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

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, Ligistic 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?

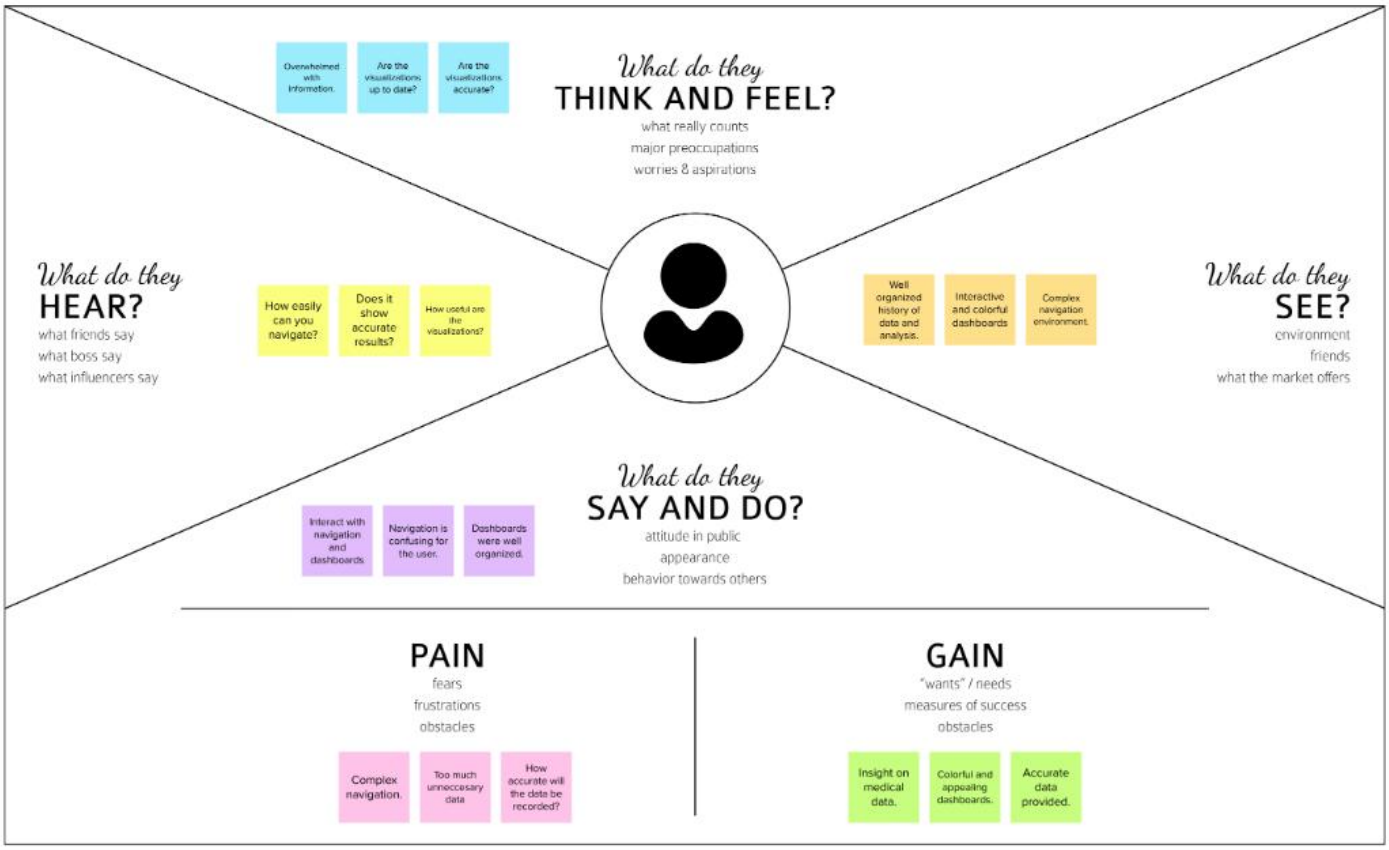
If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will increase rapidly.

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.

