

**MAHENDRA INSTITUTE OF TECHNOLOGY
(AUTONOMOUS)**

**Mahendrapuri, Mallasamudram (W), Thiruchengode,
Namakkal, Tamil Nadu 637503.**



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION
ENGINEERING**

**VISUALIZING AND PREDICTING HEART DISEASES
WITH
AN INTERACTIVE DASHBOARD**

**Project Report -2022
Team ID: PNT2022TMID17411**

Submitted by

**M. Aravind - 611619106004
M.B .Yogeshwaran - 611619106308
J.P. Muthu Aravindh - 611619106058
K. Maruthupandiyan - 611619106051**

CONTENT

1. INTRODUCTION

1. Project Overview
2. Purpose

2. LITERATURE SURVEY

1. Existing problem
2. References
3. Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

1. Empathy Map Canvas
2. Ideation & Brainstorming
3. Proposed Solution
4. Problem Solution fit

4. REQUIREMENT ANALYSIS

1. Functional requirement
2. Non-Functional requirements

5. PROJECT DESIGN

1. Data Flow Diagrams
2. Solution & Technical Architecture
3. User Stories

6. PROJECT PLANNING & SCHEDULING

1. Sprint Planning & Estimation
2. Sprint Delivery Schedule
3. Reports from JIRA

7. CODING & SOLUTIONING

8. TESTING

1. Test Cases
2. User Acceptance Testing

9. RESULTS

Performance Metrics

10.ADVANTAGES & DISADVANTAGES

11.CONCLUSION

12.FUTURE SCOPE

INTRODUCTION

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 87% for which is better than most other systems in terms of achieving accuracy quickly.

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.

LITERATURE SURVEY

PAPERSET 1

ABSTRACT:

This paper describes various methods of data mining, big data and machine learning models for predicting the heart disease. Data mining and machine learning plays an important role in building an important model for medical system to predict heart disease or cardiovascular disease. Medical experts can help the patients by detecting the cardiovascular disease before occurring.

ADVANTAGES:

Bo Jin, Chao Che et al. (2018) proposed a “Predicting the Risk of Heart Failure With EHR Sequential Data Modeling” model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the sequential nature of clinical records

DRAWBACKS:

In this paper, a literature survey of review delivers the concept of various techniques has been studied for diagnosing the cardiovascular disease. Use of big data, machine learning along with data mining can provide promising results to bring the most effective accuracy in analyzing the prediction model

PAPERSET 2

ABSTRACT:

Data Analysis is carried out to discover useful knowledge from the dataset and to drive quick and better decisions. It is also used to increase the efficiency of the work. Exploratory Data analysis is the first phase in Data Analysis. It is a method to understand the data and summarize the main features in the dataset by analyzing the data. It is also used for the visual representation of data. Visualization includes line plot, subplot, pair plot, violin plot, joint plot, swarm plot, Histograms, Box plot, Scatter plot. In this paper, Exploratory Data Analysis is done using python and implemented in Spyder IDE

ADVANTAGE:

Worldwide research shows that millions of lives lost per year because of heart disease. The healthcare sector produces massive volumes of data on heart disease that are sadly not used to locate secret knowledge for successful decision making. One of the most important aspects at this moment is detecting heart disease at an early stage. Researchers have applied distinct techniques to the UCI Machine Learning heart disease dataset. Many researchers have tried to apply some complex techniques to this dataset, where detailed studies are still missing. In this paper, Principal Component Analysis (PCA) has been used to reduce attributes. Apart from a Hybrid genetic algorithm (HGA) with k-means used for final clustering. We used the Hybrid Genetic Algorithm (HGA) for data clustering to avoid this problem. Our proposed method can predict early heart disease with an accuracy of 94.06%.

DRAWBACKS:

According to many researches that have been conducted through a period of time have found out that heart failure and heart disease has been the cruel cause of death in human beings. What aggravates this situation is that most of these diseases are being diagnosed at later stages at which it is very difficult to control

REFERENCES

PAPER SET 1

**Published in: International Research Journal of Engineering and Technology (IRJET)
Issue: 05 | May 2020**

PAPER SET 2

Published in: Md. Touhidul Islam, Sanjida Reza Rafa, et al, “Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means”, 2021.

