Visualizing and Predicting Heart Diseases with an Interactive Dashboard

PROJECT REPORT
2022

Submitted by

M. Aravind -611619106004

J.P. Muthu Aravindh - 611619106058

M.B .Yogeshwaran - 611619106308

K. Maruthupandiyan -611619106051

Team ID: PNT2022TMID17411

Content

Team ID: **PNT2022TMID17411**

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

PAPER SET 1

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

ABSTRACT :

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

ADVANTAGES :

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

DRAWBACKS :

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

PAPER SET 2

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

ABSTRACT :

Data Analysis is carried out to discover useful knowledge from the dataset and to drive quick and better decisions. It is also used to increase the efficiency of the work. Exploratory Data

analysis is the first phase in Data Analysis. It is a method to understand the data and summarize the main features in the dataset by analyzing the data. It is also used for the visual representation of data. Visualization includes line plot, subplot, pair plot, violin plot, joint plot, swarm plot, Histograms, Box plot, Scatter plot. In this paper, Exploratory Data Analysis is done using python and implemented in Spyder IDE

ADVANTAGE :

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

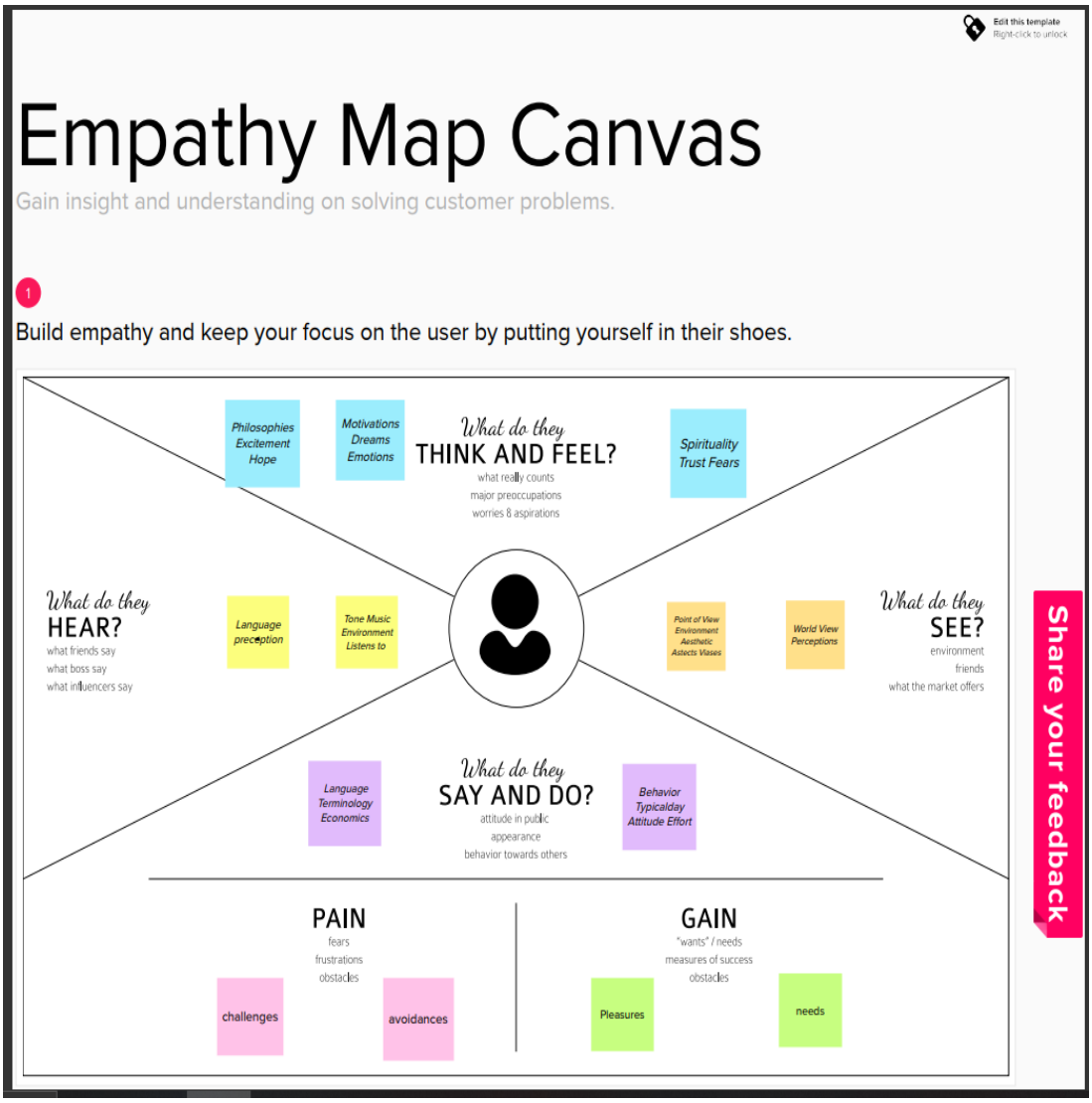
DRAWBACKS :

