



ROEVER ENGINEERING COLLEGE



Visualizing and Predicting Heart Diseases with An Interactive Dashboard

**NALAIYATHIRAN PROJECT REPORT
2022**

Team ID : [IBM-Project-41996-1660647459](#)

College Name : Roever Engineering College

Submitted by:

**Latha Nanthini P
Nanthini M
Sneha S
Selvamani S**

**(813519104022)
(813519104033)
(813519104046)
(813519104706)**

Table Of Content

1. INTRODUCTION

Project Overview

Purpose

2. LITERATURE SURVEY

Existing problem

References

Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

Empathy Map Canvas

Ideation & Brainstorming

Proposed Solution

Problem Solution fit

4. REQUIREMENT ANALYSIS

Functional requirement

Non-Functional requirements

5. PROJECT DESIGN

Data Flow Diagrams

Solution & Technical Architecture

User Stories

6. PROJECT PLANNING & SCHEDULING

Sprint Planning & Estimation

Sprint Delivery Schedule

Reports from JIRA

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

Feature 1

Feature 2

8. TESTING

Test Cases

User Acceptance Testing

9. RESULTS

Performance Metrics

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE 13. APPENDIX

Source Code

GitHub & Project Demo Link

Team ID: PNT2022TMID45847

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

1. Introduction

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.

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

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.

References

“Heart Disease Prediction using Exploratory Data Analysis” A survey :R. Indrakumari, T.Poongodi,Soumya Ranjan Jena

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.

Prediction of heart disease at early stage using data mining and big data analytics: A survey N. K. Salma Banu, Suma Swamy

Several studies have been carried out for developing prediction model using individual technique and also by combining two or more techniques. This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers. Into practice.

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?

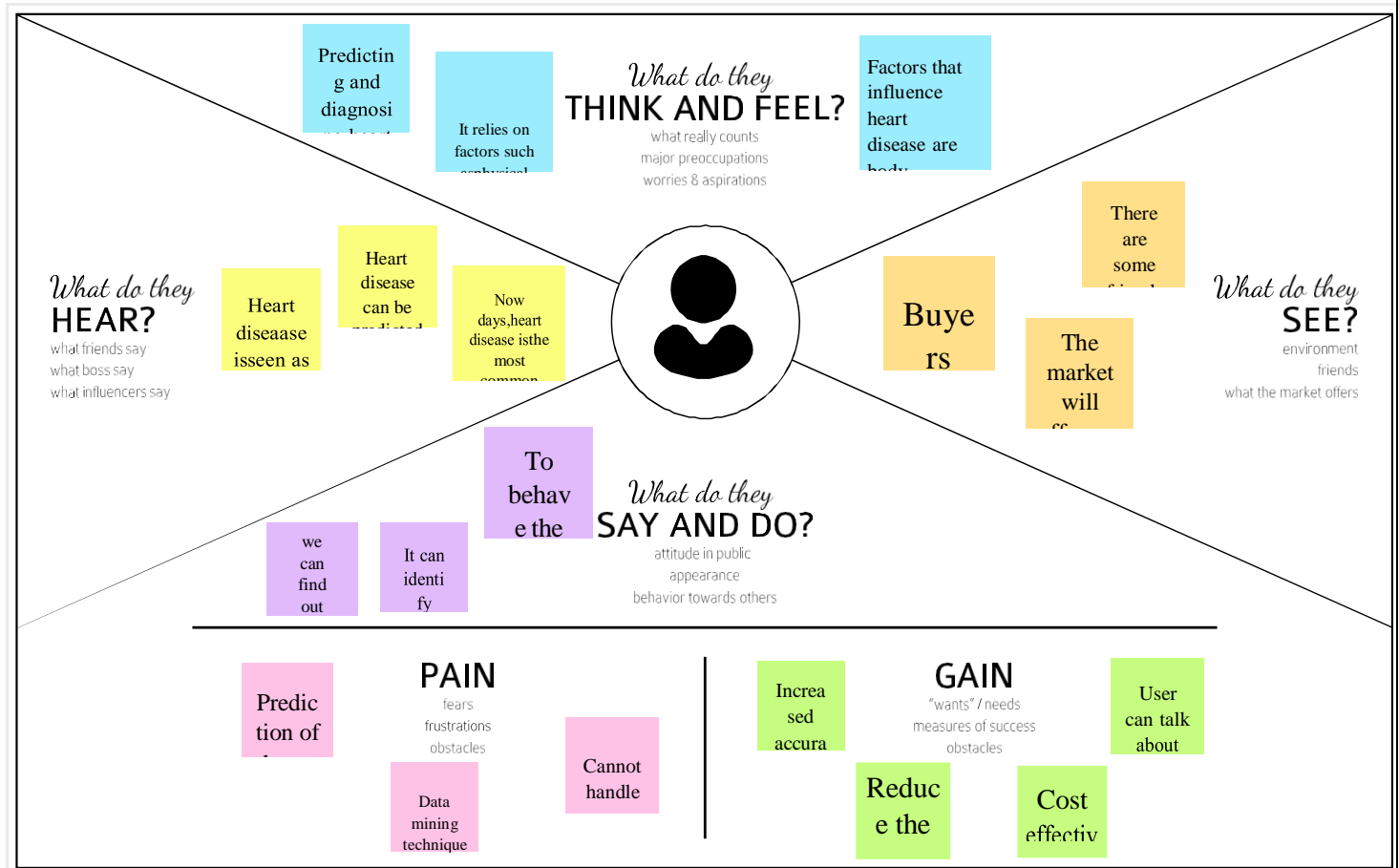
If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will increase rapidly.

