

Visualizing and Predicting Heart Diseases with an Interactive Dashboard

**NALAIYA THIRAN PROJECT REPORT
2022**

Submitted by

Keerthana S	195001132
Subha R	195001112
Swetha R	195001114
Yaswanth U	195001131

Team ID: PNT2022TMID53144

Content

- 1. INTRODUCTION**
 - 1.1 Project Overview
 - 1.2 Purpose
- 2. LITERATURE SURVEY**
 - 2.1 Existing problem
 - 2.2 References
 - 2.3 Problem Statement Definition
- 3. IDEATION & PROPOSED SOLUTION**
 - 3.1 Empathy Map Canvas
 - 3.2 Ideation & Brainstorming
 - 3.3 Proposed Solution
 - 3.4 Problem Solution fit
- 4. REQUIREMENT ANALYSIS**
 - 4.1 Functional requirement
 - 4.2 Non-Functional requirements
- 5. PROJECT DESIGN**
 - 5.1 Data Flow Diagrams
 - 5.2 Solution & Technical Architecture
 - 5.3 User Stories
- 6. PROJECT PLANNING & SCHEDULING**
 - 6.1 Sprint Planning & Estimation
 - 6.2 Sprint Delivery Schedule
 - 6.3 Reports from JIRA
- 7. CODING & SOLUTIONING (Explain the features added in the project along with code)**
 - 7.1 Feature 1
 - 7.2 Feature 2
- 8. TESTING**
 - 8.1 Test Cases
 - 8.2 User Acceptance Testing
- 9. RESULTS**
 - 9.1 Performance Metrics
- 10. ADVANTAGES & DISADVANTAGES**
- 11. CONCLUSION**
- 12. FUTURE SCOPE**
- 13. APPENDIX**
 - Source Code
 - GitHub & Project Demo Link

TEAM ID :PNT2022TMID53144

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

1. Introduction

1.1 Project Overview

Heart disease is the reason for the death of approximately 17.9 million lives each year. 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 with 13 features as variables to predict the presence/ absence of heart disease.

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. Although there are many systems that can detect heart diseases and other related issues, a system that can monitor heart diseases accurately and gives user-friendly, understandable inferences is a need of the hour.

2. Literature Survey

2.1 Existing Problem

The huge amount of data produced by the healthcare sector contains vast amounts of information that can be used to make decisions. In order to create decisions that are more accurate than intuition, a vast amount of data is used. Errors are found via exploratory data analysis (EDA). identifies pertinent information, verifies presumptions, and establishes the relationship between the explanatory variables. EDA is viewed in this context as analysing data that does not include judgments and statistical analysis Any profession needs analytics since it can predict the future. In the recent past, data analytics has been regarded as a cost-effective technology and it now plays a crucial role in healthcare, including new study discoveries, emergency circumstances, and disease outbreaks. EDA is a crucial step when analysing data, and the application of analytics in healthcare improves treatment by simplifying preventive care.

2.2 References

“Heart Disease Prediction using Exploratory Data Analysis” 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, K- means 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.

“Heart Disease Prediction Using Various Algorithms of Machine Learning” [Rati Goel](#)

This paper uses python libraries with the algorithms of machine learning for detection. In this detection, elements of biological is used such as chest pain (cp), sex, blood pressure(bp), cholesterol(chol). By using of these elements six algorithms of ML such as SVM, KNN, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression is applied for prediction of analysis and conclude that which technique is best on the basis of confusion matrix.

2.3 Problem Statement Definition

Who does the problem affect?

People with unhealthy lifestyles, stress, depression, age above 65 and people with the genes (since heart disease is hereditary).

When does the issue occur?

The issue occurs for people with unhealthy lifestyles and age above 65.

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?