IDEATION AND PROPOSED SOLUTION

EMPATHY MAP CANVAS

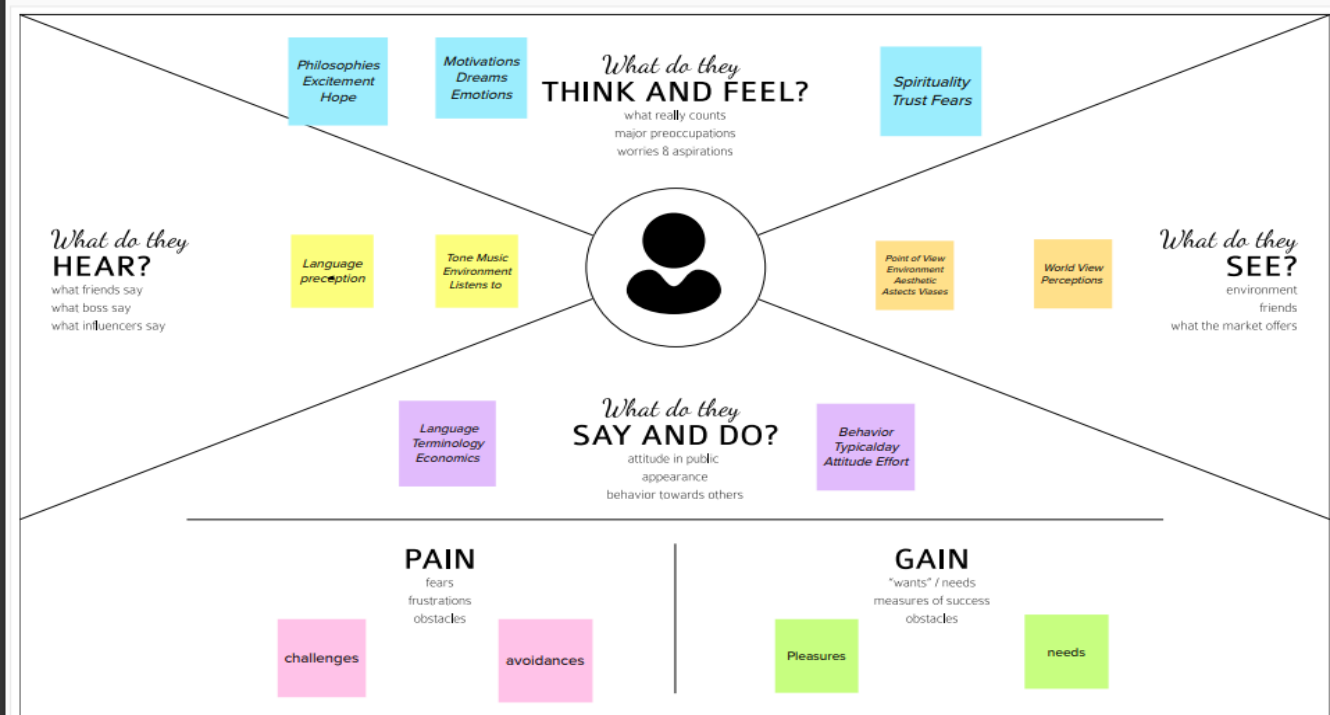
Edit this template
Right-click to unlock

Empathy Map Canvas

Gain insight and understanding on solving customer problems.

1

Build empathy and keep your focus on the user by putting yourself in their shoes.




Share your feedback

IDEATION AND BRAINSTROMING

Step-1: Team gathering, collaboration and select problem statement

Template



Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour to collaborate
- 2-6 people recommended

➔

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

1

Team gathering

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

2

Set the goal

Think about the problem you'll be focusing on solving in the brainstorming session.

3

Learn how to use the facilitation tools

Use the Facilitation Superpowers to run a happy and productive session.

Open article ➔

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

How might we [your problem statement]?

Key rules of brainstorming

To run a smooth and productive session

Stay in topic.

