PROJECT DOCUMENTATION

Visualizing and Prediction Heart Diseases with an Interactive Dashboard

Team ID – PNT2022TMID21205

Content

Chapter 1

Introduction

1.1 PROJECT OVERVIEW

The terms "heart disease" and "cardiovascular disease" are frequently used interchangeably. Heart disease is a general term that covers a wide range of heart related medical conditions. The irregular health state that directly affects the heart and all of its components is characterized by these medical conditions. In order to forecast cardiac disease, this study discusses various data mining, big data, and machine learning techniques. Building an important model for the medical system to forecast heart disease or cardiovascular illness requires the use of data mining and machine learning. Our application helps the user in finding out if they have heart disease or not. They can find out by entering details such as their heart rate, cholesterol, blood pressure etc. A dashboard is also attached along with the results for better understanding where they can compare their blood pressure and similar metrics with other users. This project focuses on Random Forest Classifier. The accuracy of our project is 82.67 % for which is better than most other systems in terms of achieving accuracy quickly.

1.2 PURPOSE

This project's goal is to determine, depending on the patient's medical characteristics—such as gender, age, chest pain, fasting blood sugar level, etc.—whether they are likely to be diagnosed with any cardiovascular heart illnesses. The leading cause of death in the developed world is heart disease. Heart disease cases are rising quickly every day, thus it's crucial and worrisome to predict any potential illnesses in advance. This diagnosis is a challenging task that requires accuracy and efficiency. Therefore, there needs to be work done to help prevent the risks of having a heart attack or stroke. It is the main factor in adult deaths. By using a person's medical history, our initiative can identify those who are most likely to be diagnosed with a cardiac condition. It can assist in identifying disease with less medical tests and effective therapies, so that patients can be treated appropriately. It can identify anyone who is experiencing any heart disease symptoms, such as chest pain or high blood pressure. Around the world, machine learning is applied in many different fields. There is no exception in the healthcare sector. Machine learning may be crucial in determining whether locomotor disorders, heart illnesses, and other conditions are present or absent. If foreseen well in advance, such information can offer valuable insights to doctors, who can then customize their diagnosis and course of care for each patient.

Chapter 2
Literature Survey

2.1 EXISTING PROBLEM

A quiet significant amount of works related to the diagnosis of heart disease using Machine Learning algorithms have been made. An efficient heart disease prediction has been made by using various algorithms some of them include Logistic Regression, KNN, Random Forest Classifier etc. It can be seen in results that each algorithm has its strength to register the defined objectives. The model incorporating IHDPS had the ability to calculate the decision boundary using the previous and new model of machine learning and deep learning. It facilitated the important and the most basic factors/knowledge such as family history connected with any heart disease. But the accuracy that was obtained in such IHDPS model was far more less than the new upcoming model such as detecting coronary heart disease using artificial neural network and other algorithms of machine and deep learning.

2.2 REFERENCES

Sno.	Paper title	Author name	Publication year	Results
1.	Cardiovascular Disease Prediction Based on Machine Learning Technology	Oleh Voloshynskyi, Victoria Vysotska, Myroslava Bublyk	2021	The purpose of this article is the development of an information system for cardiovascular disease prediction based on machine learning

				technology. The accuracy of such methods as logistic regression, random forest, support vector machine and artificial neural network is more than 85%, which is also a good result. Techniques such as k-nearest neighbours and decision tree classifier, which showed 68.65% and 78.69% accuracy, respectively, coped with the worst task.
2.	Prediction of Coronary Heart Disease using Machine Learning: An Experimental Analysis	Gurpreet Singh,Rami Mustafa A. Mohammad, Fadi Thabtah, Amanda H. Gonsalves	2019	The field of medical analysis is often referred to be a valuable source of rich information. Coronary Heart Disease (CHD) is one of the major causes of death all around the world therefore early detection of CHD can help reduce these rates. The challenge lies in the complexity of the data and correlations when

				it comes to prediction using conventional techniques. The aim of this research is to use the historical medical data to predict CHD using Machine Learning (ML) technology. The scope of this research is limited to using three supervised learning techniques namely Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT), to discover correlations in CHD data that might help improving the prediction rate. Using the South African Heart Disease dataset of 462 instances, intelligent models are derived by the considered ML techniques using 10-fold cross validation. Empirical results using different performance evaluation