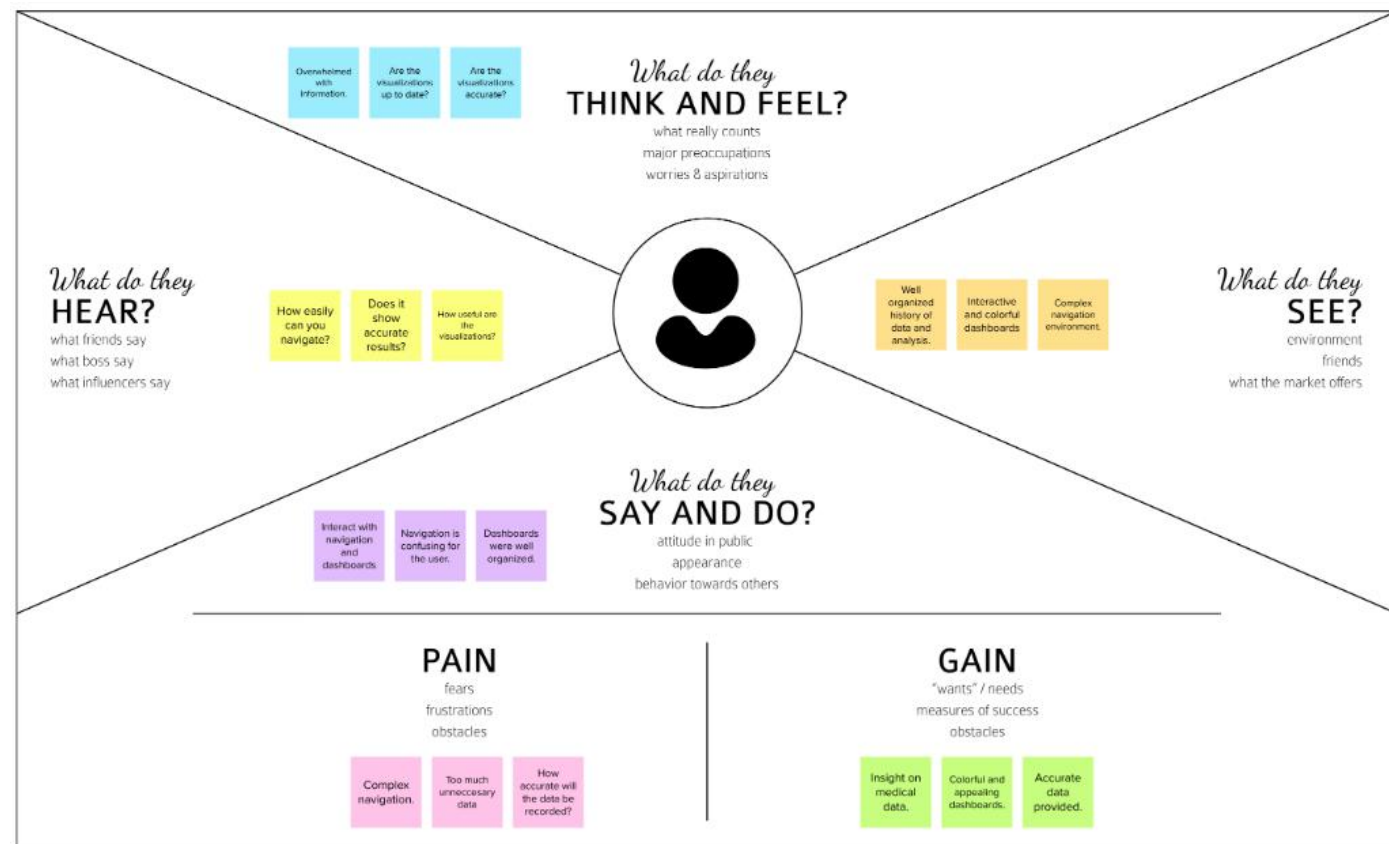
If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will increase rapidly.

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.

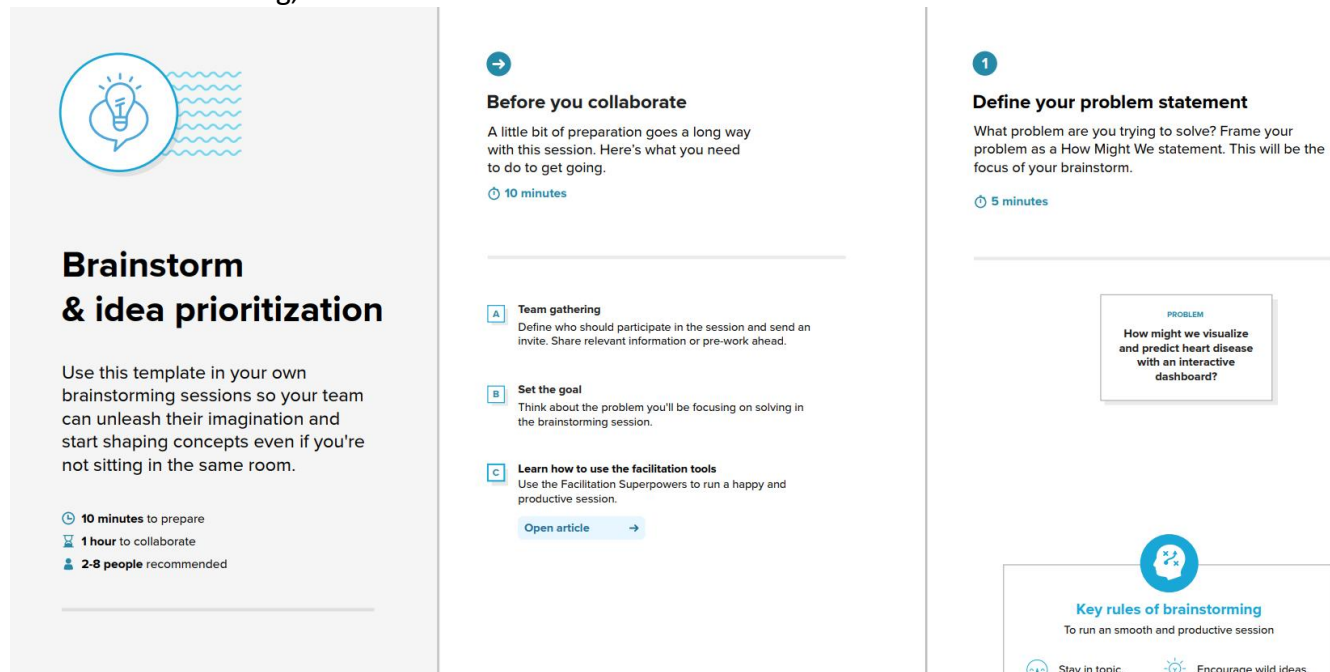
3. Ideation and proposed solution

3.1. Empathy Map Canvas



3.2. Ideation and Brainstorming

3.2.1 Team Gathering, Collaboration and Select the Problem Statement



3.2.2 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 (switch to sketch) icon to start drawing!

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

Keerthana

- A system to predict heart diseases.
- Use of large datasets.
- Early navigable dashboards.
- Use of ML techniques like random forest for prediction.
- Use of tools like IBM cognos for analytics.

Subha

- Use of mathematical and statistical analysis.
- A system to visualize heart diseases.
- User-friendly and easy-to-use dashboards.
- Proper dataset loading mechanism.
- Use of ML techniques like XGBoost for prediction.

