

# ***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

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: 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 Raza, 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%.

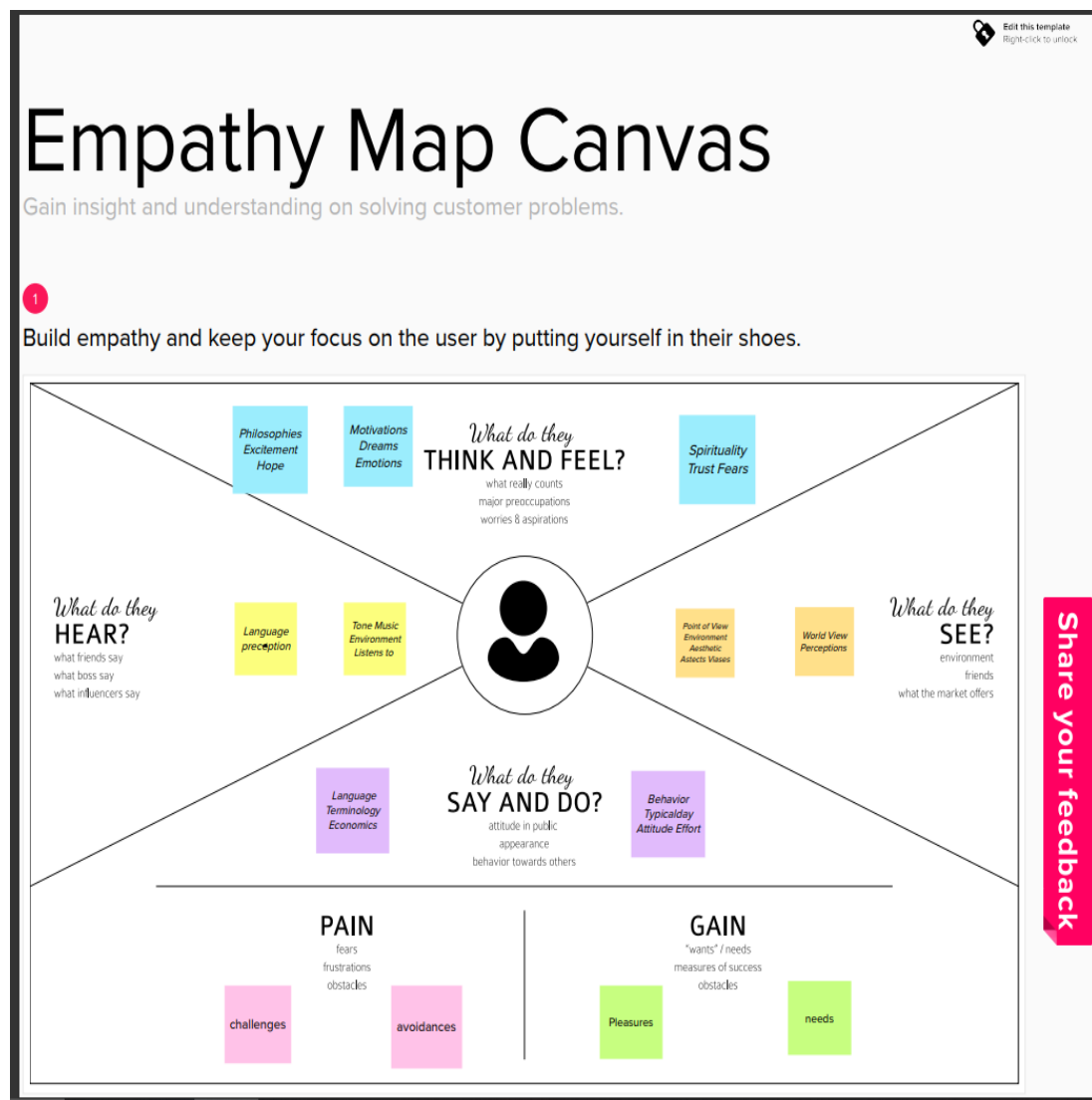
#### **DRAWBACKS :**

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

Template



## Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour to collaborate
- 3-6 people recommended

[Share template feedback](#)

2

### Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

3

Team gathering

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

4

Set the goal

Think about the problem you'll be focusing on solving in the brainstorming session.

5

Learn how to use the facilitation tools

Use the Facilitation Superpowers to run a happy and productive session.

[Open article](#) →

1

### Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

How might we [your problem statement]?

