VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

IBM-Project- 12833-1659495500
TEAM ID: PNT2022TMID00437

Submitted by

Divyasindhu P -312319205036

Desika V-312319205028

Joan Reshmi L-312319205057

Dhivyadharshini S R-312319205033

Anjhana M -312319205010

In partial fulfilment for the award of

the degree of

BACHELOR OF TECHNOLOGY in
INFORMATION TECHNOLOGY

ST. JOSEPH'S COLLEGE OF ENGINEERING

CHENNAI – 600 119

Content

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

1. Introduction

1.1 Project Overview

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 project aims to create an interactive Dashboard using IBM Cognos Tool and dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

1.2 Purpose

Heart disease (HD) is a major cause of mortality in modern society. Medical diagnosis is an extremely important but complicated task that should be performed accurately and efficiently. Cardiovascular disease is difficult to detect due to several risk factors, including high blood pressure, cholesterol, and an abnormal pulse rate. Based on the analytics we can analyze which patients are most likely to suffer from heart disease in the near future and based on the patient details we will make decisions to cure them.

2. Literature Survey

2.1 Existing Problem

Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analysing data that excludes inferences and statistical modelling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Data analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analysing data..

2.2 References

1. Multi class Prediction of Heart Disease Patients Using Big Data Analytics - Sarita Mishra, Manjusha Pandey, Siddharth Swarup Rautaray & Sabyasachi Chakraborty - 2022

The rapidly growing rate of illness and death is the result of many diseases. Another major factor is cardiovascular disease (CVD) due to heart failure. According to statistics from around

the world, the highest rate of natural death is caused by heart problems. The number of deaths resulting from this can be controlled by the early detection of heart disease chances in a person. Big data and several machine learning technologies have made it possible to discover the chances of a cardiac issue in a person in much advance. Many data scientists have successfully exploited the big data available for heart disease patients and have developed prediction models using different algorithms that are non-invasive, accurate, and appear to be very effective in analyzing patients' characteristics and detecting the presence or absence of heart disease in them. However, to provide appropriate preventive measures and appropriate treatment to patients, it is not enough to detect the presence of CVD, but the degree of impact the disease has left on a person needs to be measured. In this paper, we have compared the performance of five different machine-based algorithms (Logistic Regression, Support Vector Machine, Random Forest, KNN, and Naïve Bayes) which are used to classify the cardiovascular disease into five different classes. 0–4) with the increasing value from 0. These algorithms are used in their most common ways and in the One-vs-All method with the best performance in the latest scenario. The results of this study showed that the KNN algorithm provided 99.56% best predictive accuracy with a combination of One-vs-all and Principal Component Analysis strategies that surpassed all other algorithms.

The main idea behind this is the model gives different accuracy levels for different K values and the best K value can be identified using an analyzing error rate or accuracy rate of the model. However, the size and the features of the dataset makes a big impact on the model to come up with a good accuracy rate. Nevertheless, the KNN algorithm does not work well with large datasets and does not work well with high dimensional data because it is hard to calculate the distance between each data point which is a drawback of the algorithm.

2. A Systematic Framework for Heart Disease Prediction Using Big Data Analytics-T Poongodi, R Indrakumari, S Janarthanan, P Suresh-Internet of Things, Artificial Intelligence and Blockchain Technology- 2021

Big data in deep insight derives from heterogeneous, longitudinal, complex, voluminous, and noisy data. The significant challenges in big data lie in searching, capturing, storing, analyzing, and sharing the data. Big data analytics is emerging as a promising technology in harnessing a massive amount of data which surpasses the processing capability of conventional systems. Big data is commonly characterized by volume, velocity, value, veracity, variety, and variability. With the big data, progression in healthcare communities leads to precise analysis of physiological or digitized clinical data benefits prior to detection of diseases, patient care, and healthcare community services. Heart disease is considered as the most life-threatening and deadliest disease that humans face across the world. The vital task in heart disease management is in processing extremely larger datasets and inferring knowledge to predict, prevent, and treat such chronic diseases. The idea is to identify the potential of big data analytics in predicting heart diseases and providing appropriate medicines and treatment for the heart patients. In the healthcare industry, the knowledge that is inferred can be utilized for predicting heart diseases in the early stages. Big data is generated from the user-generated content, mobile transactions, Internet clicks, social media, and genomics data especially created through corporate transactions or sensor networks. Moreover, the advances in using genomic data assist in sharing clinical data, drug discovery, EHR processing,

patient registries, telemedicine, etc. K-means clustering algorithm is followed, and the data is visualized using Tableau Software. The diagnosis clinical parameters for heart disease prediction are age, gender, weight, chest pain, resting BP, resting ECG, cholesterol, etc. The chapter highlights the significant role of big data analytics predominantly in the healthcare industry for predicting heart diseases. Some of the challenges for implementing big data analytics in healthcare are discussed. The future directions in associating big data in healthcare for predicting heart diseases and personalizing medicine are also being investigated.

Due to lack of effective data governance procedures, capturing data is one of the biggest obstacles for healthcare organizations. To use data more efficient, it must be clean, precise, correctly formatted so that it can be used across various healthcare systems.

Most patient records are kept for fast and easy access in a centralized database these days, but the real problem lies when this information that needs to be shared with outside healthcare professionals. For most healthcare providers, data security is one of the top issues with constant hacking and security violations that need to be handled on a continuous basis.When dealing with highly sensitive data and even patient data, which is important, the healthcare industry must be very cautious. Not only can leakage of details prove costly to healthcare companies, but it is also unethical to disclose it without prior authorization.

3. Heart disease prediction using machine learning algorithms- Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain and Preeti Nagrath-2021

Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not using the medical history of the patient. We used different algorithms of machine learning such as logistic regression and KNN to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparison to the previously used classifier such as naive bayes etc. So a quiet significant amount of pressure has been lift off by using the given model in finding the probability of the classifier to correctly and accurately identify the heart disease. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format.

Logistic Regression is a statistical analysis model that attempts to predict precise probabilistic outcomes based on independent features. On high dimensional datasets, this may lead to the model being over-fit on the training set, which means overstating the accuracy of predictions on the training set and thus the model may not be able to predict accurate results on the test set. This usually happens in the case when the model is trained on little training data with lots of features. So on high dimensional datasets, Regularization techniques should

be considered to avoid over-fitting (but this makes the model complex). Very high regularization factors may even lead to the model being under-fit on the training data. Non linear problems can't be solved with logistic regression since it has a linear decision surface. Linearly separable data is rarely found in real world scenarios. So the transformation of non linear features is required which can be done by increasing the number of features such that the data becomes linearly separable in higher dimensions.