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

Empathy Map Canvas



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



Step-2: Brainstorm, Idea Listing and Grouping

2

Brainstorm

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

🕒 10 minutes

Vishwa B

| | | |
|---------------------------------------|-------------------------------|------------------------------------|
| Easy Data Retrieval | Instant Answer | Information Flaring |
| High Accuracy | Main focus on solving dataset | Diagnosis based on several factors |
| Where doctor really is not associated | Dashboard Interface | Large dataset |

Jothivel T

| | | |
|-------------------------------|-----------------------------------|------------------------------|
| Collect User feedback | Standard dataset | Accuracy Data Analysis |
| Accessing data through an app | Buy a medication | Use an visualization on data |
| Collect the required dataset | Analyze the format of the dataset | Content based filtering |

Hemachandran C

| | | |
|---------------------------|-------------------------------------|-------------------------------|
| Dashboard Design | Visualize data | User friendly |
| Easy to access | Get data report of a healthy person | Related medical data research |
| Pre define range for data | Software platform's issue | Contact Support |

Selvakumar N

| | | |
|------------------|------------------------|-------------------------|
| Interactivity | Remote Application | Usable for Doctors |
| Standard Dataset | Performance Efficiency | Content based filtering |
| Personalize | Minimize error | Instant diagnosis |