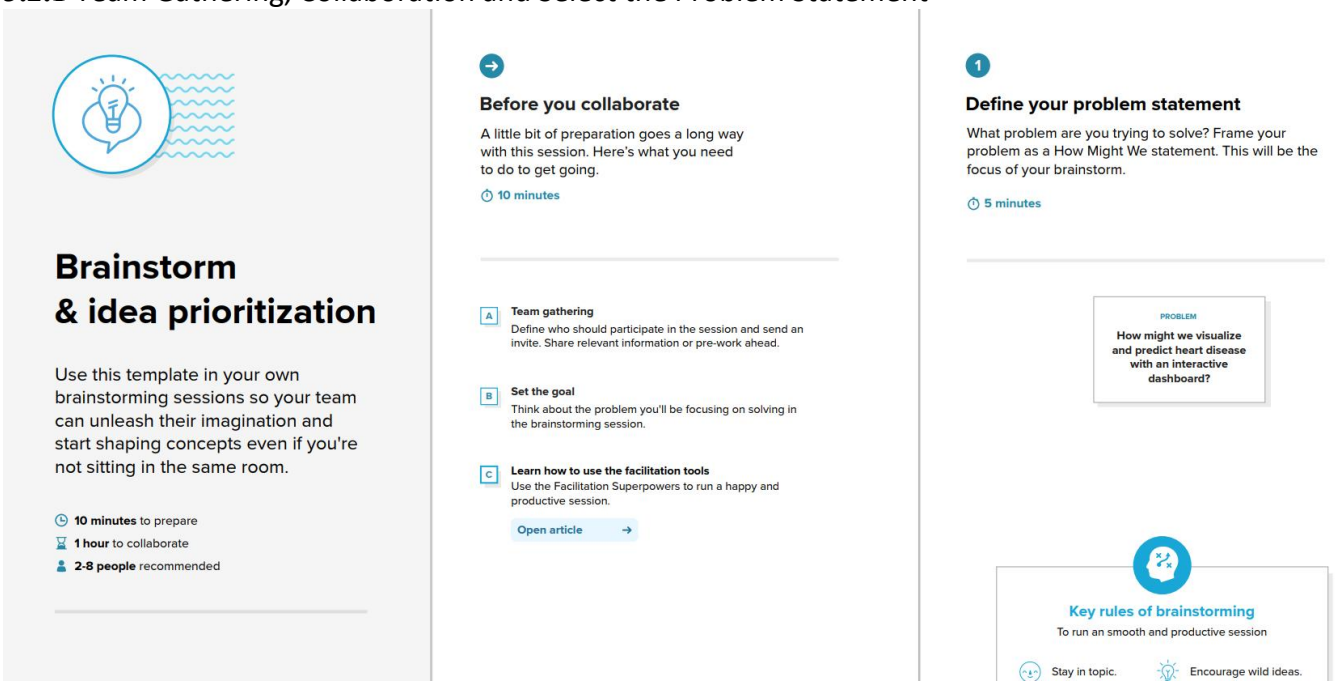
3. Ideation and proposed solution

3.1. Empathy Map Canvas

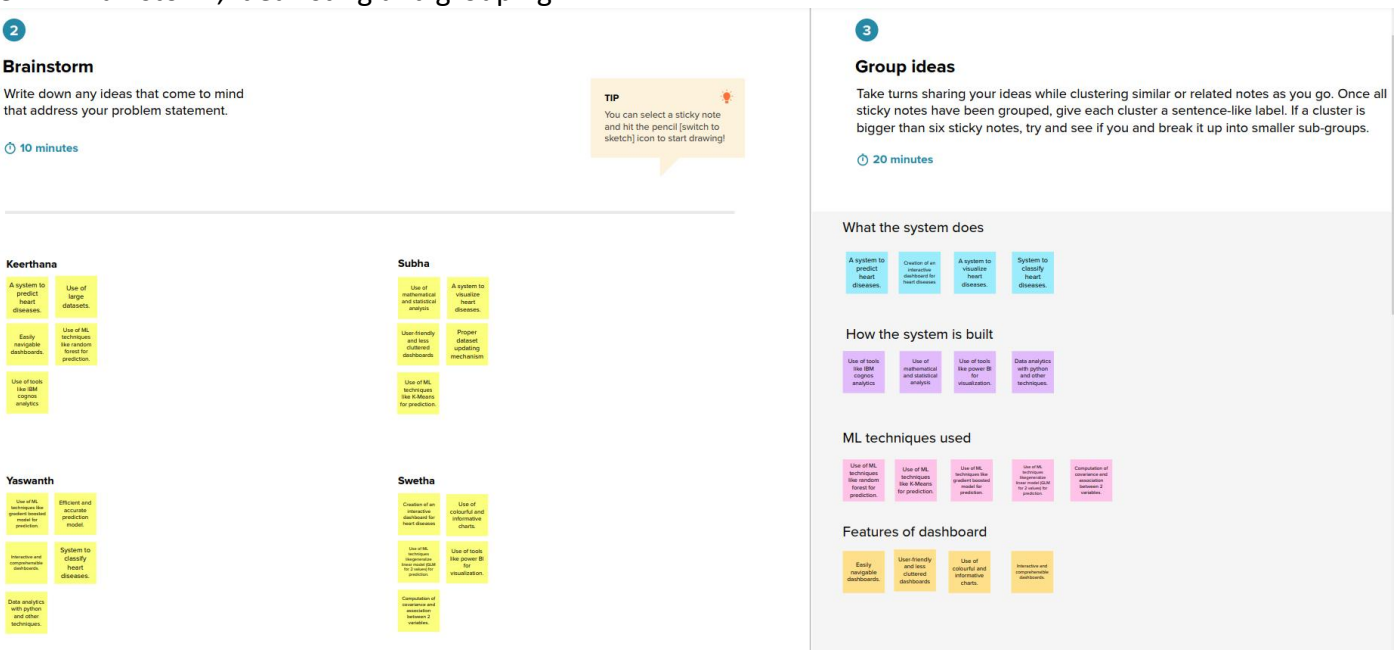


