Visualizing and Predicting Heart Diseases with an Interactive Dashboard

NALAIYA THIRAN PROJECT REPORT
2022

Submitted by

Sangeeth Kanna P	**190701094**
Subhashree B	**190701110**
Uma Maheshwari D	**190701120**
Vishnupriya G	**190701130**

Team ID: PNT2022TMID53645

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 : PNT2022TMID53645

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

Even though we have smart watches that constantly monitor various aspects like heart rate, oxygen levels, etc but it doesn't give us any conclusive results on the person's health. We also have applications that give ECG from smart watches.
Heart Disease is a complicated disease which is caused by a lot of attributes. Even though devices give us a lot of data we need to compare, analyze and interpret them to make it useful.

2.2 References

B. Dun, E. Wang, and S. Majumder, "Heart disease diagnosis on medical data using ensemble learning," 2016.

Deep learning, which belongs to a larger family of machine learning techniques, has the ability to effectively examine a lot of data. In this, overview of these machine learning techniques that may be used to improve the functionality and intelligence of an application. Determining the fundamentals of various machine learning approaches and how they can be used in a variety of real-world application areas, including cybersecurity systems, smart cities, healthcare, e-commerce, agriculture, and many more, is thus the core contribution of this work. We also discuss the difficulties and potential possibilities for future research based on our findings. Overall, this work seeks to serve as a resource for decision- makers in a range of practical scenarios and applications, including those in academia and industry.

F. Yaghouby, F. Yaghouby, A. Ayatollahi, and R. Soleimani,"Classification of cardiac abnormalities using reduced features".

In this study, a clinical decision support system (CDSS) that analyzes patients with heart failure (HF) and generates a variety of outputs, including an assessment of the severity of the HF, a prediction of the type of HF, and a management interface that contrasts the follow-ups of the various patients. The entire system is made up of an intelligent core component and an HF special-purpose management tool that also serves as an interface for training and using artificial intelligence. A machine learning strategy to put the smart intelligent functions into practice.

World Health Organization, Cardiovascular Diseases, WHO, Geneva, Switzerland, 2020.

By applying different machine learning algorithms and then using deep learning to see what difference comes when it is applied to the data, three approaches were used. In the first approach, normal dataset which is acquired is directly used for classification, and in the second approach, the data with feature selection are taken care of and there is no outliers detection.

American Heart Association, Classes of Heart Failure, American Heart Association, Chicago, IL, USA, 2020.

By providing more reliable and consistent techniques for the detection, classification, reconstruction, denoising, quantification, and segmentation of patterns in biomedical pictures, deep learning and machine learning have made significant advances in the field of biomedical image analysis.

2.3 Problem Statement Definition

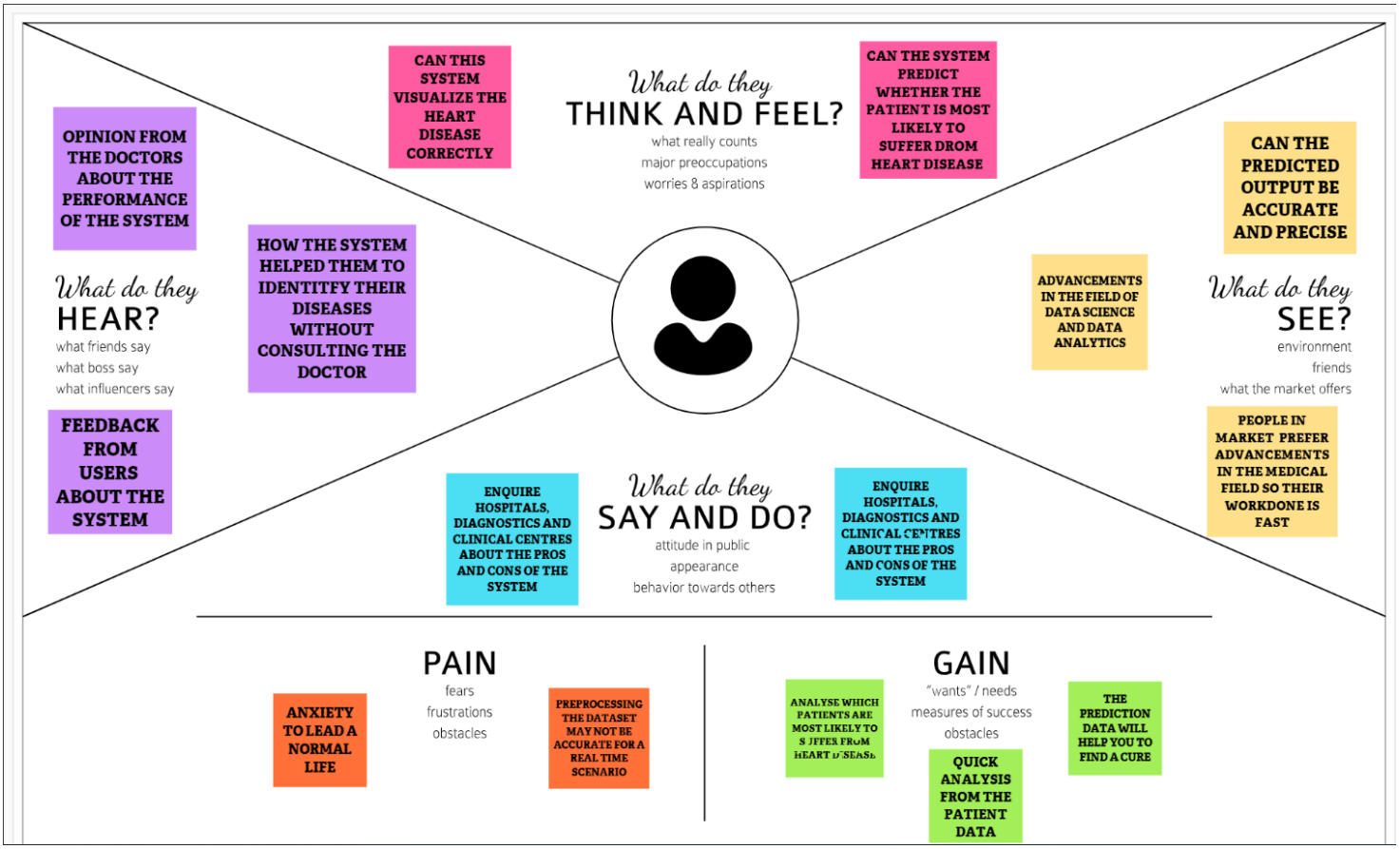
In India in 2016, CVDs (Cardiovascular Diseases) contributed to 28·1% of total deaths and 14·1% of total disability-adjusted life years (DALYs). Most persons with coronary heart disease

who pass away are 65 years of age or older. Although both sexes can get heart attacks in old age, women have a higher mortality rate (within a few weeks). Risk for heart disease can be increased by a number of medical issues, lifestyle, age, and family history. When a person is affected by heart disease, it causes side effects. Chest pain, chest tightness, chest pressure and chest discomfort Breathing difficulties, Neck, jaw, throat, upper abdomen, or back pain. Heart disease - and the conditions that lead to it - can happen at any age. High rates of obesity and high blood pressure among younger people (ages 35–64) are putting them at risk for heart disease earlier in life. CAD happens when coronary arteries struggle to supply the heart with enough blood, oxygen and nutrients. Cholesterol deposits, or plaques, are almost always to blame. These buildups narrow your arteries, decreasing blood flow to your heart. This can cause chest pain, shortness of breath or even a heart attack.

Therefore in order Predict if the patient suffers from heart disease- The health professional enters the input values from the patient's health report. The data is fed into the project model which predicts the probability of having heart disease.