				measures report that probabilistic models derived by NB are promising in detecting CHD.
3.	CARDIOVASCULAR DISEASE PREDICTION USING CLASSIFIER ALGORITHM	Ankur Sharma , Neha Arora	2020	Big data proposes vast promises for detecting interactions and nonlinearities in relationships among variables. Mobile devices, such as smartphones and tablets, and sensors, will continue to be the most indispensable tools available to deliver heart attack prediction and telecardiology services over wireless networks to reduce cardiovascular disease morbidity and mortality. The most important factor, however, in the development and application of big data, telecardiology, sensor use, mobile phone or tablet use and landline use is patient privacy and to safeguard

				the patient's ability to direct and discover the use of his or her health care information. Machine learning is getting advanced which is helpful in making correct choices and taking best decision. Machine learning has proved to be a best tool in making prediction in the healthcare sector. The study has shown that machine learning algorithms are better at predicting the absolute number of cardiovascular disease cases correctly. Through the study, we have found that awareness towards such kind of disease is important. So, correct prediction towards these will help the people and the doctors to take precautions as required. CVD is a type of disease that must be

				controlled to decrease the death ratio worldwide. In addition, given the magnitude of increase in the prevalence of heart disease cases worldwide, it is important to look at different ways of prevention. The data clearly shows that individual based risk factors evaluation is insufficient; And that social determinants of health must be looked at.
4.	A Hybrid Machine Learning Approach for Prediction of Heart Diseases	Sanchayita Dhar Krishna Roy Tanusree Dey Pritha Datta Ankur Biswas	2019	Heart diseases are the chief cause of death all over the world over the last few decades. To avoid heart disease or coronary illness and discover indications early, individuals over 55 years must have a total cardiovascular checkup. Researchers and specialists developed various intelligent techniques to improve capacity

				of the health care professionals in recognition of cardiovascular disease. In cardiovascular disease finding and treatment, single data mining strategies are giving the reasonable precision and accuracy. Nevertheless the usage data mining procedure be capable of reducing the number of test that is required to be carried out. In order to decrease the Figure of deaths from heart diseases there has to be a quick and efficient detection technique providing better accuracy and precision. The aim of this paper is to present an efficient technique of predicting heart diseases using machine learning approaches. Hence we proposed a hybrid approach for heart

				prediction using Random forest classifier and simple k-means algorithm machine learning techniques. The dataset is also evaluated using two other different machine learning algorithms, namely, J48 tree classifier and Naive Bayes classifier and results are compared. Results attained through Random forest classifier and the corresponding confusion matrix shows robustness of the methodology.
5.	GUI based Prediction of Heart Stroke Stages by finding the accuracy using Machine Learning algorithm.	Yash Prakash Kadtan, Aditya Pratap Singh Chauhan, R. Brindha	2021	Many predictive techniques are used and applied in the medical domain such as predicting occurrence, evaluating outcome of diseases and assisting clinicians to recommend treatment of diseases. Standard predictive models

				or methods, on the other hand, are incapable of simulating the complexities of feature representation in medical problem domains, and therefore are ineffective in capturing the underlying information. To address this problem, machine learning algorithms are used to apply predictive computational techniques for heart stroke on a given hospital dataset. Atrial fibrillation is a significant risk factor for cardiac attack in patients, and it shares many of the same factors that predict stroke. When a dataset is analysed using a controlled machine learning algorithm, variables such as variable recognition, univariate analysis, bivariate

				and multivariate analysis, missed value therapies, mathematical methods, and so on are all recorded. The aim of the predictive analytics model is to recognise the various stages of heart stroke in patients. Discuss the output of the provided hospital dataset, as well as the evaluation of the classification study and the uncertainty matrix. To compare supervised classification machine learning algorithms and suggest a machine learning-based method for reliably predicting heart stroke using given characteristics. Furthermore, compare and discuss the performance of different machine learning algorithms from the given healthcare department dataset

				with evaluation classification report, define the confusion matrix, and categorise data from priority, and the result depicts that the effectiveness of a graphical user interface based proposed machine learning algorithm technique can be compared with best accuracy with precision and F1 Score