3.2. Ideation and Brainstorming

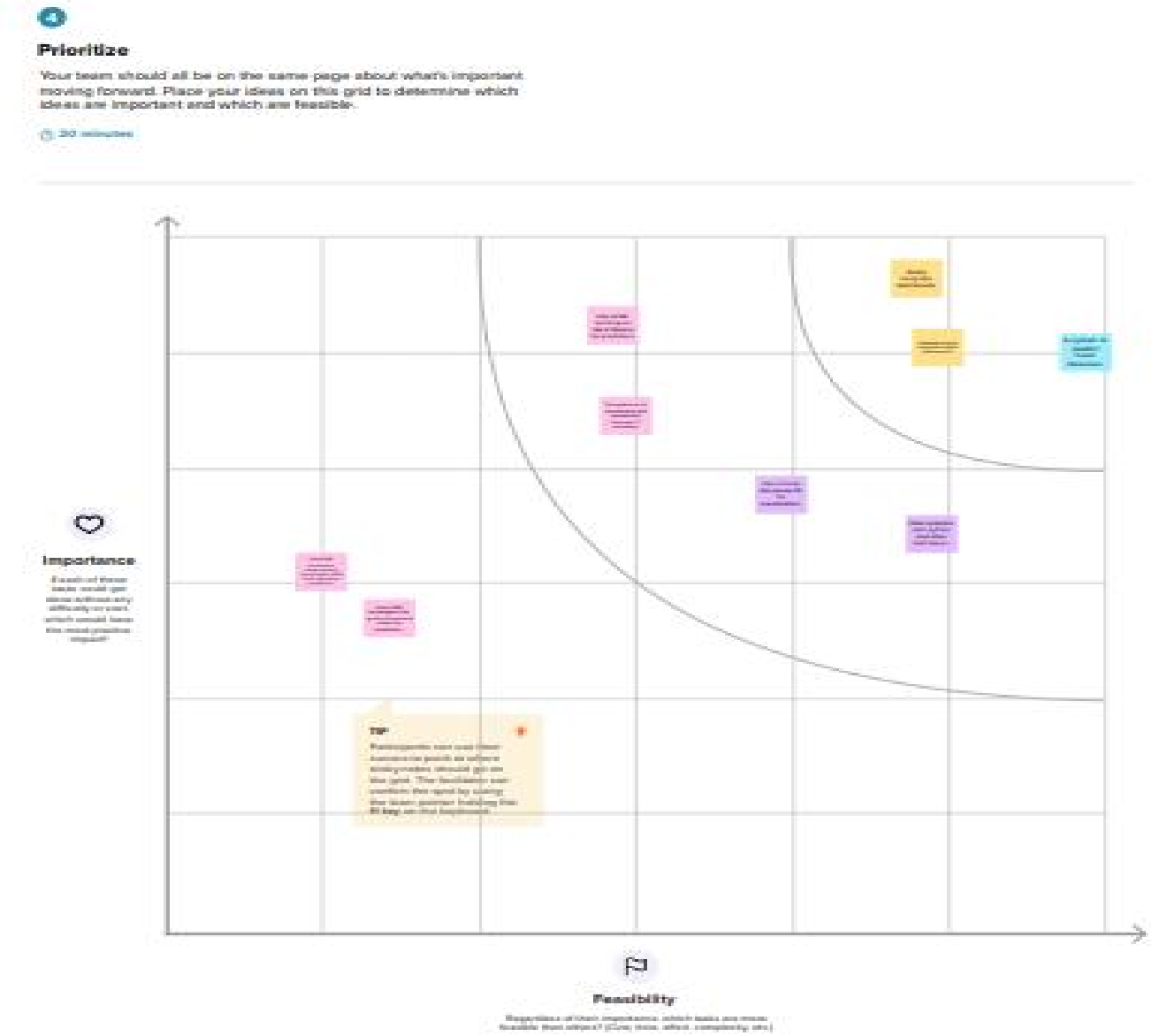
3.2.1 Team Gathering, Collaboration and Select the Problem Statement