Yaswanth

- Use of ML techniques like random forest for prediction.
- Efficient and accurate prediction model.
- Interactive and navigable dashboards.
- System to classify heart diseases.
- Data analysis with other techniques.

Swetha

- Creation of an interactive dashboard for heart diseases.
- Use of colorful and informative charts.
- User-friendly and easy-to-use dashboards.
- Use of tools like power BI for visualization.
- Comparison of accuracy and precision between 2 models.

What the system does

- A system to predict heart diseases.
- Creation of an interactive dashboard for heart diseases.
- A system to visualize heart diseases.
- System to classify heart diseases.

How the system is built

- Use of tools like IBM cognos for analytics.
- Use of mathematical and statistical analysis.
- Use of tools like power BI for visualization.
- Data analysis with other techniques.

ML techniques used

- Use of ML techniques like random forest for prediction.
- Use of ML techniques like gradient boost for prediction.
- Use of ML techniques like support vector machine for prediction.
- Use of ML techniques like decision tree for prediction.
- Comparison of accuracy and precision between 2 models.

Features of dashboard

- Early navigable dashboards.
- User-friendly and easy-to-use dashboards.
- Use of colorful and informative charts.
- Interactive and navigable dashboards.

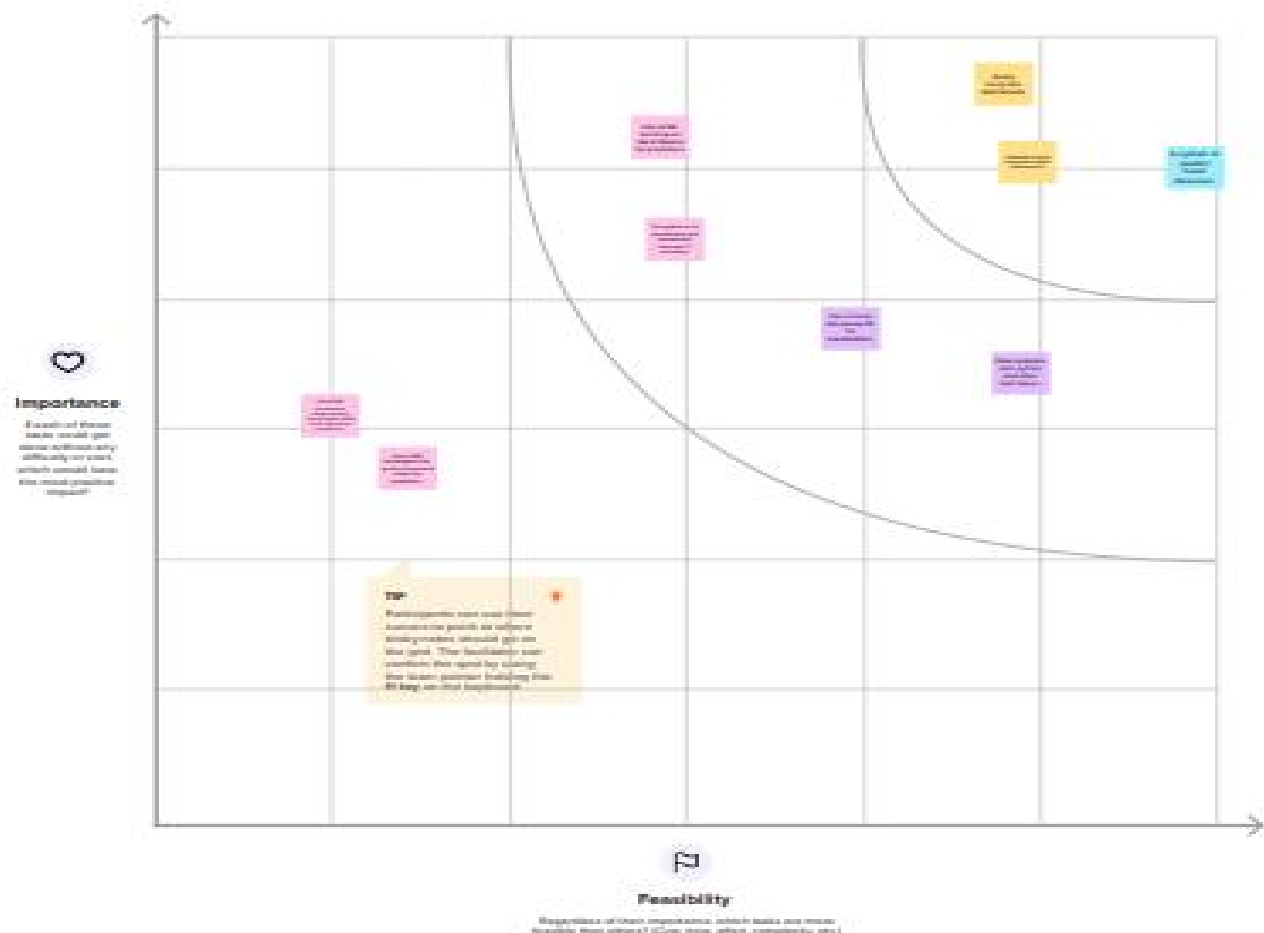
3.2.3 Idea prioritisation



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.