4. Accurate prediction of heart disease based on bio system using regressive learning based neural network classifier -A. Sheryl Oliver, Kavithaa Ganesan, S. A. Yuvaraj, T. Jayasankar, Mohamed Yacin Sikkandar, N. B. Prakash – 2021

Heart disease diagnosis is a very hard task in the medical field, so the mortality rate is increased every day. Also, the diagnosing process is implemented in recent times to predict heart disease. The method of diagnosing a disease in the medical field can be regarded not only as a new unknown situation to obtain clinical data and data collected from clinical experience, but also as a decision-making process as well as a doctor's diagnosis. The detection of heart abnormalities mainly depends on the examination of the ECG signal at the appropriate sampling period. The data is trained and tested must include more data to get the data as features. These properties are an accurate measure of the diagnosis of heart disease. The conventional system is having some problems like processing time is high, and it gives low accuracy, so the proposed Regressive Learning-Based Neural Network Classifier (RLNNC) system is implemented. The proposed system RLNNC presents a fully automated algorithm for the classification of heart disease, based on the Regressive Learning-Based Neural Network Classifier (RLNNC) and automated initial seed detection. With the advancement of machine learning and information technology, the development of an automated system. This can be predicted the same on this basis for patients with heart disease, and the drug occurs for the benefit of detecting and analyzing the heart disease. Analysis has shown that the proposed Regressive Learning-Based Neural Network Classifier (RLNNC) based techniques promote greater efficiency and higher accuracy than traditional methods.

Artificial neural networks require processors with parallel processing power, in accordance with their structure. For this reason, the realization of the equipment is dependent. Unexplained behavior of the network: This is the most important problem of ANN. When ANN produces a probing solution, it does not give a clue as to why and how. This reduces trust in the network. There is no specific rule for determining the structure of artificial neural networks. Appropriate network structure is achieved through experience and trial and error. ANNs can work with numerical information. Problems have to be translated into numerical values before being introduced to ANN. The display mechanism to be determined here will directly influence the performance of the network . This depends on the user's ability. The network is reduced to a certain value of the error on the sample means that the training has been completed. This value does not give us optimum results.

5. Artificial Intelligence-Based Ensemble Model for Rapid Prediction of Heart Disease - Navya Harika, Sita Rama Swamy & Nilima- 2021

Heart disease is the leading cause of mortality among men and women. Accurate and rapid diagnosis of heart disease will assist in saving many lives. To develop a novel ensemble framework based on heterogeneous classifiers namely support vector machine (SVM), Naïve Bayes (NB), and artificial neural networks (ANN) for rapid prediction of heart disease. The present study also verifies the most accurate algorithm among all three. Data are collected from the UCI machine learning repository. After pre-processing, the data were divided into training and test data in a ratio of 80:20. Using the training data, the three contributing algorithms were trained by providing heart disease status. The algorithms were tested with the unseen data instances and hence evaluated for accuracy. The ensemble technique uses the results from individual classifiers and yields a result based on majority voting method. The ensemble model was observed to predict heart disease with an accuracy of 87.05% followed by ANN (84.74%), NB (81.35%) and SVM (79.66%). Among the individual classifiers, ANN had the least miss-classification rate and performed best in terms of all other model diagnostics. The use of the proposed ensemble classifier is recommended to predict the heart condition to have better accuracy and least miss-classification.

Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases. This algorithm faces the 'zero-frequency problem' where it assigns zero probability to a categorical variable whose category in the test data set wasn't available in the training dataset. It would be best if we used a smoothing technique to overcome this issue. Its estimations can be wrong in some cases, so you shouldn't take its probability outputs very seriously.

6. Heart disease prediction using machine learning techniques: A systematic review- Kiranjit Kaur , Munish Saini-2020

The key task within the healthcare field is usually the diagnosis of the disease. In case, a disease is actually diagnosed at earlier stage, then many lives might be rescued. Machine learning classification techniques can considerably help the healthcare field just by offering a precise and easy diagnosis of various diseases. Consequently, saving time both formed ical professionals and patients. As heart disease is usually the most recognized killer in the present day, it might be one of the most challenging diseases to diagnose. In this paper, we provide a survey of the various machine learning classification techniques that have been proposed to assist the healthcare professionals in diagnosing the cardiovascular disease. We started by giving the overview of various machine learning techniques along with describing brief definitions of the most commonly used classification techniques to diagnose heart disease. Then, we review representable research works on employing machine learning classification techniques in this field. Furthermore, a detailed comparison table of the surveyed papers is actually presented.

The main drawback is that some datasets will have few to no values in the other classes, with the majority of the data values falling into only one or two classifications. The main drawback of the quantile classification process is that features assigned to the same class might have drastically different values, especially if the data are not distributed uniformly over the class's range. Natural breaks approach might have the drawback of producing classes with wildly disparate number ranges. Another drawback is that because the class ranges are so unique

to each dataset, it might be challenging to compare two or more maps made using the natural breaks classification approach.

7. Prediction of heart disease and classifiers' sensitivity analysis-Khaled Mohamad Almustafa-02 July 2020

Heart disease (HD) is one of the most common diseases nowadays, and an early diagnosis of such a disease is a crucial task for many health care providers to prevent their patients for such a disease and to save lives. In this paper, a comparative analysis of different classifiers was performed for the classification of the Heart Disease dataset in order to correctly classify and or predict HD cases with minimal attributes. The set contains 76 attributes including the class attribute, for 1025 patients collected from Cleveland, Hungary, Switzerland, and Long Beach, but in this paper, only a subset of 14 attributes are used, and each attribute has a given set value. The algorithms used K- Nearest Neighbor (K-NN), Naive Bayes, Decision tree J48, JRip, SVM, Ada-boost, Stochastic Gradient Decent (SGD) and Decision Table (DT) classifiers to show the performance of the selected classifications algorithms to best classify, and or predict, the HD cases.

8. A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment - Abderrahmane Ed-daoudy – 2019

A number of technologies enabled by Internet of Thing (IoT) have been used for the prevention of various chronic diseases, continuous and real-time tracking system is a particularly important one. Wearable medical devices with sensor, health cloud and mobile applications have continuously generating a huge amount of data which is often called as streaming big data. Due to the higher speed of the data generation, it is difficult to collect, process and analyze such massive data in real-time in order to perform real-time actions in case of emergencies and extracting hidden value. using traditional methods which are limited and time-consuming. Therefore, there is a significant need to real-time big data stream processing to ensure an effective and scalable solution. In order to overcome this issue, this work proposes a new architecture for real-time health status prediction and analytics system using big data technologies. The system focus on applying distributed machine learning model on streaming health data events ingested to Spark streaming through Kafka topics. Firstly, we transform the standard decision tree (DT) (C4.5) algorithm into a parallel, distributed, scalable and fast DT using Spark instead of Hadoop MapReduce which becomes limited for real-time computing. Secondly, this model is applied to streaming data coming from distributed sources of various diseases to predict health status. Based on several input attributes, the system predicts health status, send an alert message to care providers and store the details in a distributed database to perform health data analytics and stream reporting. We measure the performance of Spark DT against traditional machine learning tools including Weka. Finally, performance evaluation parameters such as throughput and execution time are calculated to show the effectiveness of the proposed architecture. The experimental results show that the proposed system is able to effectively process and predict real-time and massive amount of medical data enabled by IoT from distributed and various diseases.

