Visualizing and Predicting Heart Diseases with an Interactive Dashboard

NALAIYA THIRAN PROJECT REPORT
2022

Submitted by

Kevin Daniel P	**110819104019**
Jothi Lingam S	**110819104016**
Gunaseelan E	**110819104010**
Karthick G	**110819104018**
Karthick G	**110819104043**

Team ID: PNT2022TMID36371

Content

TEAM ID: PNT2022TMID28392

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

1. Introduction

1.1 Project Overview

Today, heart failure diseases affect more people worldwide than other autoimmune conditions. Cardiovascular Diseases (CVDs) affect the heart and obstruct blood flow through the blood vessels. Better technologies like MCGs help in detecting these diseases when they are in an early stage. This project of ours aims to create an interactive Dashboard and dataset to predict which patients are most likely to suffer from a heart disease in the near future using the variables given.

1.2 Purpose

Several risk factors for manual heart disease prediction may include inactivity in a physical form, unhealthy eating habits, or even the consumption of alcohol. With such well-defined parameters and the rise of data science, a data-driven approach can surely help in heart disease prediction using machine learning technologies. Early identification of heart disease of improved diagnosis and high-risk individuals using a prediction model can be recommended for a fatality rate reduction, and decision-making is improved for further treatment and prevention.

2. Literature Survey

2.1 Existing Problem

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

2.2 References

“Heart Disease Prediction Using Supervised Machine Learning Algorithms”
Narendra Mohan, Vinod Jain, Gauranshi Agrawal

Predicting and detecting cardiac disease has always been a difficult and time-consuming undertaking for doctors. To treat cardiac disorders, hospitals and other clinics are giving costly therapies and operations. As a result, anticipating cardiac disease in its early stages will be beneficial to people all around the world, allowing them to take required treatment before it becomes serious. Heart disease has been a major issue in recent years, with the primary causes being excessive alcohol use, tobacco use, and a lack of physical activity. Machine learning methods are utilized to forecast cardiac illnesses in this article. For training and testing, a data collection containing diverse human health parameters is used. Many AI&ML algorithms are used to predict cardiac disorders. The performance of the machine learning algorithm is compared after it has been implemented.

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

Predicting and detection of heart disease has always been a critical and challenging task for healthcare practitioners. Hospitals and other clinics are offering expensive therapies and operations to treat heart diseases. So, predicting heart disease at the early stages will be useful to the people around the world so that they will take necessary actions before getting severe. Heart disease is a significant problem in recent times; the main reason for this disease is the intake of alcohol, tobacco, and lack of physical exercise. Over the years, machine learning shows effective results in making decisions and predictions from the broad set of data produced by the health care industry.

2.3 Problem Statement Definition

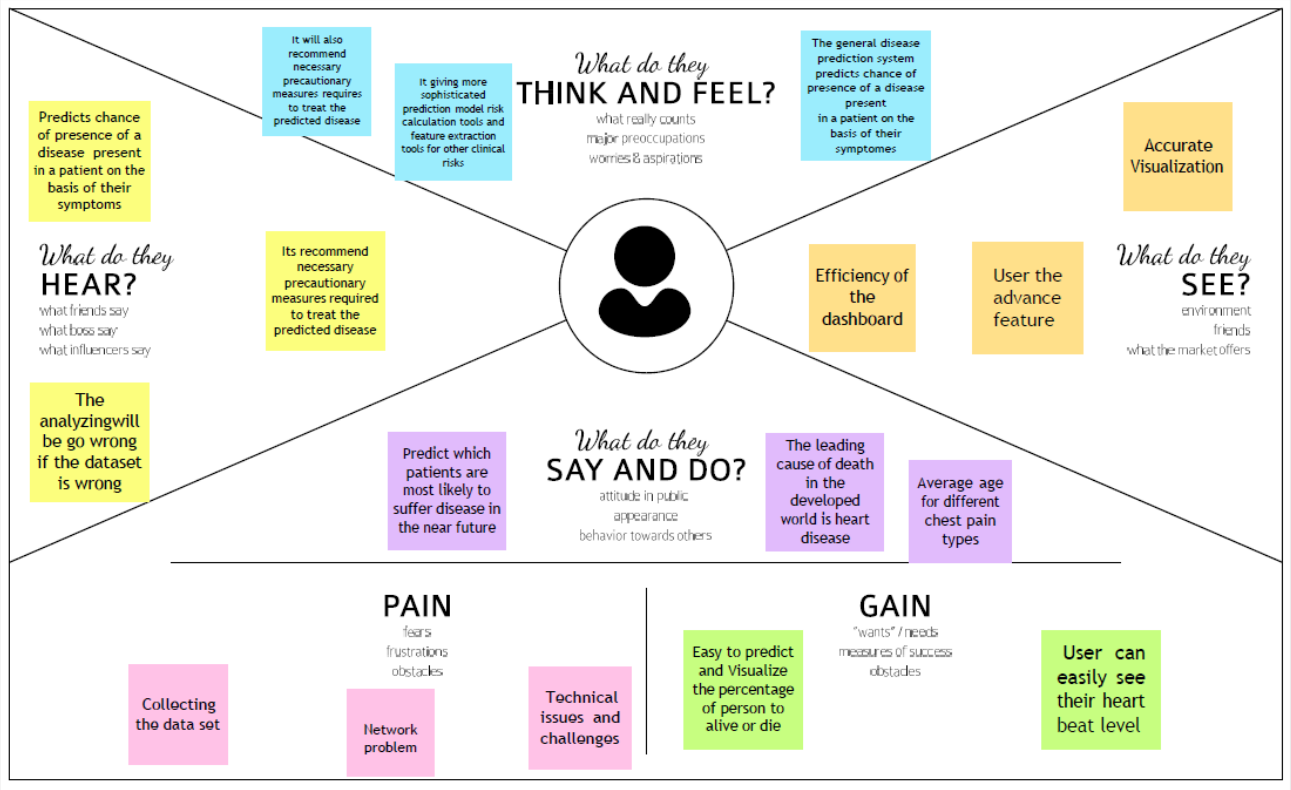
The issue occurs for people with unhealthy lifestyles and age above 40. Where is the issue occurring. The issue is originating from an unhealthy lifestyle. It mostly occurs in the blood valves of the heart. If we don’t solve the problem, many people will die at a young age. The death rate due to heart diseasewill increase rapidly. We should predict the problem before giving treatment to the patients. As the problem is predicted early, wecan solve it easily and early.