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

3. Ideation and proposed solution

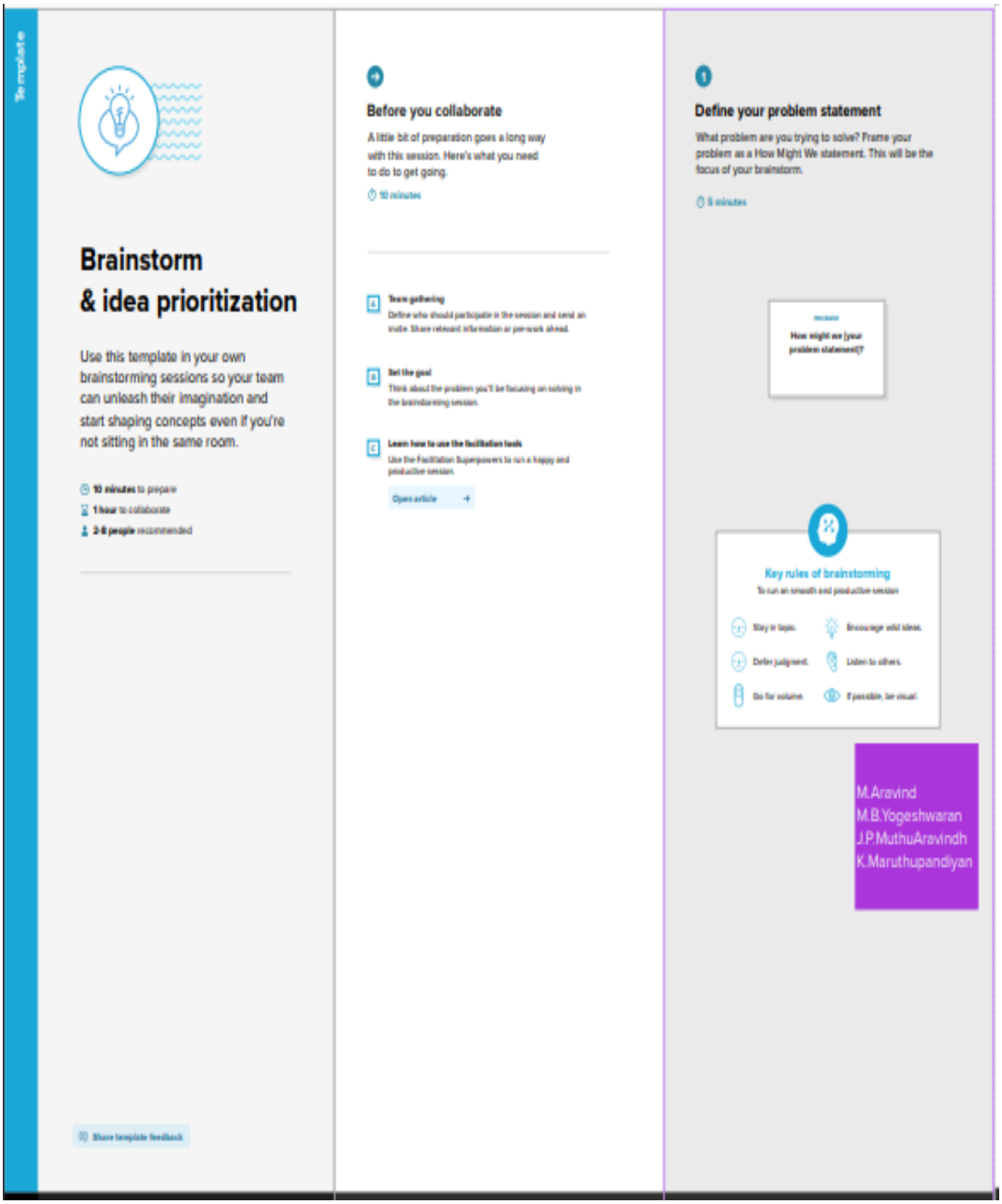
3.1 Empathy Map Canvas

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

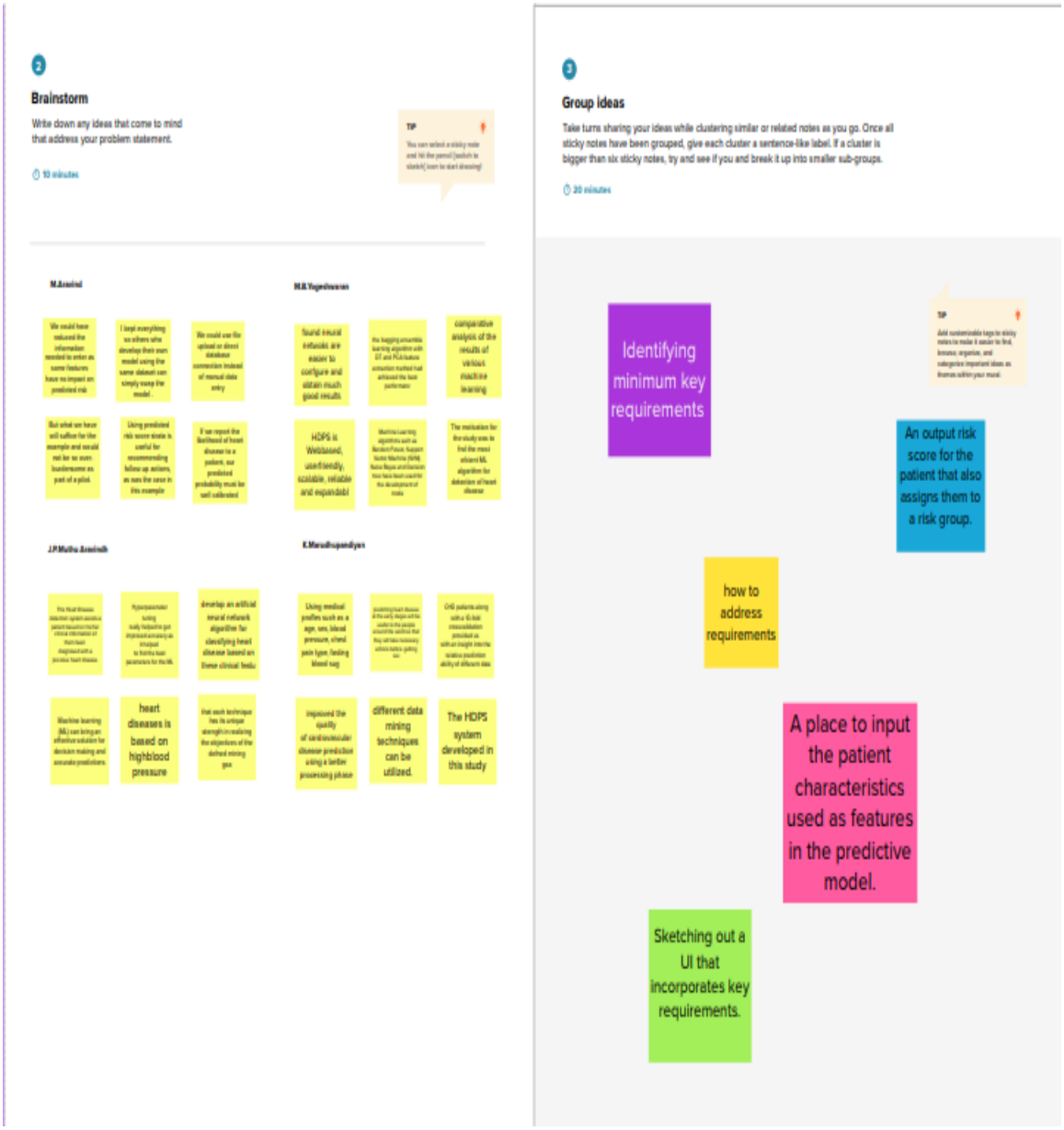


3.2 Ideation and Brainstroming

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



Step-2: Brainstrom , Idea Listing and grouping



4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

Feasibility

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

After you collaborate

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

Quick add-ons

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template →

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

Share template feedback

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	➢ To develop an interactive dashboard to predict the heart disease accurately with few tests and attributes the presence of heart disease.
2.	Idea / Solution description	➢ Analyzing data and identifying the heart disease using Cognos analysis.
3.	Novelty / Uniqueness	➢ Hoping to achieve maximum accuracy to provide prior treatment to the patients and reduce the fatality rate.
4.	Social Impact / Customer Satisfaction	➢ Saving lives, User friendly interactive dashboard. ➢ Reduces the exorbitant medical cost of the patients. ➢ Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions and experiences.
5.	Business Model (Revenue Model)	➢ Data security. ➢ Easy to use. ➢ Constant updates according to necessity.
6.	Scalability of the Solution	➢ Can be used in any platform (Windows, mac, etc.,). ➢ Adding new feature doesn't affect the performance of the system. ➢ Scalable dataset.

Project Title: Visualizing and Predicting Heart Disease with an Interactive Dashboard
Project Design Phase-I - Solution Fit Template
Team ID: PTN2022TMID17411

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

People with heart disease.

Aged Persons(Above 60)

6. CUSTOMER CONSTRAINTS CC

Avoidable medical errors. Low treatable mortality rates. Lack of transparency. Difficulty finding a good doctor. High maintenance costs. A different perspective on solving the shortage crisis.

