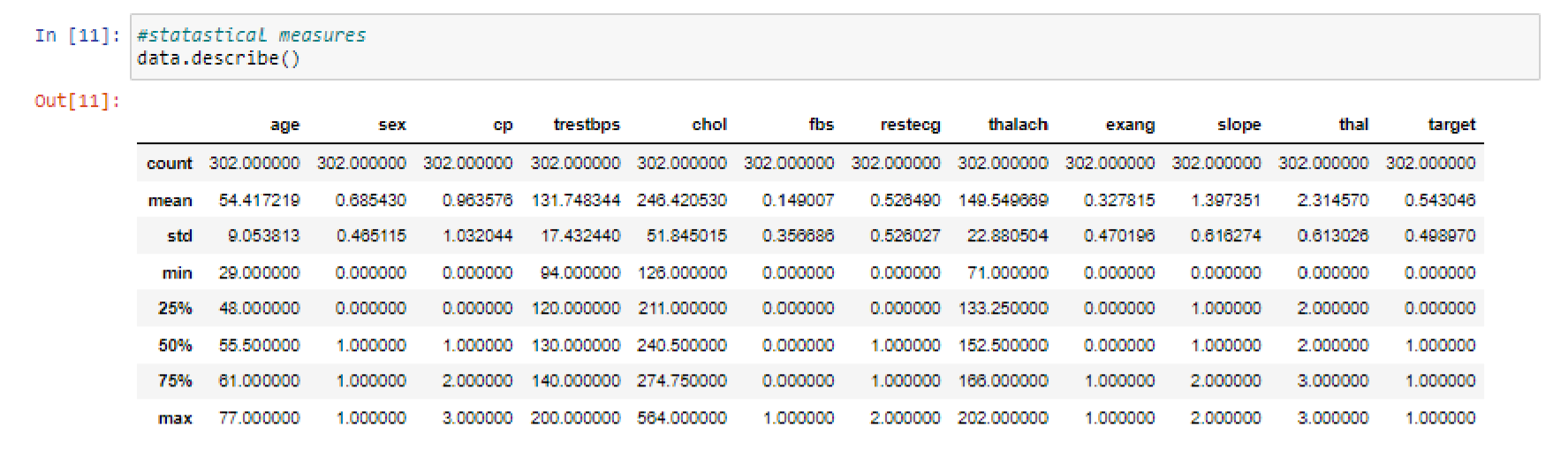
*#getting some info about data data.info()*



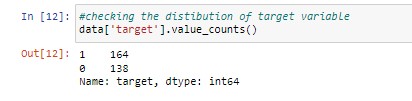
* Checking for missing values *#checking for mising values data.isnull().sum()*



* statastical measures *#statastical measures data.describe()*



* checking the distribution of target variable. Here, **1** represnts *suffering from heart disease,* **0** represents *not suffering.* *#checking the distibution of target variable data['target'].value\_counts()*