30 minutes



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heart Disease is a major cause of mortality in modern society. Medical diagnosis is extremely important but it is a complicated task that should be performed accurately and efficiently. Cardiovascular diseases are difficult to detect due to several risk factors including blood pressure, cholesterol, and an abnormal pulse rate. Although there are many systems that can detect heart diseases and other related issues, a system that can monitor heart diseases accurately and gives user-friendly, understandable inferences is a need of the hour.
2.	Idea / Solution description	<ul style="list-style-type: none">● Analyzing the history of patients records to reach accurate prediction results regarding heart diseases in an efficient manner● Making the user understand the implications with the help of an interactive, user-friendly dashboard● Complete help and guidance to help cure their disease
3.	Novelty / Uniqueness	<ul style="list-style-type: none">● Complete end-end reliable system which predicts diseases accurately, gives an overall picture of the implications, help and guidance to cure the disease● User-friendly, accurate and a real-time system to solve every patient's need● From young adults to senior citizens, one touch solution to monitor their health
4.	Social Impact / Customer Satisfaction	<ul style="list-style-type: none">● One can keep track of their own health and wellbeing● Conscious young adults and citizens who monitor their health periodically● This simple system can save many lives, help people prevent major health issues, guide them on how to cure the disease

5.	Business Model (Revenue Model)	<ul style="list-style-type: none"> Generate revenue by selling dashboards to hospitals, diagnostics & clinical centers Smartwatch companies can use this dashboard as an application
6.	Scalability of the Solution	IBM cloud can help scale the solution.

3.4 Problem Solution Fit

Project Design Phase I - Solution Fit Document

Team ID: PNT2022TMD53144

Problem-Solution Fit canvas

Project Title: Visualising and Predicting Heart Diseases with an interactive dashboard

1. CUSTOMER SEGMENT(S) <small>CS</small> Who is our customer? People who suffer from heart related issues use our system to diagnose and get results and proper cure.	6. CUSTOMER LIMITATIONS <small>CL</small> <small>EG. BUDGET, DEVICES</small> The cost of treatment, fear about the complexity of the disease limit customers from taking up an action.	5. AVAILABLE SOLUTIONS <small>AS</small> <small>PLUSES & MINUSES</small> Although there are many systems that can detect heart diseases and other related issues, a system that can monitor heart diseases accurately and gives user-friendly inferences is a need of the hour.
2. PROBLEMS / PAINS <small>PR</small> <small>+ ITS FREQUENCY</small> Heart Disease is a major cause of mortality in modern society. Medical diagnosis is extremely important but it is a complicated task that should be performed accurately and efficiently. Although there are many systems that can detect heart diseases and other related issues, a system that can monitor heart diseases accurately and gives user-friendly, understandable inferences is a need of the hour.	9. ROOT / CAUSE OF PROBLEM <small>RC</small> Cardiovascular diseases are difficult to detect due to several risk factors including blood pressure, cholesterol, and an abnormal pulse rate. Other causes include unhealthy diet, stress and depression and family history.	7. BEHAVIOR <small>BE</small> <small>+ ITS INTENSITY</small> The patient is expected to report the symptoms and issues they are facing. The system undergoes a deep analysis and gives the inference results. The patient understands the implication of the disease and is instructed with some guidelines on how to cure the disease.
3. TRIGGERS TO ACT <small>TR</small> By seeing the advanced technology providing a low cost, efficient and accurate results to their health issues.	10. YOUR SOLUTION <small>SL</small> <ul style="list-style-type: none"> Analyzing the history of patients records to reach accurate prediction results regarding heart diseases in an efficient manner Making the user understand the implications with the help of an interactive, user-friendly dashboard Complete help and guidance to help cure their disease 	8. CHANNELS of BEHAVIOR <small>CH</small> ONLINE They can check and get a complete report of the disease online, in low cost. OFFLINE They can visit a doctor with the reports collected and get treatment right away.
4. EMOTIONS <small>EM</small> <small>BEFORE / AFTER</small> While facing a health issue, one might feel lonely, depressed and insecure. But the system promises to improve their mental wellbeing.		

Problem-Solution Fit canvas is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. Designed by Denis Negropolina / ideahackers.ru - we tailor ideas to customer behaviour and increase solution adoption probability.

4. Requirement Analysis

4.1 Functional Requirements

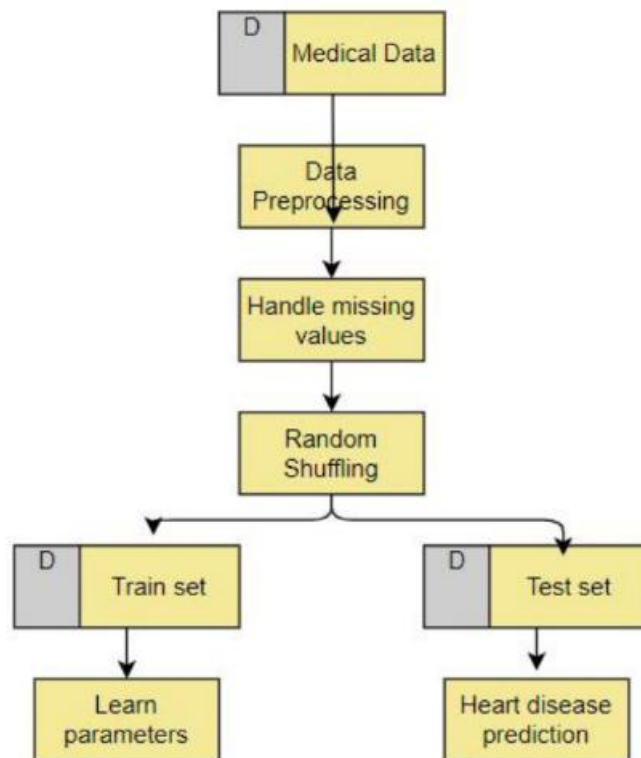
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Google Registration through Mail Registration through Form
FR-2	User Confirmation	Confirmation via Email Confirmation via SMS
FR 3	Patient details	Collection of patient details through HTML form.
FR-4	Analysis of patient data	Analyzing and Visualizing the relationship between various features of patient data.
FR-5	Feature Selection	Selecting the most relevant features and getting rid of noise in data.
FR-6	Prediction	Displaying the result of analysis.