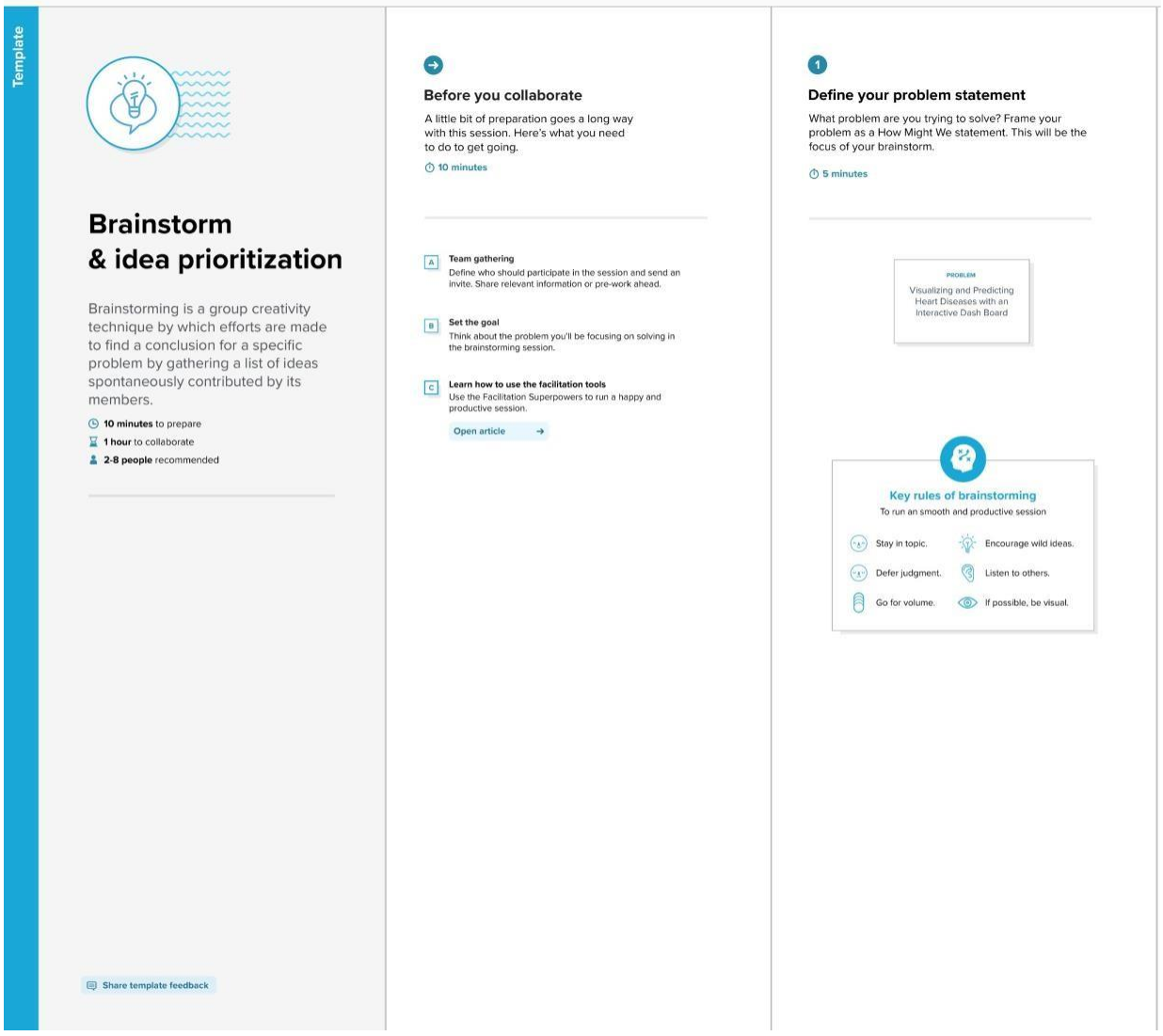
3. Ideation and Proposed Solution

3.1 Empathy Map Canvas

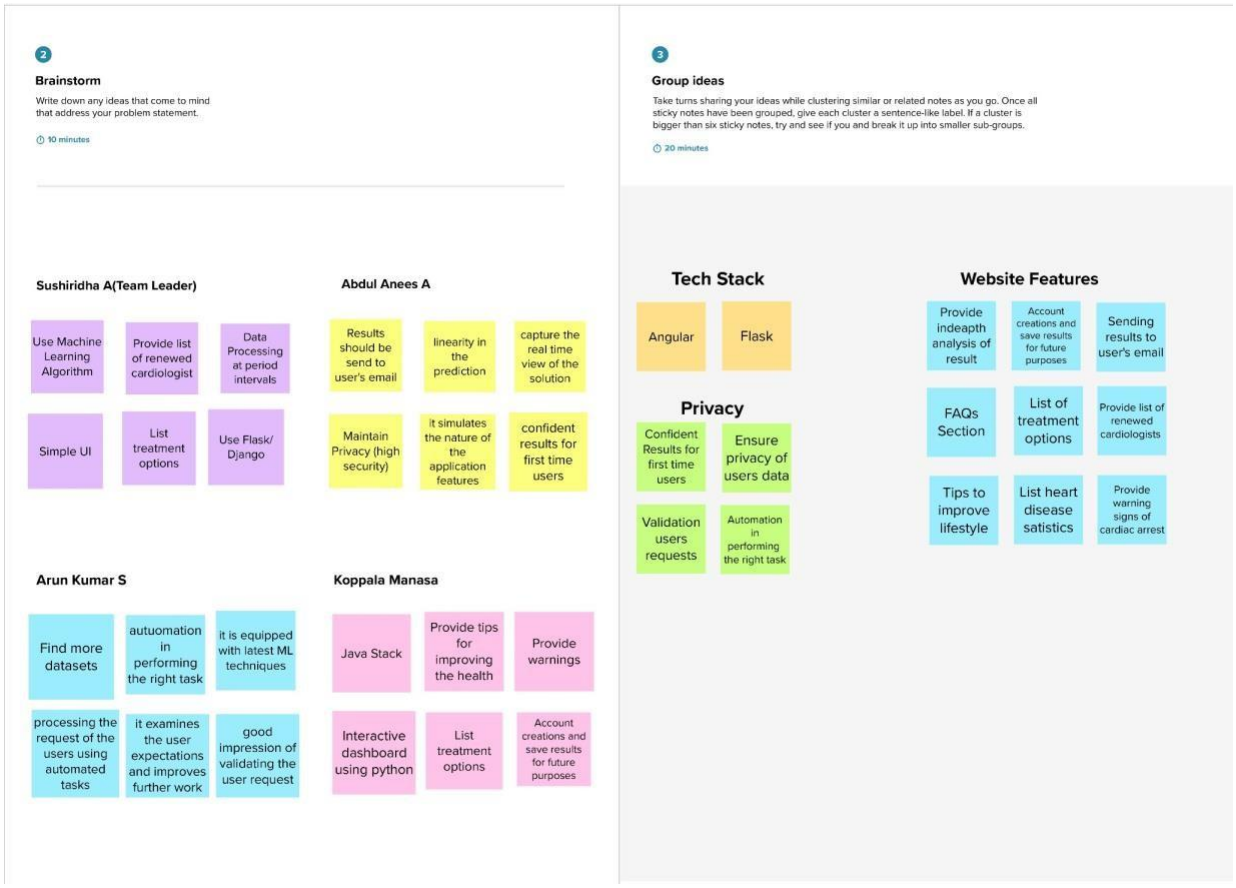


3.2 Ideation and Brainstorming

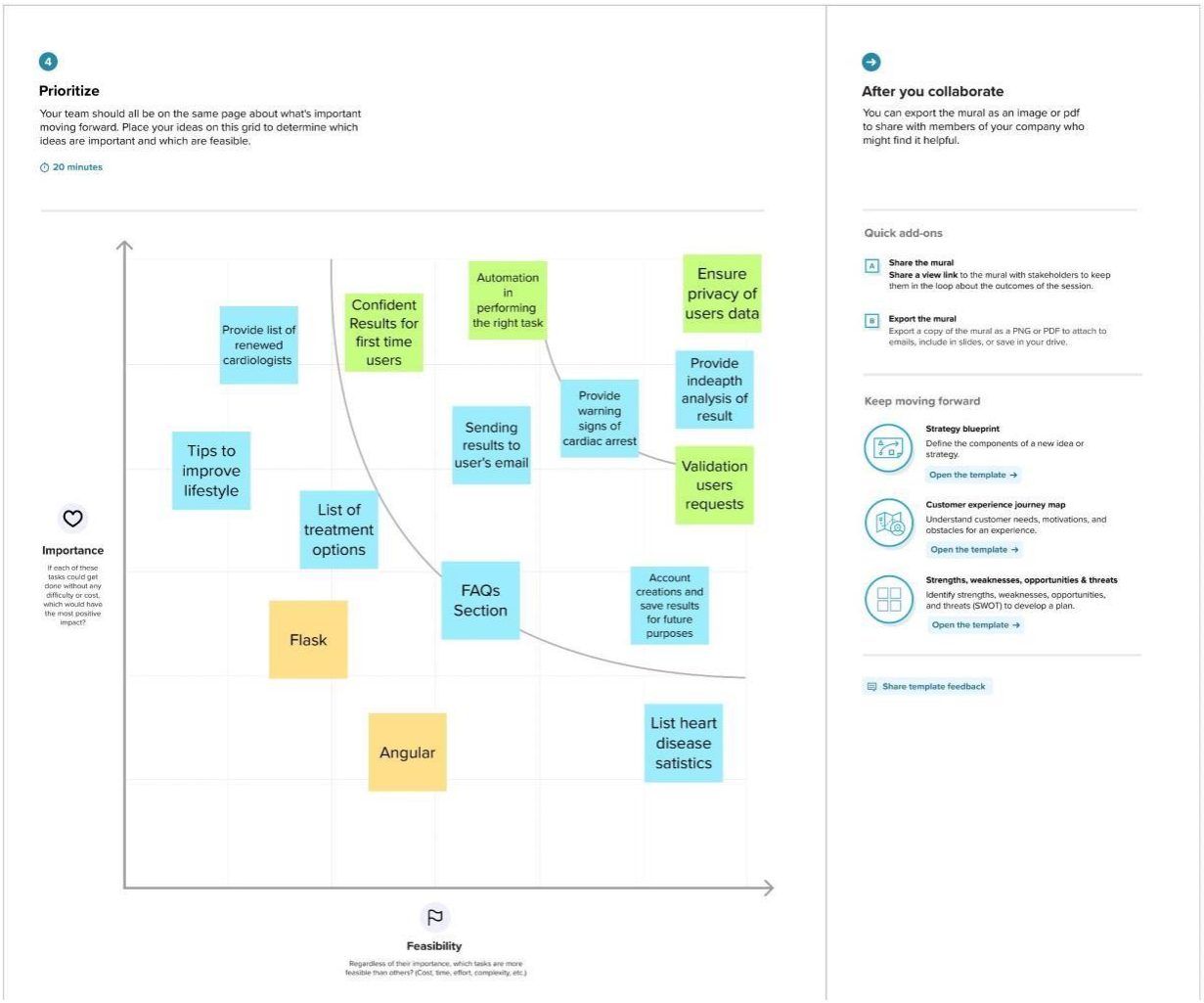
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization

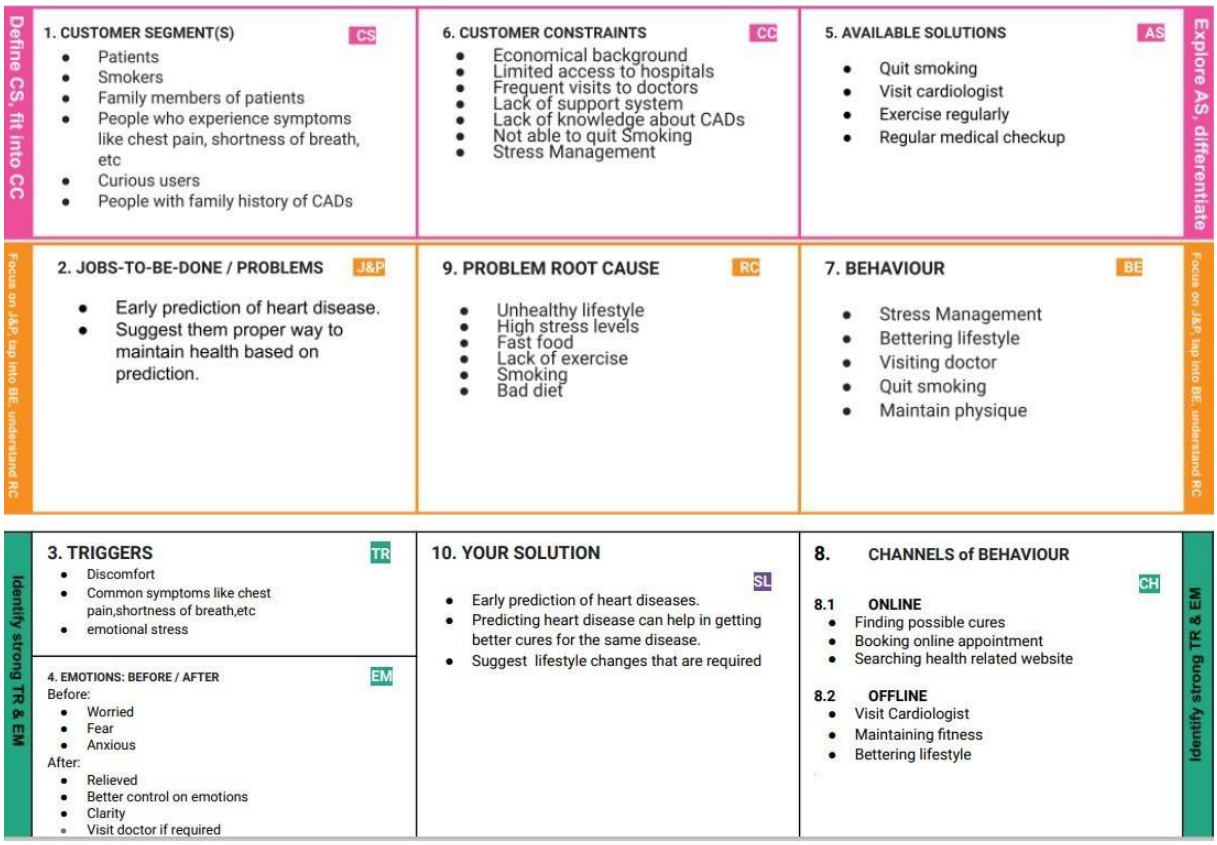


3.3 Proposed Solution

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



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	Registration through Email
FR-2	User Confirmation	Confirmation via Email
FR-3	Visualizing Data	Visualize the trends on heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generation Report	Report can be viewed by the users

