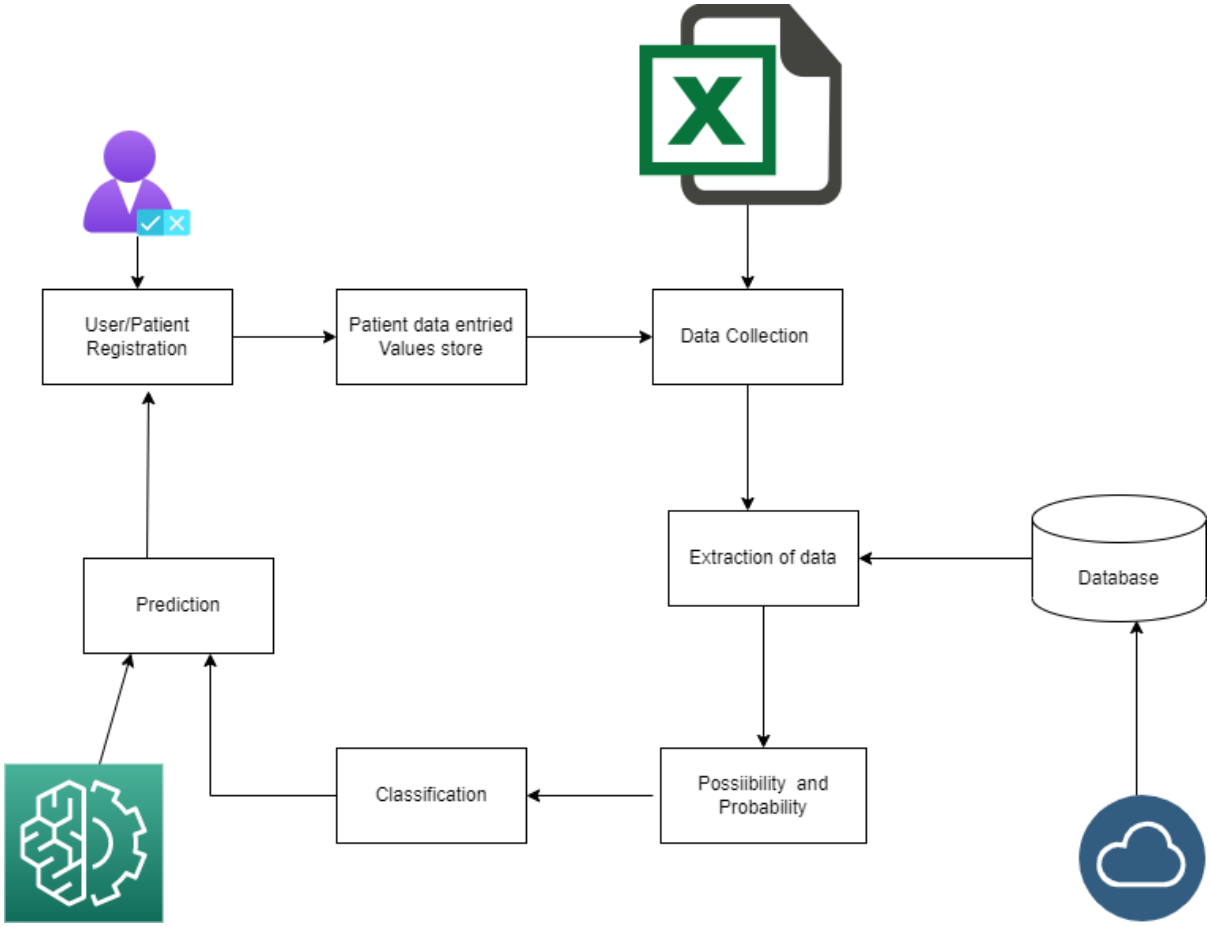
Project Design Phase-I
Solution Architecture

Date	14 November 2022
Team ID	PNT2022TMID21090
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dash Board
Maximum Marks	4 Marks



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.

The leading cause of death in the developed world is heart disease. Therefore, there needs to be work done to help prevent the risks of having a heart attack or stroke. This dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given. 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. The dataset consists of 303 individuals 'data. There are14 columns in the dataset, which are described below.

Parameters:

1.Age

2.Sex

3.Chest pain type

4.BP

5.Cholesterol

6.FBS over 120

7.EKG results

8.Max HR

9.Exercise angina

10.ST depression

11.Slope of ST

12.Number of vessels Fluro

13.Thallium

14.heart disease

- Age: displays the age of the individual.
- Sex: displays the gender of the individual using the following format:
- 1=Male
- 0=Female
- Chest-pain type: displays the type of chest-pain experienced by the individual.

- Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)
- Serum Cholesterol: displays the serum cholesterol in mg/dl (unit)
- 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)
- Resting ECG: displays resting electro cardio graphic results
- 0 = normal
- 1 = having ST- wave abnormality
- 2 = left ventricular hypertrophy
- Max heart rate achieved: displays the max heart rate achieved by an individual.
- Exercise induced angina:
- 1 = yes
- 0 = no
- ST depression induced by exercise relative to rest: displays the value which is an integer or float.

The Approach

The code is implemented in Python and different classification models are applied.

In this article I will be using the following classification models for classification:

- • SVM
- • Naive Bayes
- • Logistic Regression
- • Decision Tree
- • Random Forest
- • LightGBM
- • XGboost

Data Analysis

Let us look at the people's age who are suffering from the disease or not.

Here, target = 1 implies that the person is suffering from heart disease

and target = 0 implies the person is not suffering.