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

Data and research Work and deliverables Training and testing

Define data labels on the collected dataset

User feedback and contact support

Test the dataset and check efficiency

Compare with the healthy person's dataset and fix ideal range

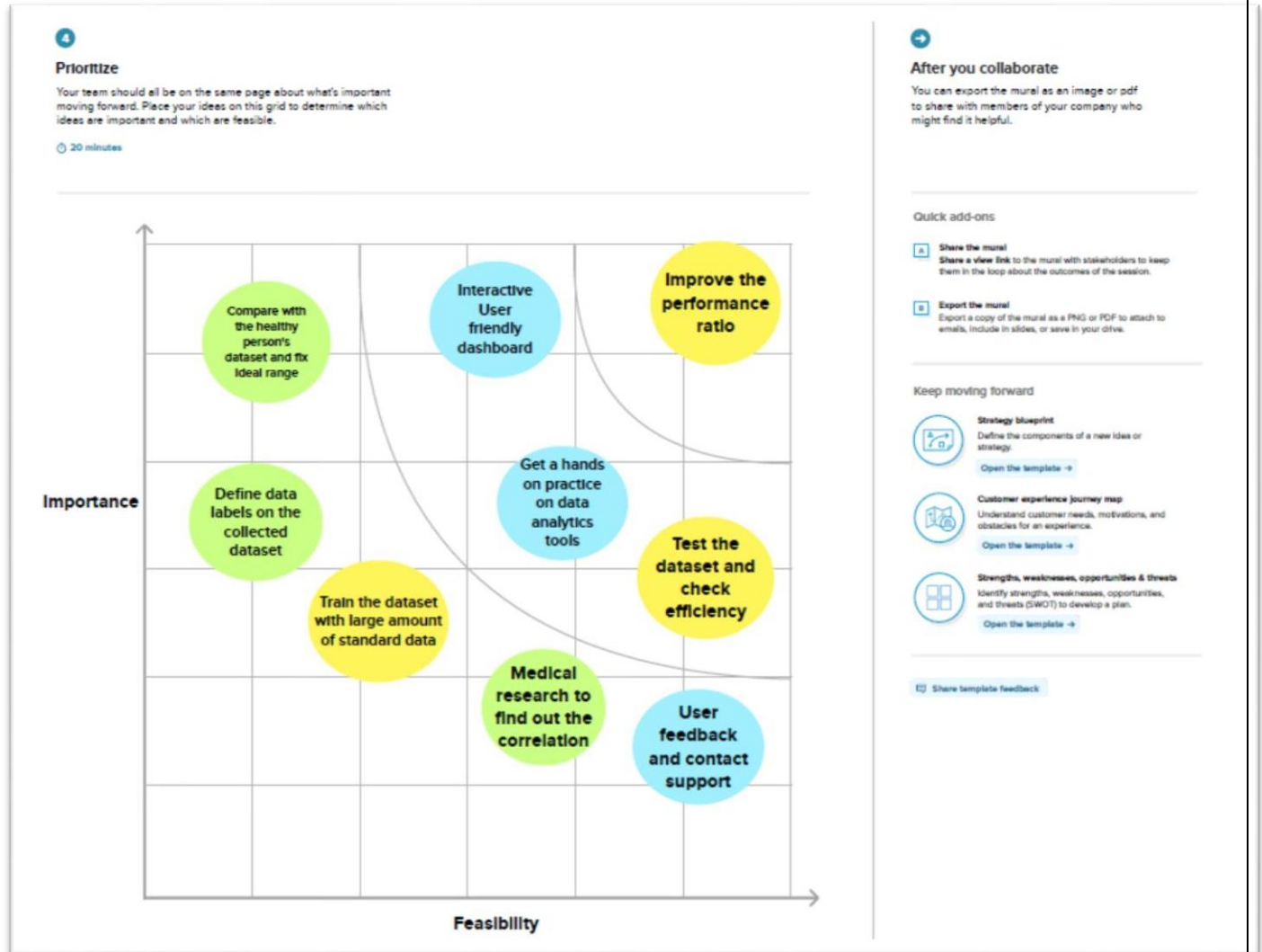
Interactive User friendly dashboard

Improve the performance ratio

Medical research to find out the correlation

Get a hands on practice on data analytics tools

Train the dataset with large amount of standard data



Proposed Solution

| S.No. | Parameter | Description |
|-------|---|---|
| 1. | Problem Statement (Problem to be solved) | The leading cause of death in the developed world is heart disease. As a result, work must be done to reduce the risks of having a heart attack or stroke. It is infeasible for a common man to frequently undergo tests for ECG and so on. Hence it requires a replacement that is both convenient and dependable. |
| 2. | Idea / Solution description | The proposed solution proposes an interactive dashboard for visualizing and forecasting heart disorders, in which the user may observe his/her |

| | | |
|----|---------------------------------------|--|
| | | medical report analysis as well as the projected end result. IBM Cognos will be used to create the dashboard. Machine learning Algorithms will be used to forecast cardiac disease. |
| 3. | Novelty / Uniqueness | Makes recommendations to the user based on that person's medical analysis. |
| 4. | Social Impact / Customer Satisfaction | It helps with disease prediction at an early stage and frequently alerts the user to their current health status. Both the user and the doctor can benefit from the system's improved decision-making regarding cardiac disease. |
| 5. | Business Model (Revenue Model) | Can be deployed by Hospitals or NFOs, so that it makes the analysis in a fast manner. |
| 6. | Scalability of the Solution | The solution can work effectively on long and small datasets. It can also be changed to predict various other diseases depending on the dataset. |

Problem Solution Fit

1. CUSTOMER SEGMENT(S)

CS

Who is your customer?

- smokers
- people who have high blood pressure
- people who have high cholesterol
- people who have high lipoprotein
- Diabetic patients
- people who have lack of regular exercise
- Thrombosis patients
- people who shortness of breath
- people who have Chest pain, chest tightness, chest pressure and chest discomfort (angina)
- people who have Pain in the neck, jaw, throat, upper belly area or back
- people who have Pain, numbness, weakness or coldness in the legs or arms if the blood vessels in those body areas are narrowed
- people who have overweight

6. CUSTOMER CONSTRAINTS

SL

What constraints prevent your customers from taking action or limit their choices of solutions?

- Lack of knowledge about heart disease.
- Negative thoughts of the customer.
- Personal characteristics and physical disability of the customer.
- Complex symptoms of heart failure.
- Psychological problems.
- Lack of support
- Lack of hope in treatment.
- Economical background is major constraints that prevent the customer from taking action.
- Medical and disease related limitations.

5. AVAILABLE SOLUTIONS

AS

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have?

There are various solutions available for the people who are affected with heart diseases. They are,

- Quit smoking
- get cholesterol test periodically
- eat plenty of fruits, vegetables and healthy foods with grains, sprouts, nuts etc.
- Exercise regularly
- Maintain a good physique

If these solutions are properly followed then the people affected with disease can be cured naturally.

- But along with these they have to go for regular medical checkup and test for any heart disease.
- If disease is found in heart they need to make arrangements under proper medications.

2. JOBS-TO-BE-DONE/PROBLEMS

J&P

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one, explore different slides.

- Lives depending on medical support
- Financial insecurity
- shortness of breath
- may feel chest pain, chest tightness, chest pressure

9. PROBLEM ROOT CAUSE

RC

What is the real reason that this problem exists? What is the backstory behind the need to do this job?

- Buildup of fatty plaques in the arteries is the most common cause of coronary artery disease.
- lack of exercise, obesity and smoking.
- Acute aortic insufficiency (AI).
- To cure the diseased patients especially to visualize the heart problems and give relief to them.
- One backstory is that many children are now affected with hole in the heart and suffer a lot than elders, so this method is initiated.
- Heart is the first formed organ when human is formed in the womb so problem in this affects the whole body.
- Thus, this visualization is made and any such heart diseases is predicted with an interactive dashboard.

7. BEHAVIOUR

BE

What does your customer do to address the problem and get the job done?

- Regular, daily physical activity can lower the risk of heart disease. Physical activity helps control your weight.
- A healthy diet can help protect the heart, improve blood pressure and cholesterol, and reduce the risk of type 2 diabetes.
- One of the best things you can do for your heart is to stop smoking or using smokeless tobacco. Even if you're not a smoker, be sure to avoid secondhand smoke.
- Maintain a healthy weight
- Get good quality sleep
- Manage stress
- High blood pressure and high cholesterol can damage the heart and blood vessels. But without testing for them, you probably won't know whether you have these conditions. Regular screening can tell you what your numbers are and whether you need to take action.

3. TRIGGERS

TR

What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.

- Lifestyle changes
- Lives depending on medical support
- need to search for heart specialist with manageable price
- need to apply for health insurance
- Financial insecurity
- Anxiety
- shortness of breath
- may feel emotional stress
- may feel chest pain, chest tightness, chest pressure
- feel for fatigue

4. EMOTIONS: BEFORE / AFTER

EM

How do customers feel when they face a problem or a job and afterwards?

i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

- Before a person knows that he/she is affected with any kind of disease, they are happy and do their work normally.
- They don't need to worry about their own body for any problems and do their work normally and comfortably.
- But, after a person comes to know about any kind of problems especially a heart disease, he/she becomes
- ill
- unhealthy
- stressed/depressed
- uncomfortable with their daily routines.
- Lifestyle becomes upside down.