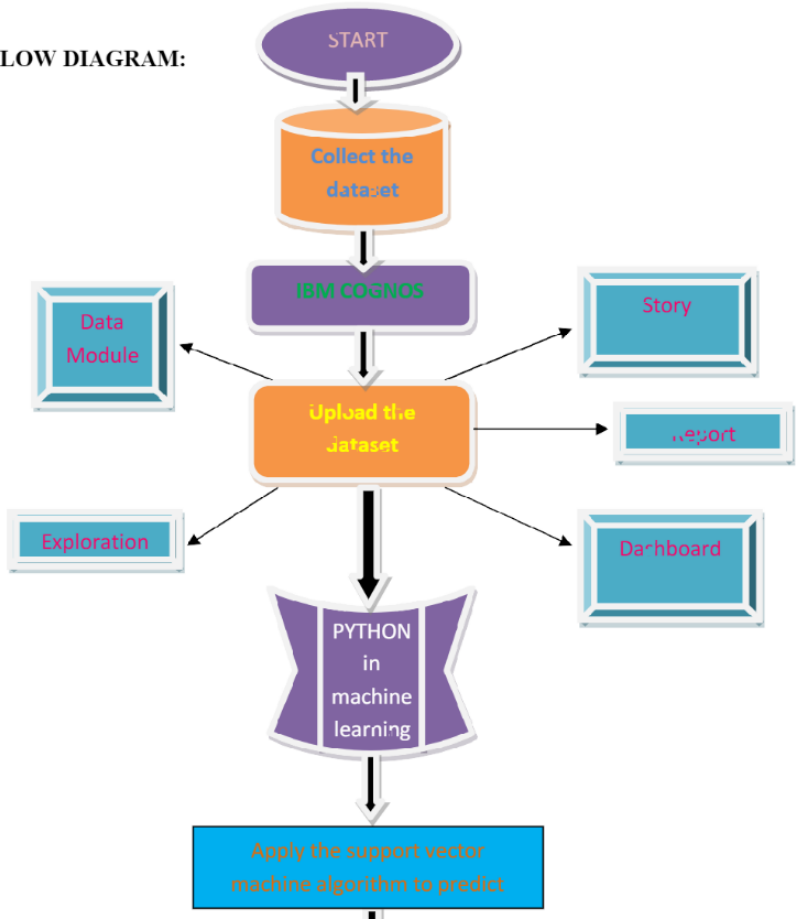
4.2 Non-Functional Requirement

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

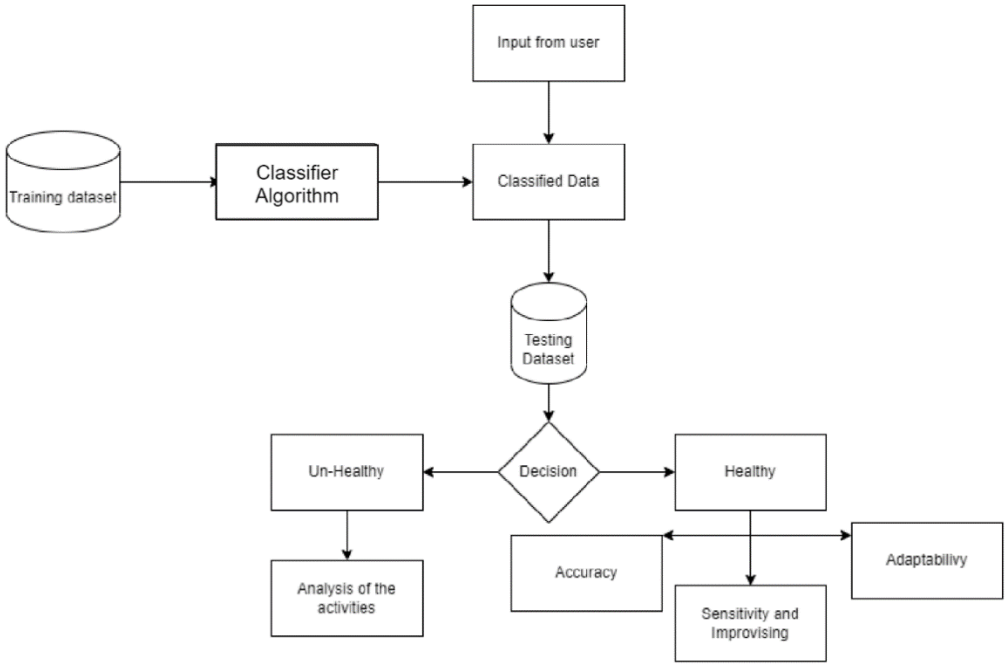
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Provide a simple UI. Actions can be easily performed by a few clicks. Features will be understandable.
NFR-2	**Security**	2 step authorization (for register) Have a backup dataset
NFR-3	**Reliability**	Error must be low.(Improve accuracy) Must work without glitches
NFR-4	**Performance**	It is affected by the implementing algorithm. Depending on the error metrics we have to choose an algorithm with high response time.
NFR-5	**Availability**	Must be available for the user 24 x 7 without interruptions
NFR-6	**Scalability**	Should withstand a high number of users and large datasets.

5. Project Design

DATA FLOW DIAGRAM:



5.1 Solution and Technical Architecture



5.2 User stories

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
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
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High
Sprint-1		USN-3	As a user, I can register for the application through Gmail	3	Medium
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email & password	6	High
Sprint-2	Dashboard	USN-5	Attractive dashboard For the Application	3	Medium
Sprint-2		USN-6	Profile - view & update your profile	5	Low
Sprint-2		USN-7	Home - Analyze your Heart problem	2	High
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
Sprint-3	Support	USN-9	Get feedback from users	10	Medium
Sprint-3		USN-10	Responds to user queries via telephone,email etc.	3	Medium
Sprint-3		USN-11	The team must respond immediately to the queries based on the priority	5	High
Sprint-4	System Requirements	USN-12	Hardware Requirement 1. Laptop or PC • i5 processor system or higher • 4 GB RAM or higher • 128 GB ROM or higher 2. Mobile • (12.0 and above)	5	Low
Sprint-4		USN-13	Software Requirement 1. Laptop or PC • Windows 10 or higher • Android Studio	8	Medium

6. Project Planning and Scheduling

6.1 Script Planning and Execution

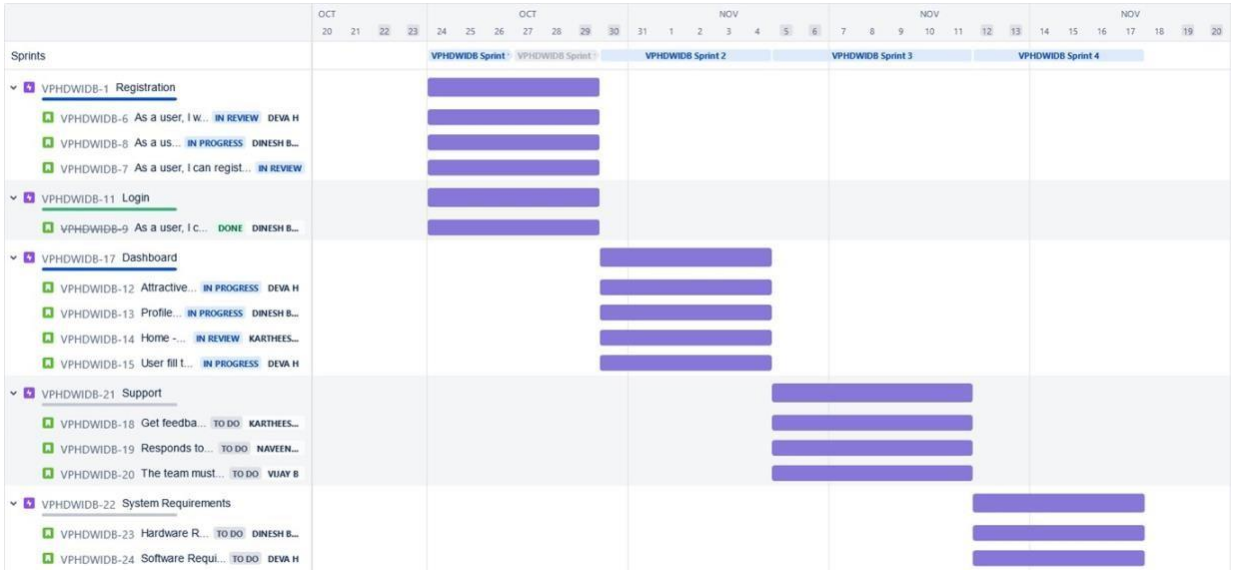
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	3	High	1
Sprint-1		USN-2	As a user, I will receive confirmation 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 Gmail	3	Medium	1
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email & password	6	High	5
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	4
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	5
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	Software Requirement 2. Laptop or PC	8	Medium	4

			• Windows 10 or higher • Android Studio			

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	30 Oct 2022	04 Nov 2022	17	04 Nov 2022
Sprint-3	20	6 Days	05 Nov 2022	11 Nov 2022	18	11 Nov 2022
Sprint-4	20	6 Days	12 Nov 2022	17 Nov 2022	19	17 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