Encourage wild ideas.

Defer judgment.

Listen to others.

Go for volume.

If possible, be visual.

M.Aravind
M.B.Yogeshwaran
J.P.MuthuAravindh
K.Maruthupandiyan

Share template feedback

Step-2: Brainstorm, Idea Listing and grouping

2

Brainstorm

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

10 minutes

TIP

You can select a sticky note and hit the pencil (edit) or trash (delete) icons to start editing!

M.Aminal

We could have reduced the information needed to enter as some features have no impact on predicted risk.

I kept everything so others who develop their own model using the same dataset can simply reuse the model.

We could use the upload or direct database connection instead of manual data entry.

But what we have will suffice for the example and could not be as much burdensome as part of a pilot.

Using predicted risk score could be useful for recommending follow up actions, so use the score in this example.

If we repeat the likelihood of heart disease in a patient, our predicted probability must be self-calibrated.

M.R.Vogelweiser

Several neural networks are easier to configure and obtain much good results.

The logging ensemble learning algorithm with DT and FL features.

comparative analysis of the results of various machine learning.

HDFS is Web-based, scalable, reliable and expandable.

Machine learning algorithms such as Random Forest, Support Vector Machine, etc.

The motivation for the study was to find the most robust ML algorithm for detection of heart disease.

J.P.Muthu Aravindh

The heart disease detection system uses a patient's medical history and clinical information of heart disease to predict heart disease.

Hyperparameter tuning helps to get improved accuracy as compared to the default parameters for the ML.

develop an artificial neural network algorithm for classifying heart disease based on these clinical tests.

Machine learning (ML) can bring an effective solution for decision making and accurate predictions.

heart diseases is based on high blood pressure.

that each technique has its unique strength in making the signatures of the dataset mining gap.

K.Murugesundari

Using medical profiles such as age, sex, blood pressure, cholesterol, pain type, smoking, blood sugar.

preprocessing phase of the early stage will be useful in the early stage of the study.

Old patients along with a 50-year-old patients are included as input features in the model.

improved the quality of cardiovascular disease prediction using a further processing phase.

different data mining techniques can be utilized.

The HDFS system developed in this study.

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

20 minutes

Identifying minimum key requirements

TIP

Add customer-like tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as they come within your mind.

An output risk score for the patient that also assigns them to a risk group.

how to address requirements

A place to input the patient characteristics used as features in the predictive model.

Sketching out a UI that incorporates key requirements.

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



➔

After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

- Share the mural**
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.
- Export the mural**
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

- Strategy blueprint**
Define the components of a new idea or strategy.
[Open the template →](#)
- Customer experience journey map**
Understand customer needs, motivations, and obstacles for an experience.
[Open the template →](#)
- Strengths, weaknesses, opportunities & threats**
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
[Open the template →](#)

[Share template feedback](#)

PROPOSED SOLUTION

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)	➤ To develop an interactive dashboard to predict the heart disease accurately with few tests and attributes the presence of heart disease.
2.	Idea / Solution description	➤ Analyzing data and identifying the heart disease using Cognos analysis.
3.	Novelty / Uniqueness	➤ Hoping to achieve maximum accuracy to provide prior treatment to the patients and reduce the fatality rate.
4.	Social Impact / Customer Satisfaction	➤ Saving lives, User friendly interactive dashboard. ➤ Reduces the exorbitant medical cost of the patients. ➤ Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions and experiences.
5.	Business Model (Revenue Model)	➤ Data security. ➤ Easy to use. ➤ Constant updates according to necessity.
6.	Scalability of the Solution	➤ Can be used in any platform (Windows, mac, etc.,). ➤ Adding new feature doesn't affect the performance of the system. ➤ Scalable dataset.