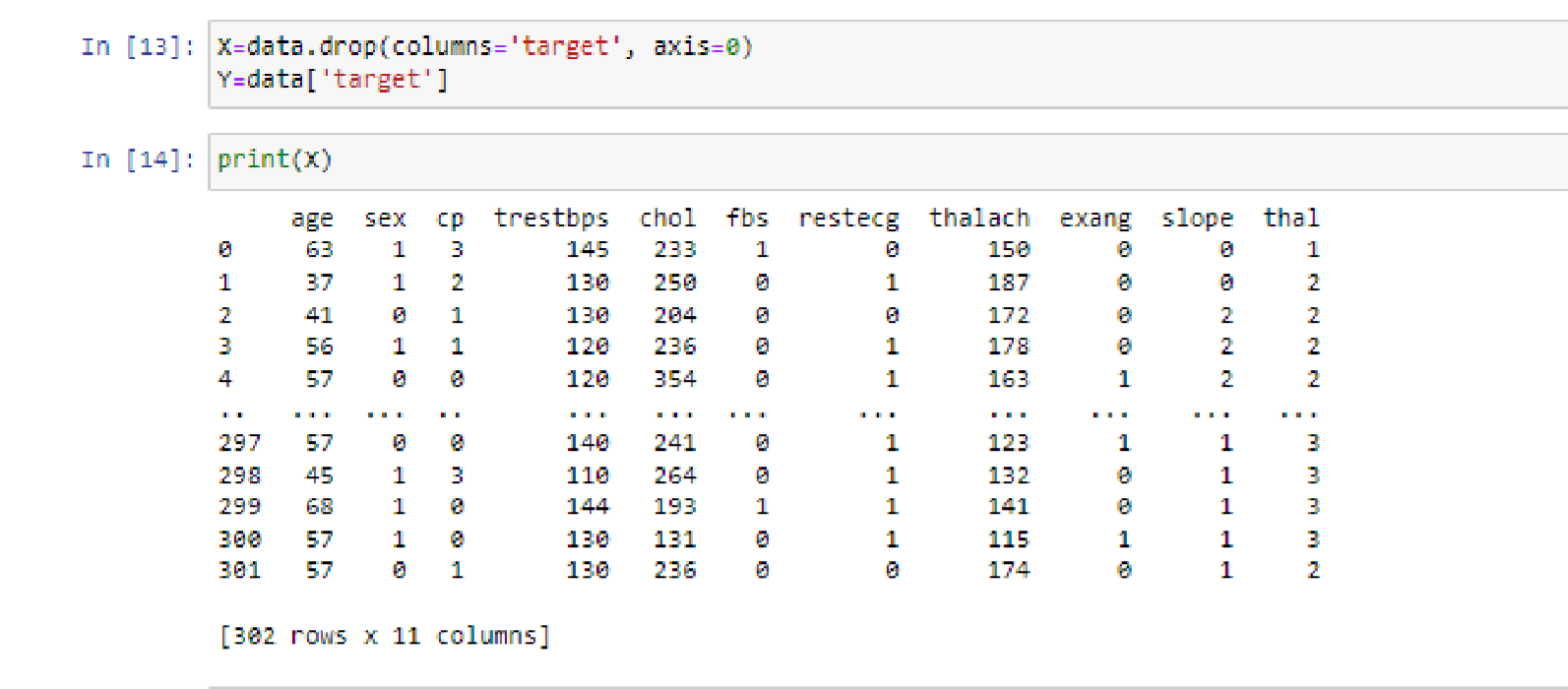


STEP-3: Splitting the features and target

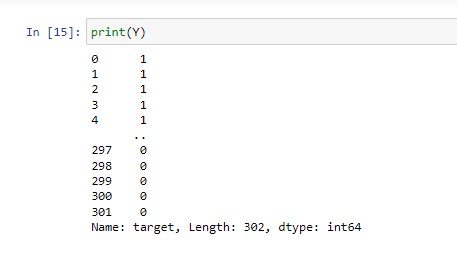
**Code:**

*X=data.drop(columns='target', axis=0) Y=data['target'] print(X)*

**Output:**



*print(Y)*

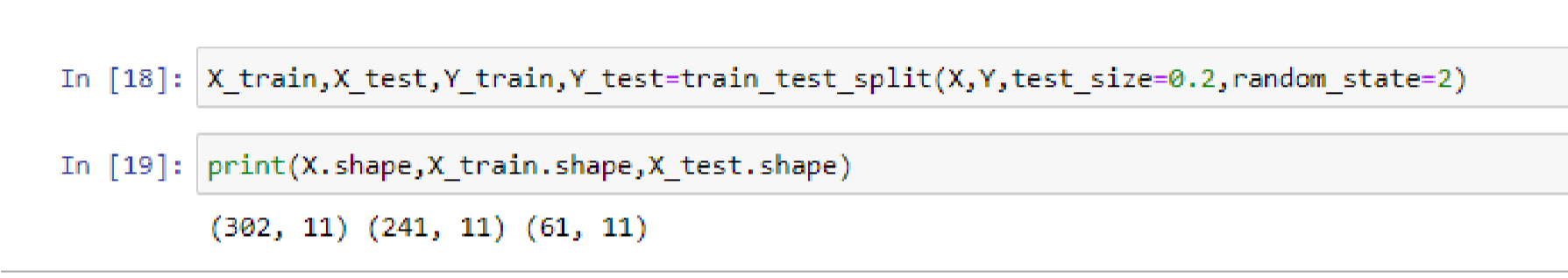


STEP-4: Splitting the data into training and testing data

**Code:**

*X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X,Y,test\_size=0.2,rando m\_state=2)*

*print(X.shape,X\_train.shape,X\_test.shape)* **Output:**



STEP-5: Model Training.

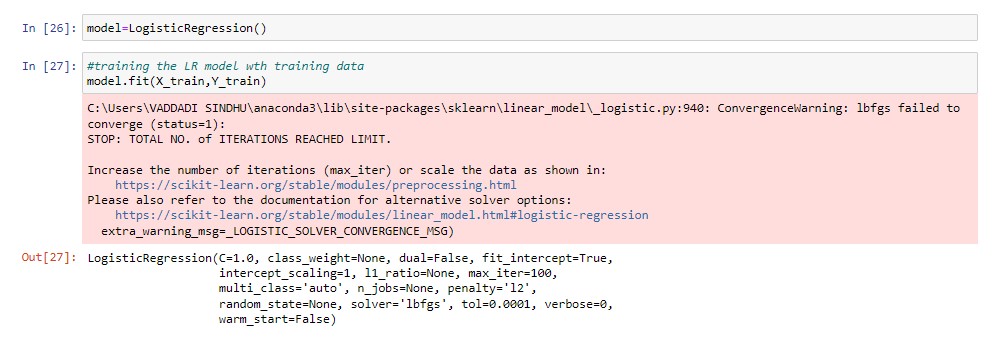
Training the Logistic Regression with the training data.