3.2.2 Brainstorm, Idea listing and grouping



3.2.3 Idea prioritisation

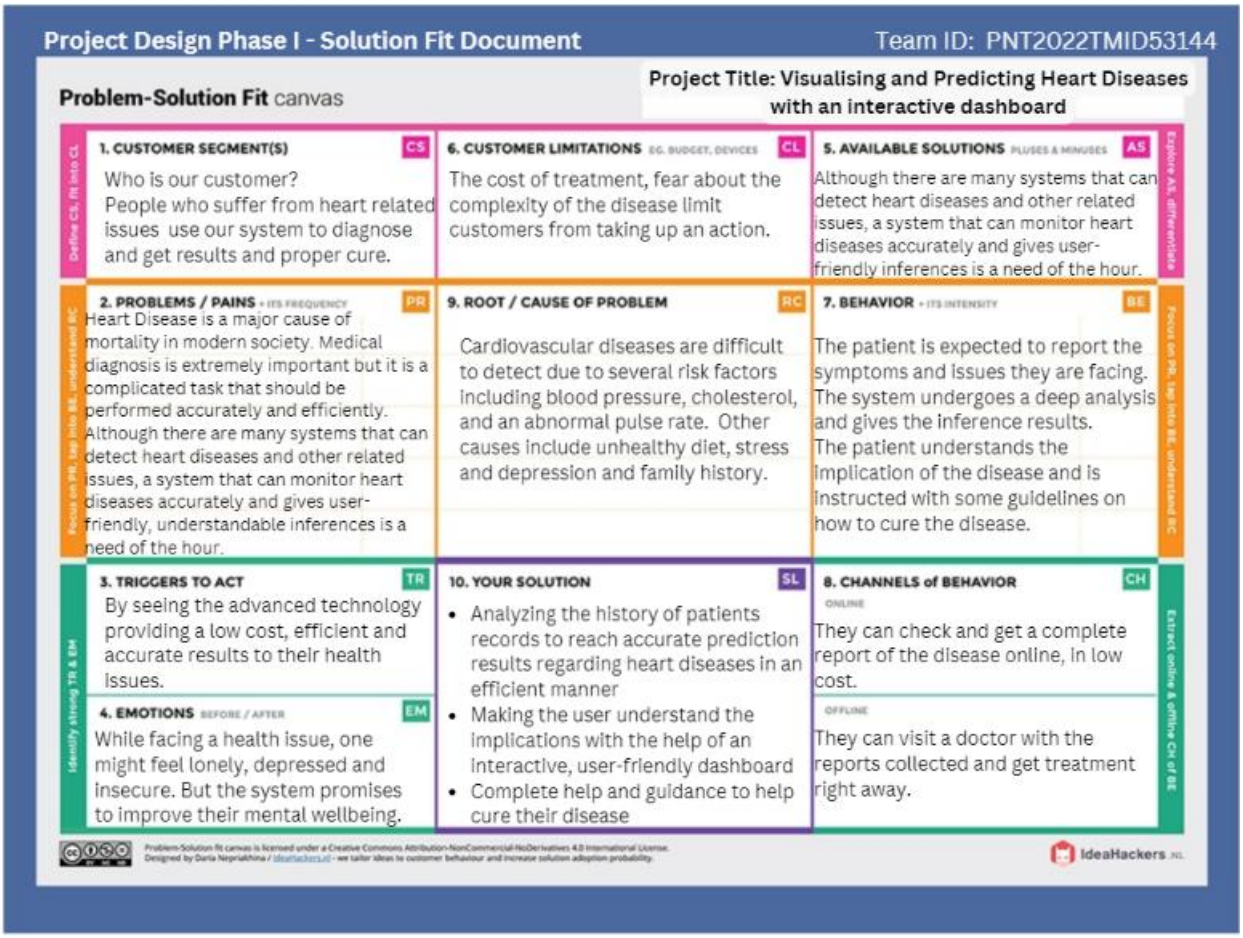


3.3 Proposed Solution

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	• 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	• 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	• 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)	• 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



4. Requirement Analysis

4.1 Functional Requirements

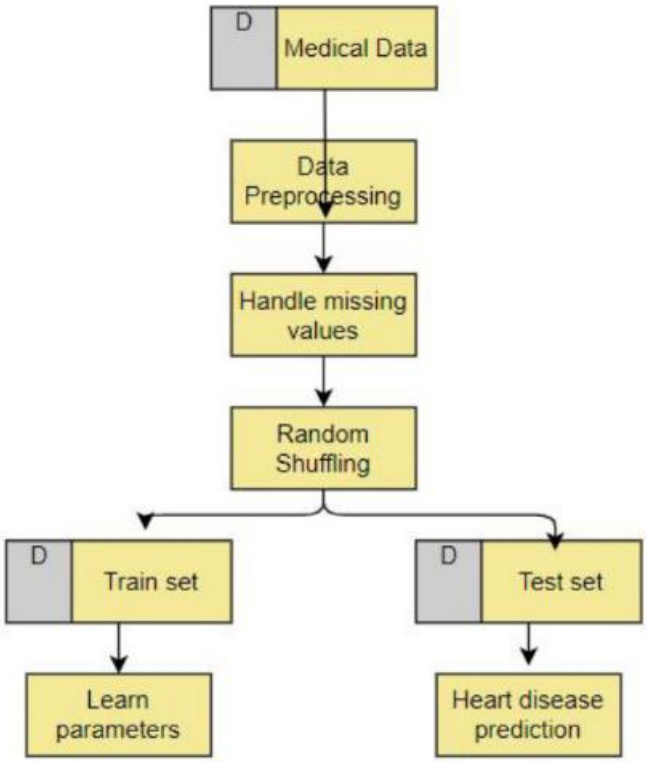
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Google Registration through Mail Registration through Form
FR-2	User Confirmation	Confirmation via Email Confirmation via SMS
FR 3	Patient details	Collection of patient details through 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