10. YOUR SOLUTION

SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.

If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.

- Heart disease treatment depends on the cause and type of heart damage. Healthy lifestyle habits — such as eating a low-fat, low-salt diet, getting regular exercise and good sleep, and not smoking — are important part of treatment.
- If lifestyle changes alone don't work, medications may be needed to control heart disease symptoms and to prevent complications. The type of medication used depends on the type of heart disease.
- Some people with heart disease may need a procedure or surgery. The type of procedure or surgery will depend on the type of heart disease and the amount of damage to the heart.

8. CHANNELS of BEHAVIOUR

CH

8.1 ONLINE

What kind of actions do customers take online? Extract online channels from #7

- Online appointments with doctors.
- Research about the heart disease they are diagnosed with.
- Finding possible natural cures.

8.2 OFFLINE

What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

- Maintaining proper diet and eating healthy food.
- Having adequate amount of sleep.
- Maintaining a calm and relaxed mind state.
- Following the suggestions made by the doctors.
- Doing exercise and maintaining fitness.
- Taking the right doses of pills at the right time mentioned by doctors.

4. Requirement Analysis

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 |

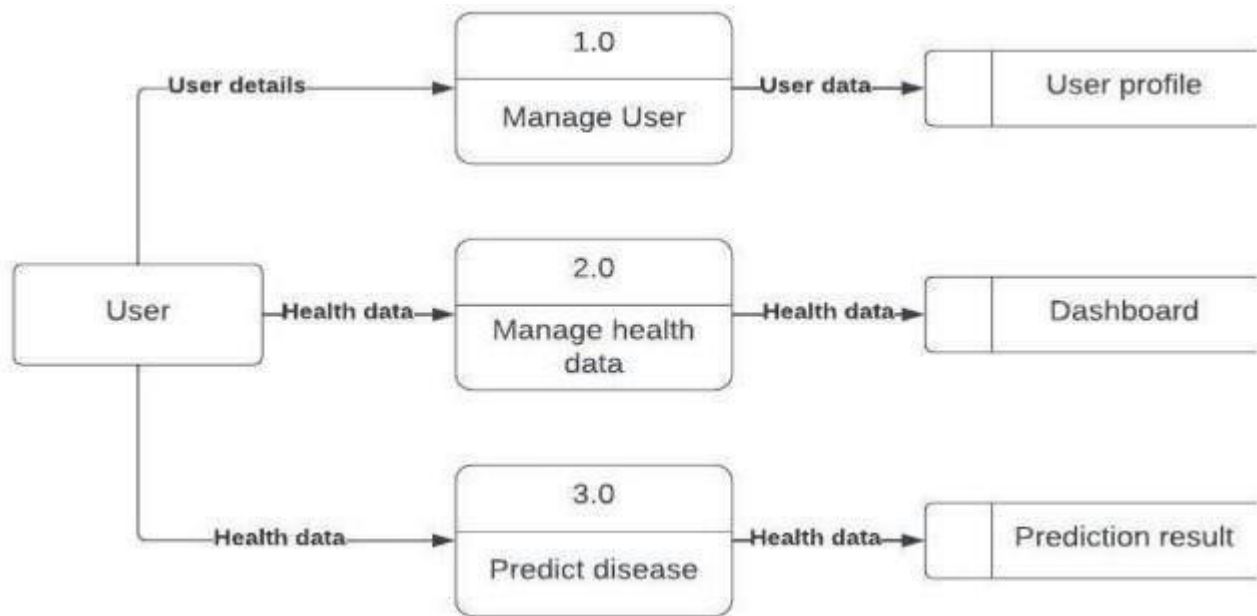
Non-Functional Requirement

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

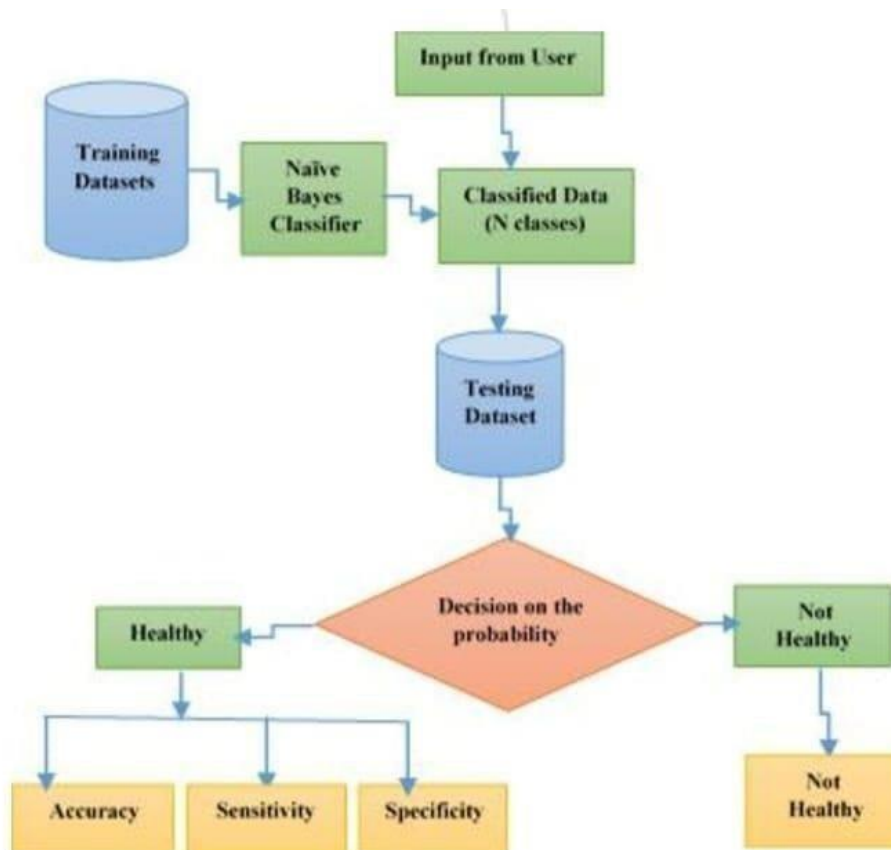
| FR No. | Non-Functional Requirement | Description |
|--------|----------------------------|--|
| NFR-1 | Usability | The application will have a simple and user friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed 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. Incase 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 |