9. Prediction of Coronary Heart Disease using Machine Learning: An Experimental Analysis - Amanda H. Gonsalves – July 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.

10. Predictive System: Comparison of Classification Techniques for Effective Prediction of Heart Disease - Debjani Panda & Satya Ranjan Dash– 27 September 2019

Today's world is challenging to most of its people with major concerns for keeping up a good health. Among these challenges, one of the most haunting ones is heart disease. Worldwide, the maximum number of deaths is related to heart diseases. Most of the affected people suffering from heart-related diseases are unaware of their health conditions, and cases are reported at a very later stage, which becomes challenging for doctors to advise them proper treatment and medication with lifestyle changes. This research work aims in comparing classification techniques in finding out which is the most efficient one to predict the disease in less time. Mining important factors and analyzing the relativity between them help in predicting if the patient is having heart disease. The classification techniques used are SVM, Decision Tree, Naïve Bayes, KNN, Random Forest, Ensemble Classification (Extra Trees) and Logistic Regression.

10. Learning Classifier Methods in Prediction of Heart Disease -D Parvathinathan - 2018

Cardiovascular Disease (CVD) refers to any condition involving narrow or blocked blood vessels, which can lead to a heart attack, chest pain or stroke. CVD is a chronic illness and the leading cause of death for both men and women. CVD can attack a person instantly resulting in high healthcare costs and, in some cases, can result in death. A serious and important challenge facing medical practitioners is the ability to accurately diagnose patients with CVD early on. In recent years, medical practitioners have sought the help of computer scientists in order to apply advanced data mining techniques, which can facilitate decision support and help accurately diagnose CVD soon. In this thesis, three data mining techniques are evaluated for their accuracy in predicting CVD. The techniques implemented and analyzed are Logistical Regression, Naive Bayes and Artificial Neural Networks using Multi-Layer Perceptron (MLP).

Results show that Logistic regression predicted the presence of CVD with an accuracy of 88.6%, while Naive Bayes predicted CVD with an accuracy rate of 83% and Neural Networks with an accuracy of 80%.

In this paper, the risk factors that causes heart disease is considered and predicted using K-means algorithm and the analysis is carried out using a publicly available data for heart disease. The dataset holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain. To predict the heart disease, Kmeans clustering algorithm is used along with data analytics and visualization tool. The paper discusses the pre-processing methods, classifier performances and evaluation metrics. In the result section, the visualized data shows that the prediction is accurate.

2.3 Problem Statement Definition

Who does the problem affect?

People with unhealthy lifestyles, stress, depression, age above 40 and when their ancestors got heart disease (since heart disease is hereditary).

When does the issue occur?

The issue occurs for people with unhealthy lifestyles and age above 40. Where is the issue occurring? The issue is originating from an unhealthy lifestyle. It mostly occurs in the blood valves of the heart.

What would happen if we didn't solve the problem?

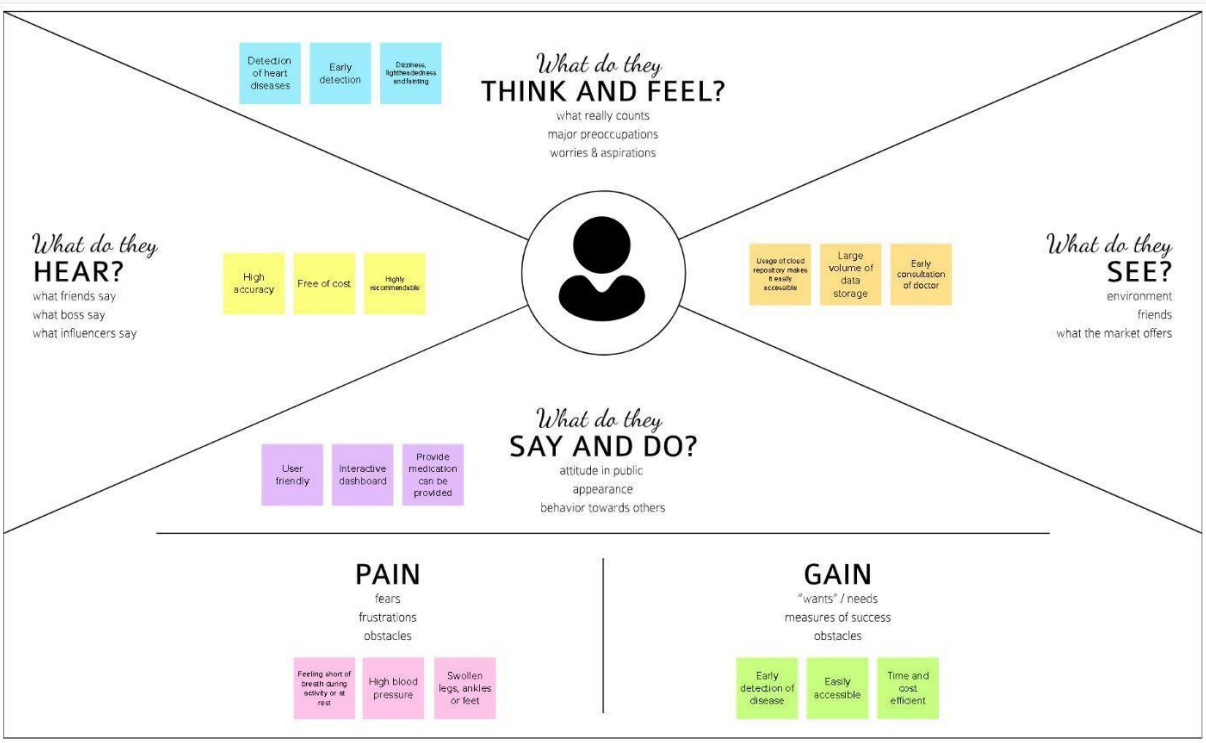
If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will increase rapidly.

Why is it important to fix the problem?

We should predict the problem before giving treatment to the patients. As the problem is predicted early, we can solve it easily and early.

