

Visualizing and Predicting Heart Diseases with an Interactive Dashboard

**NALAIYA THIRAN PROJECT REPORT
2022**

Submitted by

MANASA K	722819205020
KOUSHIKA S	722819205025
SAI HAREESH V	722819205037
SARANYA DEVI A	722819205040

Team ID: PNT2022TMID19718

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

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

1. Introduction

1.1 Project Overview

The leading cause of death in the developed world is heart disease. 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 to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

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. Based on the analytics we can analyze which patients are most likely to suffer from heart disease in the near future and based on the patient details we will make decisions to cure them.

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 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 304 records with 14 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, Kmeans 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.

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

Several studies have been carried out for developing prediction model using individual technique and also by combining two or more techniques. This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers. Into practice.

2.3 Problem Statement Definition

Who does the problem affect?

People with unhealthy lifestyles, stress, depression, age above 40 and when their ancestors got heart disease (since heart disease is hereditary).

When does the issue occur?

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.

What would happen if we didn't solve the problem?

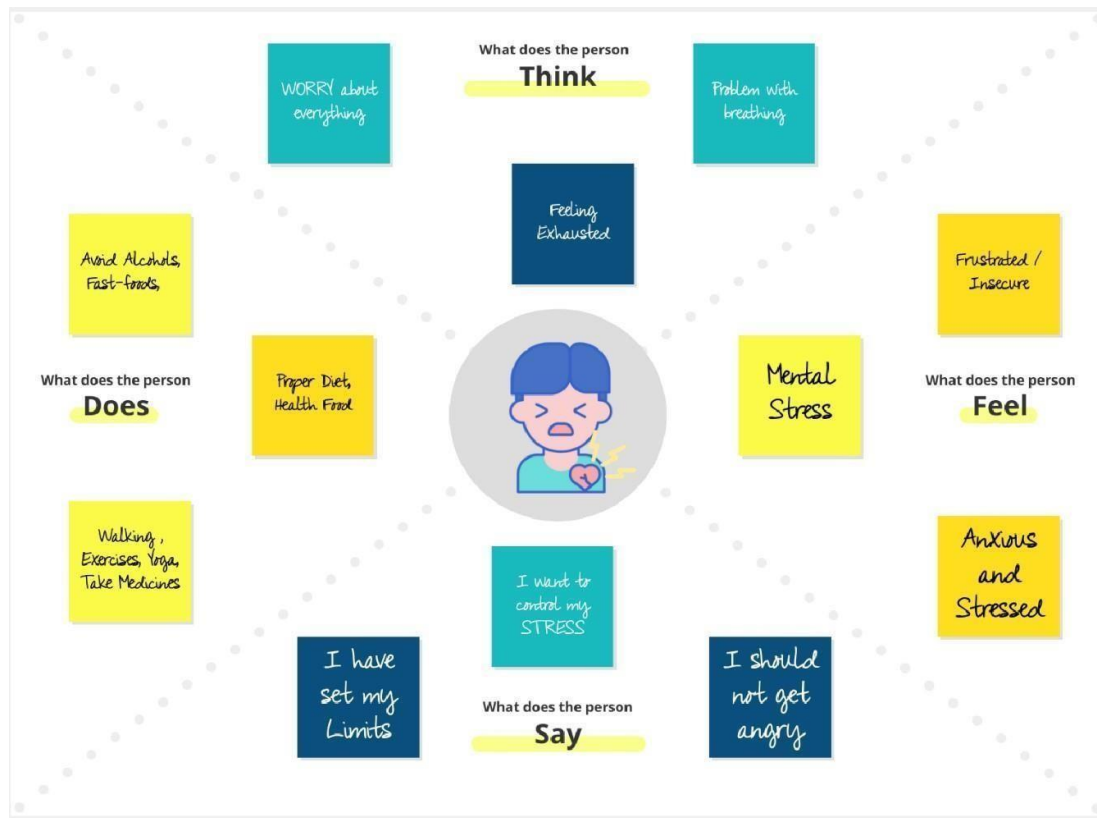
If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will increase rapidly.