4.2. Non Functional Requirement

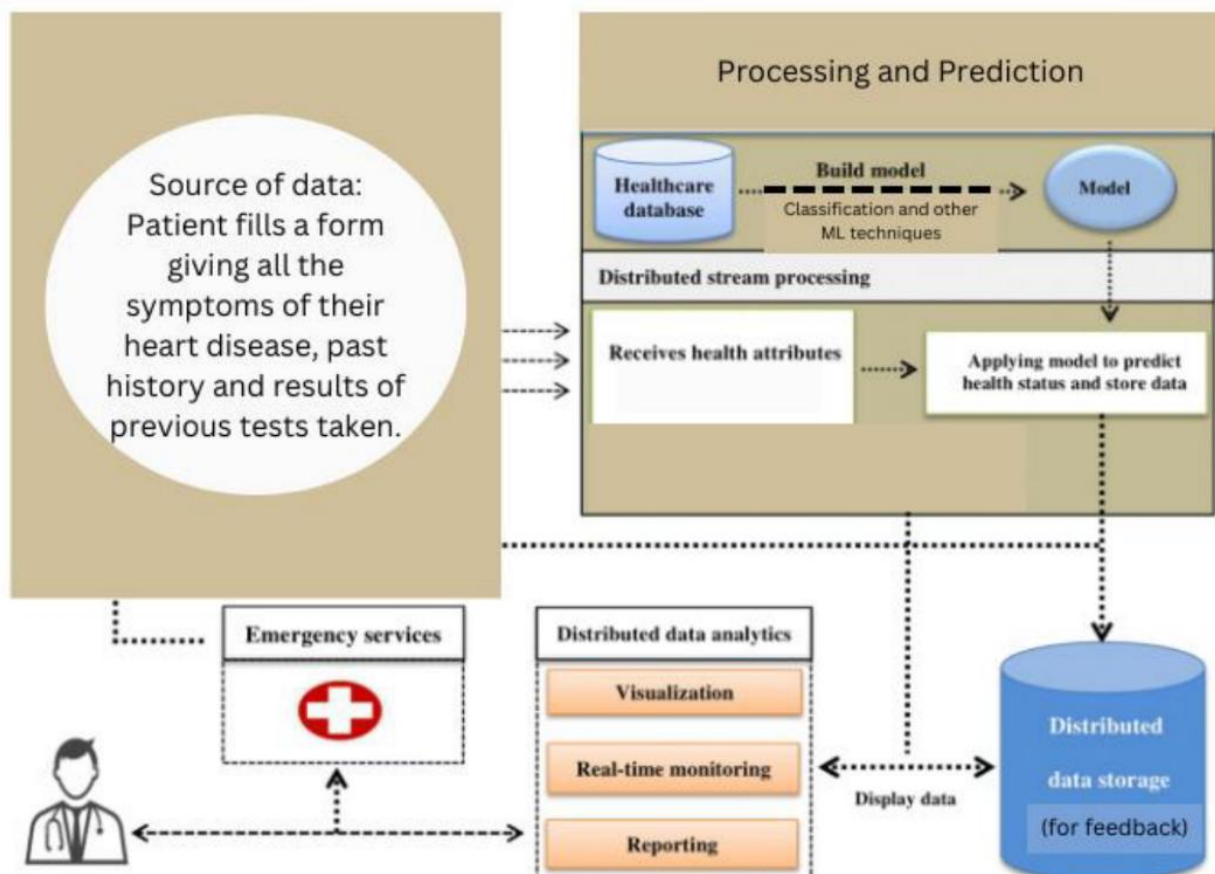
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge like the relationships between medical factors related to heart disease and patterns to be established.
NFR-2	Security	When it comes to health-related issues, we should offer greater security services. While developing the software or application, there shouldn't be any errors, lagging, or bases of data for patient profiles.
NFR-3	Reliability	Support vector machine (SVM), Gaussian Naive Bayes, logistic regression, and random forest algorithm have been employed for developing heart disease risk prediction model and obtained the accuracy as 80.32%, 78.68%, 80.32% and 88.5%, respectively
NFR-4	Performance	The relaying performance needs to be quick. . To enable improved accessibility and make a significant advancement in the provision of high-quality, reasonably priced healthcare, this prediction system should be made available on the cloud.
NFR-5	Availability	By setting up An Application Performance Monitoring (APM) system that helps to monitor the availability of application. Consistent performance monitoring and optimization help you to tackle issues as quickly as they show up. Our app is designed in such a way that to emphasize availability by spreading data across clusters so that if one fails the entirety of the data is not lost..
NFR - 6	Scalability	A scalable app can easily accommodate double, triple, or even ten times its current amount of users by withstanding no crashes, no downtime, Fast loading speeds, Top -notch security. We're gonna make our app more scalable by using right Tech stack & Infrastructure scaling to process millions of data with bug free , multiple database servers that accommodate millions of user to secure our app's fail -safe performance, using caching and stateless approach to reduce the load, Content Delivery Networks (CDN) to minimal response time.