PROBLEM SOLUTION FIT

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS <div> People with heart disease. Aged Persons(Above 60) </div>	6. CUSTOMER CONSTRAINTS CC <div> Avoidable medical errors. Low treatable mortality rates. Lack of transparency. Difficulty finding a good doctor. High maintenance costs. A different perspective on solving the shortage crisis. </div>	5. AVAILABLE SOLUTIONS AS <div> <ul style="list-style-type: none"> • Avoid smoking • Take healthy foods • Visit cardiologist in case of any symptoms • Maintaining healthy exercise </div>	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS J&P <div> Coronary artery disease is a common heart condition that affects the major blood vessels. Cholesterol deposits (plaques) in the heart arteries are usually the cause of coronary artery disease. </div>	9. PROBLEM ROOT CAUSE RC <div> A buildup of fatty plaques in the arteries (atherosclerosis) is the most common cause of coronary artery disease. Risk factors include a poor diet, lack of exercise, obesity and smoking </div>	7. BEHAVIOUR BE <div> Chest pain or discomfort, Shortness of breath, Slow heartbeat, Lightheadedness, Swelling in the legs, belly area or areas around the eyes. </div>	

Identify strong TR & EM	3. TRIGGERS TR <div> Insufficient ways of handling huge amounts of datasets and inferring the root cause of the heart disease cannot be found out. Similarity of heart disease has not been identifiable. </div>	10. YOUR SOLUTION SL <div> With the notable technology of AI/ML we are able to visualize and predict heart diseases and related diseases, by the ultimate power Cognos Analytics Tool we will be able to properly create a dashboard for the customers to work with and visualize and analyze the heart disease on their work with limited knowledge. </div>	8. CHANNELS of BEHAVIOR CH <div> 8.1 ONLINE Visualizing the datasets. Exploration of data. 8.2 OFFLINE Cleansing of datasets. Collection and noting the datasets. </div>	Identify strong TR & EM
	4. EMOTIONS: BEFORE / AFTER EM <div> Before -> It creates a huge ambiguity in knowing the proper or accurate reasons for a heart disease. After -> There is a large chance understanding of the heart disease and root cause of it. which makes a better solution and finding a preventive way over it. </div>			

REQUIREMENT ANALYSIS

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 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 Congo's Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

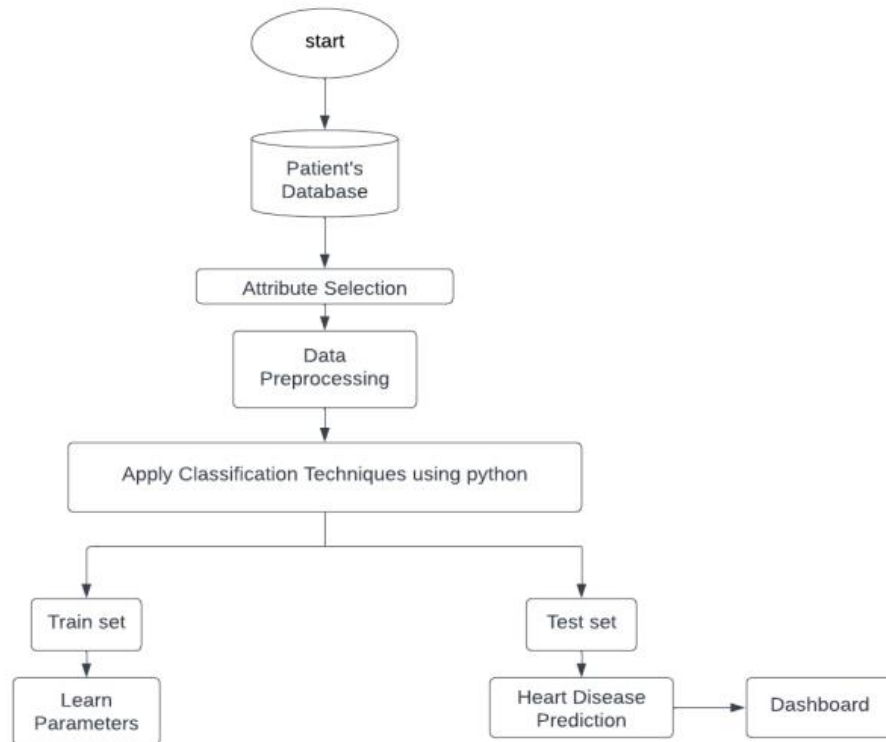
Non-functional Requirements:

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

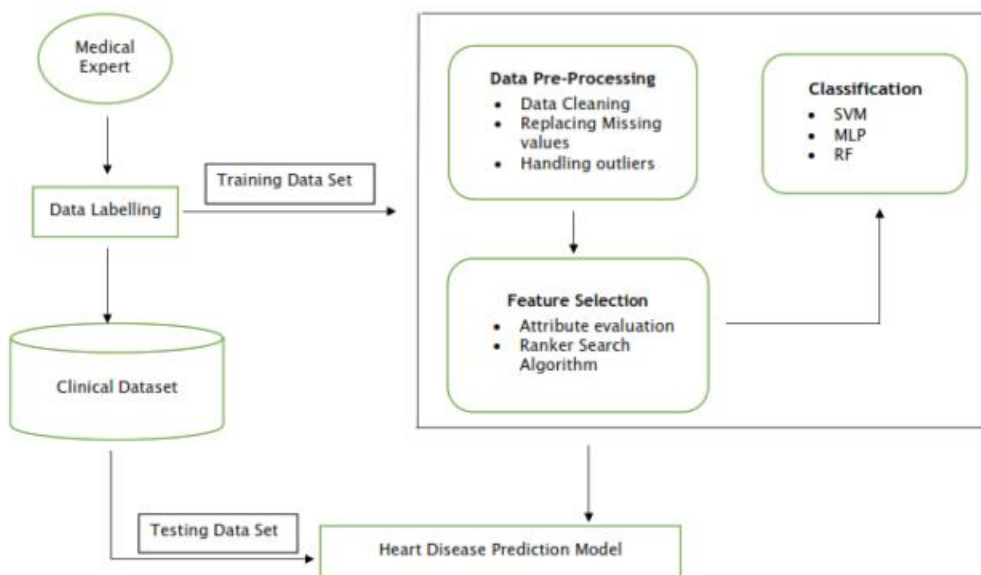
NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and user-Friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be Performed with just a few clicks.
NFR-2	Security	For security of the application the technique Known as database replication should be used so that all the important data should be kept safe. In case of crash, the system should be able to backup and recover the data.
NFR-3	Reliability	The application has to be consistent at every scenario and has to work without failure in any environment
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 has to be available 24 x 7 for users without any interruption
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

DATA FLOW DIAGRAM



SOLUTION AND TECHNICAL ARCHITECTURE



PROJECT PLANNING AND SCHEDULING

SCRIPT PLANNING EXECUTION

Product Backlog, Sprint Schedule, and Estimation

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.	1	High	Aravind M Yogeshwaran M B Muthu Aravindh J P Maruthupandiyan K
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password	1	High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-2	Dashboard	USN-4	User can view his/her complete medical analysis and accuracy of disease prediction	2	High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-2		USN-5	User can view the accuracy of occurrence of heart disease	2	High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-3	Helpdesk	USN-6	As a customer care executive, he/she can view the customer queries.	2	Medium	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-3		USN-7	As a customer care executive, he/she can answer the customer queries.		High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-4	User Profile	USN-8	As an admin, he/she can update the health details of users.		High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-4		USN-9	As an admin, he/she can add or delete users.		High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P
Sprint-4		USN-10	As an admin, he/she can manage the user details.		High	Aravind M Yogeshwaran M B Maruthupandiyan K Muthu Aravindh J P

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	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