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

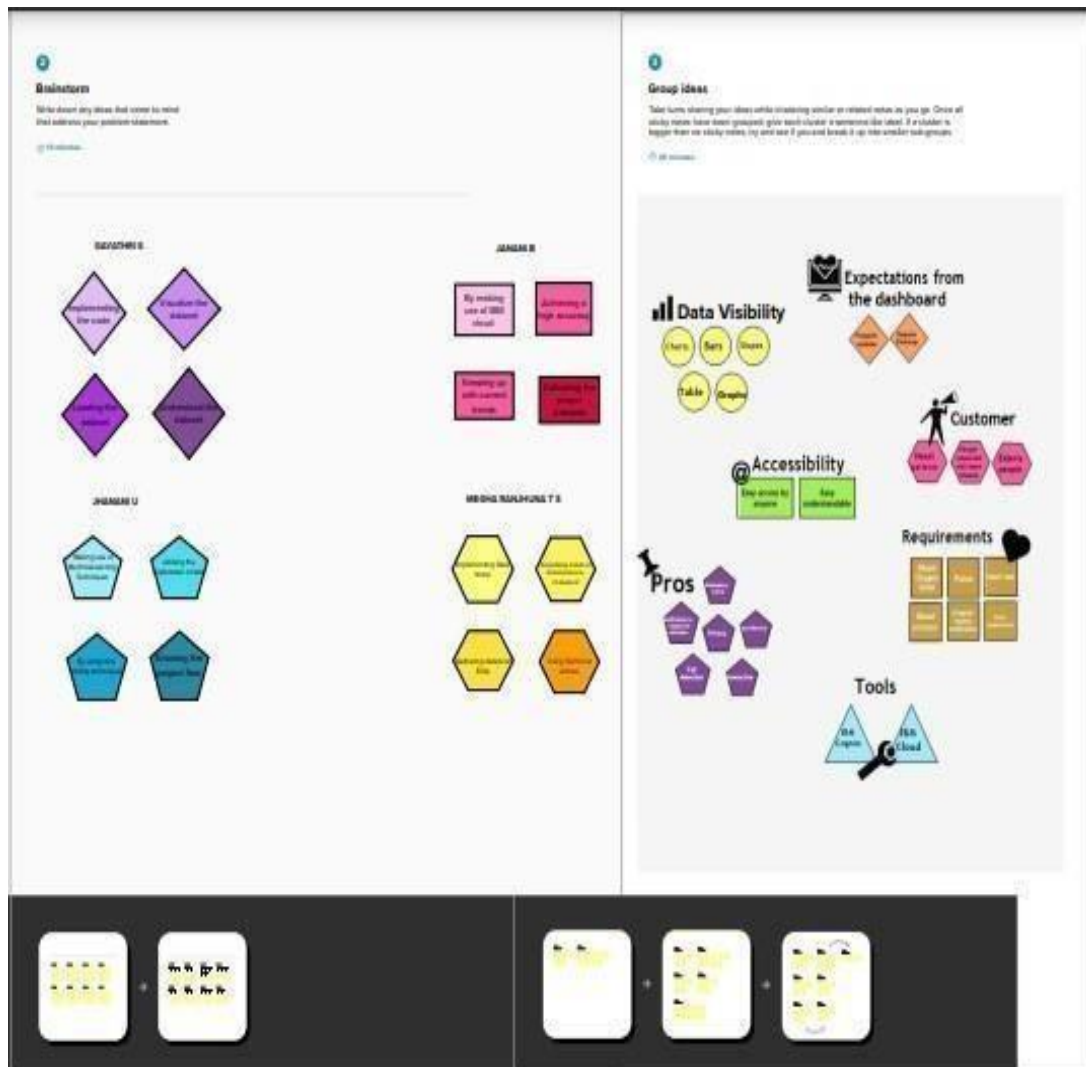


3.2 Ideation and Brainstorming

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



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization

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
How much does your idea matter? How much effort will it take to implement?

Feasibility
How easy is it to implement? How much effort will it take to implement?

After you collaborate

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

Quick add-ons

- Share the mural**
Share a new link to the mural with collaborators 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 share to email, include in slides, or save in your drive.

Keep moving forward

- Strategy blueprint**
Define the components of a new design strategy.
[Open the template](#)
- Customer experience journey map**
Understand customer needs, motivations, and obstacles for an experience.
[Open the template](#)
- Strengths, weaknesses, opportunities & threats (SWOT) to develop a plan**
[Open the template](#)











[Share template feedback](#)