2.3 PROBLEM STATEMENT DEFINITION

As we discussed earlier about the problem that is occurring now can be broadly be taken into different perspective from Health care sector side, and for patients' self-care itself. People are dying from various heart diseases due to unawareness of their heart issues or they just ignore of being cautious with their issues. From other side, doctor have to check all the old reports and records of patients and diagnose them for further proceedings with the treatment. Predicting exact diagnosis is difficult for even an experienced doctor. So, the manually predicting is very difficult task so it must be done either other sources or using technical knowledge. Perhaps, data analytics and AI domain will help with the problem to provide useful solution to the society.

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

How might we predict if someone has a heart disease?

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

Aneerudh M

- Use Machine Learning Algorithms
- MERN stack
- Sending results to the user's email
- Provide a list of renowned cardiologists
- Use Tableau Dashboard
- Provide worldwide and weekly mortality reports

Mahesh Aravind V

- Create interactive dashboard using Python
- Account creation to save results for later use
- FAQs section
- Use Django/ Flask
- Provide risk factors for cardiac arrest
- List treatment options

Arun M

- Find more datasets
- Table to show normal range of HR, BP and Cholesterol
- Perform data exploration for better understanding
- MEAN stack
- List reputed scan centres
- Latest news regarding cardiology

Prithika S

- Java Stack
- Provide tips on leading a healthy lifestyle
- Provide in-depth analysis of results
- Provide information on warning signs of Cardiac Arrest
- List heart disease statistics
- List the causes of the various types of heart diseases

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

Tech Stack

- Java Stack
- MEAN stack
- Use Django/ Flask
- MERN stack

Website Sections

- Provide information on warning signs of Cardiac Arrest
- Provide tips on how to lead a healthy lifestyle
- Table to show normal range of HR, BP and Cholesterol
- Provide a list of renowned cardiologists
- FAQs section
- Provide risk factors for cardiac arrest
- List the causes of the various types of heart diseases
- List heart disease statistics
- List treatment options
- List reputed scan centres
- Latest news regarding cardiology
- Provide worldwide and weekly mortality reports

Data Analysis

- Use Machine Learning Algorithms
- Find more datasets
- Perform data exploration for better understanding
- Use Tableau Dashboard

Website Features

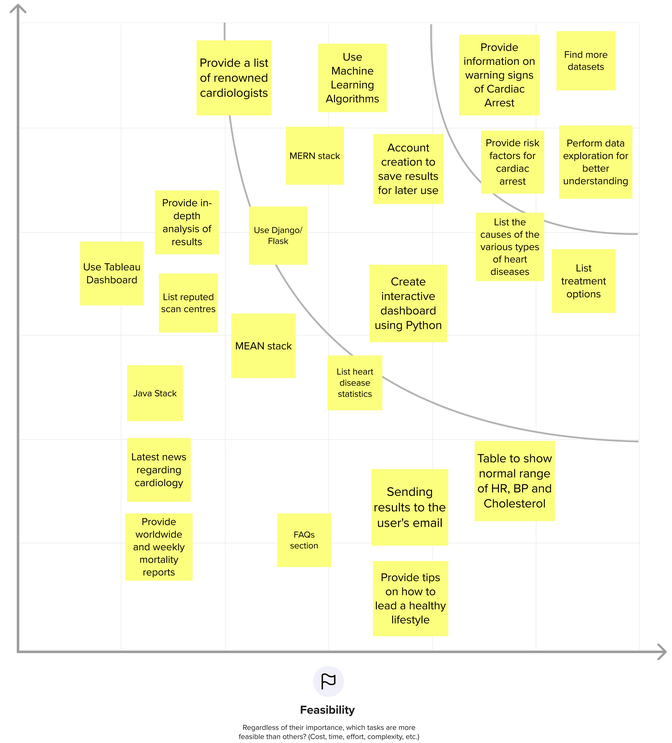
- Provide in-depth analysis of results
- Create interactive dashboard using Python
- Account creation to save results for later use
- Sending results to the user's email