5. Project Design

5.1. Data Flow Diagram



5.2. Solution and Technical Architecture



6. Project Planning and scheduling

6.1. Script planning and execution

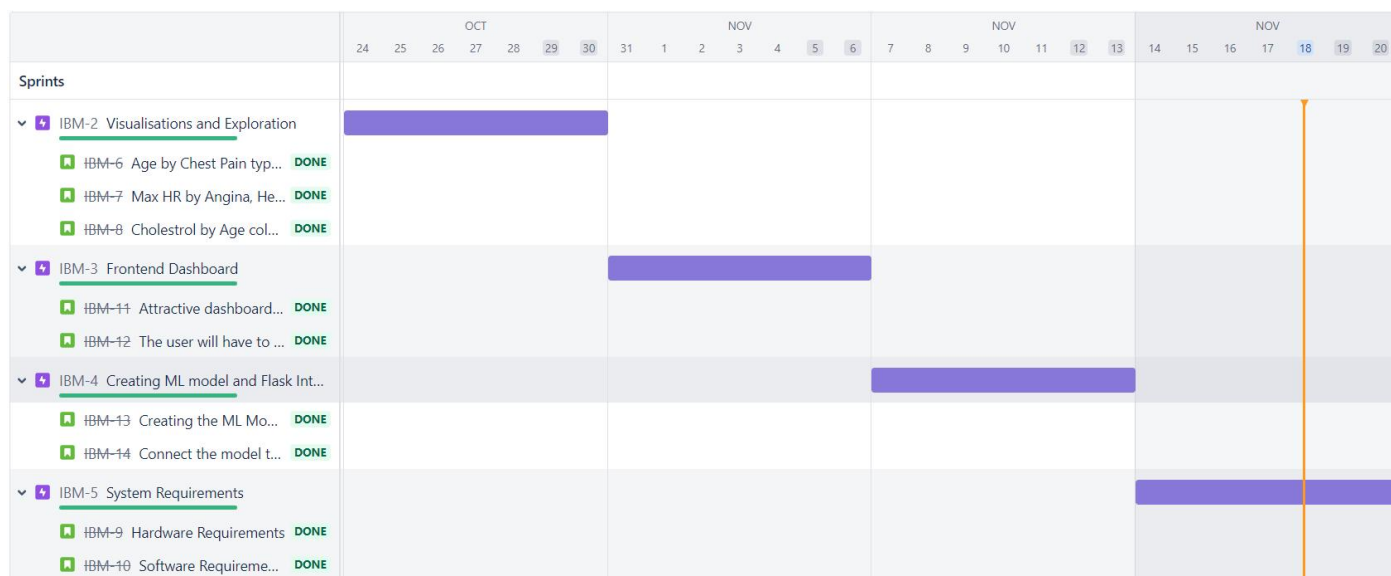
Sprint	Functional Requirement	User Story number	User story/task	Story points	Priority	Team Members
Sprint 1	Visualizations and Exploration	USN-1	Need to see visualisations of Age by Chest pain type, BP by age, Exercise angina by chest pain type.	3	High	2
Sprint 1		USN-2	Visualisations for Max HR by Angina, Heart disease for chest pain type and sex.	3	High	2
Sprint 1		USN-3	Visualisations for cholesterol by age colored by sex, Max HR and angina by heart disease.	3	High	2
Sprint 2	Front-end Dashboard	USN-4	Attractive dashboard for the application	3	Medium	2
Sprint 2		USN-5	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	Creating ML Model and Flask Interface	USN-6	Creating the ML model	6	High	1
Sprint 3		USN-7	Connect the model to the interface using Flask	6	High	1
Sprint 4	System requirements	USN-8	Hardware Requirement 3. Laptop or PC • i5 processor system or higher	5	Low	2

			<ul style="list-style-type: none"> • 4 GB RAM or higher • 128 GB ROM or higher 4.Mobile <ul style="list-style-type: none"> • (Android 12.0 and above) 			
Sprint 4		USN-9	Software requirement: Laptop or PC <ul style="list-style-type: none"> • Windows 10 or higher • Android Studio 	5	Medium	4

6.2 Sprint delivery schedule

Sprint	Total Delivery points	Duration	Sprint start date	Sprint end date	Story points completed	Sprint release date
Sprint 1	9	1 week	24 th Oct, 2022	30 th Oct, 2022	9	30 th Oct, 2022
Sprint 2	10	1 week	31 st Oct, 2022	06 th Nov, 2022	10	06 th Nov, 2022
Sprint 3	12	1 week	07 th Nov, 2022	13 th Nov, 2022	12	13 th Nov, 2022
Sprint 4	10	1 week	14 th Nov, 2022	20 th Nov, 2022	10	20 th Nov, 2022

6.3 Jira Report



7 Coding And Solutioning

7.1 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.