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The leading cause of death in the developed world is heart disease. As a result, work must be done to reduce the risks of having a heart attack or stroke. It is infeasible for a common man to frequently undergo tests for ECG and so on. Hence, it requires a replacement that is both convenient and dependable.
2.	Idea / Solution description	The proposed solution proposes an interactive dashboard for visualizing and forecasting heart disorders, in which the user may observe his/her

		medical report analysis as well as the projected end result. IBM Cognos will be used to create the dashboard. Machine learning Algorithms will be used to forecast cardiac disease.
3.	Novelty / Uniqueness	Makes recommendations to the user based on that person's medical analysis.
4.	Social Impact / Customer Satisfaction	It helps with disease prediction at an early stage and frequently alerts the user to their current health status. Both the user and the doctor can benefit from the system's improved decision-making regarding cardiac disease
5.	Business Model (Revenue Model)	Can be deployed by Hospitals or NFOs, so that it makes the analysis in a fast manner.
6.	Scalability of the Solution	The solution can work effectively on long and small datasets. It can also be changed to predict various other diseases depending on the dataset

3.4 Problem Solution Fit

Project Title: Visualizing and Predicting Heart Diseases with an Interactive Dash Board			Project Design Phase-4 - Solution Fit Template			Team ID: PNT2022TMD09615		
Define CS, M, Info, CC	1. CUSTOMER SEGMENT(S) 		4. CUSTOMER CONSTRAINTS 		5. AVAILABLE SOLUTIONS 		Explore AS, differentiate	
	<ul style="list-style-type: none">HospitalsClinicsWHOAny medical related agencies those prepare medicines or any kind of solutions inferring over the data of diseases.		The unawareness over the AI/ML technologies, collaborative dashboards, network connection, lack of data.		The customers can prefer over a manual data visualization and prediction, which is very tedious job and requires the knowledge over the technologies of AI/ML. Hard mathematical formulae were created and the results were being calculated manually.			
Time to do CS, M, Info, CC	2. JOBS-TO-BE-DONE / PROBLEMS 		6. PROBLEM ROOT CAUSE 		7. BEHAVIOUR 		Time to do JS, RC, BE, CS, M, Info, CC	
	<p>Quality of Data:</p> <p>The quality of data should be accurate and reliable. Obviously, the outcome will solely depend on the data we put into the prediction. If the data is skewed, then the prediction which is dependent on it, will be skewed as well.</p>		<ul style="list-style-type: none">Difficulty of predicting a heart disease.Will not have a proper idea of relation between similar heart diseases.There is a chance of identifying every heart diseases as same.Reason of increase in heart disease will not be rootly identified.		<ul style="list-style-type: none">Generation of legitimate and reliable datasets.Customers need to collect more number of datasets in order to obtain more accurate result.Must obtain knowledge of difference between datasets that is used for comparison.			
Identify along TR & CH	3. TRIGGERS 		8. YOUR SOLUTION 		9. CHANNELS of BEHAVIOR 		Identify along TR & CH	
	<ul style="list-style-type: none">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.		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>9.1 ONLINE Visualizing the datasets. Exploration of data.</p> <p>9.2 OFFLINE Cleaning of datasets. Collection and noting the datasets.</p>			
4. EMOTIONS: BEFORE / AFTER 								
<p>Before -> It creates a huge ambiguity in knowing the proper or accurate reasons for a heart disease.</p> <p>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.</p>								

4.

4. Requirement Analysis

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 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 Cognos Analytics
FR-4	Generation Report	User can view his/her health report and can make decisions accordingly

4.2 Non-Functional Requirement

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

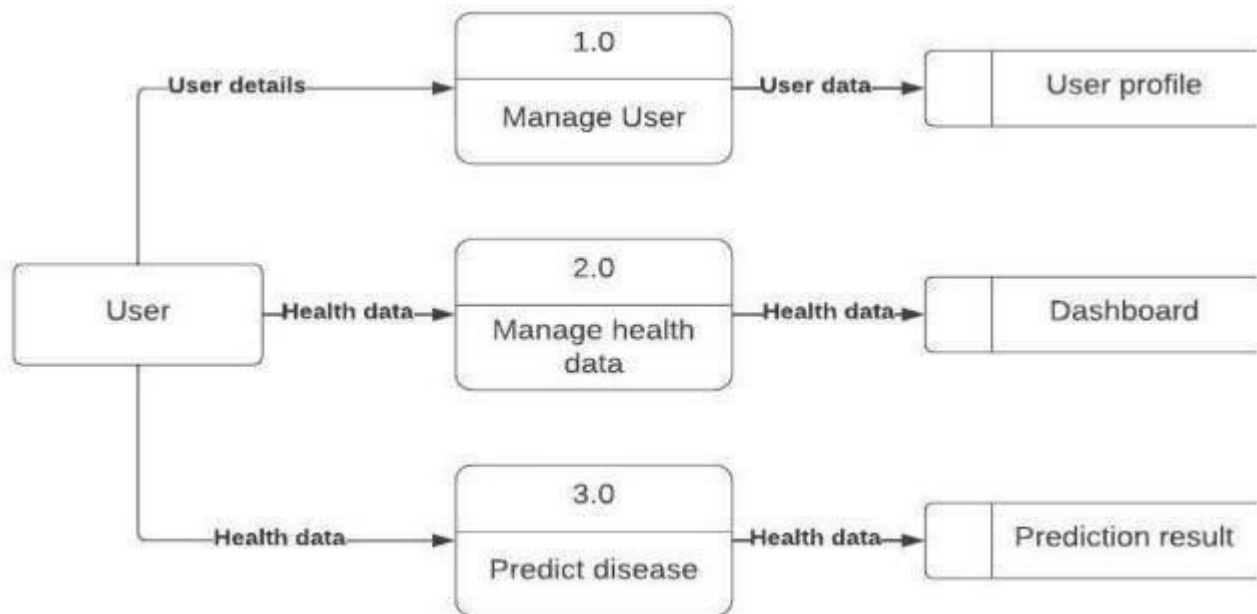
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and userfriendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	Security	For security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