5. Project Design

Data Flow Diagram



Solution and Technical Architecture 5.3



6.Project Planning and Scheduling

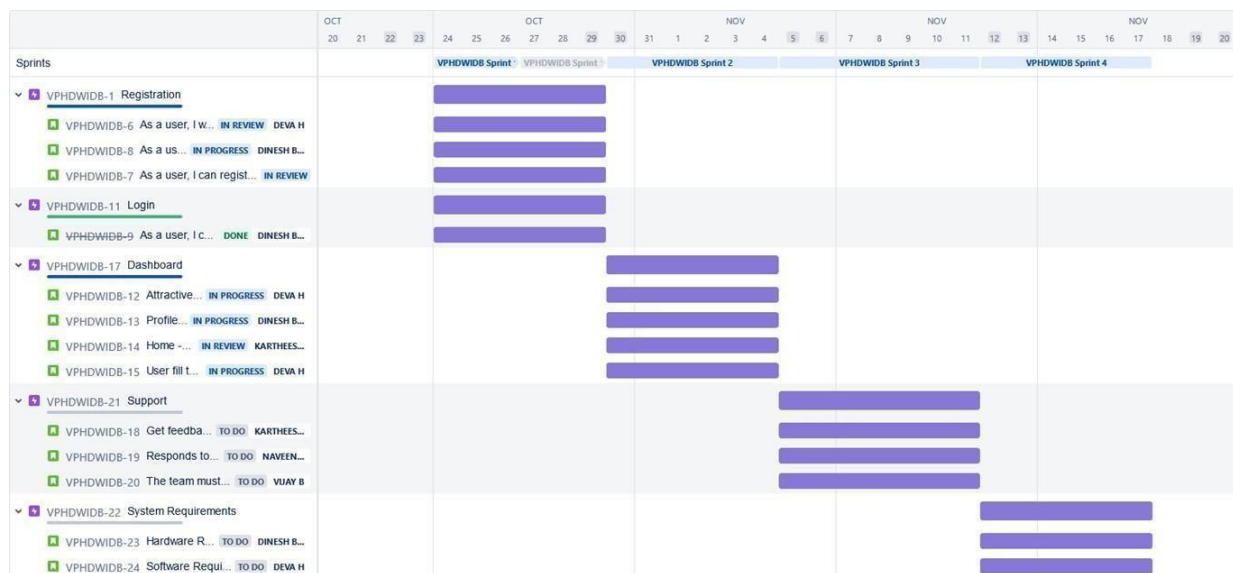
Script Planning and Execution

| 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 | 3 | High | 1 |
| Sprint-1 | | USN-2 | As a user, I will receive confirmation email once I have registered for the application | 3 | High | 3 |
| Sprint-1 | | USN-3 | As a user, I can register for the application through Gmail | 3 | Medium | 1 |
| Sprint-1 | Login | USN-4 | As a user, I can log into the application by entering email & password | 6 | High | 5 |
| Sprint-2 | Dashboard | USN-5 | Attractive dashboard For the Application | 3 | Medium | 3 |
| Sprint-2 | | USN-6 | Profile - view & update your profile | 5 | Low | 2 |
| Sprint-2 | | USN-7 | Home - Analyze your Heart problem | 2 | High | 4 |
| Sprint-2 | | USN-8 | 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 -Exercise Induced Angina -Trust Blood Pressure | 7 | High | 2 |
| Sprint-3 | Support | USN-9 | Get feedback from users | 10 | Medium | 3 |
| Sprint-3 | | USN-10 | Responds to user queries via telephone,email etc. | 3 | Medium | 2 |
| Sprint-3 | | USN-11 | The team must respond immediately to the queries based on the priority | 5 | High | 5 |

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

Machine Learning

Learning which model is best for the given Dataset

Out[]:

| | Estimators | Accuracy |
|---|------------------------|----------|
| 0 | Linear Regression | 0.565830 |
| 3 | K-Nearest Neighbor | 0.729167 |
| 4 | Random Forest | 0.854167 |
| 5 | Bagging Decision Tree | 0.854167 |
| 6 | Hard coting classifier | 0.854167 |
| 2 | Gaussian Naive Bayes | 0.875000 |
| 1 | Logistic Regression | 0.895833 |

From the above result we can conclude that Logistic Regression has the hisgest accuracy for this particular dataset.