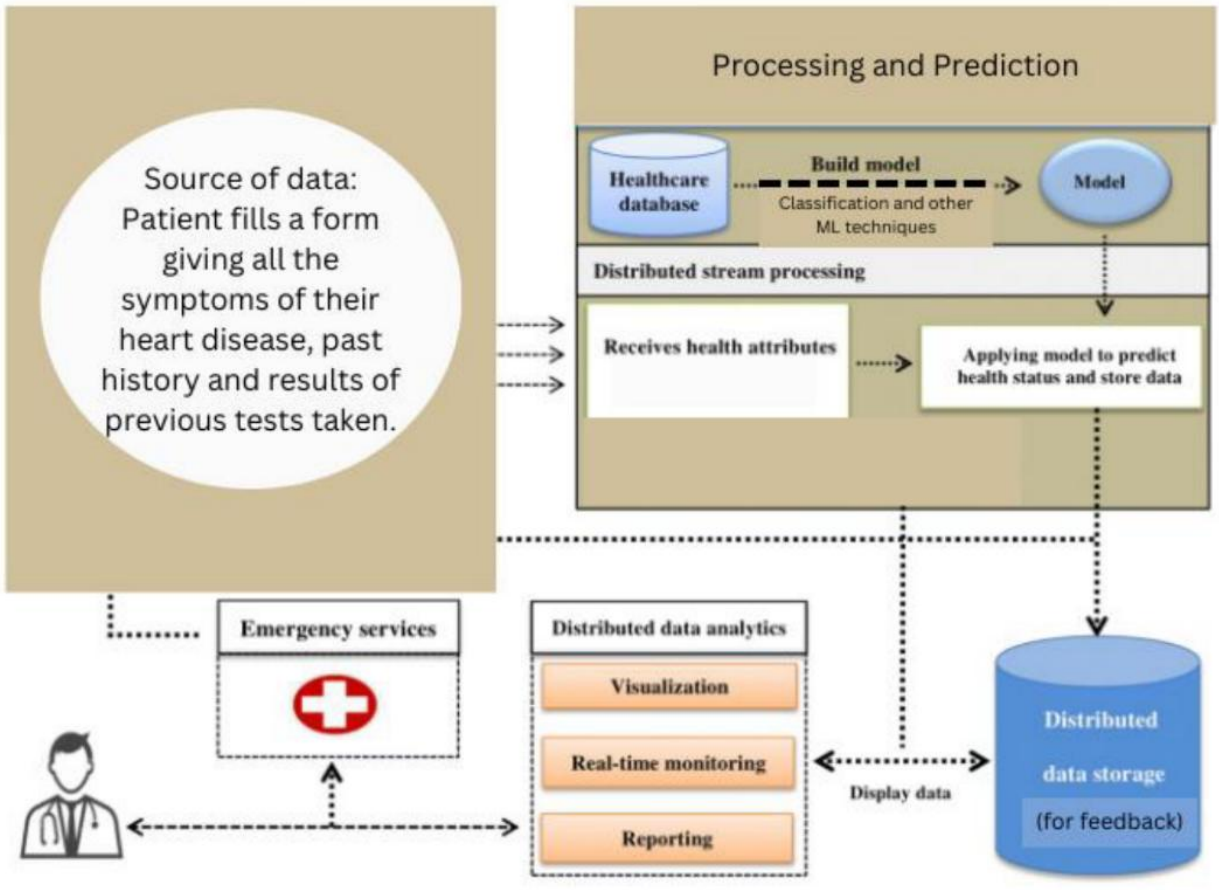
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge like the relationships between medical
		factors related to heart disease and patterns to be established.
NFR-2	**Security**	When it 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

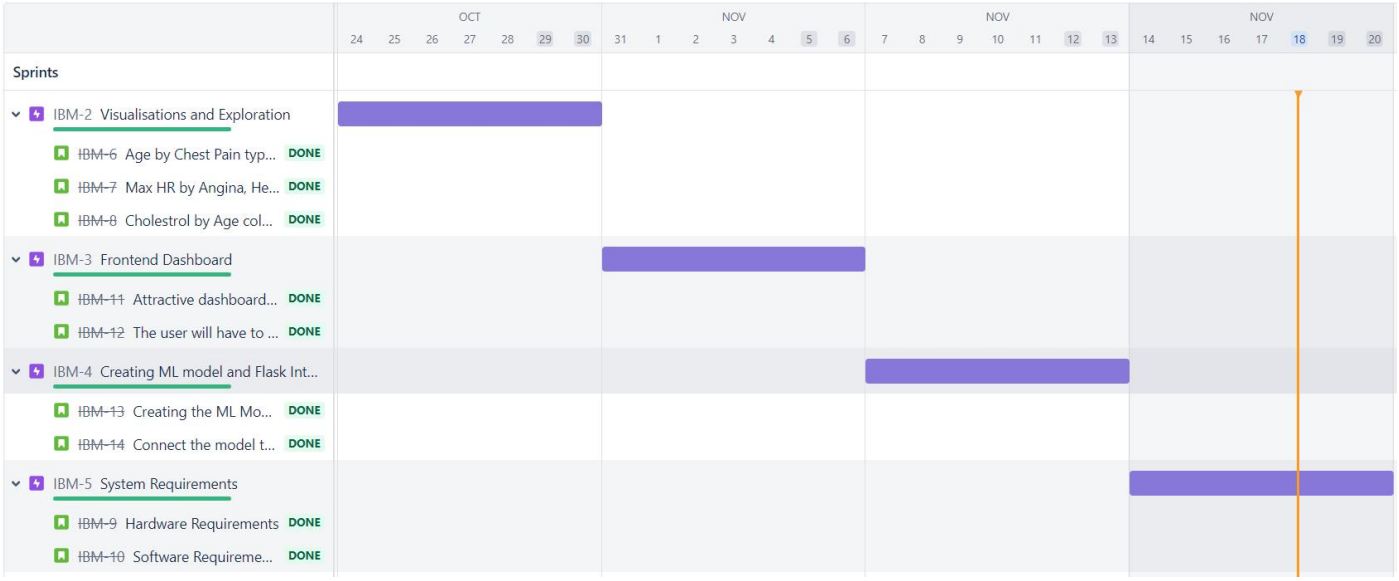
Sprint	Functional Requirement	User Story number	User story/task	Story points	Priority	Team Members
Sprint 1	Visualizations and Exploration	USN-1	Need to see visualisations of Age by Chest pain type, BP by age, Exercise angin by chest pain type.	3	High	2
Sprint 1		USN-2	Visualisations for Max HR by Angina, 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