3. Ideation and Proposed Solution

3.1 Empathy Map

Visualizing and Predicting Heart Diseases with an Interactive Dash Board

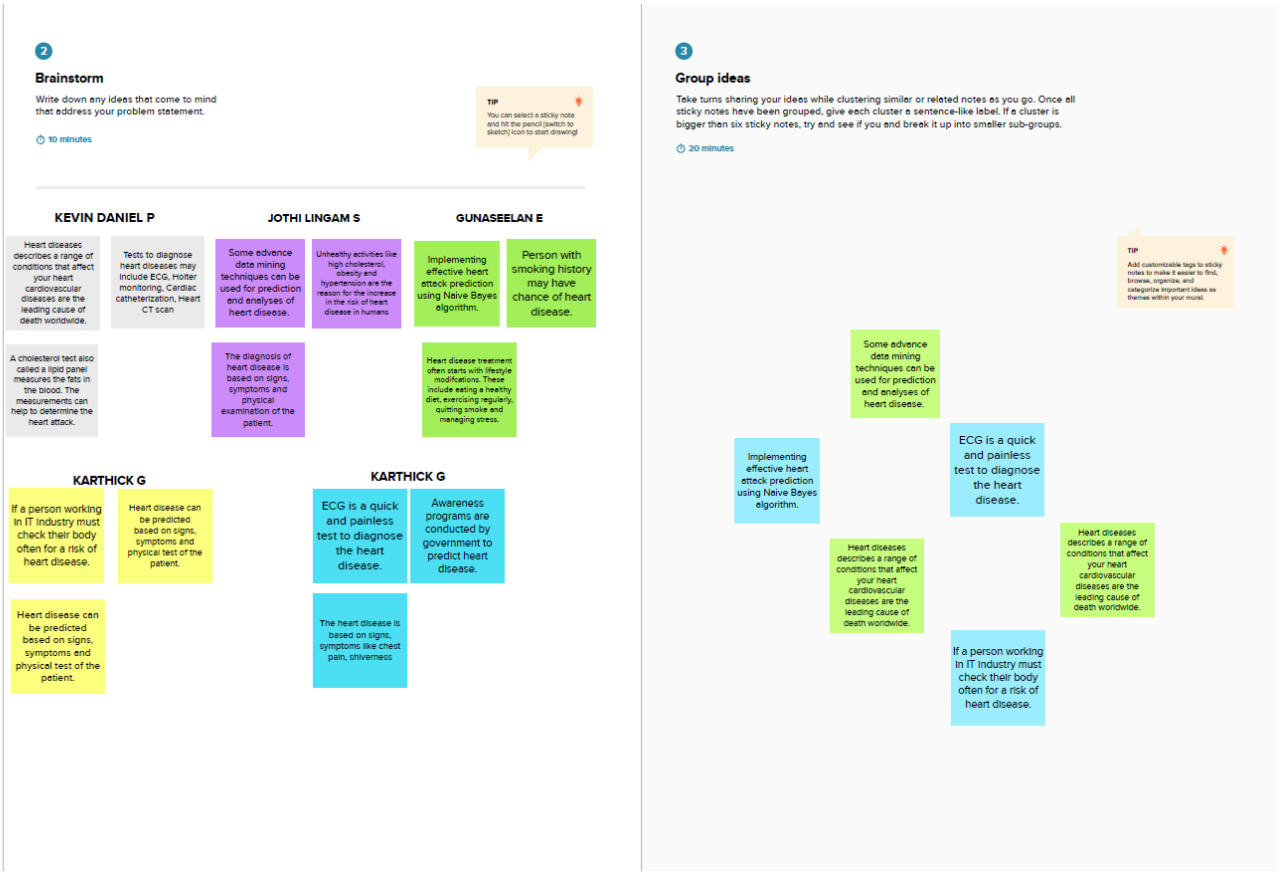
Team ID: PNT2022TMID36371



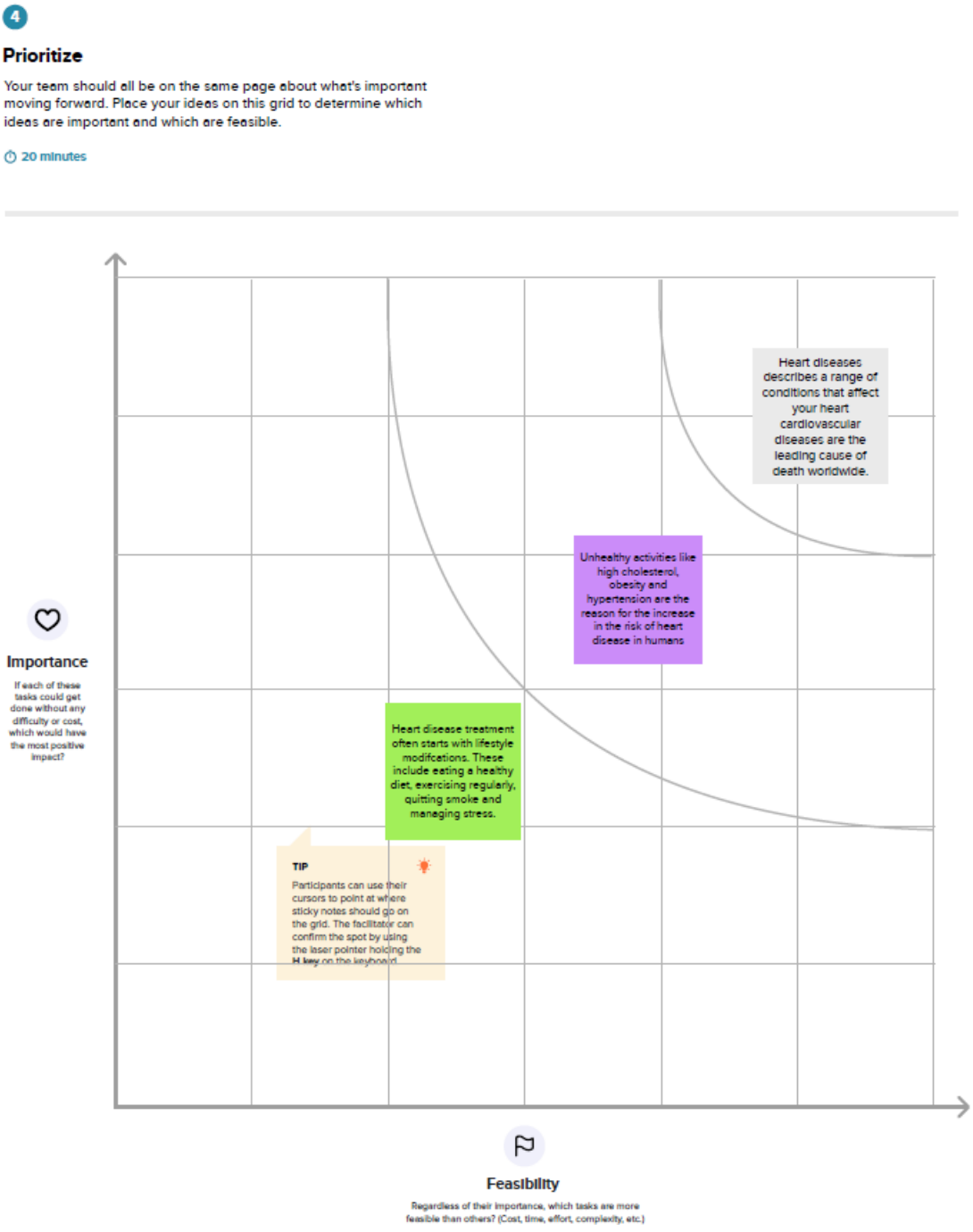
3.2 Ideation and Brainstorming

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

Step-2: Brainstorm, Idea Listing and Grouping



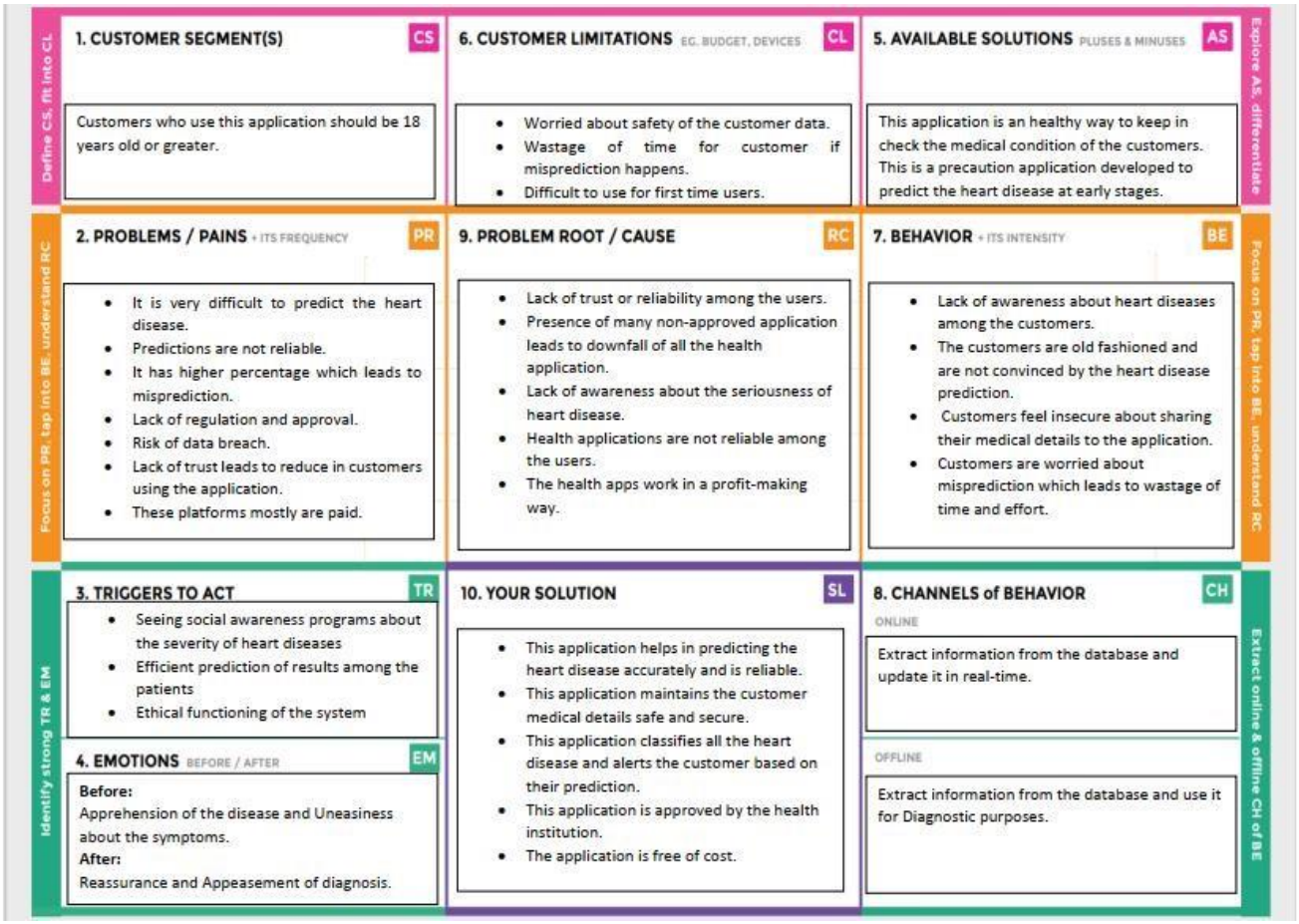
Step-3: Idea Prioritization



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs.
2.	Idea / Solution description	The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre-processed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system.
3.	Novelty / Uniqueness	The primary purpose of this study is to give clinicians a tool to detect cardiac problems at an early stage.