Key rules of brainstorming

To run a smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

M.Aravind  
M.B.Yogeshwaran  
J.P.MuthuAravindh  
K.Maruthupandian

## Step-2: Brainstrom , Idea Listing and grouping

2

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

10 minutes

**M.Ansaid**

We could have reduced the information needed to enter as some features have no impact on predicted risk.

**I kept everything**  
so others who develop their own model using the same dataset can simply copy the model.

**We could use the**  
optimal or almost optimal recommendation instead of manual data entry.

**Sound clinical**  
inference can make it easier to configure and obtain much good results.

**The logging provides**  
learning algorithm with all the information needed to adjust the best parameters.

**comparative**  
analysis of the results of various models is learning.

**But what we have**  
will suffice for the example and would not be as easy to implement as part of a pilot.

**Using predicted**  
risk score seems to be useful for recommending follow-up actions, as can be seen in this example.

**If we repeat the**  
test of heart disease in a patient, our predicted probability must be self-adjusted.

**HDP is**  
Wellness, uniformly, scalable, reliable and expandable.

**Machine learning**  
algorithms such as Naïve Bayes, Support Vector Machine, etc. have been used to predict the likelihood of risk.

**The motivation for**  
the study was to find the most robust ML algorithm for detection of heart disease.

**J.P.Mulla Ansaidh**

Two heart disease detection system based on clinical information of heart disease patients and a predictive model.

**Hyperparameter**  
tuning using random search or grid search to find the best parameters for the ML model.

**develop an artificial**  
neural network algorithm for classifying heart disease based on some clinical tests.

**Using medical**  
profiles such as a age, sex, blood pressure, cholesterol, pain type, feeling blood sugar.

**predictor variables**  
such as age, sex, blood pressure, cholesterol, pain type, feeling blood sugar.

**CRF patients along**  
with a 10-fold cross-validation procedure as well as model metrics to evaluate the ability of different data.

**Machine learning**  
ML can bring an effective solution for decision making and accurate predictions.

**heart**  
disease is based on high blood pressure.

**that each technique**  
has its unique strength in making the diagnosis of the disease using gen.

**Improved the**  
quality of cardiovascular disease prediction using a better processing phase.

**different data**  
mining techniques can be utilized.

**The HDP**  
system developed in this study.

2

**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

**Identifying**  
minimum key requirements

**how to**  
address requirements

**A place to input**  
the patient characteristics used as features in the predictive model.

**Sketching out a**  
UI that incorporates key requirements.

**An output risk**  
score for the patient that also assigns them to a risk group.

**TIP**  
Add understandable tags to sticky notes to make it easier to find, organize, organize, and categorize important ideas as themes within your model.

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



5

## 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 to 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	<ul style="list-style-type: none"><li>➤ Saving lives, User friendly interactive dashboard.</li><li>➤ Reduces the exorbitant medical cost of the patients.</li><li>➤ Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions and experiences.</li></ul>
5.	Business Model (Revenue Model)	<ul style="list-style-type: none"><li>➤ Data security.</li><li>➤ Easy to use.</li><li>➤ Constant updates according to necessity.</li></ul>
6.	Scalability of the Solution	<ul style="list-style-type: none"><li>➤ Can be used in any platform (Windows, mac, etc.,).</li><li>➤ Adding new feature doesn't affect the performance of the system.</li><li>➤ Scalable dataset.</li></ul>

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	<b>1. CUSTOMER SEGMENT(S)</b> <span>CS</span> <p>People with heart disease. Aged Persons(Above 60)</p>	<b>6. CUSTOMER CONSTRAINTS</b> <span>CC</span> <p>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.</p>	<b>5. AVAILABLE SOLUTIONS</b> <span>AS</span> <ul style="list-style-type: none"> <li>• Avoid smoking</li> <li>• Take healthy foods</li> <li>• Visit cardiologist in case of any symptoms • Maintaining healthy exercise</li> </ul>	Explore AS, differentiate
	<b>2. JOBS-TO-BE-DONE / PROBLEMS</b> <span>J&amp;P</span> <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.</p>	<b>9. PROBLEM ROOT CAUSE</b> <span>RC</span> <p>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</p>	<b>7. BEHAVIOUR</b> <span>BE</span> <p>Chest pain or discomfort, Shortness of breath, Slow heartbeat, Lightheadedness, Swelling in the legs, belly area or areas around the eyes.</p>	
Identify strong TR & EM	<b>3. TRIGGERS</b> <span>TR</span> <p>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.</p>	<b>10. YOUR SOLUTION</b> <span>SL</span> <p>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.</p>	<b>8. CHANNELS of BEHAVIOR</b> <span>CH</span> <p><b>8.1 ONLINE</b> Visualizing the datasets. Exploration of data.</p> <p><b>8.2 OFFLINE</b> Cleansing of datasets. Collection and noting the datasets.</p>	Identify strong TR & EM
	<b>4. EMOTIONS: BEFORE / AFTER</b> <span>EM</span> <p><b>Before</b> -&gt; It creates a huge ambiguity in knowing the proper or accurate reasons for a heart disease.</p> <p><b>After</b> -&gt; 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.</p>			