5. AVAILABLE SOLUTIONS AS

- Avoid smoking
- Take healthy foods
- Visit cardiologist in case of any symptoms • Maintaining healthy exercise

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

Coronary artery disease is a common heart condition that affects the major blood vessels Cholesterol deposits (plaques) in the heart arteries are usually the cause of coronary artery disease.

9. PROBLEM ROOT CAUSE RC

A buildup of fatty plaques in the arteries (atherosclerosis) is the most common cause of coronary artery disease. Risk factors include a poor diet, lack of exercise, obesity and smoking

7. BEHAVIOUR BE

Chest pain or discomfort, Shortness of breath, Slow heartbeat, Lightheadedness, Swelling in the legs, belly area or areas around the eyes.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

Insufficient ways of handling huge amounts of datasets and inferring the root cause of the heart disease cannot be found out.

Similarity of heart disease has not been identifiable.

10. YOUR SOLUTION SL

With the notable technology of AI/Ml we are able to visualize and predict heart diseases and related diseases, by the ultimate power Cognos Analytics Tool we will be able to properly create a dashboard for the customers to work with and visualize and analyze the heart disease on their work with limited knowledge.

8. CHANNELS of BEHAVIOR CH

8.1 ONLINE

Visualizing the datasets.

Exploration of data.

8.2 OFFLINE

Cleansing of datasets.

Collection and noting the datasets.

Identify strong TR & EM

4. EMOTIONS: BEFORE / AFTER EM

Before -> It creates a huge ambiguity in knowing the proper or accurate reasons for a heart disease.

After -> There is a large chance understanding of the heart disease and root cause of it.

which makes a better solution and finding a preventive way over it.

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Congo's Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

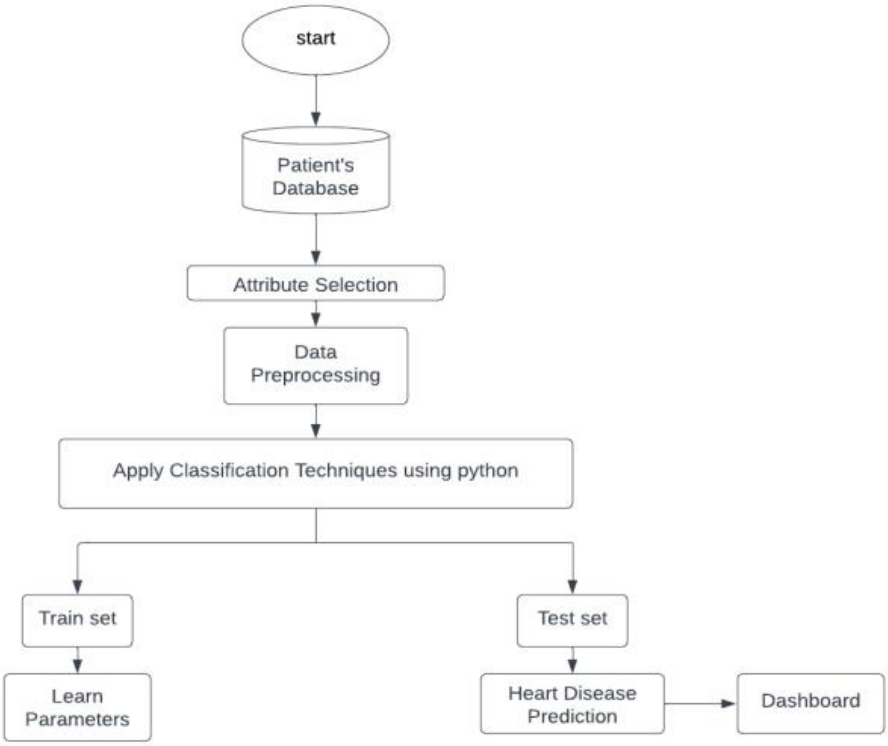
Non-functional Requirements:

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

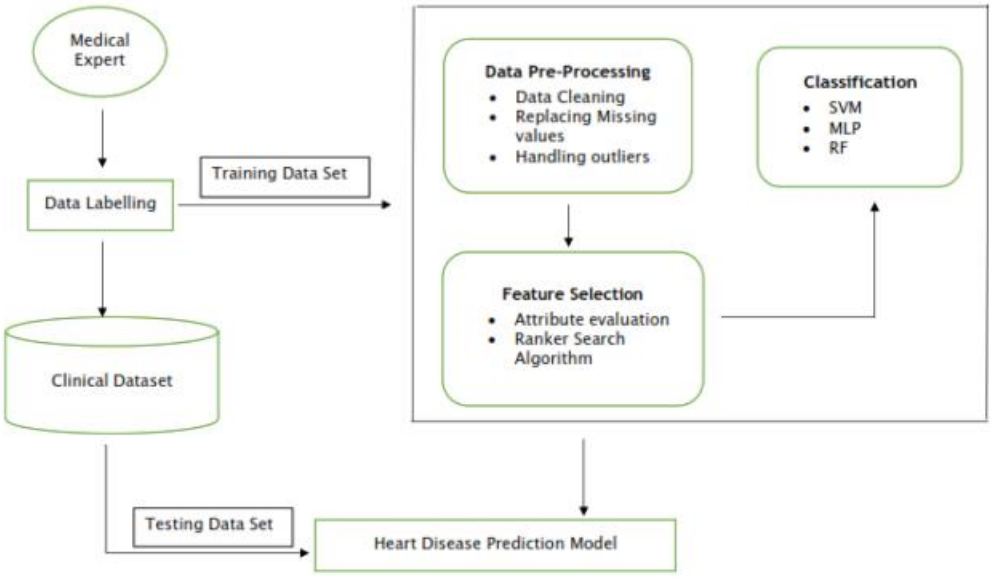
NFR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user-Friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be Performed with just a few clicks.
NFR-2	**Security**	For security of the application the technique Known as database replication should be used so that all the important data should be kept safe. In case of crash, the system should be able to backup and recover the data.
NFR-3	**Reliability**	The application has to be consistent at every scenario and has to work without failure in any environment
NFR-4	**Performance**	Performance of the application depends on the Response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
NFR-5	**Availability**	The application has to be available 24 x 7 for users without any interruption
NFR-6	**Scalability**	The application can withstand the increase in the no. of users and has to be able to develop higher versions