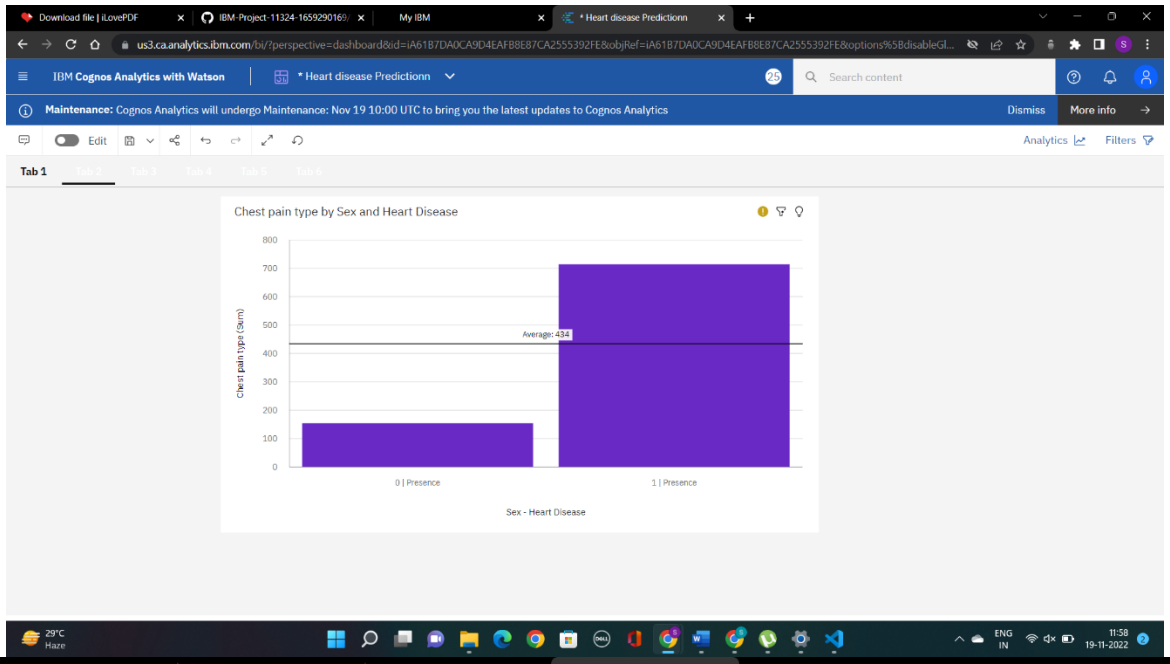
From the above result we can conclude that Logistic Regression has the hisgest accuracy for this particular dataset.

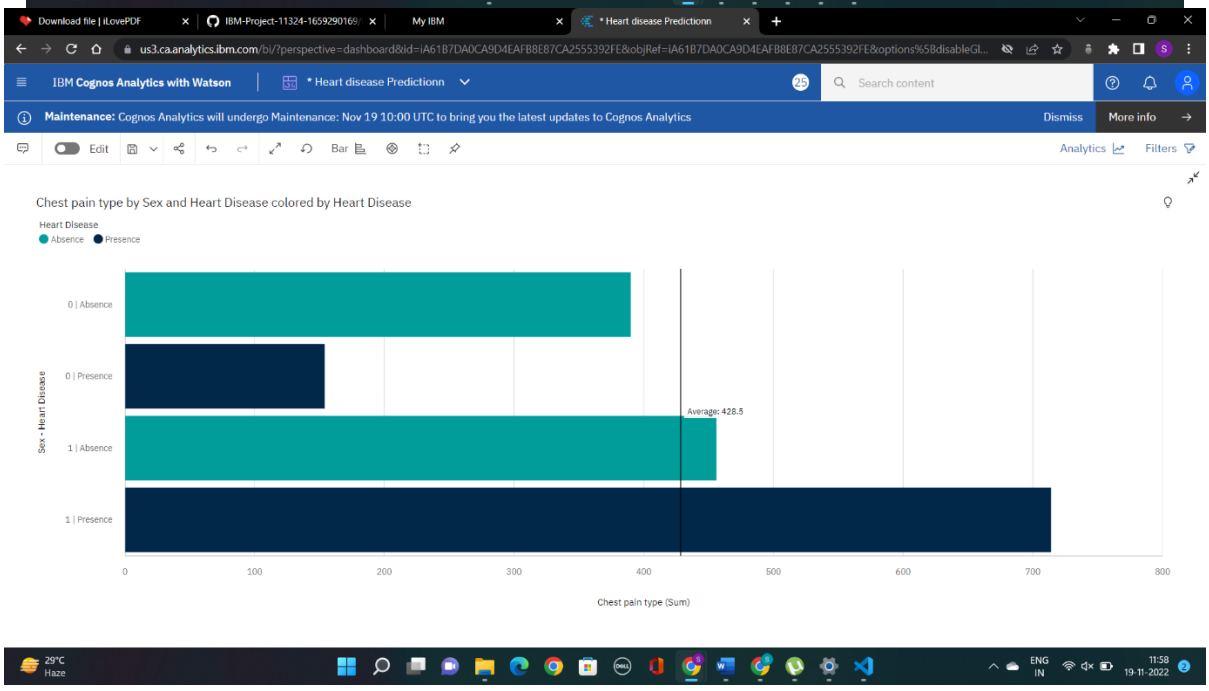
Comparing it with the accuracy gotten from Decision Tree:

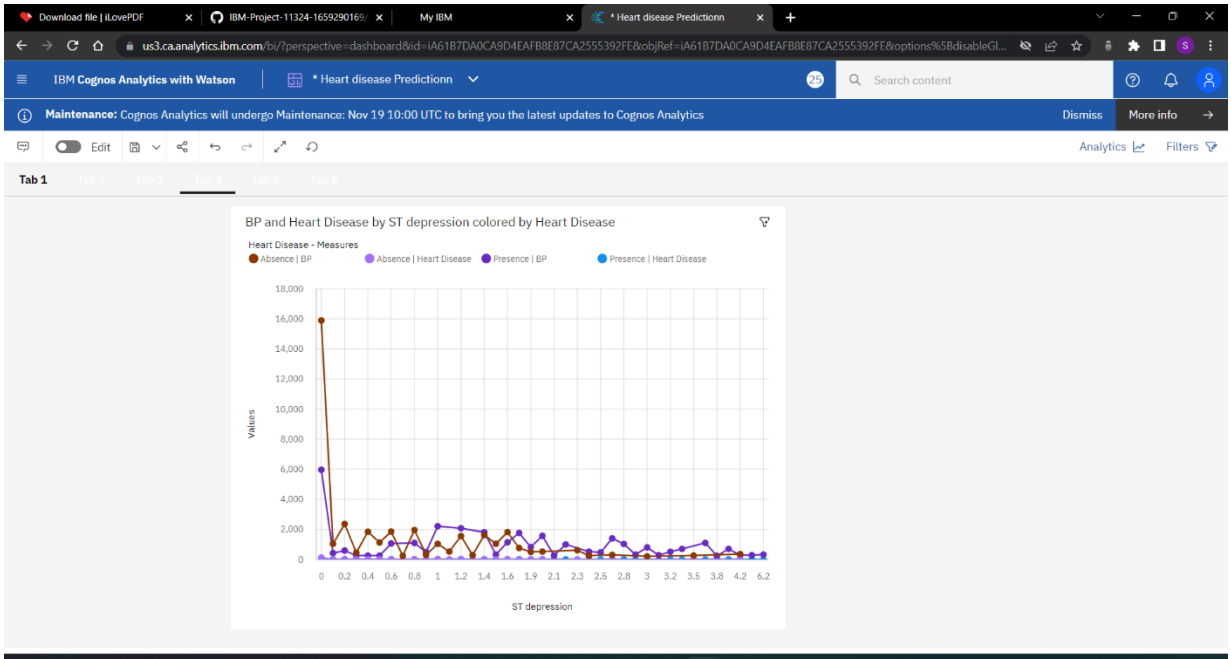
```
TP=cm[0][0] #cm=Confusion Matrix
TN=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN)))
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))
print('Testing Precision for Decision Tree:',(TP/(TP+FP)))
```