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



Project Design Phase-I
Proposed Solution Template

Date	31 October 2022
Team ID	PNT2022TMID21205
Project Name	Project - Visualizing and Predicting Heart Disease with an Interactive Dashboard
Maximum Marks	2 Marks

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Predicting the heart disease of a patient by analyzing past or Historical records. Where patient can get cure or take necessary treatment before the disease is affected.
2.	Idea / Solution description	Our idea is to predict the heart disease at the beginning stage and provide treatment for speedy recovery.
3.	Novelty / Uniqueness	Comparing other models the prediction will vary. But our model will predict accurately and give effective results.
4.	Social Impact / Customer Satisfaction	By this project people can able to diagnose the heart disease at initial stage by themselves.
5.	Business Model (Revenue Model)	By subscription technique, one user will be allowed to predict the disease.
6.	Scalability of the Solution	In future, some more health associated prediction will be added with the same interactive dashboard.

PROJECT DESIGN PHASE-I
PROBLEM – SOLUTION FIT

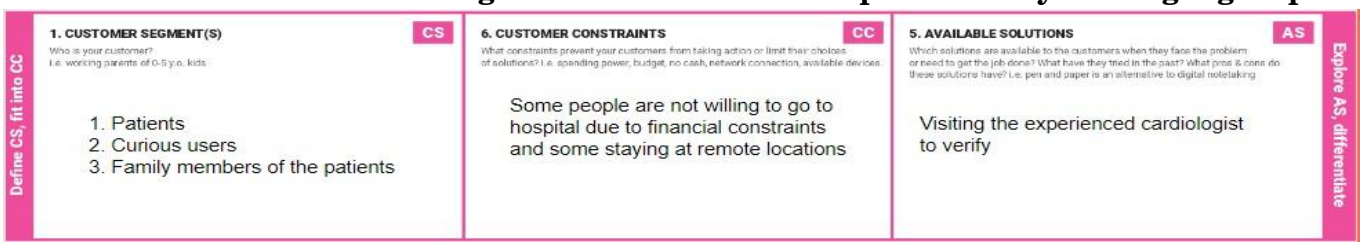
Date	31 October 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dashboard
Maximum Marks	2 Marks

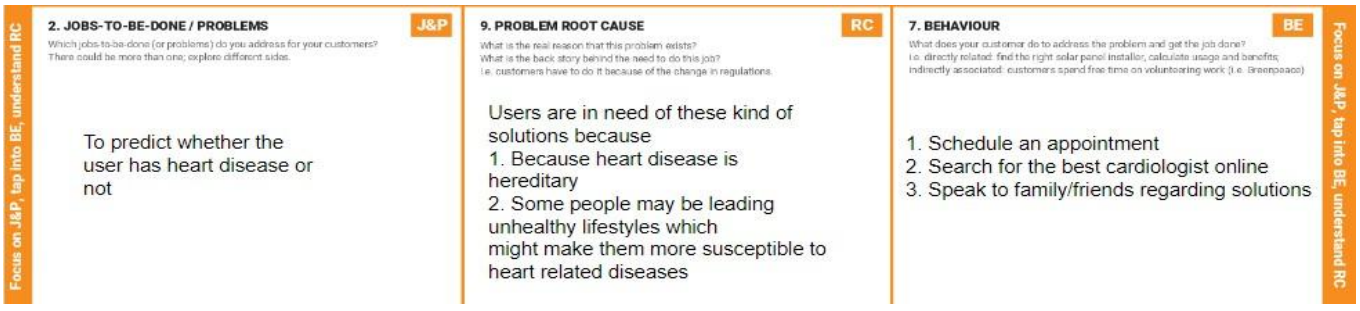
Problem – Solution Fit Template:

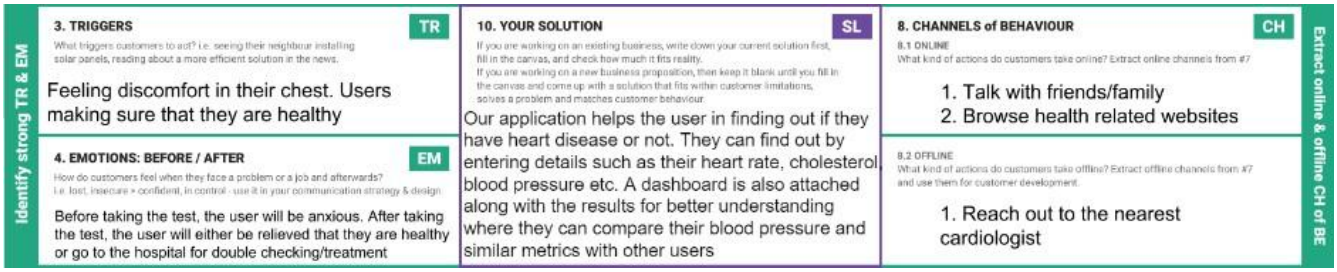
The Problem-Solution Fit simply means that we have found a problem with our customer and that the solution we have realized for it actually solves the customer's problem. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns and recognize what would work and why.

Purpose:

Solve complex problems in a way that fits the state of your customers.
Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behavior.
Sharpen your communication and marketing strategy with the right triggers and messaging.
Increase touch-points with your company by finding the right problem-behavior fit andbuilding trust by solving frequent annoyances, or urgent or costly problems.
Understand the existing situation in order to improve it for your target group.







Project Design Phase-II
Solution Requirements (Functional & Non-functional)

Date	13 November 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dashboard
Maximum Marks	4 Marks

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration**	Enables the user to make registration for using the application.
FR-2	**User Confirmation**	During registration, the user will get confirmation mail for authentication purpose.
FR-3	**Visualizing Data**	User can visualize the records on heart disease through the Dashboard created using IBM Cognos Analytics.
FR-4	**Generating Report**	User can view the health report and can come to an conclusion.

Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

NFR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user-friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	**Security**	Security of the application should be higher as it handles user data. For this database replication technique should be used as the important data can be kept safe. So that in case of crash, the system can be able to backup and recover the data.
NFR-3	**Reliability**	The application has to be reliable and consistent at every situation and has to run without failure.

NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the calculation on data. The calculation time of the application depends on the efficiency of algorithm used.
NFR-5	**Availability**	The application should to be available 24 x 7 for users without any kind of interruption.
NFR-6	**Scalability**	The application can withstand in increase of no. of users and has to be able to upgrade to higher versions.

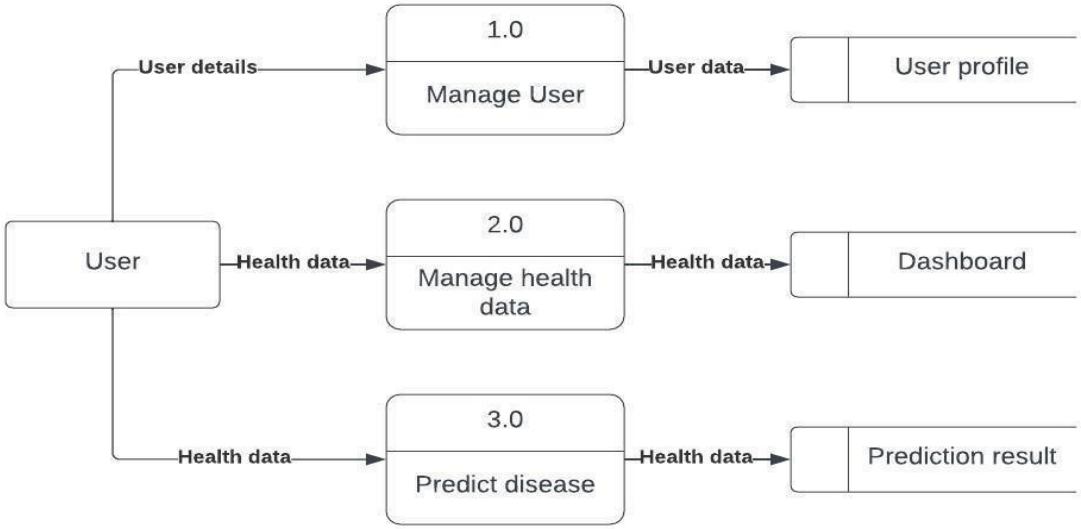
Project Design Phase-II
Data Flow Diagram

Date	13 November 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dashboard
Maximum Marks	4 Marks

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Data Flow Diagram for Heart Disease Prediction Dashboard:



Flow:

1) User creates an account in the application.
2) User enters the medical records in the dashboard.
3) User can view the visualizations of trends in the form of graphs and charts for his/her medical records with the trained dataset.
4) User can view the accuracy of probability of occurrence of heart disease in the dashboard.