4.	Social Impact / Customer Satisfaction	This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease by using these computer aided techniques we can predict the patient fast and better and the cost can be reduced very much.
5.	Business Model (Revenue Model)	Can be deployed by Hospitals or other Health organizations, so that the success rate of prediction is higher.
6.	Scalability of the Solution	It is the property of a system to handle a growing amount of work by adding resources to the system. In order to stay at top positions in market the scalability and adaptability towards changing needs must be met.

3.3 Problem Solution Fit



4.RequirementAnalysis

4.1 Functional Requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Enables user to make registration for the application through Google Account
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 and see their deliverables.
FR-4	Generation Report	User can view his/her health report and can get treated accordingly

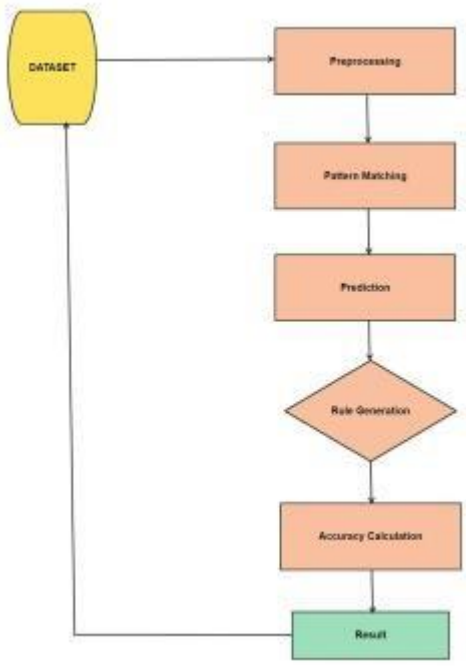
4.2 Non-Functional Requirement

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

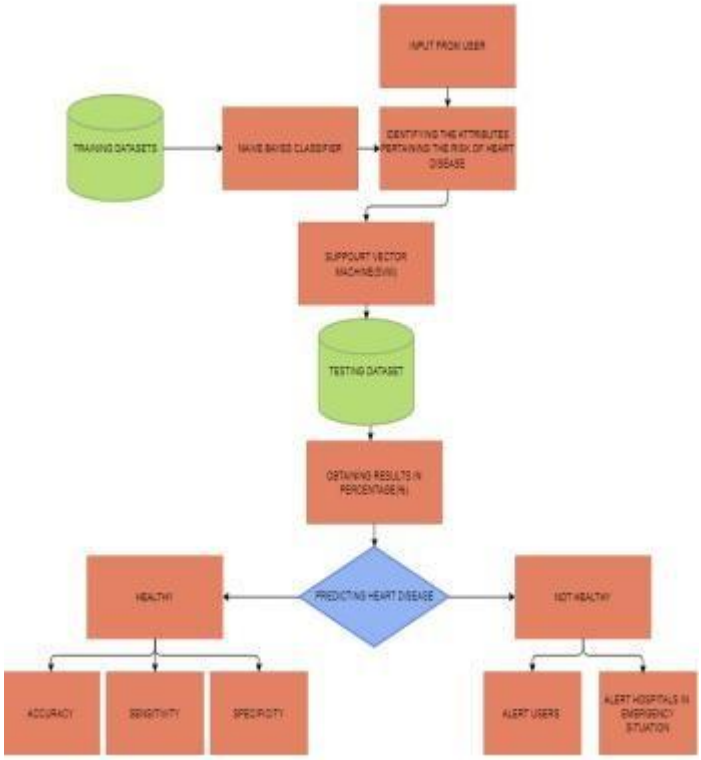
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user-friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	**Security**	to overcome that the data we collected and processed are encrypted with standards that cannot be broken. So, this makes our system to ensure privacy rights to the users.
NFR-3	**Reliability**	The application has to be consistent at every scenario and has to work without failure in any Environment
NFR-4	**Performance**	Making the data available in times need is important. We have enclosed a feature of sharing the patients record to authorized facilities in case of emergencies so that right medications can be given on time.
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