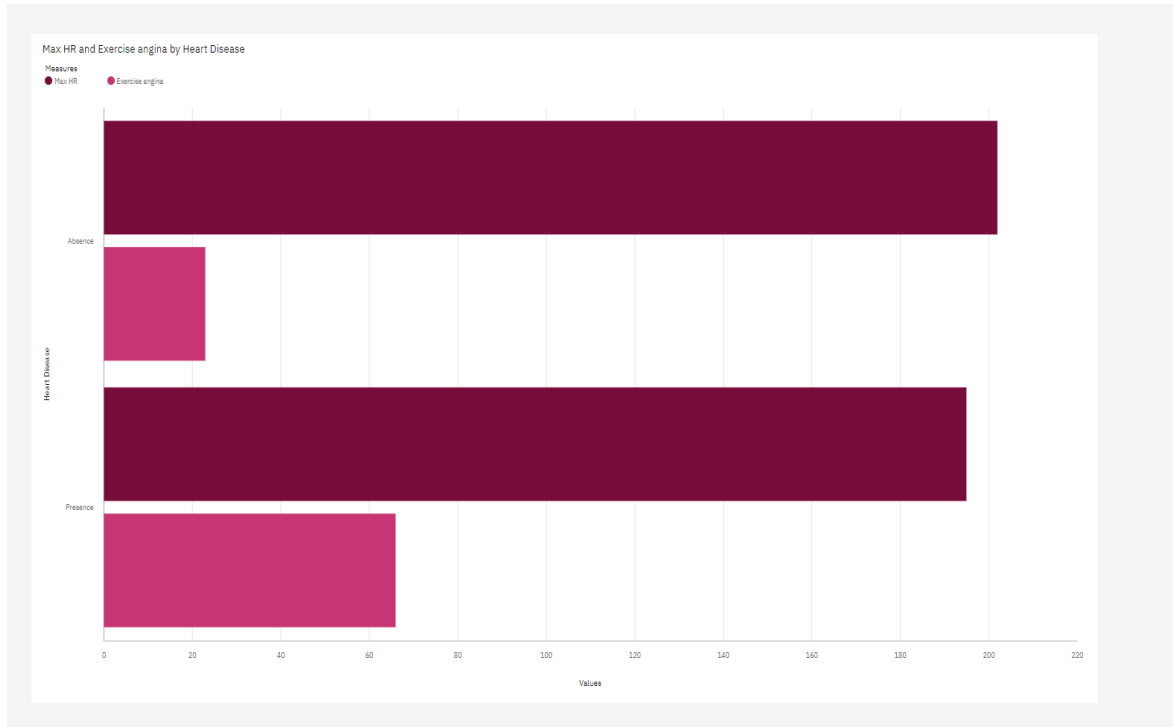
Comparing it with the accuracy gotten from Decision Tree:

```
TP=cm[0][0]
#cm=Confusi
on 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)))
```

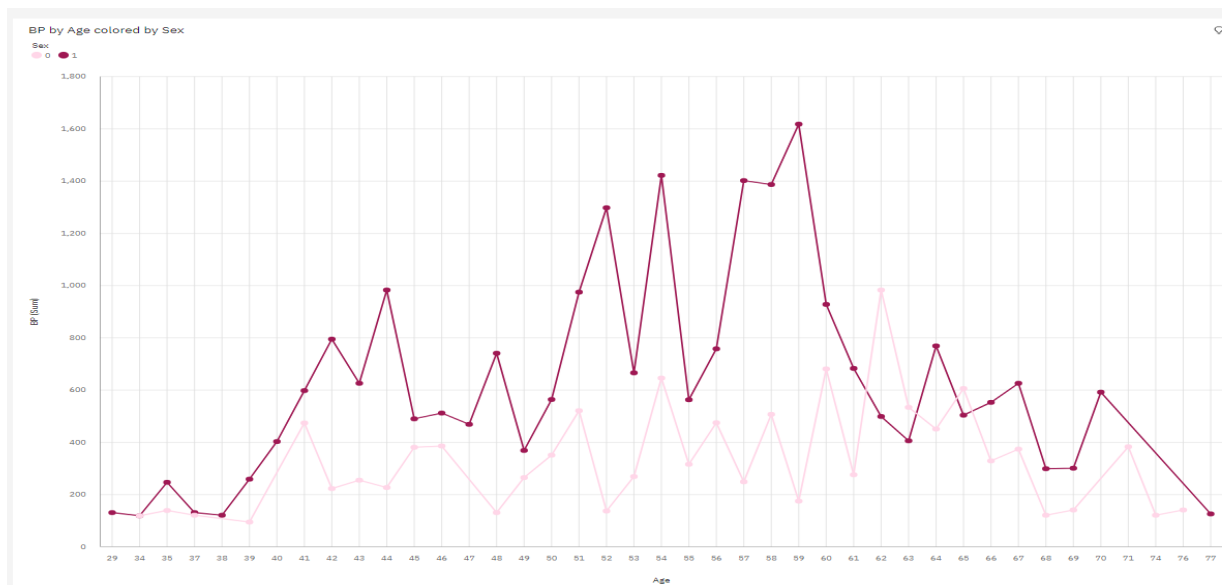
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

7.2 Dashboard

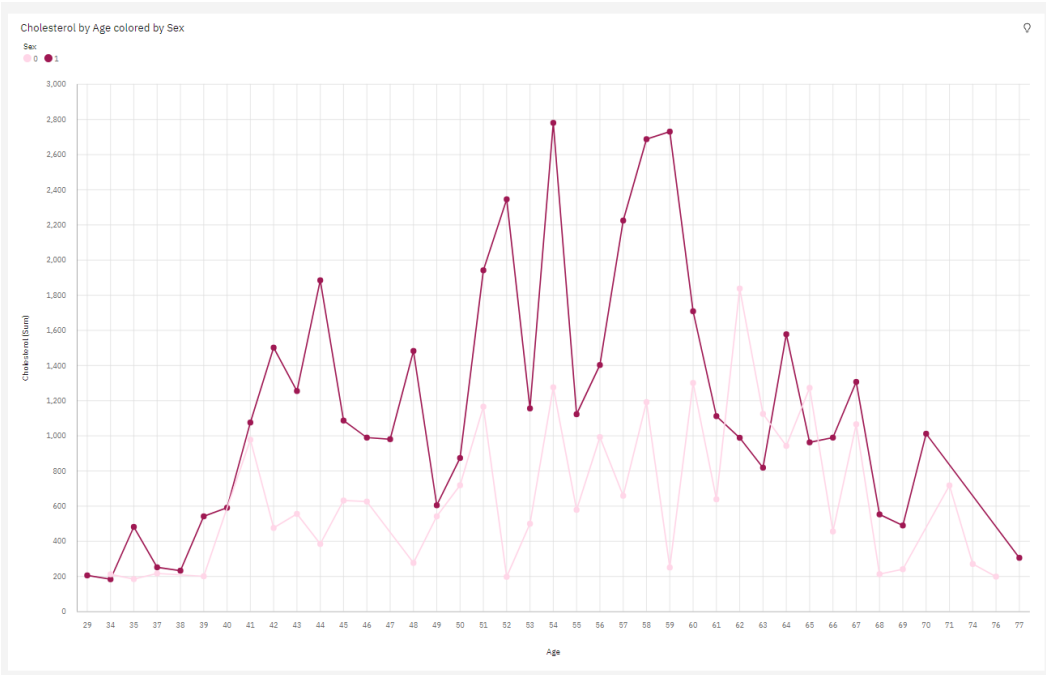
Max Heart Rate and Exercise Angina by Heart Disease