			• 4 GB RAM or higher • 128 GB ROM or higher 4.Mobile • (Android 12.0 and above)			
Sprint 4		USN-9	Software requirement: Laptop or PC • 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	24th Oct, 2022	30th Oct, 2022	9	30th Oct, 2022
Sprint 2	10	1 week	31st Oct, 2022	06th Nov, 2022	10	06th Nov, 2022
Sprint 3	12	1 week	07th Nov, 2022	13th Nov, 2022	12	13th Nov, 2022
Sprint 4	10	1 week	14th Nov, 2022	20th Nov, 2022	10	20th 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.

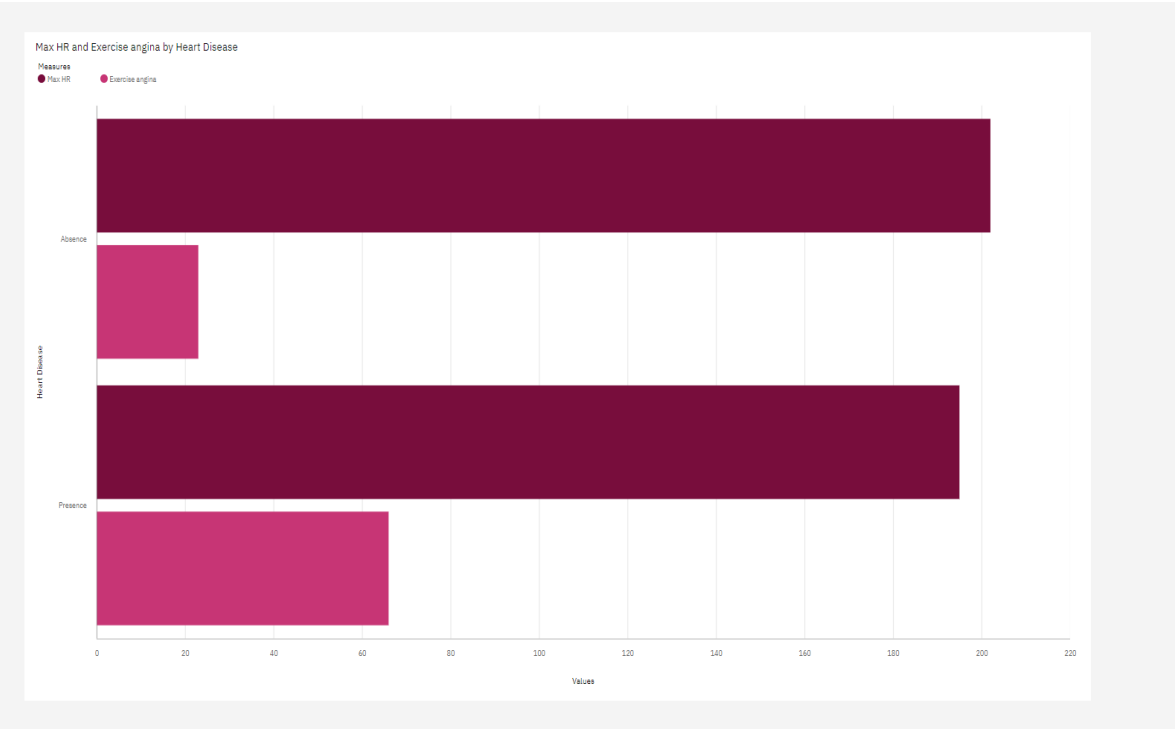
Comparing it with the accuracy gotten from Decision Tree:

```
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)))
```

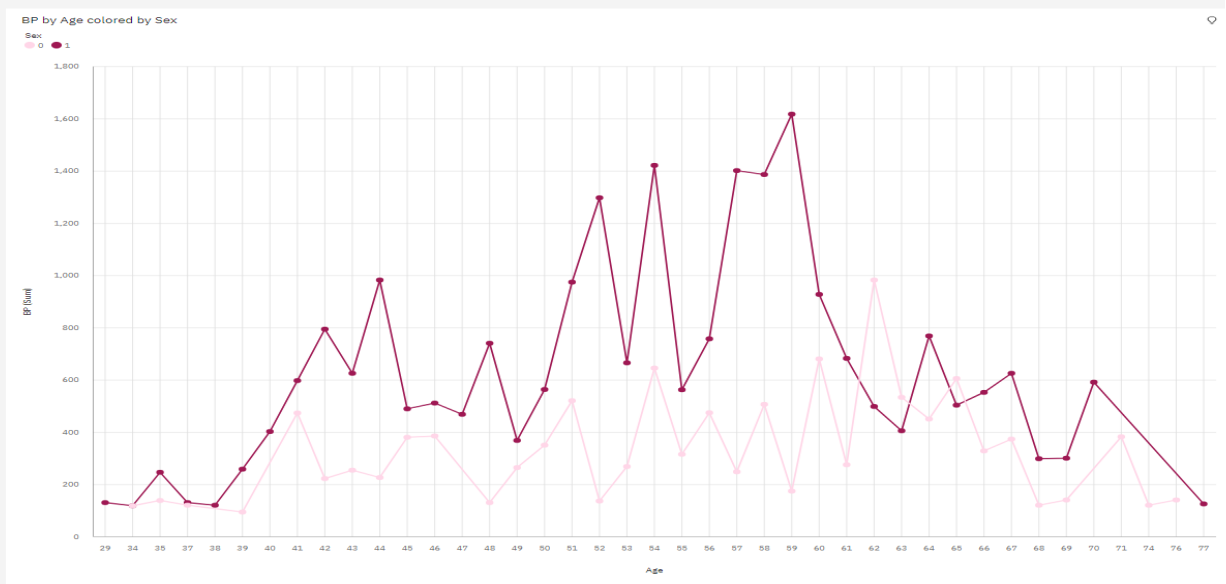
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