5.

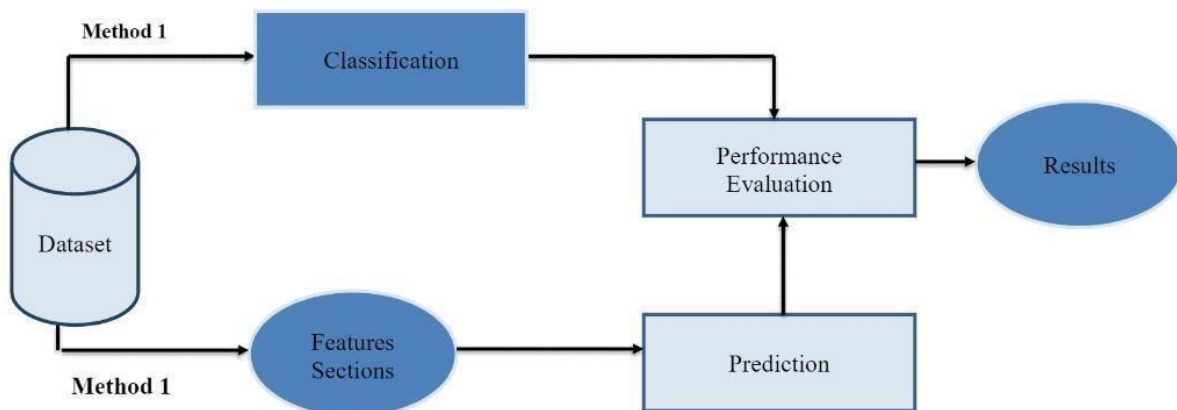
NFR-6	Scalability	The application can withstand the increase in the no. of users and has to be able to develop Higher versions
-------	-------------	--

5. Project Design

5.1 Data Flow Diagram



5.2 Solution and Technical Architecture



6.