3. Ideation and Proposed Solution

3.1 Empathy Map Canvas



3.2 Ideation and Brainstorming

Step-1: Team Gathering, Collaboration and Select the Problem Statement

Step-2: Brainstorm, Idea Listing and Grouping

Divyasindhu

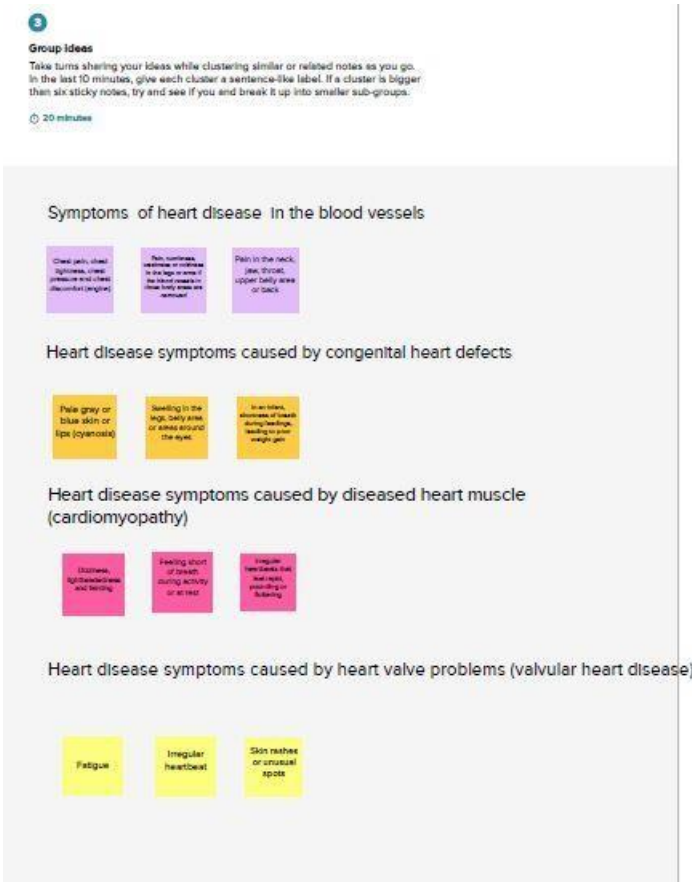
Joan Reshmi

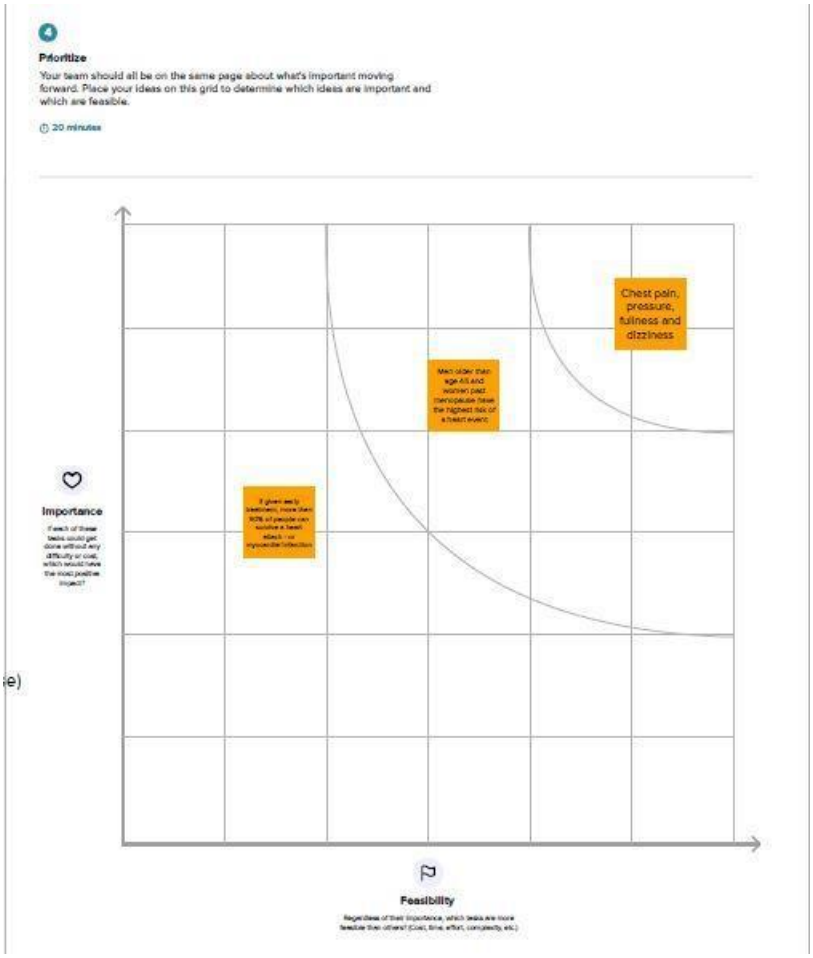
Desika

Dhivyadharshini

Anjhana

Step-3: Idea Prioritization



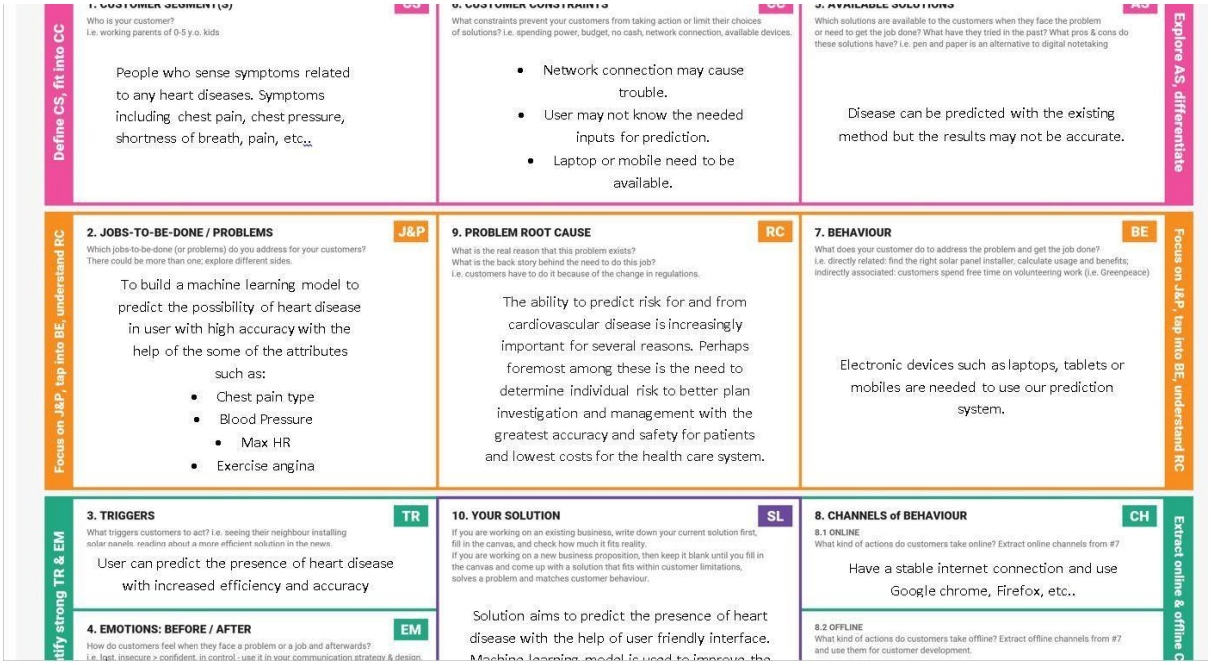


3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Early and Automatic prediction of Heart diseases and visualising the data using interactive dash board.
2.	Idea / Solution description	The IBM Cognos Analytic tool was used, which provides a predictive data visualization and analysis service that can be used to determine patterns, relationships, associations and meaning of a large set of data quickly and in a timely manner. Likewise, prediction is done with the help of Random Forest Classifier.
3.	Novelty / Uniqueness	After evaluating the results from the existing methodologies, we have used python and pandas operations to perform heart disease classification for the data obtained from the UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the

		predictive analytics. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, classification of modelling performance evaluation. Random forest technique is used to improve the accuracy of the result. Several reduction methods may also be used to improve the random forest classification algorithm's accuracy.