CODING & TESTING

app.py

```
Import numpy as np
import pickle
import sklearn
from flask import Flask, render_template, request, redirect, url_for, flash
import sqlite3
model = pickle.load(open('models.pkl', 'rb'))
app = Flask(__name__)
app.secret_key = "7847541"

def get_db():
    conn = sqlite3.connect('user_details.db')
    conn.row_factory = sqlite3.Row
    return conn
@app.route('/')
def index():
    return render_template('index.html', title='Home')

@app.route('/about')
def about():
    return render_template('about.html', title='About')

@app.route('/signin', methods=('GET', 'POST'))
def signin():
    error = None
    if request.method == 'POST':
        name = request.form['name']
        password = request.form['password']
        db = get_db()
        user = db.execute(
            'SELECT name FROM user_details WHERE password = ?', (password, )
        ).fetchone()

    if user is None:
        error = 'Incorrect Username/Password.'
```

```

        if error is None:
            return render_template('index.html', title="Home", succ="login successfull!")
        flash(error)
        db.close()

    return render_template('signin.html', title='Sign In', error=error)

@app.route('/signup', methods=('POST', 'GET'))
def signup():
    if request.method == 'POST':
        name = request.form['name']
        email = request.form['email']
        password = request.form['password']
        db = get_db()
        curr = db.cursor()
        curr.execute(
            'INSERT INTO user_details (name, email, password) VALUES (?, ?, ? );',
            (name, email, password)
        )
        db.commit()
        curr.close()
        db.close()
        return render_template('index.html', title="Home", succ="Registration Successfull!")
    return render_template('signup.html', title='Sign Up')

@app.route('/Heart_Disease_Classifier')
def Heart_Disease_Classifier():
    return render_template('Heart_Disease_Classifier.html')

@app.route('/predict', methods=['POST'])
def predict():

    features = [float(i) for i in request.form.values()]
    #Convert features to array
    array_features = [np.array(features)]
    #Predict features
    prediction = model.predict(array_features)
    output = prediction
    if output == 1:
        return render_template('Heart_Disease_Classifier.html', result = 'The patient is not likely

```

```
        to have heart disease!')
    else:
        return render_template('Heart_Disease_Classifier.html', result = 'The patient is likely to have heart
disease!')

if __name__ == '__main__':
    debug(True)
```

TEST CASES TESTING THE DATA MODEL FOR VARIOUS INPUT VALUES.

```
In [ ]: from sklearn.metrics import accuracy_score
input=(53,1,3,145,200,150,98,0,0,0,0,0)
input_as_numpy=np.asarray(input)
input_resaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resaped)
print(pre1)
a1=accuracy_score(pre1,model1.predict(input_resaped))*100
print(a1)

['Absence']
100.0

In [ ]: from sklearn.metrics import accuracy_score
input=(70,1,4,130,322,0,2,100,0,2,4,2,3,3)
input_as_numpy=np.asarray(input)
input_resaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resaped)
print(pre1)
a1=accuracy_score(pre1,model1.predict(input_resaped))*100
print(a1)

['Presence']
100.0
```

USER ACCEPTANCE TESTING A CASE WHERE USER HAS HEART DISEASE

Heart Disease Test

Heart Disease Test Form

Age	Sex		
18	Male		
Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholesterol in mg/dl	Fasting Blood Sugar > 120 mg/dl
Atypical Angina	145	146	True
Resting ECG Results	Maximum Heart Rate	Exercise Induced Angina	ST Depression Induced
Normal	136	-- Select an Option --	
Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Fluoroscopy	Thalassemia	
-- Select an Option --	-- Select an Option --	-- Select an Option --	

Result

The patient is likely to have heart disease!

RESULTS

Performance Metrics

Hours worked: 50 hours

Stick to Timelines: 100%

Stay within budget: 100%

Consistency of the product: 85%

Efficiency of the product: 85%

Quality of the product: 85%

ADVANTAGES & DISADVANTAGES:

ADVANTAGES:

- Smooth User Interface
- Accuracy is achieved quickly

DISADVANTAGES:

Random forest can be used for both classification and regression tasks, built is not more suitable for Regression tasks

CONCLUSION

This overview of the project conveys the idea that numerous methods have been investigated for diagnosing cardiovascular disease. Big data, machine learning, and data mining can be used to great success to analyse the prediction model with the highest degree of accuracy. The primary goal of this project is to diagnose cardiovascular disease or heart disease utilizing a variety of techniques and procedures to obtain a prognosis.

FUTURE SCOPE

A future update shall comprise of section for viewing renowned cardiologists and scan centres in their city. The obtained output can be further processed and sent to smart devices to provide necessary assistance. Constant monitoring can provide necessary data to recommend to consult a doctor in case of an emergency.