### 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

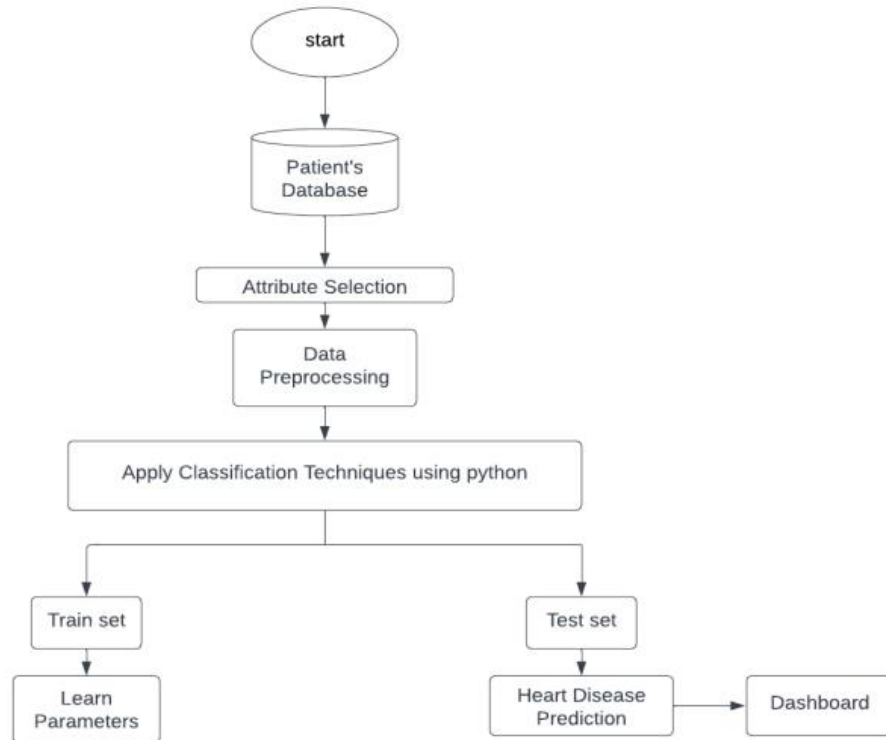
### Non-functional Requirements:

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

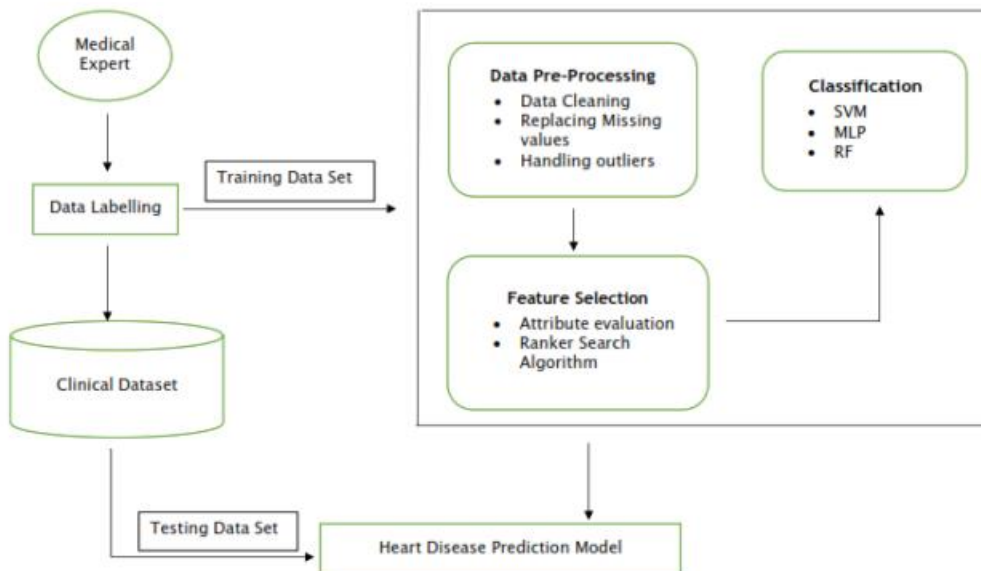
NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and user-Friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be Performed with just a few clicks.
NFR-2	Security	For security 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