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 (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	7	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 Facebook	3	Medium	1
Sprint-1		USN-4	As a user, I can register for the application through Gmail	3		1
Sprint-2	Login	USN-5	As a user, I can log into the application by entering email & password	20	High	4
Sprint-3	Dashboard	USN-6	As a user, I can see current health status, personalize the interface etc.	10	High	3
Sprint-3		USN-7	As a user, I can update my user profile, medical details and have access to all the medical records.	5	Medium	1
Sprint-3		USN-8	As a user, I can add or customize all the health fit tracking devices	5	Medium	1
Sprint-4	Notification Manager	USN-9	As a user, I can get to know about my health condition at regular intervals via monitoring system	4	High	3
Sprint-4	Appointments	USN-10	As a user, I can book appointments with nearby health care specialist for consultation along with the current medical status report of the user	3	Medium	1
Sprint-4	SOS	USN-11	R In emergency scenario, the automated system must intimate the nearby emergency facility for assistance and also notify their guardian	10	High	4
Sprint-4	Community support service	USN-12	As a patient, I can connect with people of the community to share thoughts and embrace self-confidence to fight against the disease	3	Medium	3

6.2 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	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

7.1 Coding and Solution

Exploratory Data Analysis(EDA)

In [3]:
```
data_set.head()
```

Out[3]:

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium	Heart Disease
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	Presence
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	Absence
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	Presence
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	Absence
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	Absence

In [4]:
```
data_set.describe()
```

Out[4]:

	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope of ST	Number of vessels fluro	Thallium
count	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.00000	270.000000	270.000000	270.000000
mean	54.433333	0.677778	3.174074	131.344444	249.659259	0.148148	1.022222	149.677778	0.329630	1.05000	1.585185	0.670370	4.696296
std	9.109067	0.468195	0.950090	17.861608	51.686237	0.355906	0.997891	23.165717	0.470952	1.14521	0.614390	0.943896	1.940659
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.00000	1.000000	0.000000	3.000000
25%	48.000000	0.000000	3.000000	120.000000	213.000000	0.000000	0.000000	133.000000	0.000000	0.00000	1.000000	0.000000	3.000000
50%	55.000000	1.000000	3.000000	130.000000	245.000000	0.000000	2.000000	153.500000	0.000000	0.80000	2.000000	0.000000	3.000000
75%	61.000000	1.000000	4.000000	140.000000	280.000000	0.000000	2.000000	166.000000	1.000000	1.60000	2.000000	1.000000	7.000000
max	77.000000	1.000000	4.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.20000	3.000000	3.000000	7.000000

In [33]:
```
accuracy = []

classifiers = ['Decision Trees', 'Logistic Regression', 'Naive Bayes', 'Random Forests','Support Vector Machine','XGBOOST ALGORITHM']

models = [DecisionTreeClassifier(max_depth=3, random_state=0), LogisticRegression(),
        GaussianNB(), RandomForestClassifier(n_estimators=100, random_state=0), SVC(kernel = 'rbf', random_state = 0),XGBClassifier() ]

for i in models:
    model = i
    model.fit(X_train, y_train)
    score = model.score(X_test, y_test)
    accuracy.append(score)
```

In [34]:
```
summary = pd.DataFrame({'Accuracy':accuracy}, index=classifiers)
summary
```

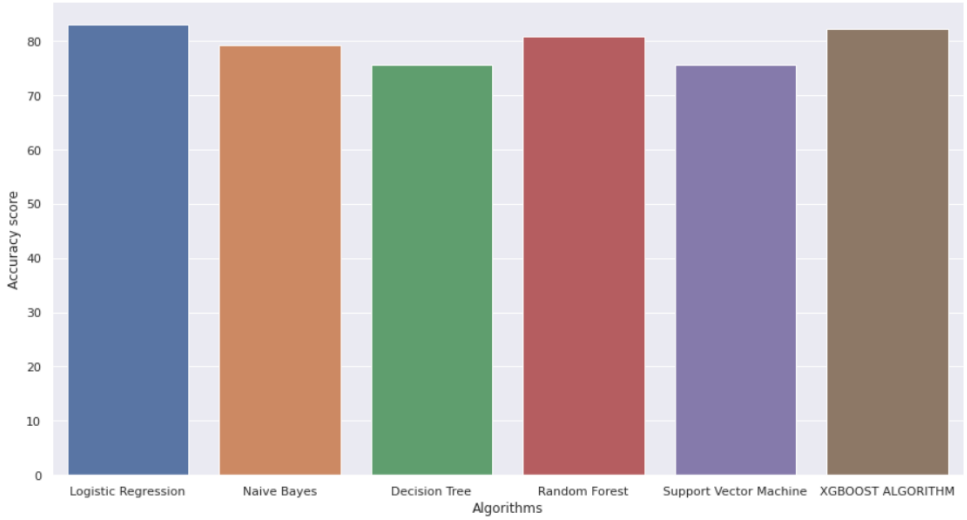
Out[34]:

	Accuracy
Decision Trees	0.755556
Logistic Regression	0.829630
Naive Bayes	0.792593
Random Forests	0.807407
Support Vector Machine	0.711111
XGBOOST ALGORITHM	0.822222

In [37]:

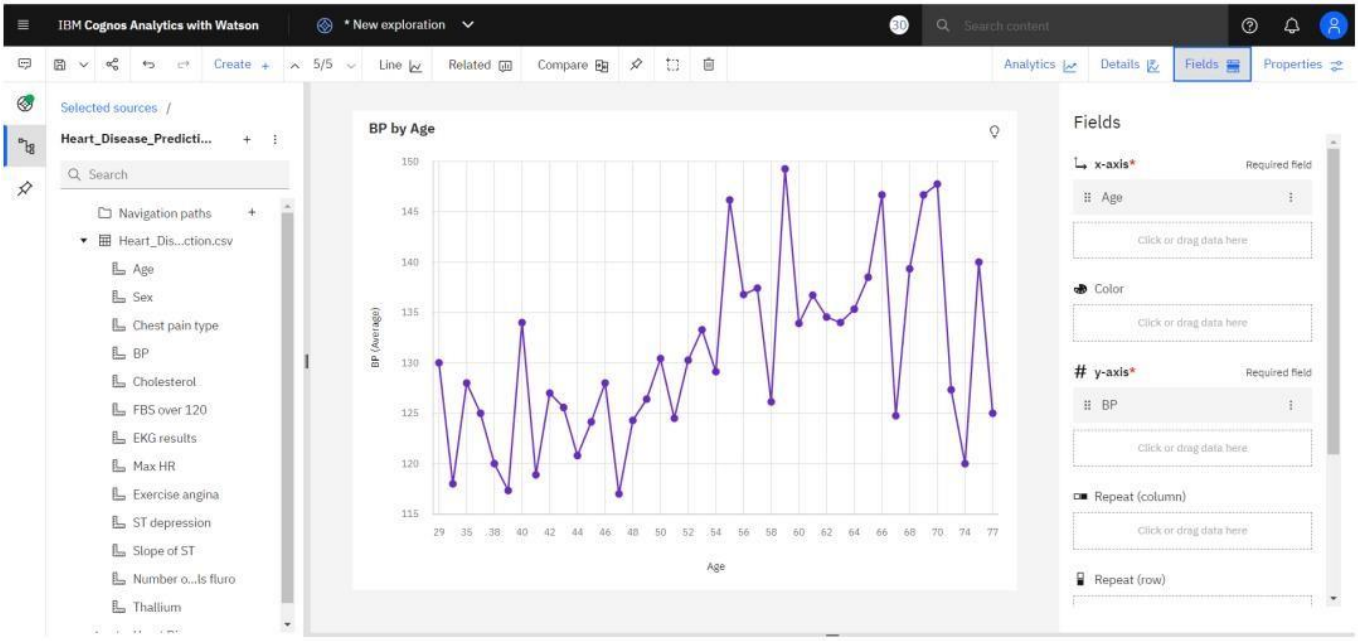
```
scores = [score_lr,score_nb,score_dt,score_rf,score_svm,accuracy1*100]
algorithms = ["Logistic Regression","Naive Bayes","Decision Tree","Random Forest","Support Vector Machine",'XGBOOST ALGORITHM']
sns.set(rc={'figure.figsize':(15,8)})
plt.xlabel("Algorithms")
plt.ylabel("Accuracy score")

sns.barplot(algorithms,scores)
```

Out[37]:

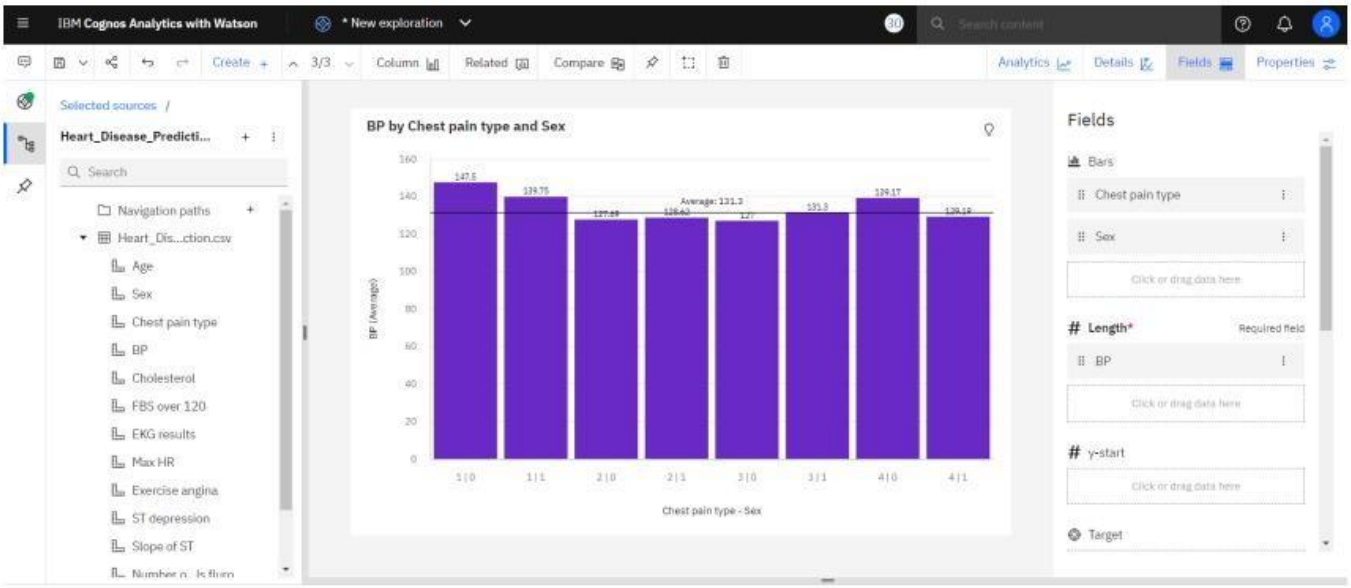


7.2 Dashboard

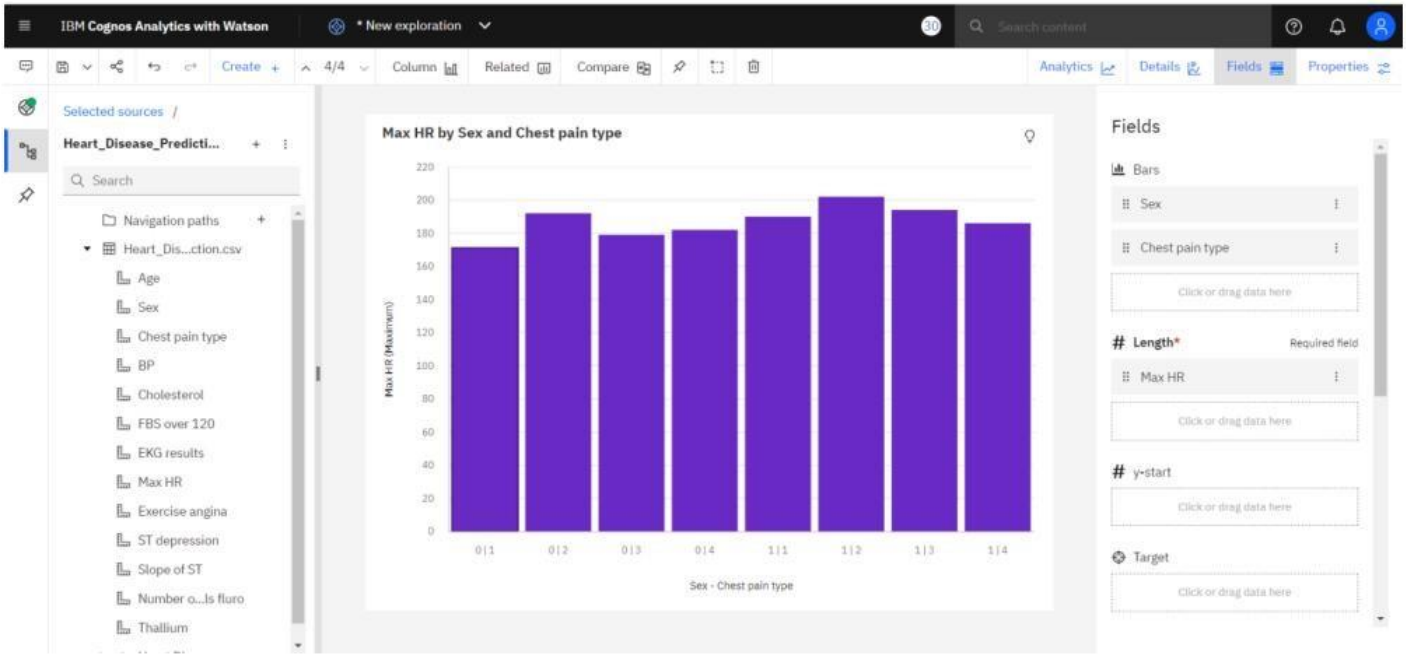
Average BP during chest pain



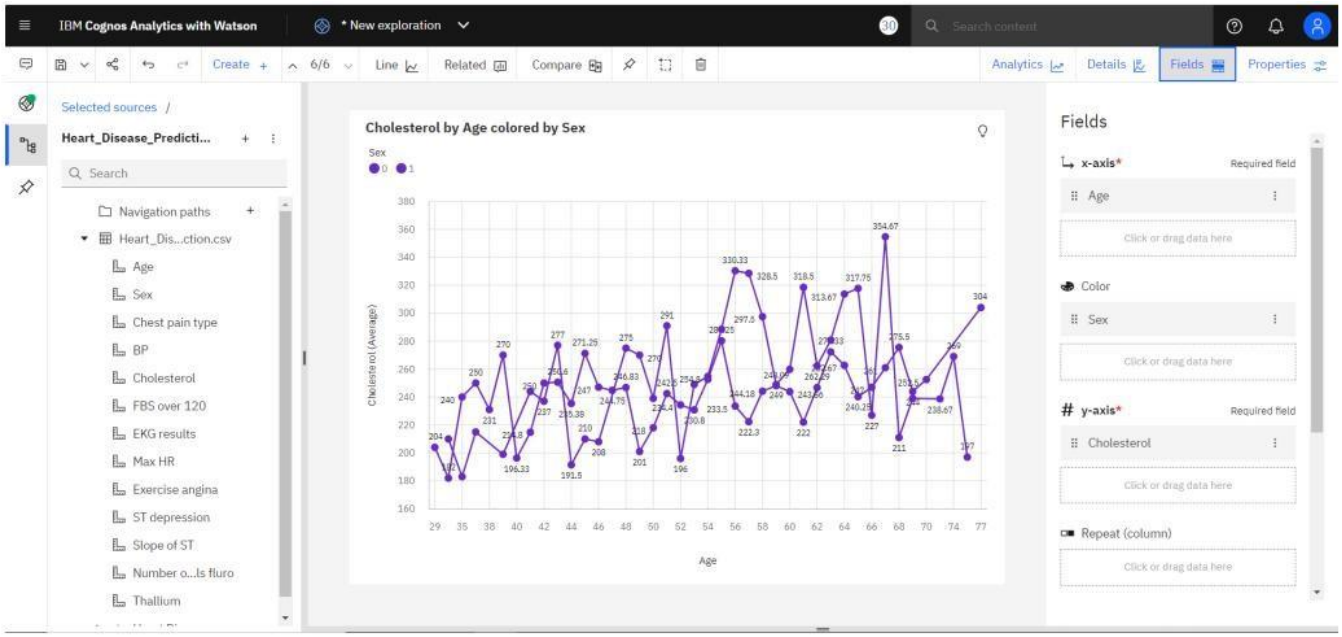
Exploration Of BP vs Chest pain Type and Gender:



Exploration Of Max Heart Rate During the Chest Pain:



Exploration Of Cholesterol by age and Gender:



8. Testing

Performance Testing

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs - 12
2.	Data Responsiveness	Instant Response
3.	Amount Data to Rendered (DB2 Metrics)	12
4.	Utilization of Data Filters	4
5.	Effective User Story	No of Scene Added - 5
6.	Descriptive Reports	No of Visualizations / Graphs - 12

8.2 User acceptance Testing

Testing a case where user has heart disease

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the Visualizing and Predicting Heart Diseases with an Interactive Dashboard project at the time of the release to User Acceptance Testing (UAT).

2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	7	5	1	1	14
Duplicate	1	0	2	0	3
External	2	1	1	1	5
Fixed	7	5	1	1	12
Not Reproduced	0	0	0	0	0
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	17	11	4	3	34

3. Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
View Home Page	10	0	0	10
View Prediction Page	8	0	0	8
Enter the Scores	5	0	0	5
Click Submit Button	5	0	0	5
Prediction Output	5	0	0	5
View Dashboard Page	4	0	0	4
Version Control	10	0	0	10

9. Result

9.1 Performance Metrics

Decision Tree

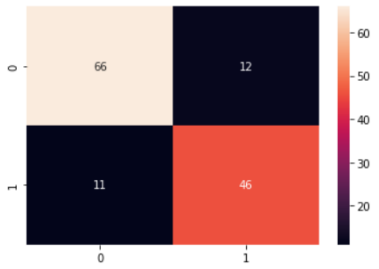
In [25]:
```
dt = DecisionTreeClassifier(max_depth=3, random_state=0)

dt.fit(X_train, y_train)

y_pred_dt = dt.predict(X_test)

score_dt = round(accuracy_score(y_pred_dt,y_test)*100,2)

print("The accuracy score achieved using Decision Tree is: "+str(score_dt)+" %")
```

```
The accuracy score achieved using Decision Tree is: 75.56 %
```

In [26]:
```