4.	Social Impact / Customer Satisfaction	People can predict the heart diseases at a very early stage and improve the quality of living. They can take proper precautions and lead a healthy and safe life.
5.	Business Model (Revenue Model)	It is cost efficient as it is a Software as a Service Platform. People need not spend much money to detect the disease.
6.	Scalability of the Solution	Better execution in accuracy, sensitivity, and specificity as well as in system design flexibility.

3.4. Problem Solution Fit



Requirement Analysis

4.1 Functional Requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generation Report	User can view his/her health report and can make decisions accordingly

4.2 Non-Functional Requirement

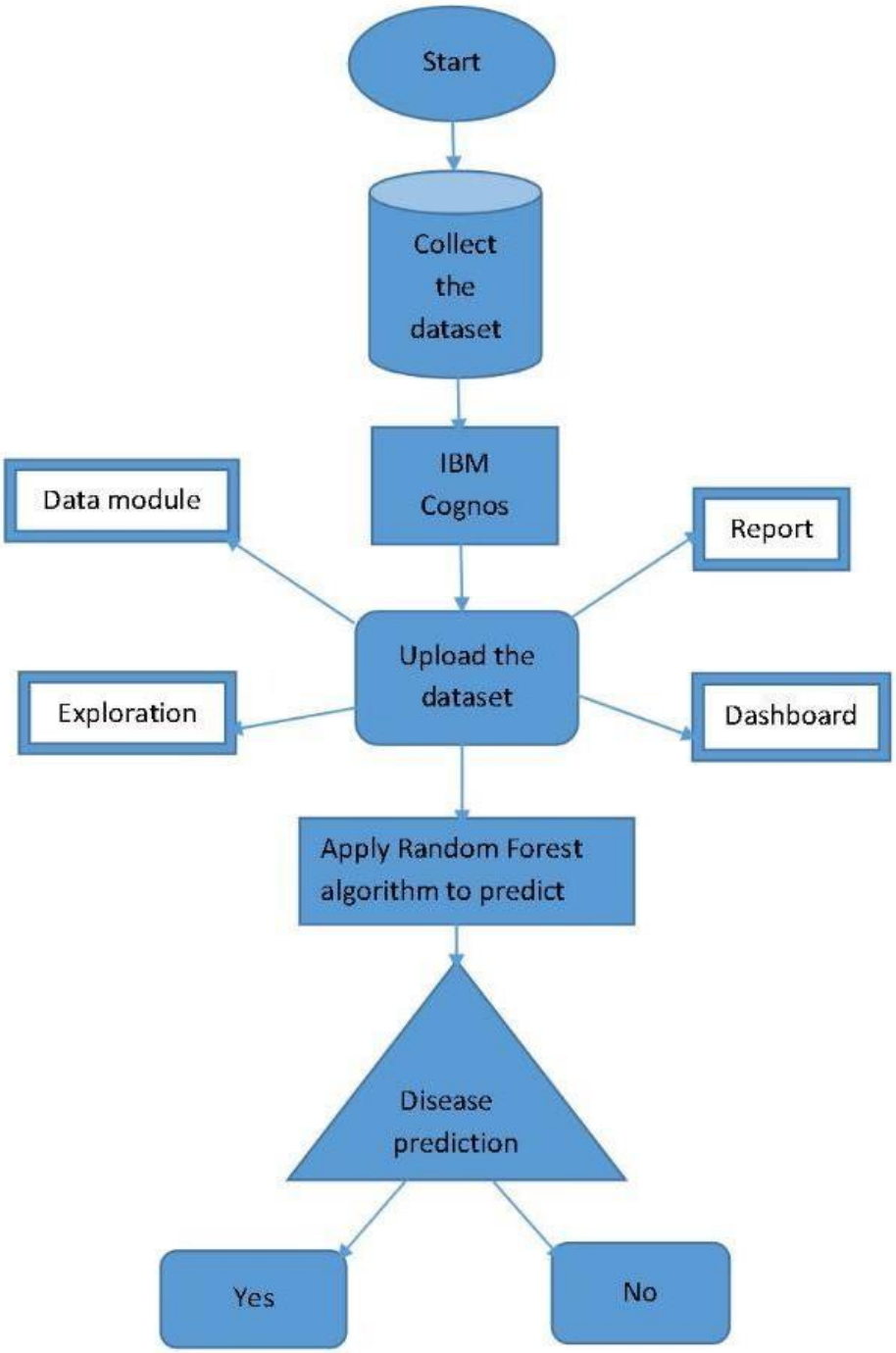
Following are the non-functional requirements of the proposed solution.

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**	When it comes to customers' data, we should provide security. There should be no errors or leakage of data. The data should be retrieved in emergency situation such as crash. The important data should be kept safe.
NFR-3	**Reliability**	The data is complete and accurate, and it is a crucial foundation for building data trust across the organization. Ensuring data reliability is one of the main objectives of
		data integrity initiatives, which are also used to maintain data security, data quality, and regulatory compliance.
NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
NFR-5	**Availability**	The application is functioning properly and usable to meet the requirements of an individual or business.