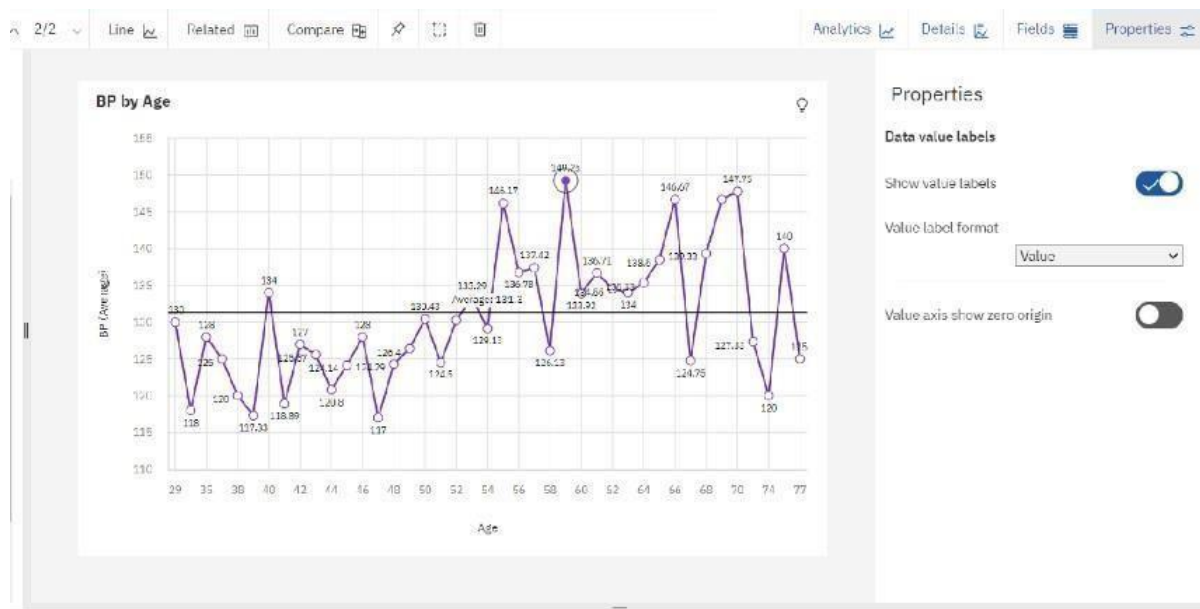
Comparing it with the accuracy gotten from Decision Tree:

```
TP=cm[0][0] #cm=Confusion Matrix
TN=cm[1][1]
FN=cm[1][0] FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN))) print('Testing
Specificity for Decision Tree:',(TN/(TN+FP))) print('Testing Precision for
Decision Tree:',(TP/(TP+FP)))
```

```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

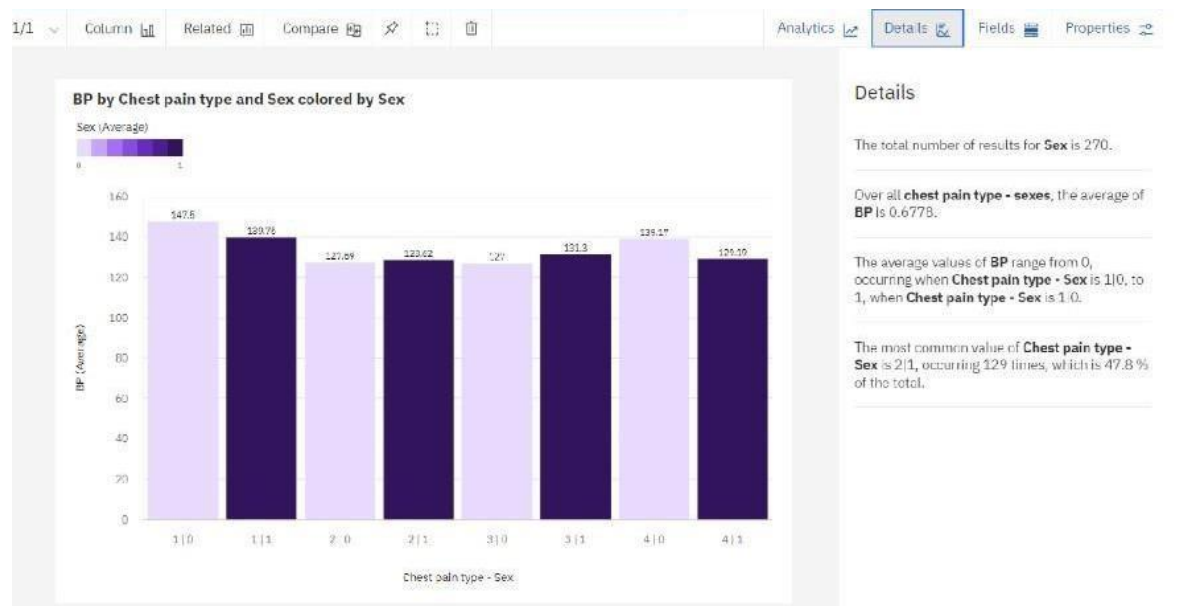
7.2 Dashboard

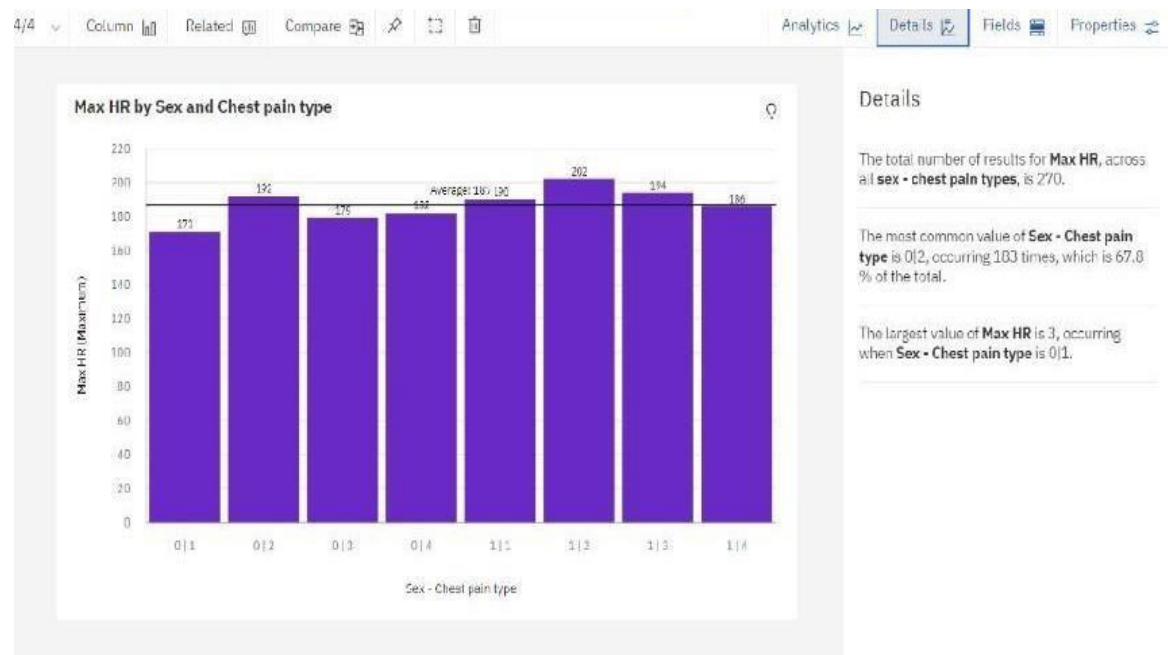
Average BP during chest pain



Exploration Of BPvsChestPainType And Gender:

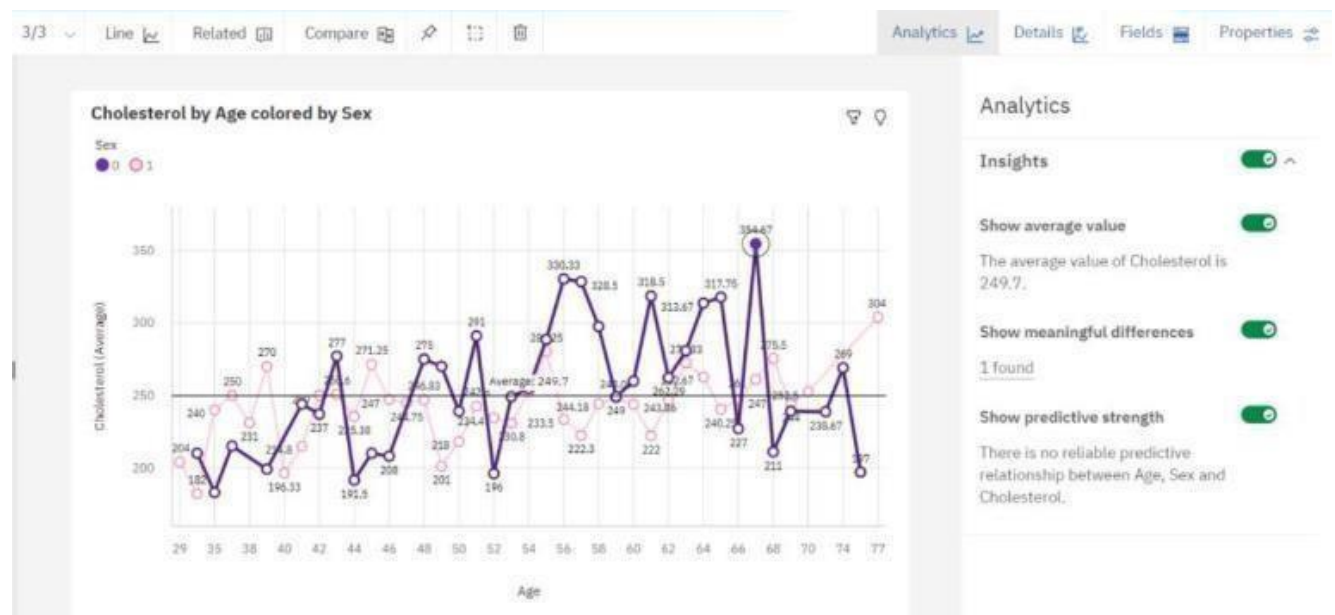
Exploration Of Max Heart Rate During The Chest Pain:





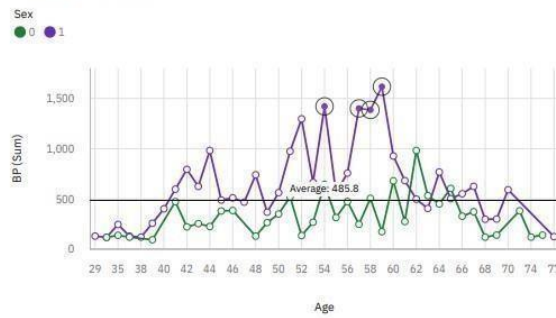
Exploration Of Cholesterol by age and Gender:

Dashboard Showing Different Types Of Visuals:

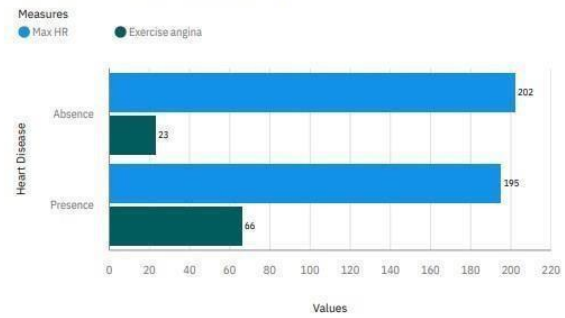


Tab 8

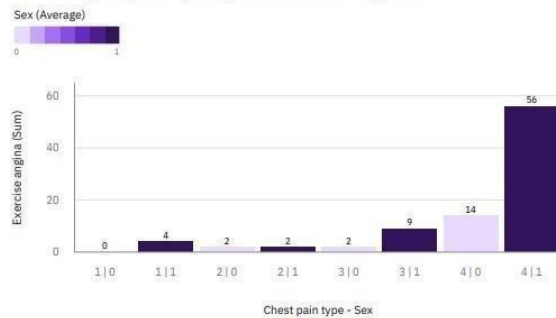
BP by Age colored by Sex



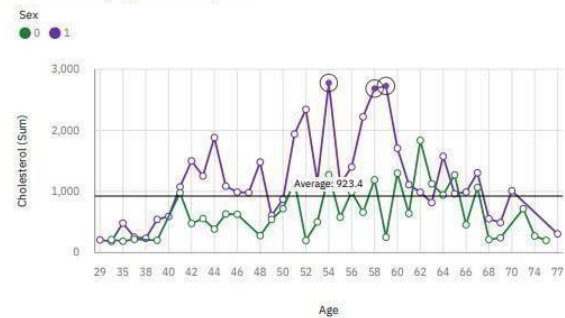
Max HR and Exercise angina by Heart Disease



Exercise angina by Chest pain type and Sex colored by Sex



Cholesterol by Age colored by Sex



8. Testing

Test Cases

Testing the data model for various input values.