5. Project Design

5.1 Data flow Diagram



Solution and Technical Architecture



6.project planning and scheduling

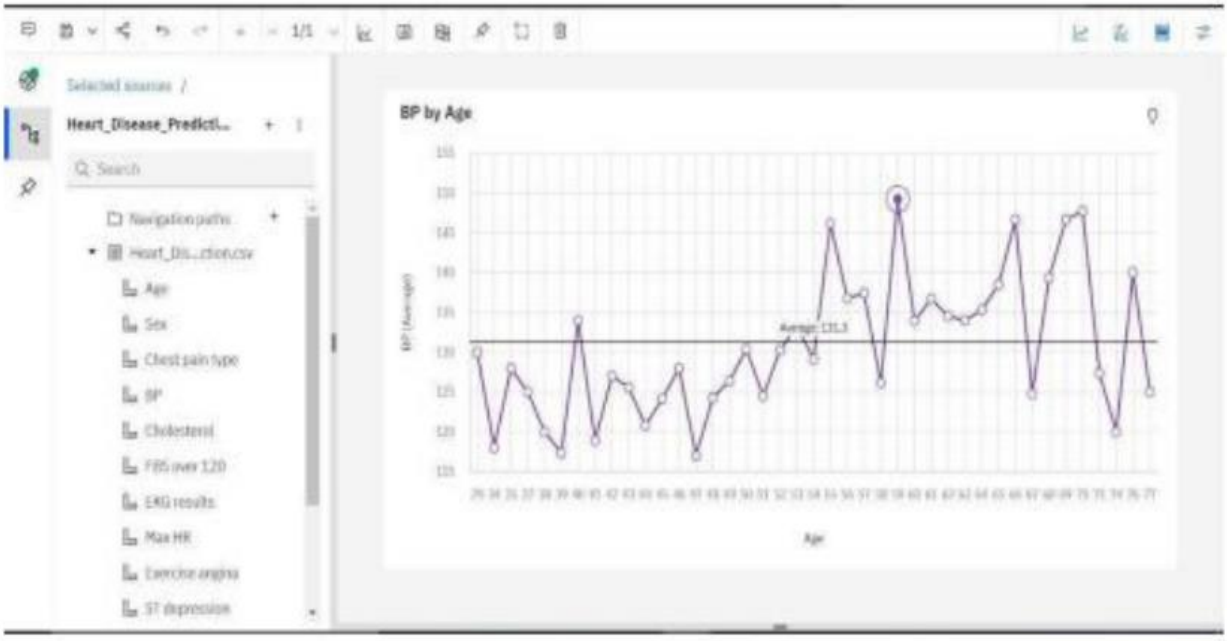
6.1 Iscript planning Execution

Product Backlog, Sprint Schedule, and Estimation

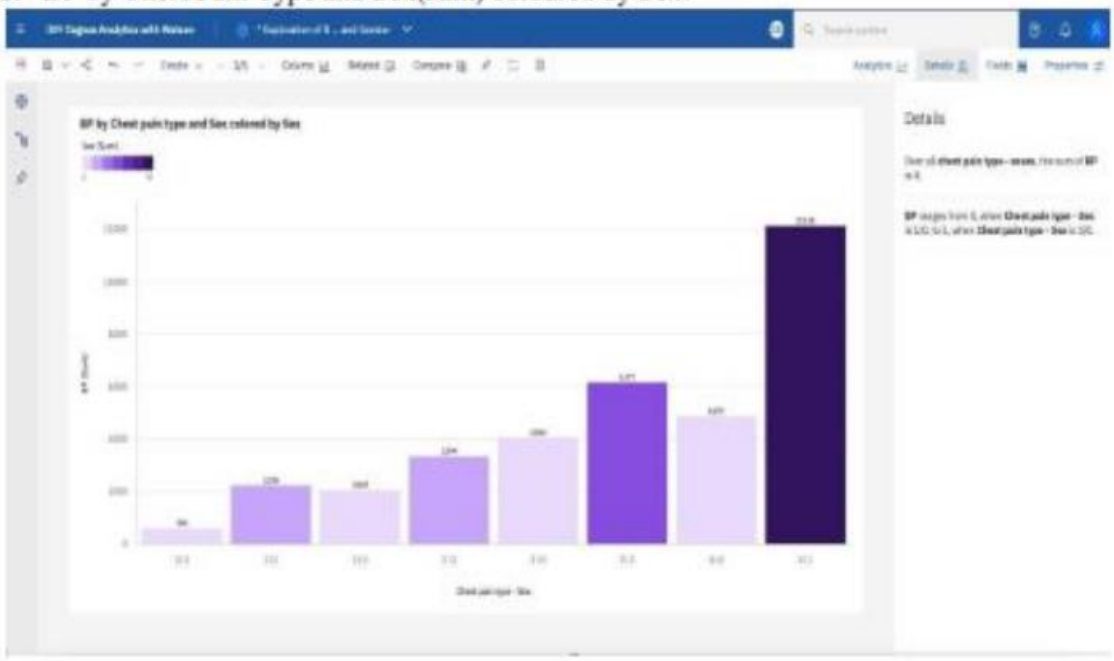
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	1	High	Aravind M Yogeshwaran M B Muthu Aravindh J P Maruthupandiyan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