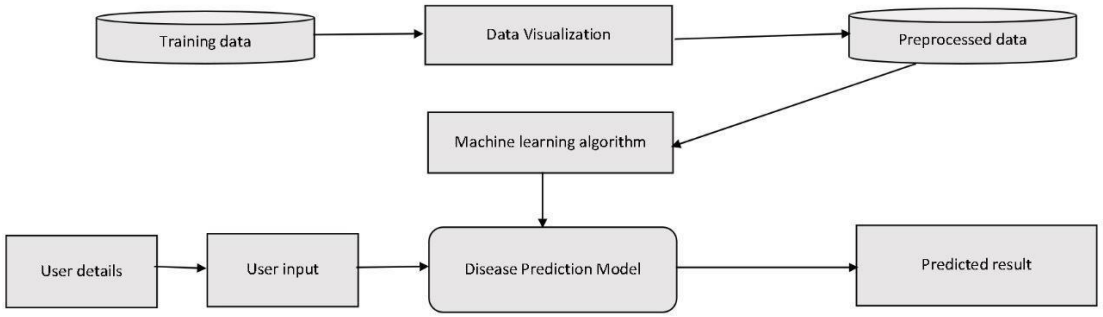
NFR-6	**Scalability**	The application can withstand the increase in the no. of users and has to be able to develop higher versions

Project Design

5.1 Data Flow Diagram



5.2 Solution and Technical Architecture



5.3. 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
	Login	USN-2	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-3	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-4	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-5	As a customer care executive, he/she can view the customer queries.	I can post my queries in the dashboard	Medium	Sprint-3

		USN-6	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-7	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-8	As an admin, he/she can add or delete users.	I can access my account / Dashboard when logged in	High	Sprint-4
		USN-9	As an admin, he/she can manage the user details.	I can view the organized data of myself.	High	Sprint-4

6. Project Planning and Scheduling

6.1 Script Planning and Execution

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	1
Sprint-2		USN-2	As a user, I will receive confirmation email once I have registered for the application	2	Low	4
Sprint-1		USN-3	As a user, I can register for the application through Gmail	2	Medium	3
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email & password	1	High	2

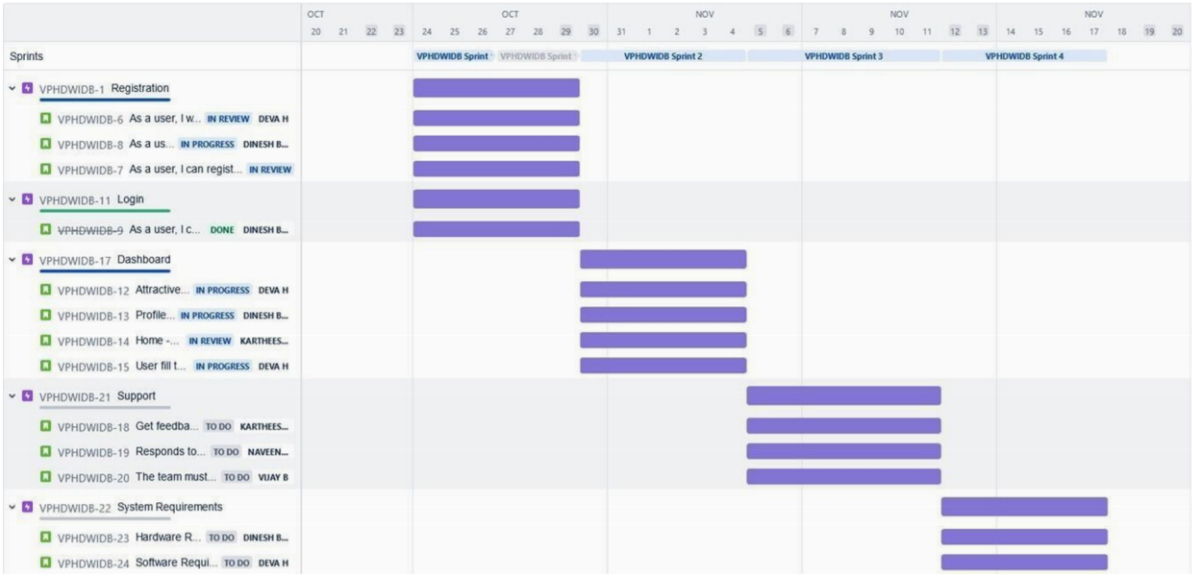
Sprint-2	Dashboard	USN-5	Profile - view & update your profile	2	High	5
Sprint-2		USN-6	Home - Analyze your Heart	2	High	5
Sprint-3		USN-7	The user will have to fill in the below 13 fields for the system to predict a disease -Age in Year -Gender -Chest Pain Type -Fasting Blood Sugar -Resting Electrographic Results(Restecg) -Exercise Induced Angina(Exang) -The slope of the peak exercise ST segment -CA -Number of major vessels colored by fluoroscopy -Thal -Trest Blood Pressure -Serum Cholesterol -Maximum heart rate achieved(Thalach) ST depression induced by exercise(Oldpeak)	2	High	5
		USN-8	User can view the accuracy of occurrence of heart disease	1	Medium	4
Sprint-3	System Requirement	USN-9	I. Hardware Requirement i. Laptop or PC □ I5 processor system or higher	2	High	2
			□ 4 GB RAM or higher □ 128 GB ROM or higher			
Sprint-3		USN-10	II. Software Requirement iii. Laptop or PC • Windows 10 or higher	2	Medium	2

Sprint-4	Administrator	USN-11	As an administrator, he/she can view the customer queries.	1	High	1
		USN-12	Administrator can view the ratings of the customer	2	Medium	2
		USN-13	As a customer care executive, he/she can answer the customer queries.	2	High	2
		USN-14	As an admin, he/she can add or delete users.	1	High	2
		USN-15	Customer Feedback – customers can send feedback to the Admin	2	Medium	3

6.2 Sprint Delivery Schedule

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-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	30 Oct 2022	04 Nov 2022	17	04 Nov 2022
Sprint-3	20	6 Days	05 Nov 2022	11 Nov 2022	18	11 Nov 2022
Sprint-4	20	6 Days	12 Nov 2022	17 Nov 2022	19	17 Nov 2022