```
In [ ]: from sklearn.metrics import accuracy_score
input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
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=(70,1,4,130,322,0,2,109,0,2,4,2,3,3)
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

Testing a case where user has heart disease

localhost:4200

Exercise angina (exercise induced angina (1 = yes; 0 = no))

0

ST depression

2.4

Slope of ST

2

Number of vessels fluro (number of major vessels (0-3) colored by flourosopy)

3

Thallium: 3 = normal; 6 = fixed defect; 7 = reversable defect

3

Submit

localhost:4200 says
The patient has increased risk of heart diseases

OK

Testing a case where user does not have heart disease

localhost:4200

Max HR (maximum heart rate achieved)

160

Exercise angina (exercise induced angina (1 = yes; 0 = no))

0

ST depression

1.6

Slope of ST

2

Number of vessels fluro (number of major vessels (0-3) colored by flourosopy)

0

Thallium: 3 = normal; 6 = fixed defect; 7 = reversable defect

7

Submit

localhost:4200 says
The patient has no risk of heart diseases

OK

9. Result

Performance Metrics

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

```

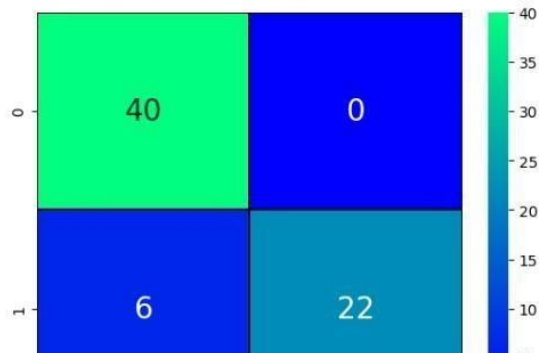
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

tree_model = DecisionTreeClassifier(max_depth=5,criterion='entropy')
cv_scores = cross_val_score(tree_model, x, y, cv=10, scoring='accuracy')
m=tree_model.fit(x, y)
prediction=m.predict(X_test)
cm= confusion_matrix(y_test,prediction)
sns.heatmap(cm, annot=True,cmap='winter',linewidths=0.3, linecolor='black',annot_kws={"size": 20})
print(classification_report(y_test, prediction))

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Absence | 0.87 | 1.00 | 0.93 | 40 |
| Presence | 1.00 | 0.79 | 0.88 | 28 |
| accuracy | | | 0.91 | 68 |
| macro avg | 0.93 | 0.89 | 0.91 | 68 |
| weighted avg | 0.92 | 0.91 | 0.91 | 68 |

Testing Accuracy for Decision Tree: 0.9117647058823529
 Testing Sensitivity for Decision Tree: 0.8695652173913043
 Testing Specificity for Decision Tree: 1.0
 Testing Precision for Decision Tree: 1.0



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: <https://github.com/IBM-EPBL/IBM-Project-41996-1660647459/tree/main/Final%20Deliverables/Source%20Code>

Demolink: <https://github.com/IBM-EPBL/IBM-Project-41996-1660647459/tree/main/Final%20Deliverables>