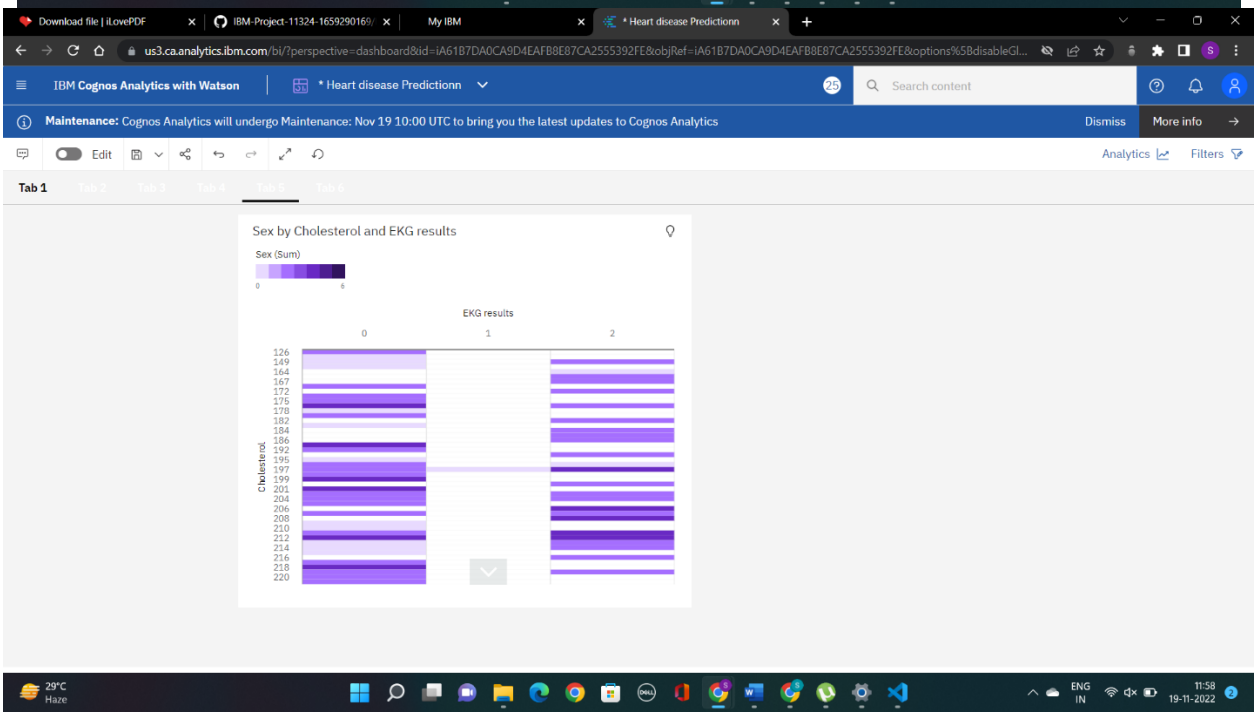
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