**Code:**

*model=LogisticRegression()*

*#training the LR model with training data model.fit(X\_train,Y\_train)*

**Output:**



STEP-5: Model Evaluation

* Accuracy score

**Code:**

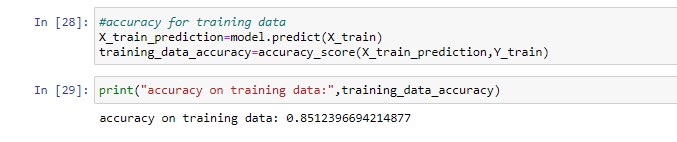
*#accuracy for training data*

*X\_train\_prediction=model.predict(X\_train)*

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

*print("accuracy on training data:",training\_data\_accuracy)*

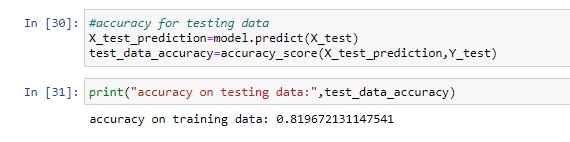
**Output:**



*#accuracy for testing data*

*X\_test\_prediction=model.predict(X\_test)*

*test\_data\_accuracy=accuracy\_score(X\_test\_prediction,Y\_test) print("accuracy on training data:",test\_data\_accuracy)*



STEP-6: Building a predictive system

**Code:**

*print("enter details") a=int(input("enter your age:")) b=int(input("Gender(Male-1,Female-0):")) c=int(input("chest pain type(0=typical angina\n1= atypical angina\n2=non—anginal pain\n3=asymptotic):")) d=int(input("Resting bloodpressure(in mmHg):")) e=int(input("cholestrol(mg/dl ):")) f=int(input("fasting blood sugar>120mg/dl(true-1,false-0):")) g=int(input("resting ecg(normal-0\n1 = having ST-T wave abnormality\n2 = left ventricular hyperthrophy)")) h=int(input("Max heart rate achieved:")) i=int(input("Exercise induced angina(1 = yes,0 = no):")) j=int(input("Peak exercise ST segment(0 = upsloping,1 = flat,2 = downsloping):")) k=int(input("thalassemia(1=normal,2=fixed defect,3=reversible defect):")) prediction=model.predict([[a,b,c,d,e,f,g,h,i,j,k]]) print(prediction)*

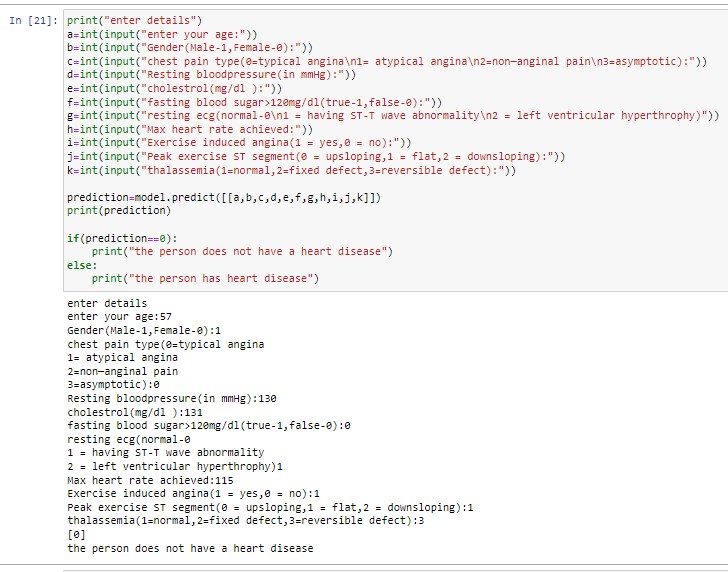
*if(prediction==0):*

*print("the person does not have a heart disease")*

*else:*

*print("the person has heart disease")*

**Output:**



RESOURCES & REFERENCES: sources taken from kaggle.

*Github repository link:* <https://github.com/sindhuvaddadi/DSP_PROJECT.git>

PROJECT DONE BY

**V.SINDHU**

**S180151**

**CSE 2A**

**s180151@rguktsklm.ac.in**