7.2 Dashboard

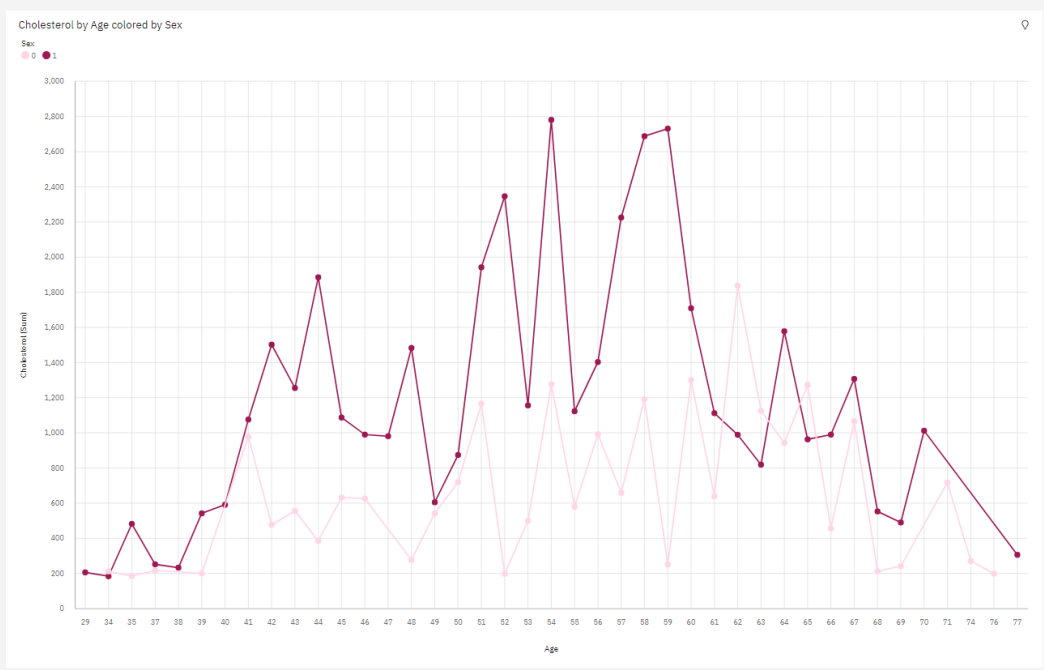
Max Heart Rate and Exercise Angina by Heart Disease