6. Project Planning and Scheduling

6.1 Script Planning and Execution

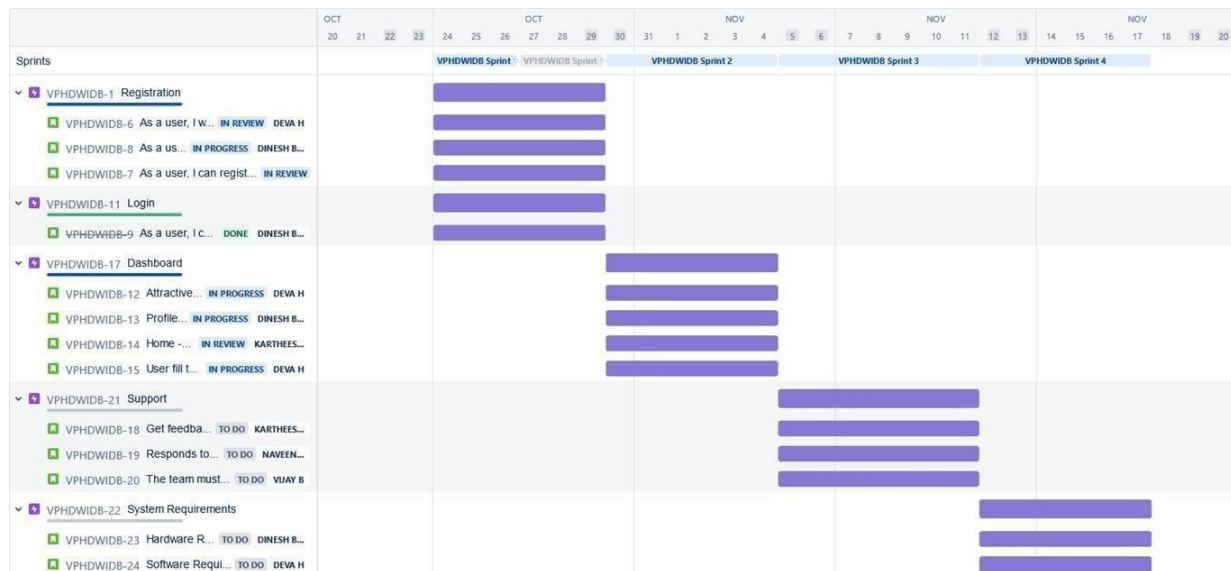
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password	3	High	2
Sprint-1		USN-2	As a user, I will receive confirmation 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 Gmail	3	Medium	1
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email & password	6	High	2
Sprint-2	Dashboard	USN-5	Attractive dashboard For the Application	3	Medium	3
Sprint-2		USN-6	Profile - view & update your profile	5	Low	2
Sprint-2		USN-7	Home - Analyze your Heart problem	2	High	1
Sprint-2		USN-8	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	Support	USN-9	Get feedback from users	10	Medium	3
Sprint-3		USN-10	Responds to user queries via telephone,email etc.	3	Medium	2
Sprint-3		USN-11	The team must respond immediately to the queries based on the priority	5	High	3

Sprint-4	System Requirements	USN-12	Hardware Requirement 3. Laptop or PC i5 processor system or higher 4 GB RAM or higher 128 GB ROM or higher 4. Mobile (12.0 and above)	5	Low	2
Sprint-4		USN-13	<ul style="list-style-type: none"> • Software Requirement • Laptop or PC • Windows 10 or higher • Android or IOS 	8	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	28 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	02 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	09 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	16 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.

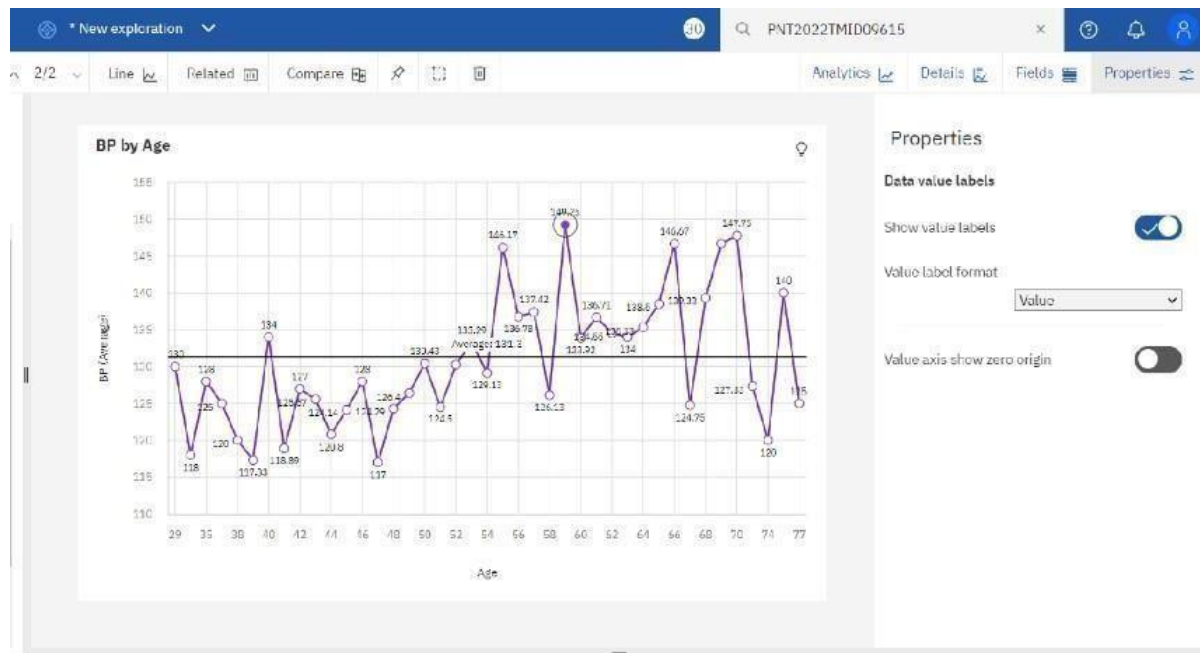
Comparing it with the accuracy gotten from Decision Tree:

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

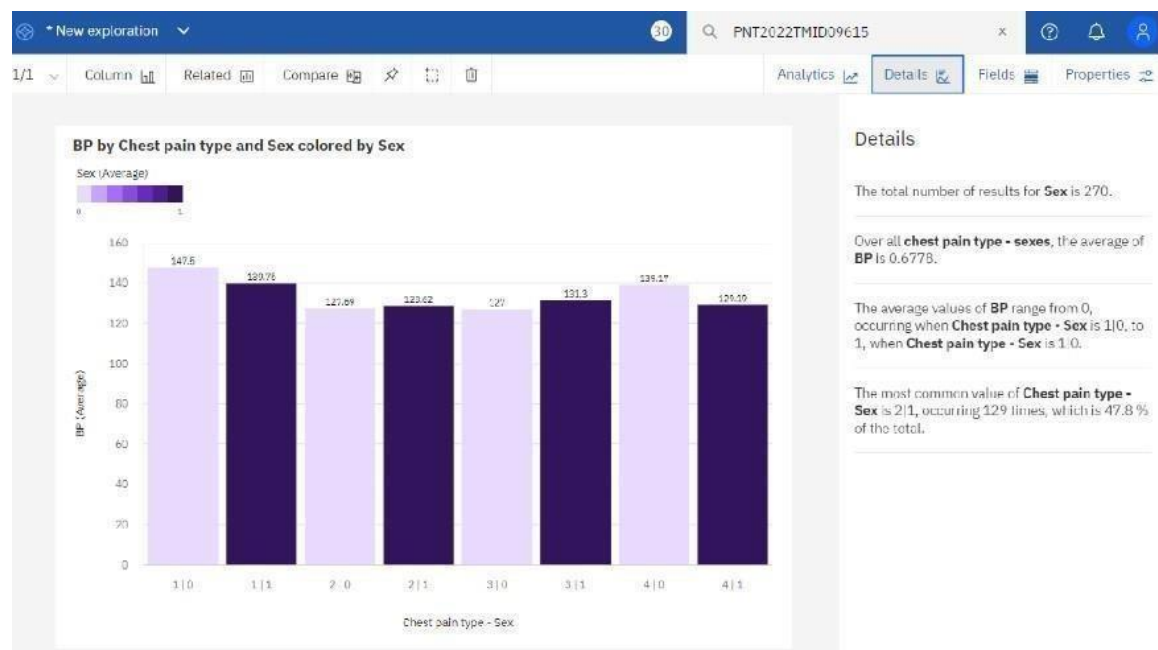
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