6.3 Jira Report



7 Coding And Solutioning

7.1 Machine Learning

Learning which model is best for the given Dataset

7. CODING & SOLUTIONING

7.1. Feature 1

Home Page

Heart disease prediction app is a responsive web application which detects whether the user have heart disease using the data given by the user. The home page contains information about the application. It also contains the tab to upload the input of the user.

```
<!DOCTYPE html>
<html lang="en">

<head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <link rel="stylesheet" href="/static/style/home.css">
   <link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">
   <title>Health Disease Detector</title>
</head>
<body>
<div class="overlay" id="myNav">
      <div class="topnav">
         <a class="active" href="/"><i class="fa fa-fw fa-home"></i> Home</a>
         <a href="heart-disease-predictor"><i class="fa fa-fw fa-heartbeat"></i> Heart Disease Diagnosis</a>
         <a href="contact"><i class="fa fa-fw fa-envelope"></i> Contact</a>
         <a href="about"><i class="fa fa-fw fa-info-circle"></i> About</a>
      </div>
   </div>
   <div class="information">
```

```
    <div class="info">
        The Website is developed for all the user who want to test there risk of getting Heart disease. We are using
        multiple machine learning approch to provide as good solution as possible. However, it is entirely dependent
        on the accuracy of the data provide by the user.
    </div>
    <div class="heart"><i class="fa fa-fw fa-heartbeat"></i></div>
  </div>
  <script src="static/script/home.js"></script>
</body>
</html>
```

7.2. Feature 2

Prediction page:

The user will add the data asked in the prediction page and press the predict button. Then the page will redirect to a new page and provide information whether the person has the possibility of having heart disease or not.

app.py

```
from flask import Flask
from flask import Flask, render_template, request
import sklearn
import pickle
from assembleModel.modelcombine import combineModel
model1 = pickle.load(open("static/model/model.sav", 'rb'))
```

```
sc = pickle.load(open('static/model/scaler.pkl', 'rb'))

app = Flask(__name__)



@app.route("/")
def home():
    return render_template("home.html")



@app.route("/about")
def about():
    return render_template("about.html")



@app.route("/heart-disease-predictor", methods=['POST', 'GET'])
def heartDiseasePredictor():
    if request.method == 'POST':
        result = request.form.to_dict()
        age = int(result['age'])
        gender = int(result['gender'])
        chest_pain_type = int(result['chest-pain-type'])
        fasting_blood_sugar = int(result['fasting-blood-sugar'])
        induced_agina = int(result['induced-agina'])
        resting_blood_pressure = int(result['resting-blood-pressure'])
        peak_exercise_st = int(result['peak-exercise-st'])
        st_depression_value = float(result['st-depressed-value'])
        heart_rate_value = int(result['heart-rate-value'])
        serum_cholestrol_value = int(result['serum-cholestrol-value'])
        resting_ecg = int(result['resting-ecg'])
```

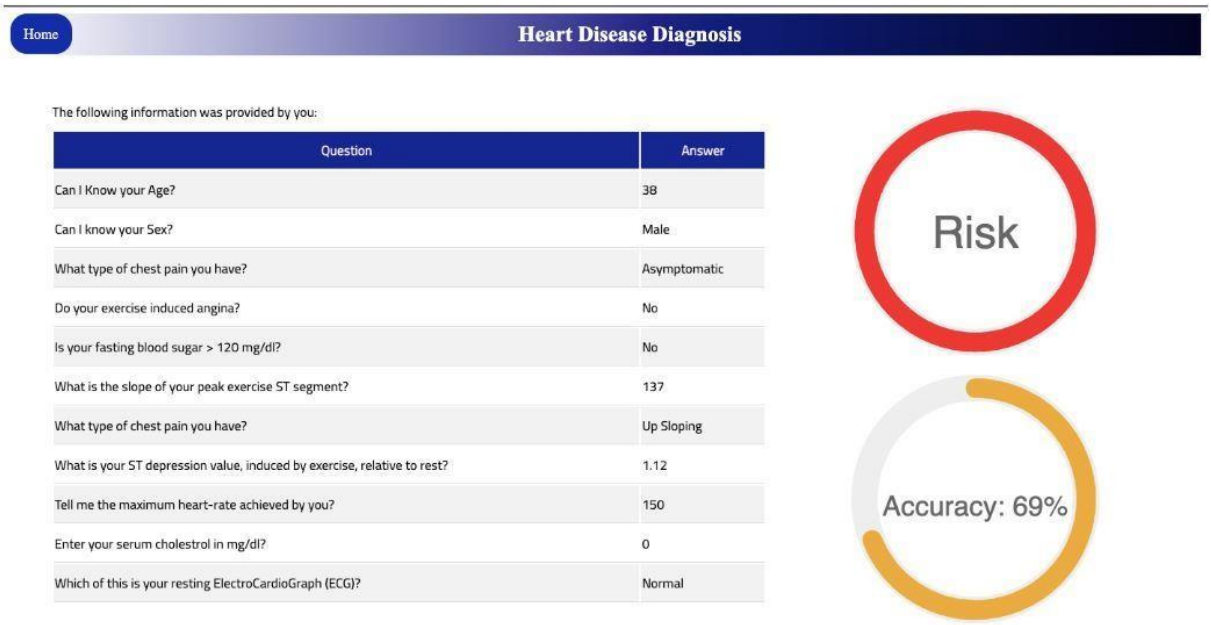
```
        element = sc.transform([[age, gender, chest_pain_type, resting_blood_pressure,
serum_cholestrol_value,
                                fasting_blood_sugar, resting_ecg, heart_rate_value, induced_agina,
st_depression_value, peak_exercise_st]])
        prediction,prediction_prob = combineModel(models=[model1],element=element)
        k = 0
        if prediction>=0.5:
           k=1
        result['prediction'] = k
        result['prediction-prob']=prediction_prob


        return render_template("result.html", results=result)
    return render_template("heart_disease.html")


@app.route("/contact")
def contact():
    return render_template("contact.html")
if __name__== "_main_":
    app.run()
```