Average BP by Age



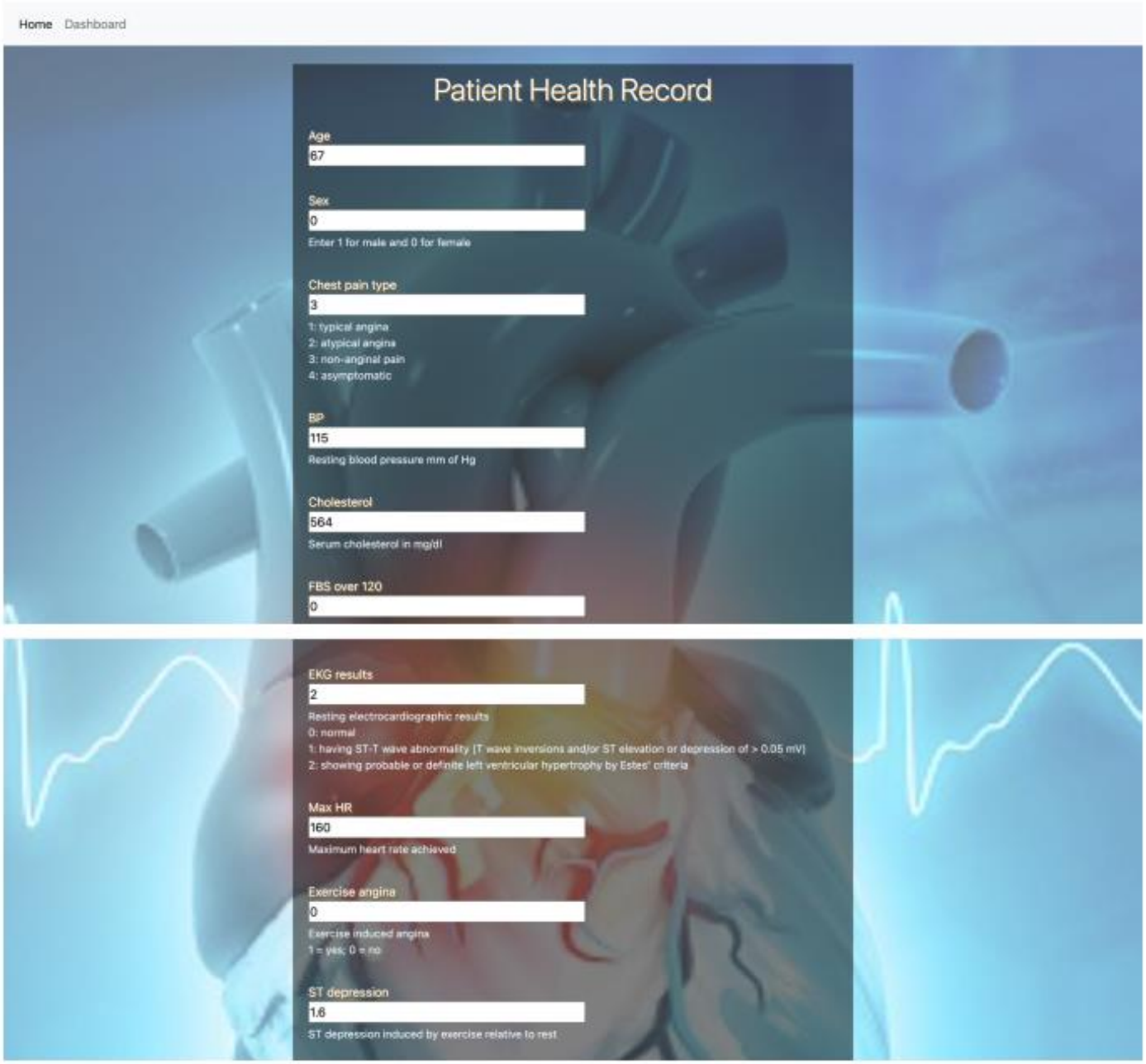
Cholesterol by Age



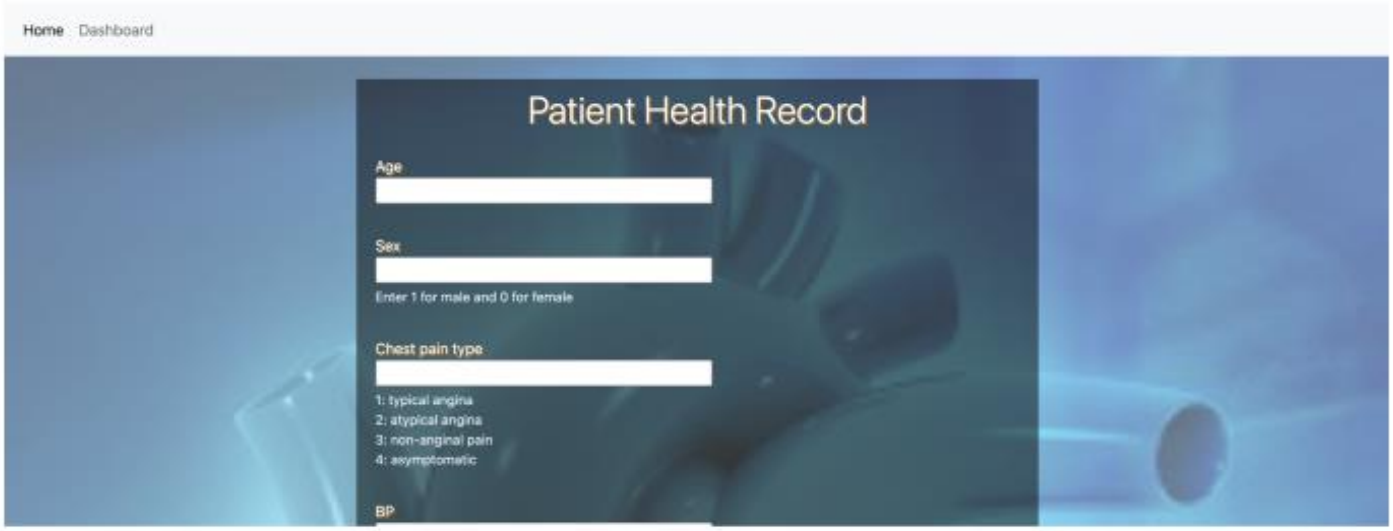
8. Test cases

8.1 User acceptance testing

Case 1: Person is not likely to have heart disease



The patient is not likely to have heart disease!



Case 2: Person is likely to have heart disease

The patient is likely to have heart disease!

Home Dashboard

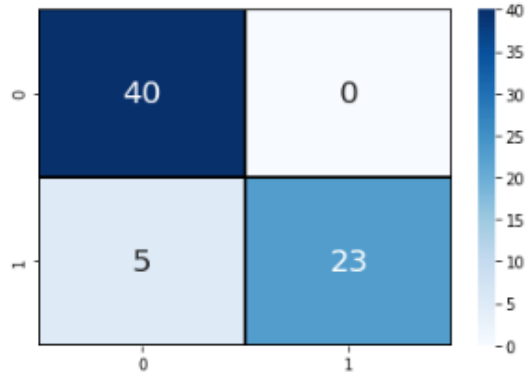
9.Result

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