User Stories:

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / Dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email & password	I can access my account / Dashboard when logged in	High	Sprint-1
Customer (Web user)	Dashboard	USN-4	User can view his/her complete medical analysis and accuracy of disease prediction	I can view my medical analysis in the dashboard	High	Sprint-2
		USN-5	User can view the accuracy of occurrence of heart disease	I can view the accuracy of heart disease in the dashboard	High	Sprint-2
Customer Care Executive	Helpdesk	USN-6	As a customer care executive, he/she can view the customer queries.	I can post my queries in the dashboard	Medium	Sprint-3
		USN-7	As a customer care executive, he/she can answer the customer queries.	I can get support from helpdesk	High	Sprint-3
Administrator	User Profile	USN-8	As an admin, he/she can update the health details of users.	I can view my updated health details.	High	Sprint-4

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
		USN-9	As an admin, he/she can add or delete users.	I can access my account / Dashboard whenlogged in	High	Sprint-4
		USN-10	As an admin, he/she can manage the user details.	I can view the organized data of myself.	High	Sprint-4

Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	13 November 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and Predicting Heart Diseases with an Interactive Dash Board
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Technology Architecture for Heart Diseases prediction

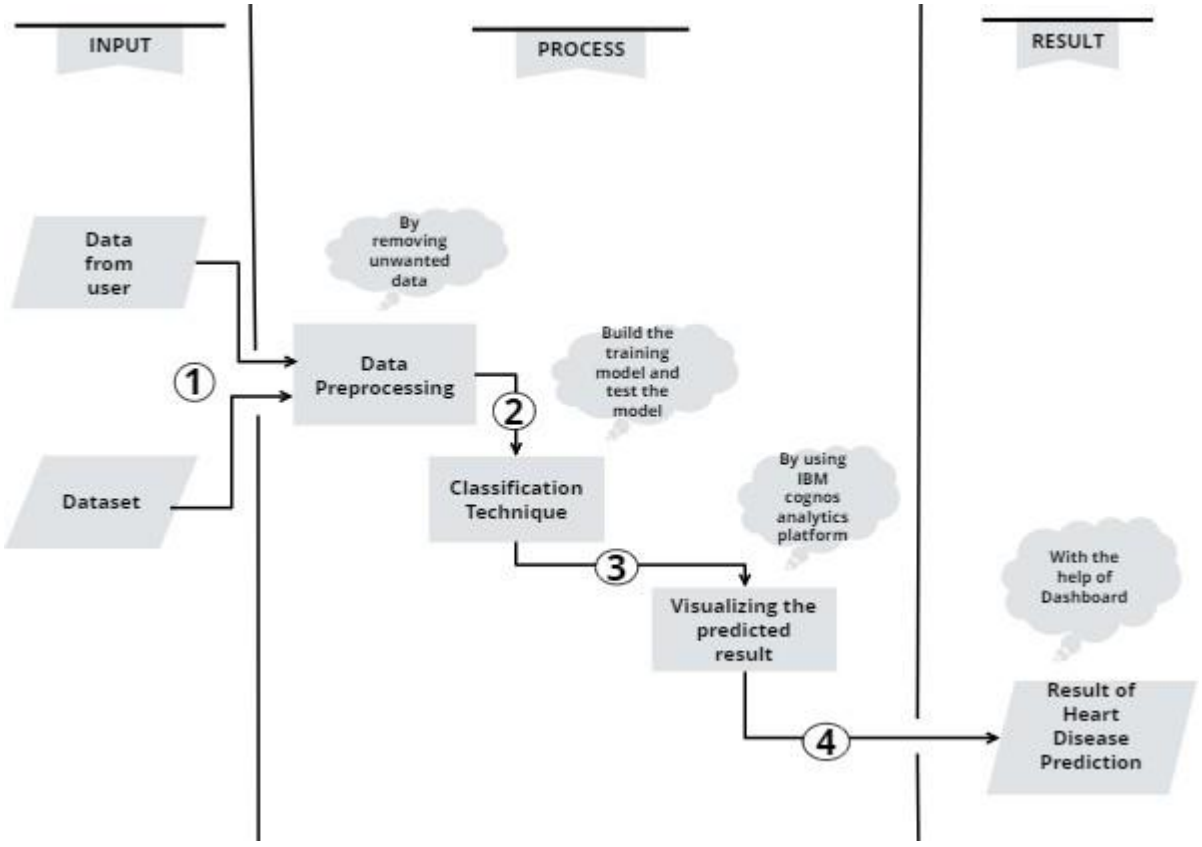


Table-1: Component & Technologies:

S.No	Component	Description	Technologies
1	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript, React JS
2	Application Logic	Logic for a process in the application	Python/Nodejs
3	Database	Data Type, Configurations etc.	MySQL/PostgreSQL
4	Cloud Database	Database Service on Cloud	IBM DB2
5	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
6	Machine Learning Model	To train and test the model for prediction	k-means, Decision Tree, Naïve Bayes
7	Infrastructure (server/cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration	Local, Cloud Foundry

Table-2: Application Characteristics:

S. No	Characteristics	Descriptions	Technology
1	Open-Source Framework	List the open-source frameworks used	IBM Cognos Analytics
2	Security Implementation	List all the security / access controls implemented, use of firewalls etc.	Security provided by cloud
3	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	IBM Cognos Analytics