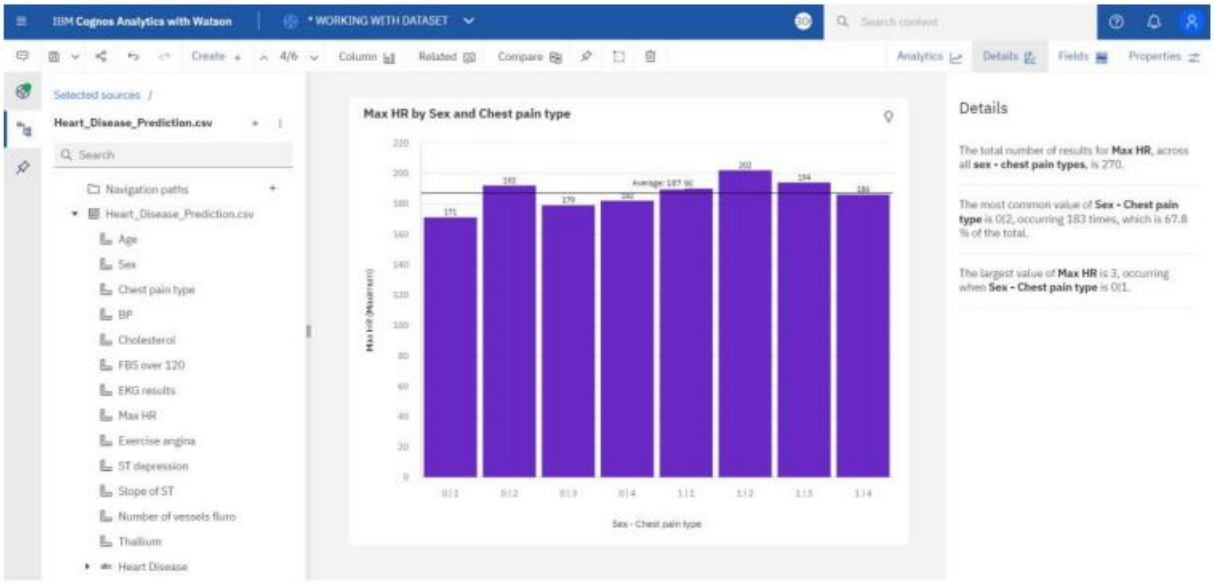
6.2 sprint Delivery Schedule

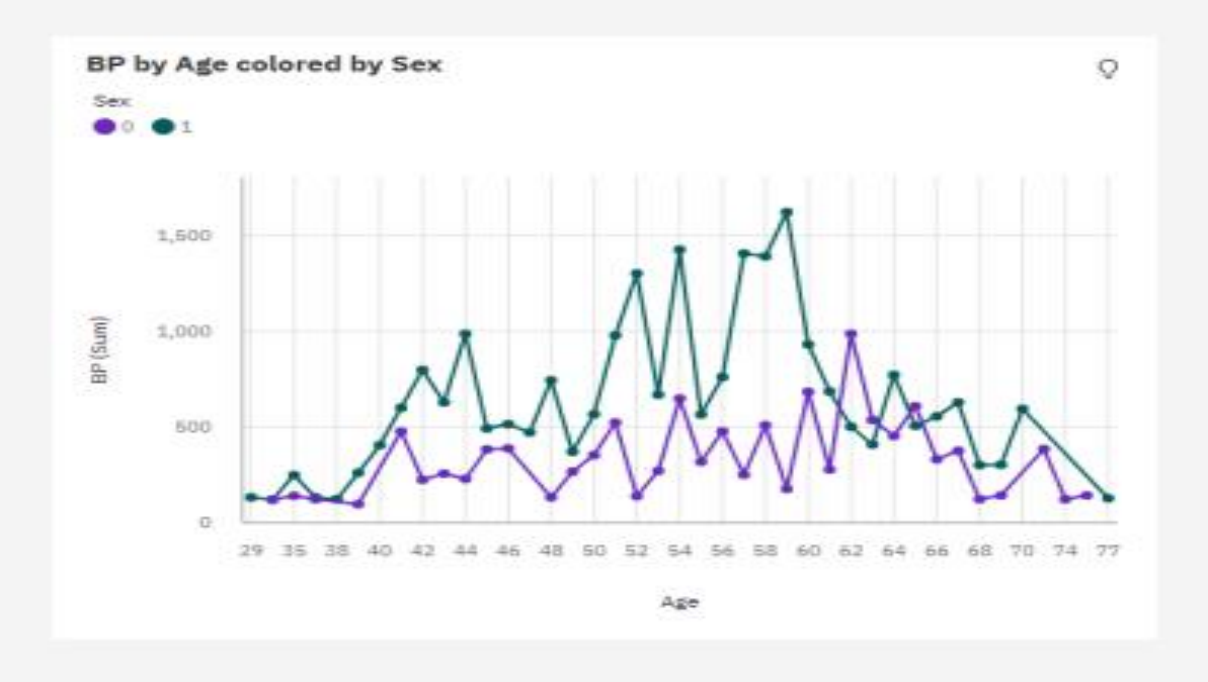
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022



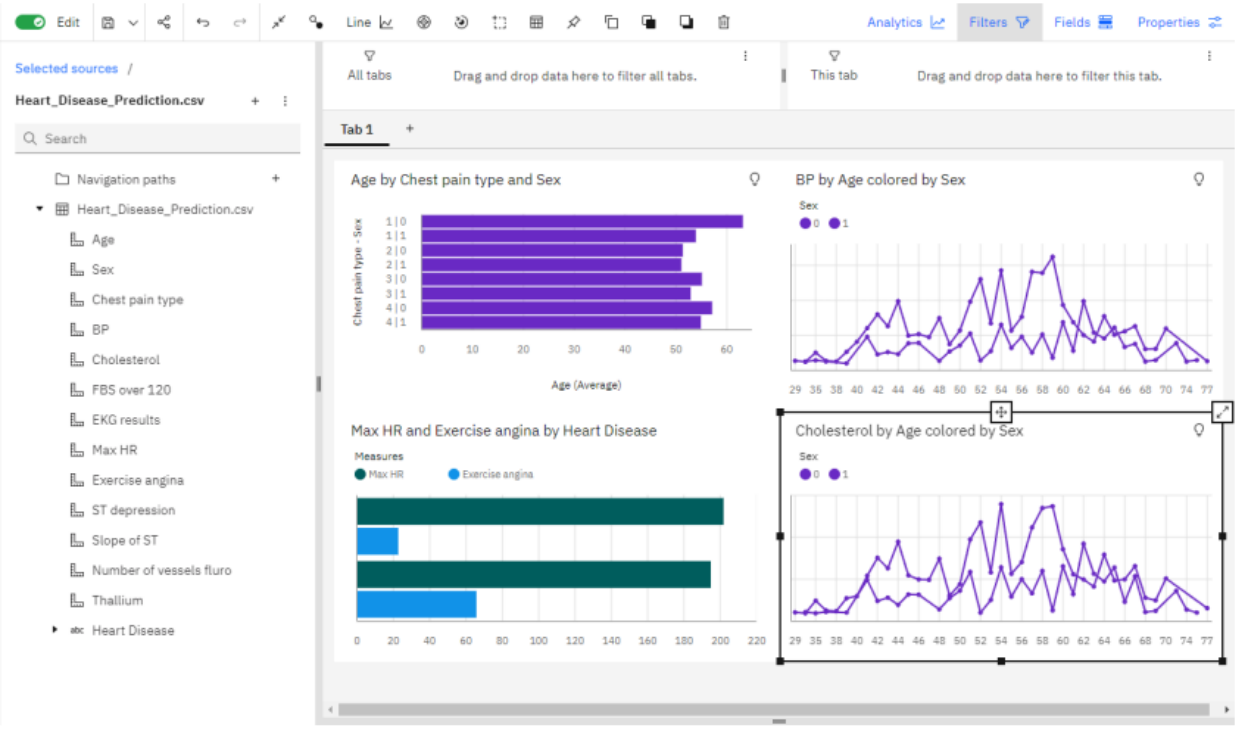
3. BP by Chest Pain Type and Sex(sum) coloured by Sex:





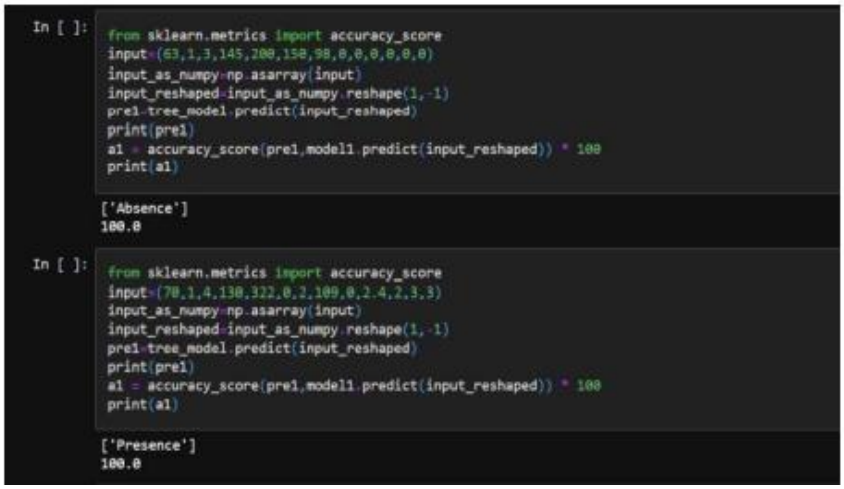


Dashboard showing Different types of Visuals:



8. Testing

8.1 Test Cases Testing the data model for various input values.



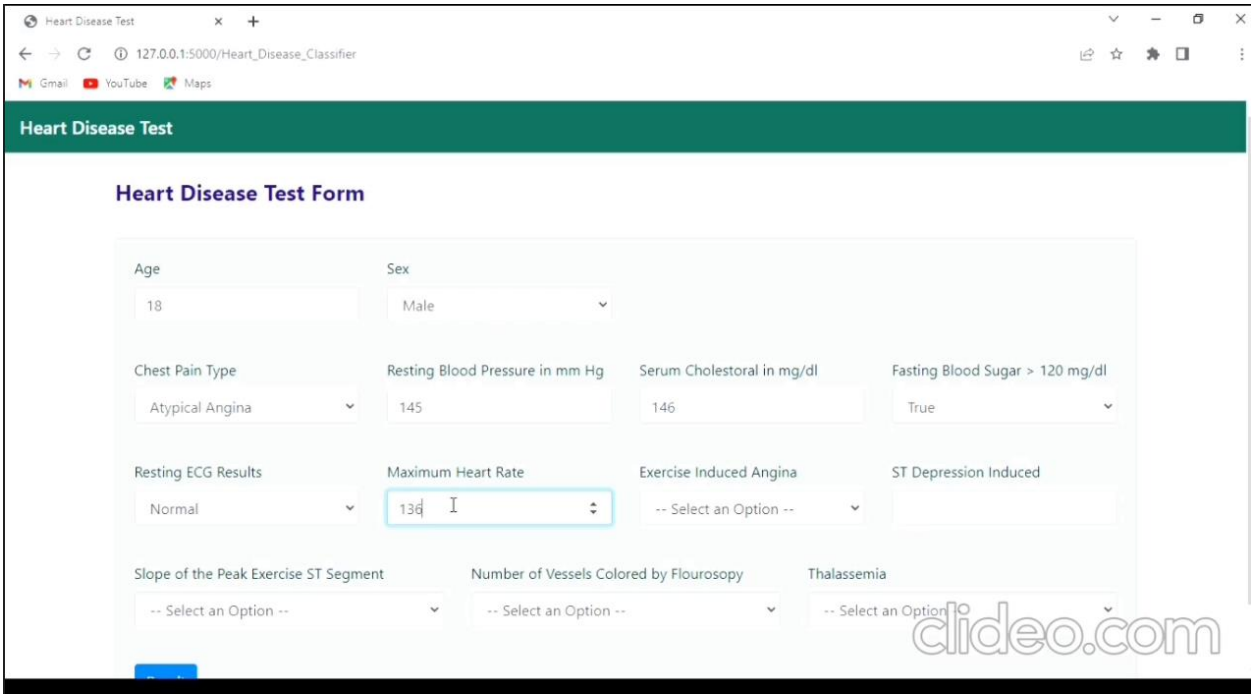
```
In [ ]: from sklearn.metrics import accuracy_score
        input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped)) * 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_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped)) * 100
        print(a1)

        ['Presence']
        100.0
```

8.2 User acceptance Testing Testing a case where user has heart disease




```
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))

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

     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