7.2 Dashboard

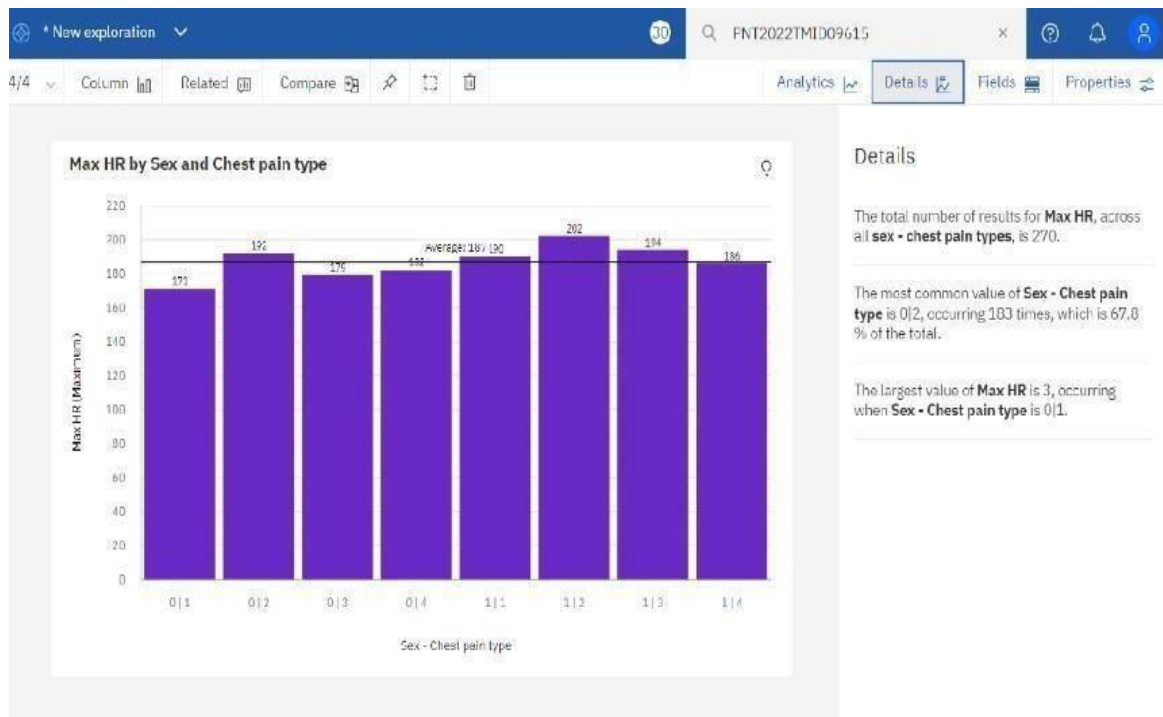
Average BP during chest pain



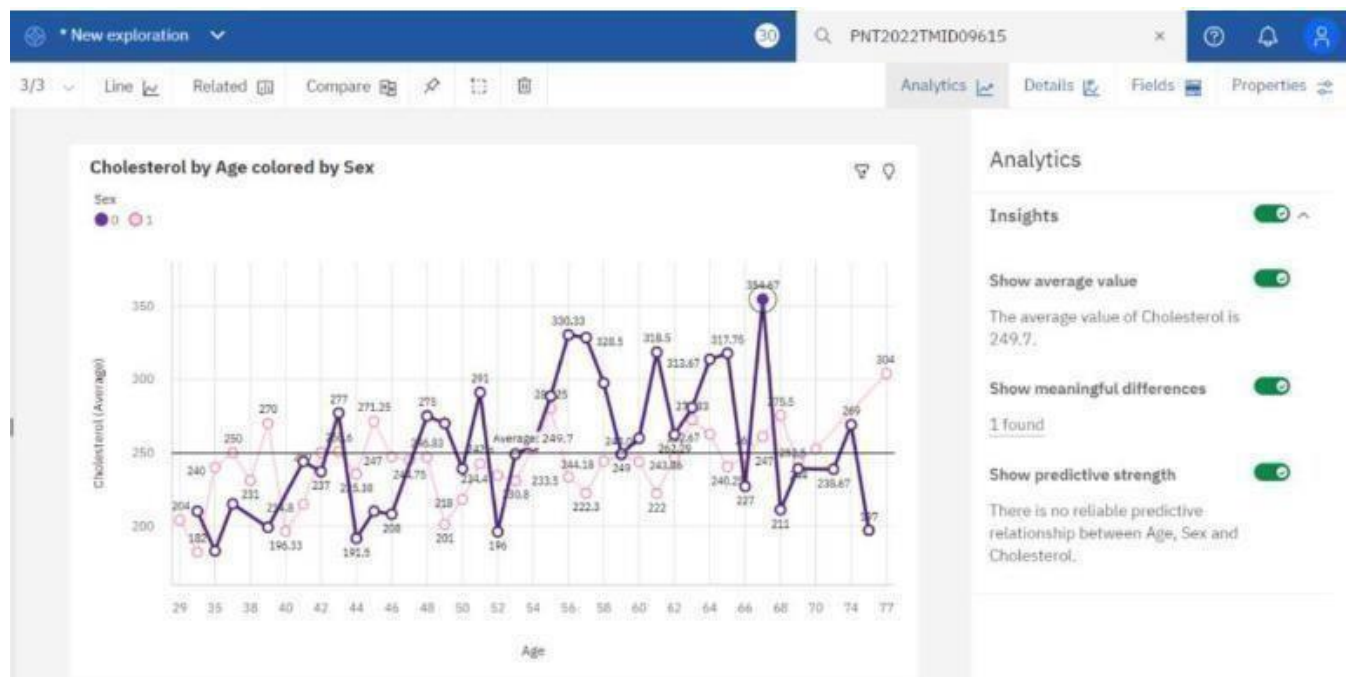
Exploration Of BPvsChestPainType And Gender:



Exploration Of Max Heart Rate During The Chest Pain:



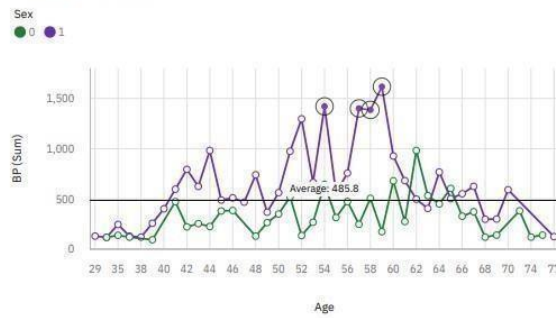
Exploration Of Cholesterol by age and Gender:



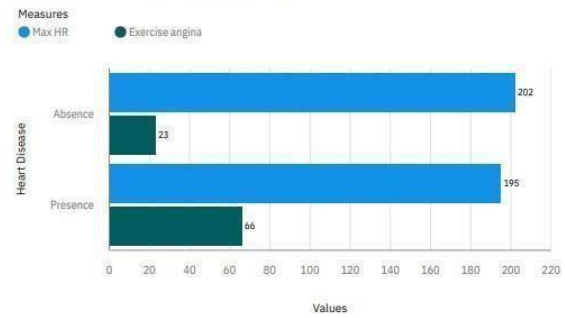
Dashboard Showing Different Types Of Visuals:

Tab 8

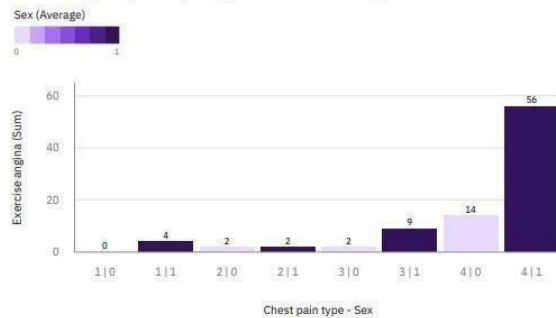
BP by Age colored by Sex



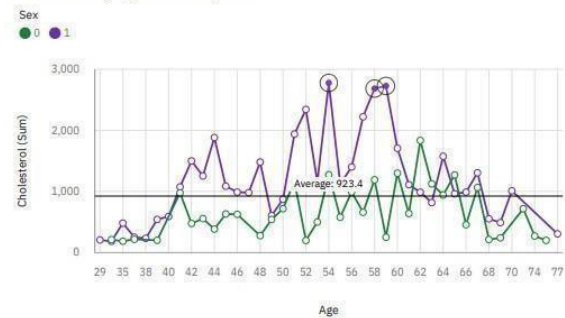
Max HR and Exercise angina by Heart Disease



Exercise angina by Chest pain type and Sex colored by Sex



Cholesterol by Age colored by Sex



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_resaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resaped)
print(pre1)
a1 = accuracy_score(pre1,model1.predict(input_resaped)) * 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_resaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resaped)
print(pre1)
a1 = accuracy_score(pre1,model1.predict(input_resaped)) * 100
print(a1)

['Presence']
100.0
```

8.2 User acceptance Testing

Testing a case where user has heart disease

localhost:4200

Exercise angina (exercise induced angina (1 = yes; 0 = no))

0

ST depression

2.4

Slope of ST

2

Number of vessels fluro (number of major vessels (0-3) colored by flourosopy)

3

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

3

Submit

localhost:4200 says
The patient has increased risk of heart diseases

OK

Testing a case where user does not have heart disease

localhost:4200

Max HR (maximum heart rate achieved)

160

Exercise angina (exercise induced angina (1 = yes; 0 = no))

0

ST depression

1.6

Slope of ST

2

Number of vessels fluro (number of major vessels (0-3) colored by flourosopy)

0

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

7

Submit

localhost:4200 says
The patient has no risk of heart diseases

OK

9. Result

9.1 Performance Metrics

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

```

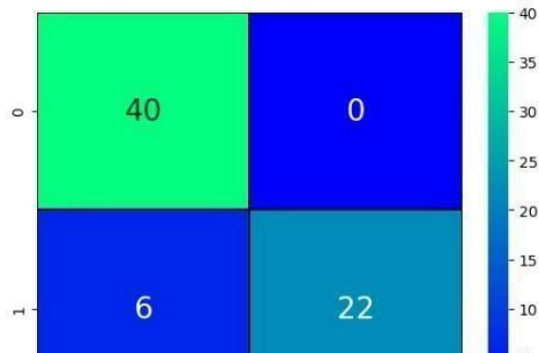
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

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

```

	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

13. Appendix

Source Code:

https://github.com/IBM-EPBL/IBM-Project-5292-1658755860/blob/main/IBM_WORKS/FINAL%20DELIVERABLES/Project%20Report/Project%20Report.pdf

Demo video link:

<https://www.dropbox.com/s/dowswwjuz0rgj6p/D.V.mp4?dl=0>