8. Testing

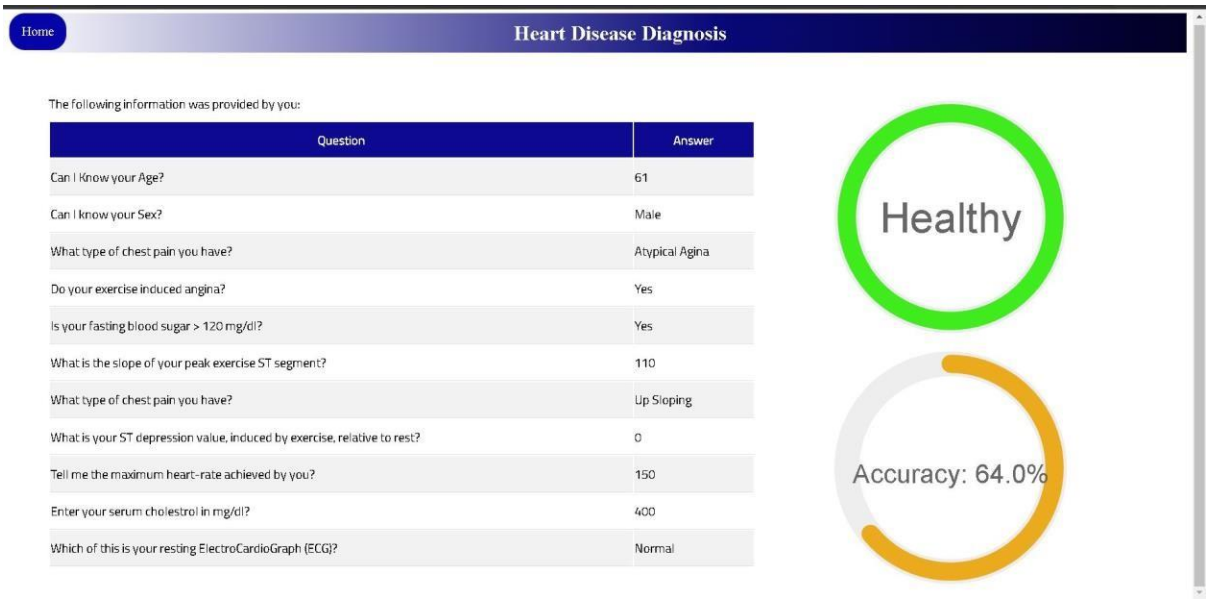
8.1. User Acceptance testing

				Date	03-Nov-22								
				Team ID	PNT2022TMID00437								
				Project Name	Project - Visualisation and Prediction of heart di								
				Maximum Marks	4 marks								
Test case ID	**Feature Type**	**Component**	**Test Scenario**	**Pre-Requisite**	**Steps To Execute**	**Test Data**	**Expected Result**	**Actual Result**	**Status**	**Comments**	**TC for Automation(Y/N)**	**BUG ID**	**Executed By**
HomePage_TC_001	UI	Home Page	Verify all the UI elements in Home page rendered properly		1.Enter URL and click go 2.Verify all the UI elements displayed or not	-	All the UI elements rendered properly	Working as expected	Pass		N		Divyasindhu
HomePage_TC_002	Functional	Home Page	Verifiy the Data Entry page can be reachable.		1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the Predict button	-	User should navigate to Data Entry Page	Working as expected	Pass		N		Desika
DataEntryPage_TC_001	UI	Data Entry Page	Verify all the UI elements in Data Entry page rendered properly		1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the Predict button in the home page 4. Verify all the UI elements displayed or not	-	All the UI elements rendered properly	Working as expected	Pass		N		Joan Reshmi
DataEntryPage_TC_002	Functional	Data Entry Page	Verify user is able to enter all values		1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the Predict button in the home page 4. Verify all the UI elements displayed or not 5. Verify if all values can be entered	67 Male 1 142 242 1 0 125 0 2.5	User should be able to enter all values in data entry page	Working as expected	Pass		N		Anjhana
DataEntryPage_TC_003	Functional	Data Entry Page	Verifiy the Output Display page can be reachable.		1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the Predict button in the home page 4. Verify all the UI elements displayed or not 5. Verify if all values can be entered 6. Press the submit Button	-	User should navigate to Output Display Page	Working as expected	Pass		N		Dhivyadharshini
OutputDisplayPage_TC_001	UI	Output Display Page	Verify all the UI elements in Output Display page rendered properly		1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the predict button in the home page 4. Verify all the UI elements displayed or not 5. Verify if all values can be entered 6. Press the submit Button 7. Verify all the UI elements displayed or not	-	All the UI elements rendered properly	Working as expected	Pass		N		Desika
OutputDisplayPage_TC_002	Functional	Output Display Page	Verify user is able to get predicted result		"1.Enter URL and click go 2.Verify all the UI elements displayed or not. 3.Press the predict button in the home page 4. Verify all the UI elements displayed or not 5. Verify if all values can be entered 6. Press the submit Button 7. Verify all the UI elements displayed or not 8. Verify if the predicted value is displayed or not"	-	Predited Car Resale Value is displayed on the page	Working as expected	Pass		N		Divyasindhu

Testing a case where user has heart disease



Testing a case where user does not have heart disease



9. Result

9.1 Performance Metrics

The confusion matrix below shows the performance metrics of the machine learning model.

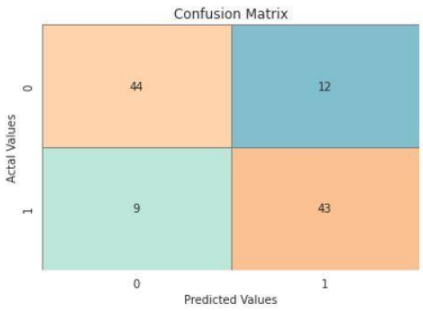
```
cm = confusion_matrix(y_test,pred)
print(classification_report(y_test,pred))
```

```
              precision    recall  f1-score   support

           0       0.83      0.79      0.81        56
           1       0.78      0.83      0.80        52

    accuracy                           0.81       108
   macro avg       0.81      0.81      0.81       108
weighted avg       0.81      0.81      0.81       108
```

In [24]:

```
sns.heatmap(cm, annot = True, fmt = 'g', cbar = False, cmap = 'icefire', linewidths= 0.5, linecolor= 'grey')
plt.title('Confusion Matrix')
plt.ylabel('Actal Values')
plt.xlabel('Predicted Values')
```

Out[24]: Text(0.5, 15.0, 'Predicted Values')



In [25]:

```
print("Accuracy Score = {}".format(round(accuracy_score(y_test,pred),5)))
```

```
Accuracy Score = 0.80556
```

10. Advantages Disadvantages

Advantages:

- This is one of the fastest ways to determine if a person is likely to suffer from a heart disease or not.
- Useful for medical practitioners to easily classify their patients.
- User Friendly
- Easy to understand
- Secure
- Dashboard provides insightful informations

Disadvantages:

- Needs work
- Users need to know all the fields
- Does Not take null value as input
- Does not provide suggestions to user

11. Conclusion

Complications of heart disease include heart attack and stroke. You can reduce the risk of complications with early diagnosis and treatment. So the suggestion that

we get from the website might help save patients. It is always to get treated in the early stages of heart disease.

12. Future Scope

Like the saying goes “Prevention is better than cure”. We have to look into methods to prevent heart diseases altogether other than just predicting it in early stages.
To use this website we need to take a lot of tests beforehand. So it would be better if we require less attributes and still give an effective result.

13.Appendix

SOURCE CODE
(https://github.com/IBM-EPBL/ IBM-Project-12833-1659495500)
GITHUB & PROJECT DEMO LINK
(https://github.com/IBM-EPBL/ IBM-Project-12833-1659495500**)**
DEMO LINK
(https://youtu.be/CMdGdXBtlME**)**