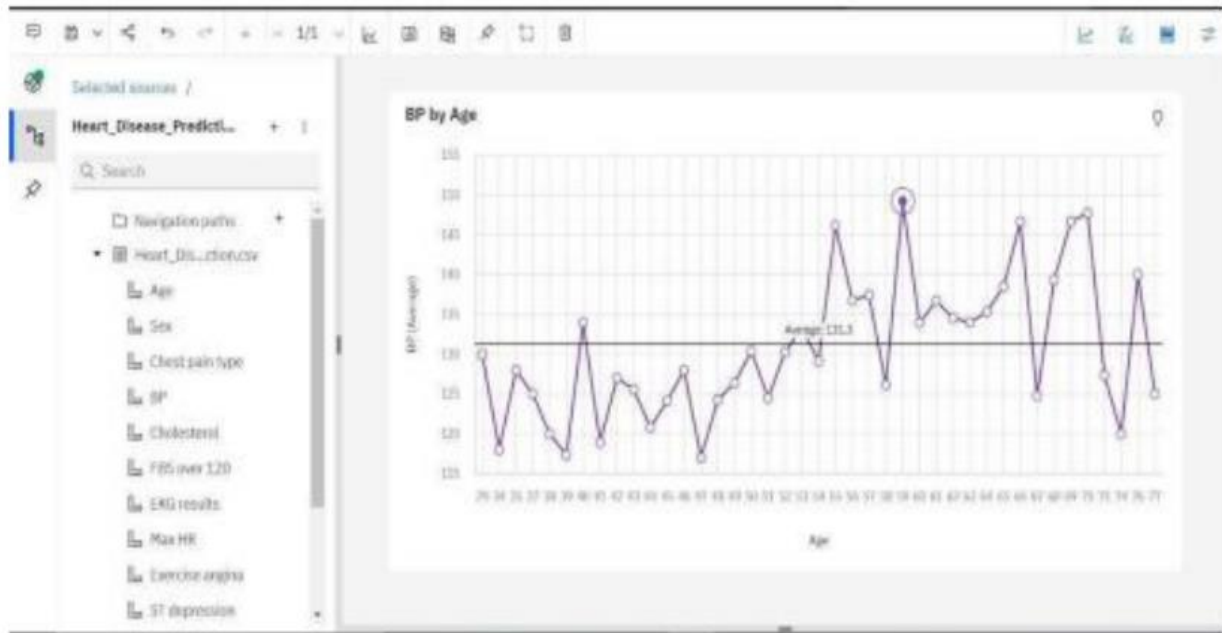
### 6.1 Iscript planning Execution

**Product Backlog, Sprint Schedule, and Estimation**

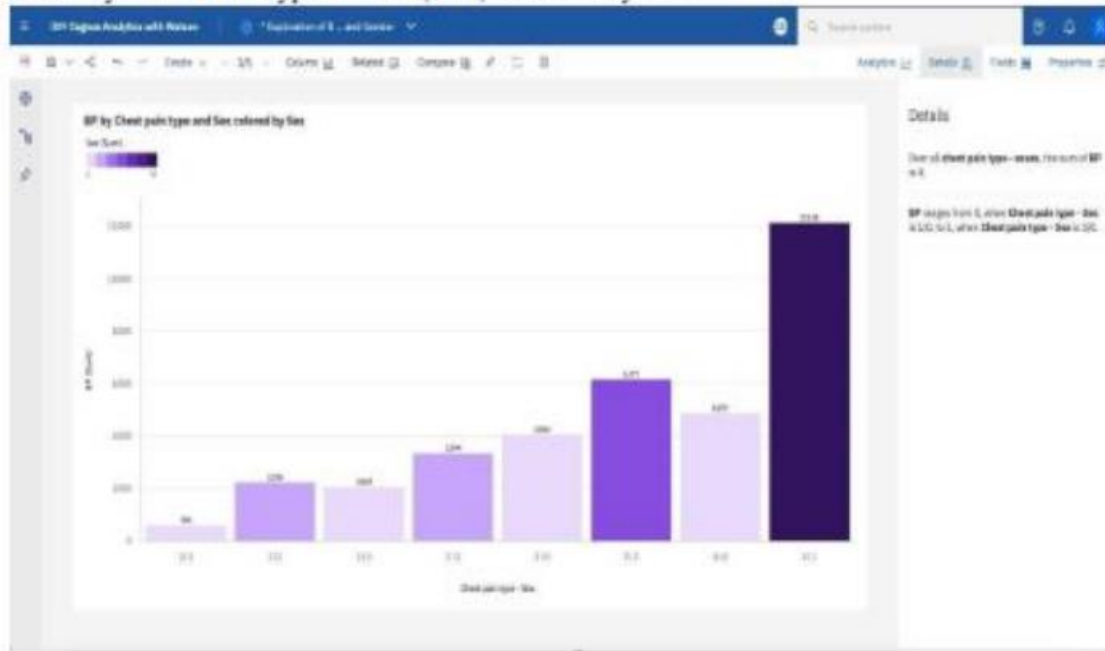
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	1	High	Aravind M Yogeshwaran M B Muthu Aravindh J P Maruthupandiyan 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