Average BP by Age



Cholesterol by Age



8. Test cases

8.1 User acceptance testing

Case 1: Person is not likely to have heart disease

Home Dashboard

Patient Health Record

Age
67

Sex
0
Enter 1 for male and 0 for female

Chest pain type
3
1: typical angina
2: atypical angina
3: non-anginal pain
4: asymptomatic

BP
115
Resting blood pressure mm of Hg

Cholesterol
564
Serum cholesterol in mg/dl

FBS over 120
0

EKG results
2
Resting electrocardiographic results
0: normal
1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
2: showing probable or definite left ventricular hypertrophy by Estes' criteria

Max HR
160
Maximum heart rate achieved

Exercise angina
0
Exercise induced angina
1 = yes; 0 = no

ST depression
1.8
ST depression induced by exercise relative to rest



Slope of ST

Slope of the peak exercise ST segment
 1: upsloping
 2: flat
 3: downsloping

Number of vessels fluro

Number of major vessels (0-3) colored by flourescopy

Thallium

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

The patient is not likely to have heart disease!

[Home](#) [Dashboard](#)



Patient Health Record

Age

Sex

Enter 1 for male and 0 for female

Chest pain type

1: typical angina
 2: atypical angina
 3: non-anginal pain
 4: asymptomatic

BP

Case 2: Person is likely to have heart disease

[Home](#) [Dashboard](#)



Patient Health Record

Age

Sex

Enter 1 for male and 0 for female

Chest pain type

1: typical angina
 2: atypical angina
 3: non-anginal pain
 4: asymptomatic

BP

Resting blood pressure mm of Hg

Cholesterol

Serum cholesterol in mg/dl

FBS over 120

Fasting blood sugar > 120 mg/dl
1 = true; 0 = false

EKG results

Resting electrocardiographic results
0: normal
1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
2: showing probable or definite left ventricular hypertrophy by Estes' criteria

Max HR

Maximum heart rate achieved

Exercise angina

Exercise induced angina
1 = yes; 0 = no

ST depression

ST depression induced by exercise relative to rest

Slope of ST

Slope of the peak exercise ST segment
1: upsloping
2: flat
3: downsloping

Number of vessels fluro

Number of major vessels (0-3) colored by flourosocopy

Thallium

3 = normal; 6 = fixed defect; 7 = reversible defect

The patient is likely to have heart disease!

[Home](#) [Dashboard](#)

Patient Health Record

Age

Sex

Enter 1 for male and 0 for female

9.Result

9.1 Performance Metrics

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

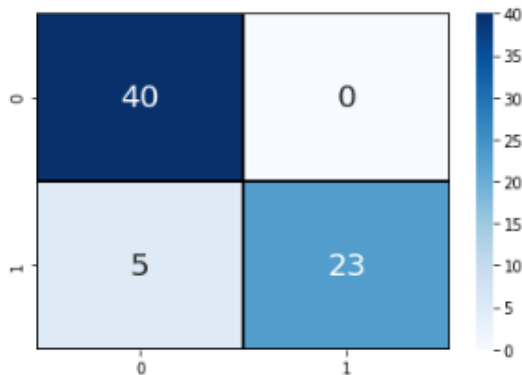
```
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

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="Blues",linewidths=0.3, linecolor='black',annot_kws={"size": 20})
print(classification_report(y_test, prediction))

TP=cm[0][0]
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)))
```

	precision	recall	f1-score	support
0	0.89	1.00	0.94	40
1	1.00	0.82	0.90	28
accuracy			0.93	68
macro avg	0.94	0.91	0.92	68
weighted avg	0.93	0.93	0.93	68

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



10 Advantages Disadvantages

Advantages:

- User Friendly
- Easy to understand
- Secure
- 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
- Dashboard provides insightful informations

Disadvantages:

- Needs lots of information and effort from the user
- Users need to know all the fields
- Does Not take null value as input
- Does not provide suggestions for treatment 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 result that users get from the website might help save patients as it is always better 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. Also we could provide suggestions for the treatment of the disease. We could also try to reduce the number of features required as it may be vexing to the user.

13 Appendix

Source Code:

<https://github.com/IBM-EPBL/IBM-Project-17191-1659630221/tree/main/Final%20Deliverables>

Demo video link:

<https://drive.google.com/file/d/1M6c4CXYzWokwbpGo9x4PsfGKzsthcHRK/view?usp=sharing>