4	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	IBM Cloud Service
5	Performance	Design consideration for the performance of the application	Reliable Data Classification Models

Project Planning Phase

Project Milestone and Activity list

Date	12 November 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and predicting heart disease with an Interactive dashboard
Maximum Marks	4 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-I	Registration	USN-I	As a user, I can register for the application by entering my email, password, and confirming my password.	8	High	Aneerudh M Arun M
		USN-2	As a user, I will receive confirmation email once I have registered for the application	8	High	Mahesh Aravind V Prithika
	Login	USN-3	As a user, I can log into the application by entering email & password	4	Medium	Arun M
Sprint-2	Working with the dataset	USN-4	To work on the dataset, understand and load the dataset	10	High	Arun M Aneerudh M
		USN-5	Exploration of BP vs chest pain type and gender, maximum heart rate during the chest pain	5	High	Mahesh Aravind Prithika
		USN-6	BP by age, Cholesterol by agent gender	5	High	Aneerudh Prithika

Sprint-3	Data Visualization	USN-7	Visualization of average age for chest pain types, average exercise angina during chest pain	2	Medium	Arun M
		USN-8	BP variation with respect to age, Effect of existing heart disease on average of Exercise Angina	6	High	Mahesh Aravind Aneerudh Prithika
		USN-9	Average age for different types of chest pain in existing heart disease, serum cholesterol levels vs age	6	High	Arun Aneerudh Mahesh Prithika
		USN-10	Maximum heart rate in Existing heart disease by Exercise Angina	6	High	Prithika Arun
Sprint-4	Dashboard Creation	USN-11	Dashboard showing different types of visualization	20	High	Aneerudh M Mahesh Aravind V Prithika Arun M

Project Planning Phase

Project Sprint Delivery Plan

Date	12 November 2022
Team ID	PNT2022TMID21205
Project Name	Visualizing and predicting heart disease with an Interactive dashboard
Maximum Marks	4 Marks

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date	Sprint Release Date (Actual)
Sprint-I	20	1 Days	13 Nov 22	13 Nov 22	20	13 Nov 22
Sprint-2	20	1.5 Days	14 Nov 22	15 Nov 22	20	15 Nov 22
Sprint-3	20	1.5 Days	15 Nov 22	16 Nov 2022	20	16 Nov 2022
Sprint-4	20	2 Days	17 Nov 2022	18 Nov 2022	20	18 Nov 2022

Velocity:

Imagine we have a average 1.5-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

AV= Sprint duration/Velocity=20/1.5=13.33

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.