print(classification_report(y_test, y_pred_lr))


matrix= confusion_matrix(y_test, y_pred_lr)
sns.heatmap(matrix,annot = True, fmt = "d")
```

```
              precision    recall  f1-score   support

           0       0.86      0.85      0.85        78
           1       0.79      0.81      0.80        57

    accuracy                           0.83       135
   macro avg       0.83      0.83      0.83       135
weighted avg       0.83      0.83      0.83       135
```

Out[26]:



Random Forest

In [27]:
```
randfor = RandomForestClassifier(n_estimators=100, random_state=0)

randfor.fit(X_train, y_train)

y_pred_rf = randfor.predict(X_test)

score_rf = round(accuracy_score(y_pred_rf,y_test)*100,2)

print("The accuracy score achieved using Random Forest is: "+str(score_rf)+" %")
```

```
The accuracy score achieved using Random Forest is: 80.74 %
```

In [28]:
```
print(classification_report(y_test, y_pred_rf))

matrix= confusion_matrix(y_test, y_pred_rf)
sns.heatmap(matrix,annot = True, fmt = "d")
```

10. Advantages Disadvantages

Advantages:

- One of the fastest and easiest way to determine if a person is likely to suffer from a heart disease or not.
- Useful for all the medical officers to know about the patient's history
- User Friendly
- Easy to understand ● Secure

Disadvantages:

- Need a more dynamic User interface
- Users need to know all the fields
- Does Not take null value as input ● Does not provide suggestions to the user.

11. Conclusion

There are several complications when it comes to heart diseases. We can reduce the risk of complications with early diagnosis and treatment. The solution that we provide from the website will help the patients. It is always to get treated in the early stages when it comes to heart disease.

12. Future Scope

There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research.