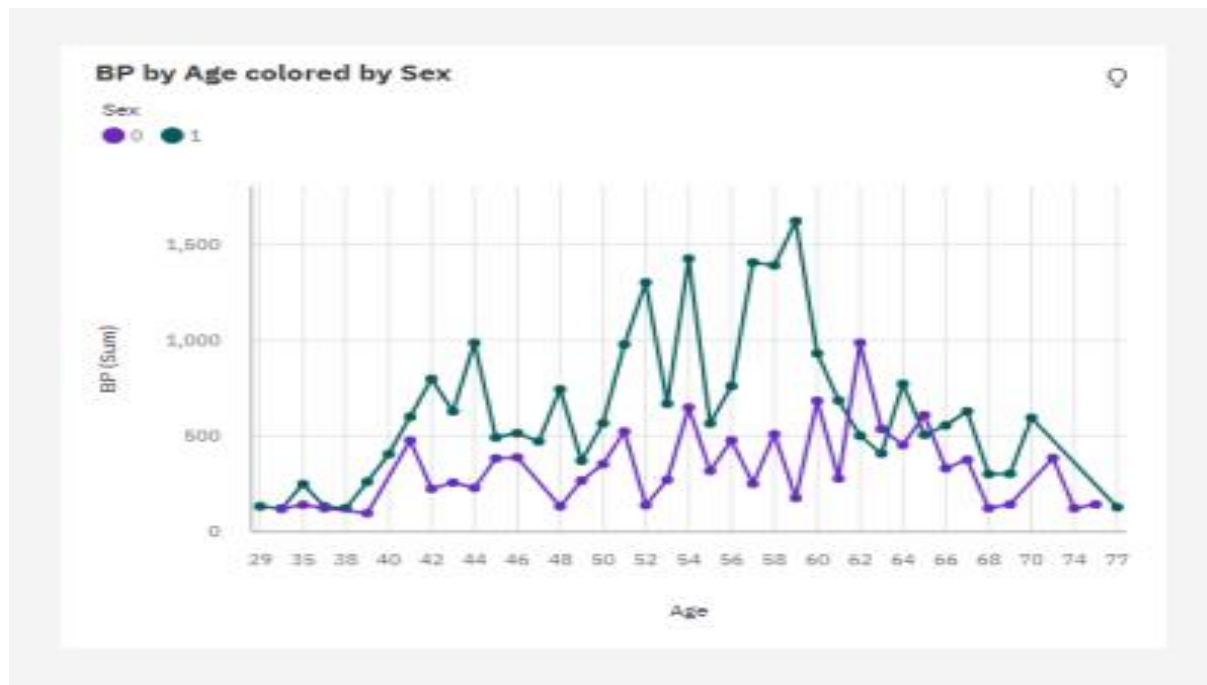
## 6.2 sprint Delivery Schedule

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

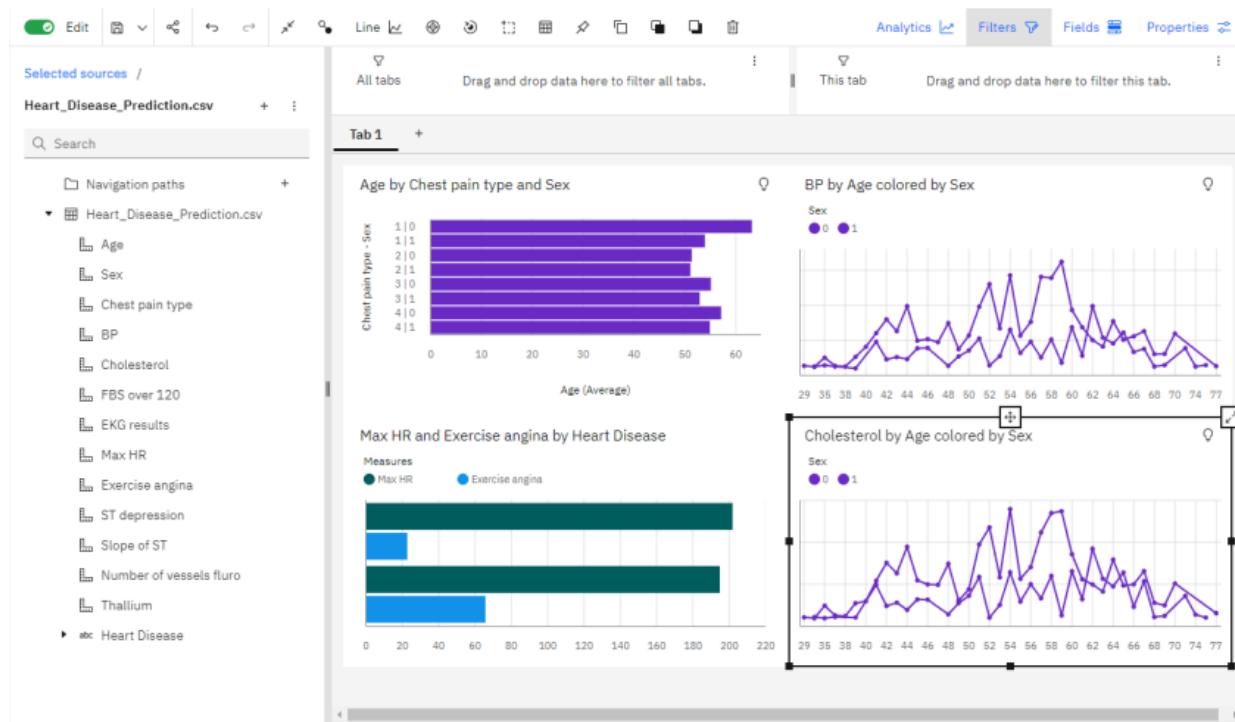


### 3. BP by Chest Pain Type and Sex(sum) coloured by Sex:





## 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_resshaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resshaped)
print(pre1)
a1=accuracy_score(pre1,model1.predict(input_resshaped))*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_resshaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resshaped)
print(pre1)
a1=accuracy_score(pre1,model1.predict(input_resshaped))*100
print(a1)

['Presence']
100.0
```

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

Heart Disease Test

### Heart Disease Test Form

Age	Sex		
18	Male		
Chest Pain Type	Resting Blood Pressure in mm Hg	Serum Cholesterol in mg/dl	Fasting Blood Sugar > 120 mg/dl
Atypical Angina	145	146	True
Resting ECG Results	Maximum Heart Rate	Exercise Induced Angina	ST Depression Induced
Normal	136	-- Select an Option --	
Slope of the Peak Exercise ST Segment	Number of Vessels Colored by Flourosopy	Thalassemia	
-- Select an Option --	-- Select an Option --	-- Select an Option --	

clideo.com

Heart Disease Test

Age  Sex

Chest Pain Type  Resting Blood Pressure in mm Hg  Serum Cholesterolal in mg/dl  Fasting Blood Sugar > 120 mg/dl

Resting ECG Results  Maximum Heart Rate  Exercise Induced Angina  ST Depression Induced

Slope of the Peak Exercise ST Segment  Number of Vessels Colored by Flourosopy  Thalassemia

**Result**

**The patient is likely to have heart disease!**

clideo.com

```
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=0.5, 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