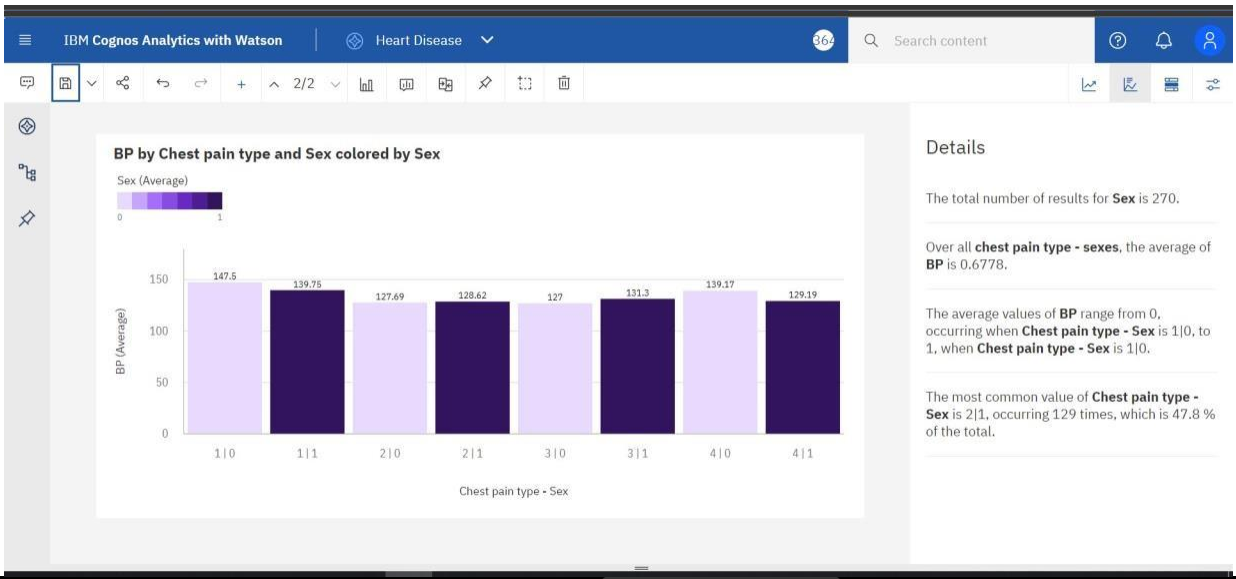
7.2 Dashboard

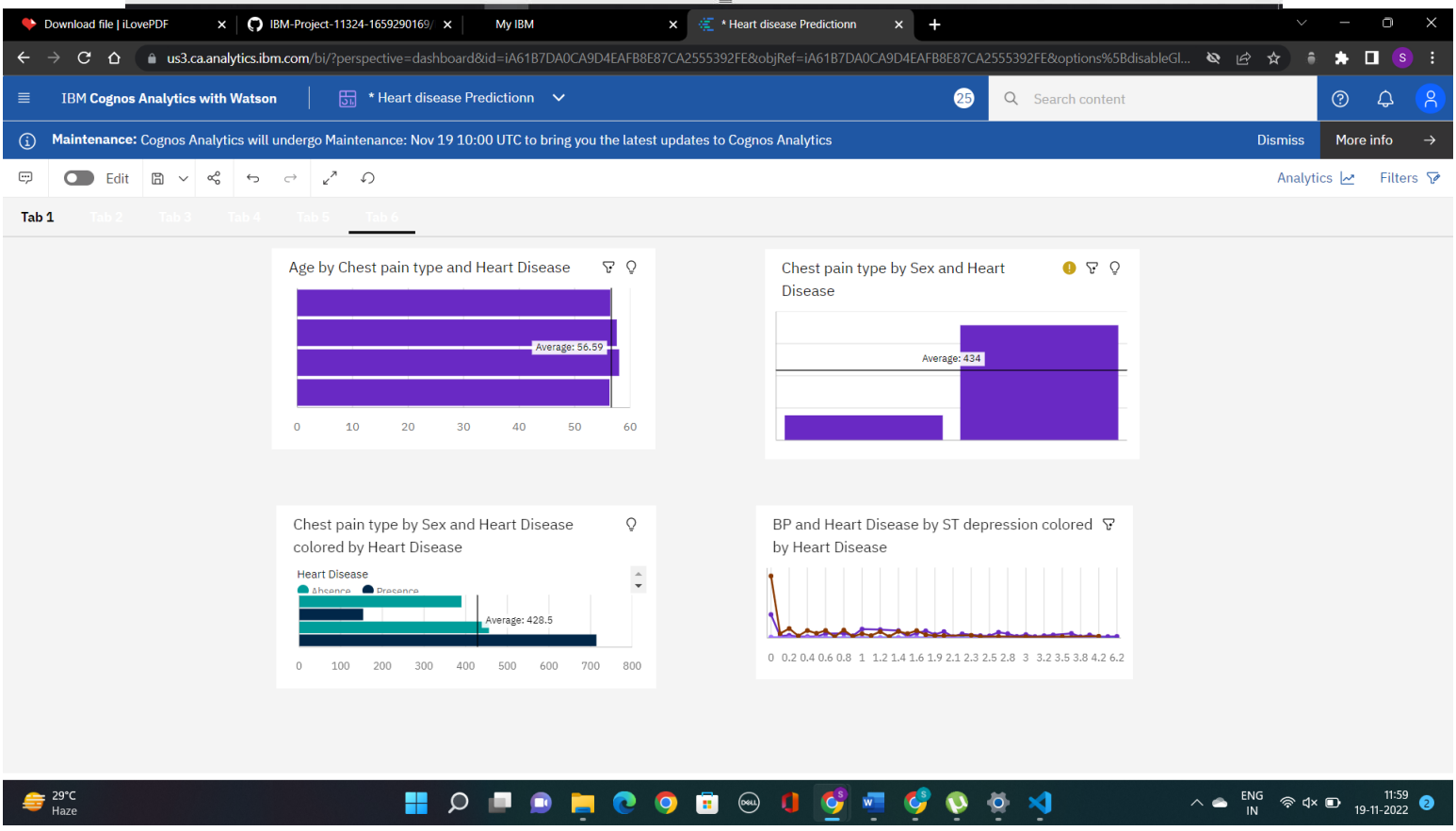





Download file | iLovePDF
IBM-Project-11324-1659290169
My IBM
* Heart disease Predictionn
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iA61B7DA0CA9D4EAFB8E87CA2555392FE&objRef=iA61B7DA0CA9D4EAFB8E87CA2555392FE&options%5BdisableGl...
IBM Cognos Analytics with Watson
* Heart disease Predictionn
Search content
Maintenance: Cognos Analytics will undergo Maintenance: Nov 19 10:00 UTC to bring you the latest updates to Cognos Analytics
Dismiss
More info
Edit
Analytics
Filters
Tab 1
BP and Heart Disease by ST depression colored by Heart Disease
Heart Disease - Measures
Absence | BP
Absence | Heart Disease
Presence | BP
Presence | Heart Disease
Values
ST depression
29°C
Haze
ENG
IN
11:58
19-11-2022


Download file | iLovePDF
IBM-Project-11324-1659290169
My IBM
* Heart disease Predictionn
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iA61B7DA0CA9D4EAFB8E87CA2555392FE&objRef=iA61B7DA0CA9D4EAFB8E87CA2555392FE&options%5BdisableGl...
IBM Cognos Analytics with Watson
* Heart disease Predictionn
Search content
Maintenance: Cognos Analytics will undergo Maintenance: Nov 19 10:00 UTC to bring you the latest updates to Cognos Analytics
Dismiss
More info
Edit
Analytics
Filters
Tab 1
Sex by Cholesterol and EKG results
Sex (Sum)
EKG results
Cholesterol
29°C
Haze
ENG
IN
11:58
19-11-2022


IBM Cognos Analytics with Watson
Heart Disease
Search content
BP by Chest pain type and Sex colored by Sex
Sex (Average)
BP (Average)
147.5
139.75
127.69
128.62
127
131.3
139.17
129.19
1|0
1|1
2|0
2|1
3|0
3|1
4|0
4|1
Chest pain type - Sex
Details
The total number of results for Sex is 270.
Over all chest pain type - sexes, the average of BP is 0.6778.
The average values of BP range from 0, occurring when Chest pain type - Sex is 1|0, to 1, when Chest pain type - Sex is 1|0.
The most common value of Chest pain type - Sex is 2|1, occurring 129 times, which is 47.8 % of the total.


Download file | iLovePDF
IBM-Project-11324-1659290169/
My IBM
* Heart disease Predictionn
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iA61B7DA0CA9D4EAFB8E87CA2555392FE&objRef=iA61B7DA0CA9D4EAFB8E87CA2555392FE&options%5BdisableGl...
IBM Cognos Analytics with Watson
* Heart disease Predictionn
Search content
Maintenance: Cognos Analytics will undergo Maintenance: Nov 19 10:00 UTC to bring you the latest updates to Cognos Analytics
Dismiss
More info
Edit
Analytics
Filters
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Age by Chest pain type and Heart Disease
Average: 56.59
0
10
20
30
40
50
60
Chest pain type by Sex and Heart Disease
Average: 434
Chest pain type by Sex and Heart Disease colored by Heart Disease
Heart Disease
Absence
Presence
Average: 428.5
0
100
200
300
400
500
600
700
800
BP and Heart Disease by ST depression colored by Heart Disease
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.9 2.1 2.3 2.5 2.8 3 3.2 3.5 3.8 4.2 6.2
29°C
Haze
ENG
IN
11:59
19-11-2022

Dashboard Showing Different Types Of Visuals:

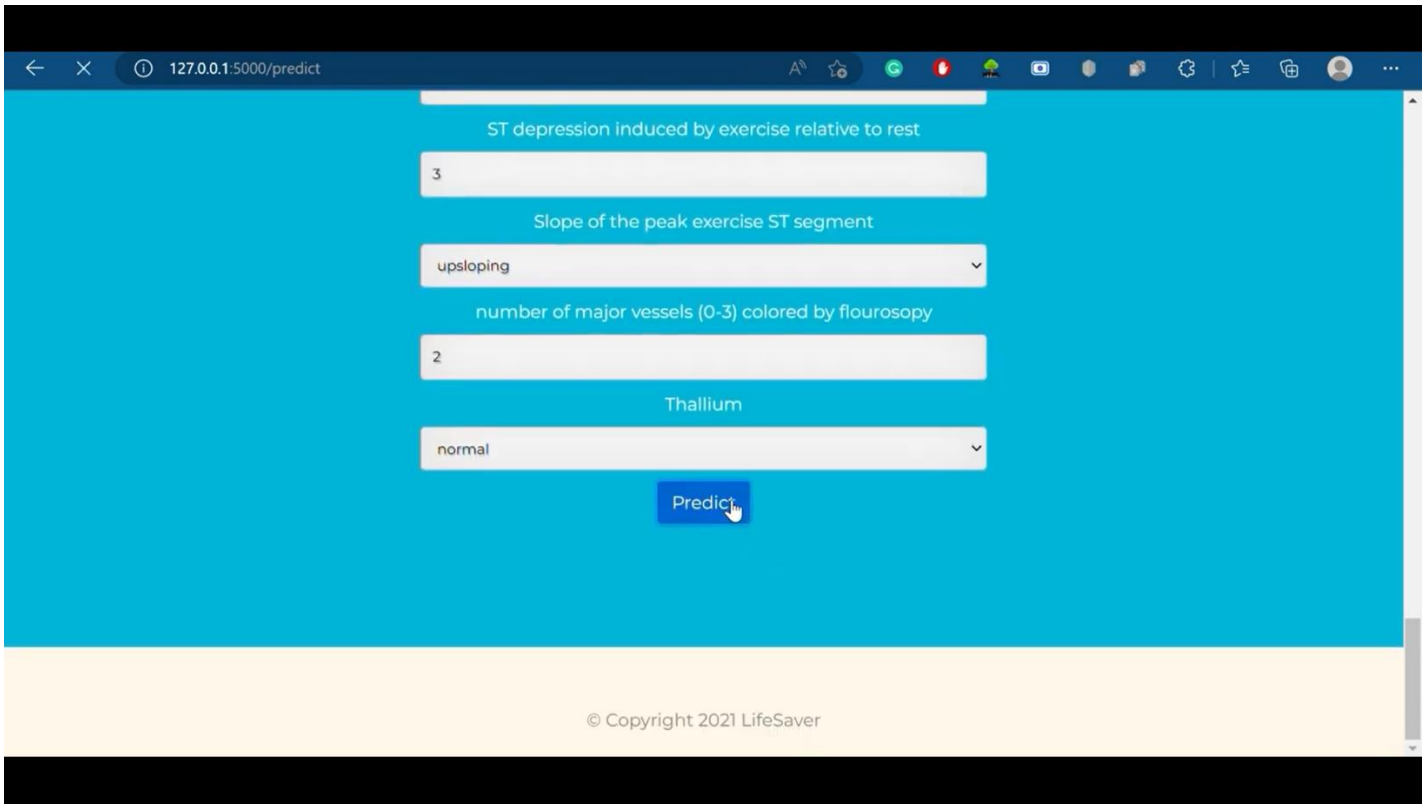
8. Testing

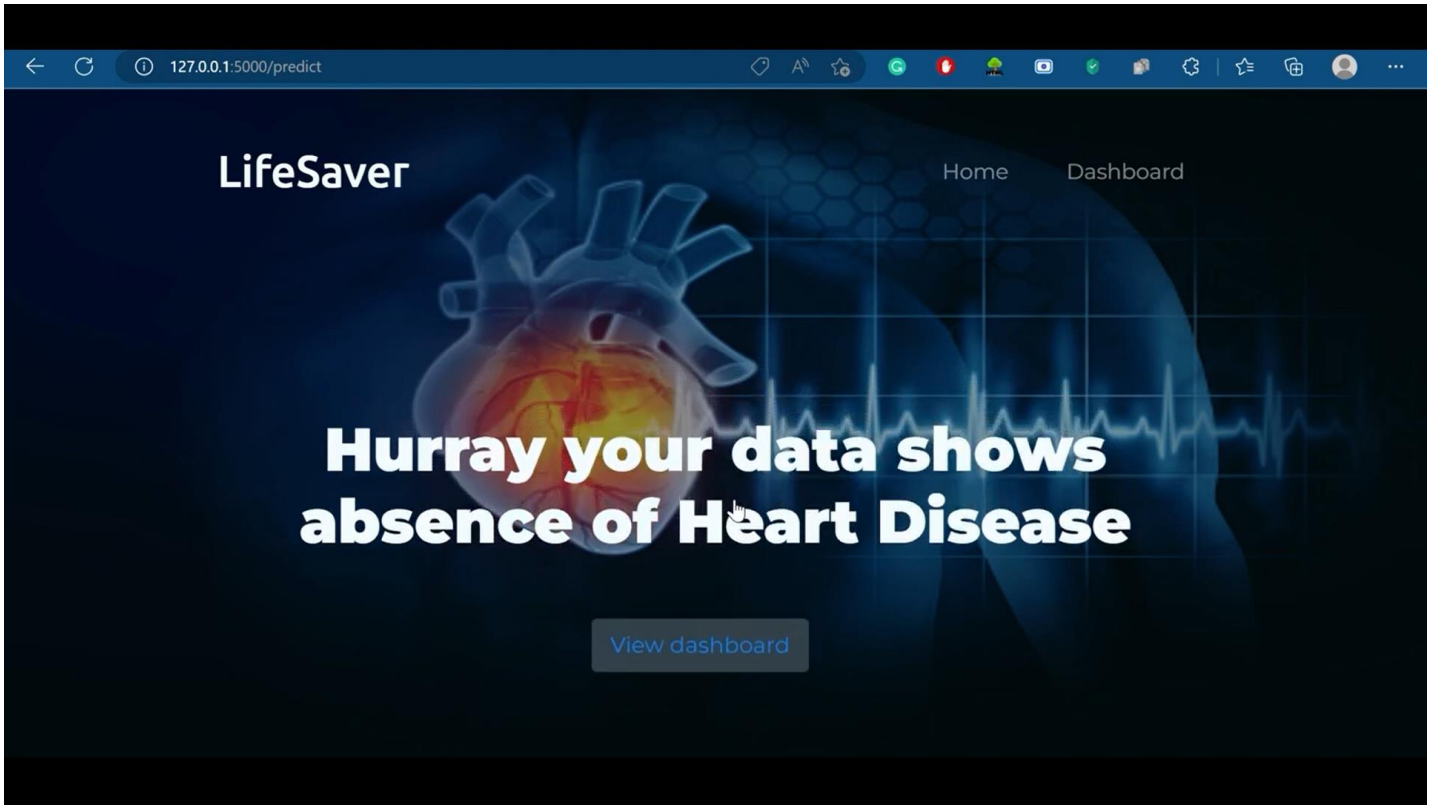
8.1 Test Cases

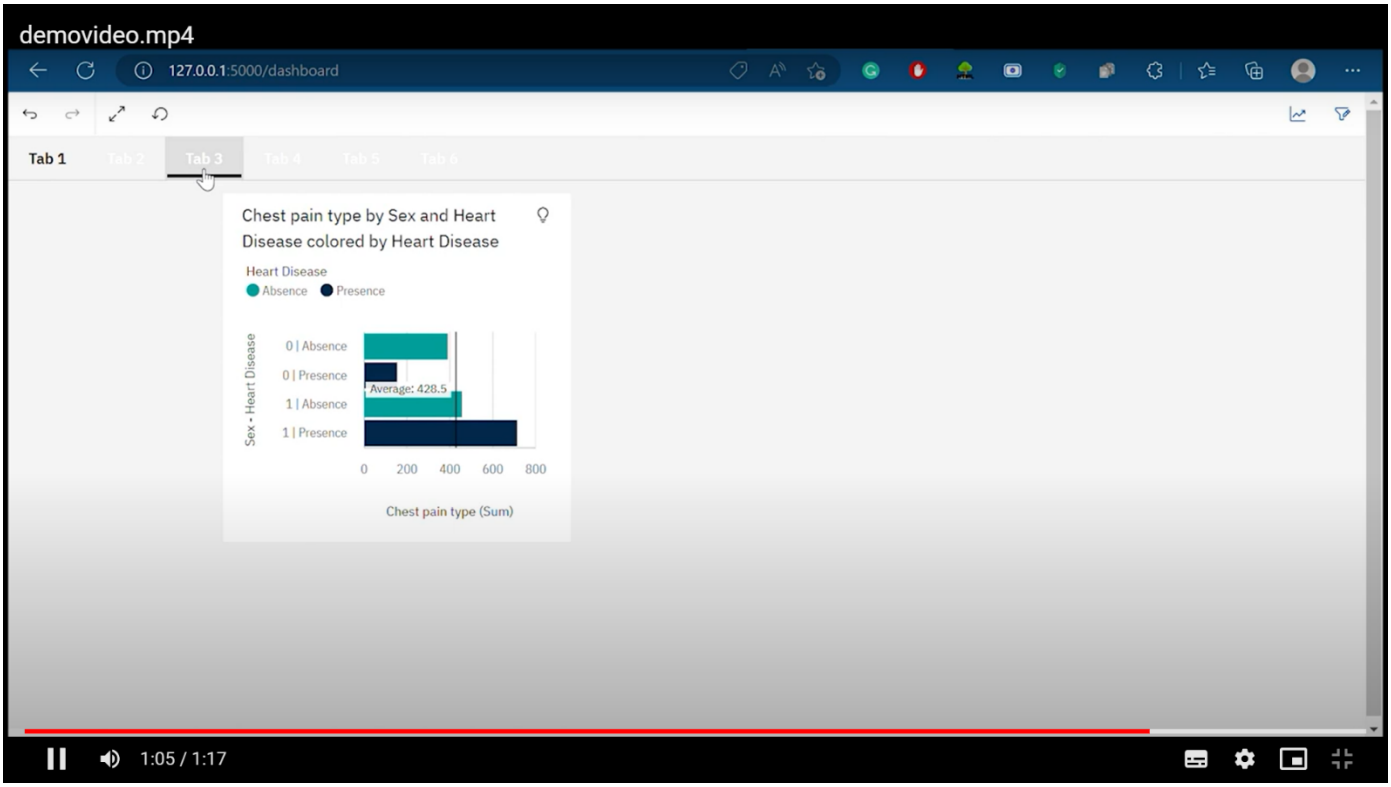
Testing the data model for various input values.

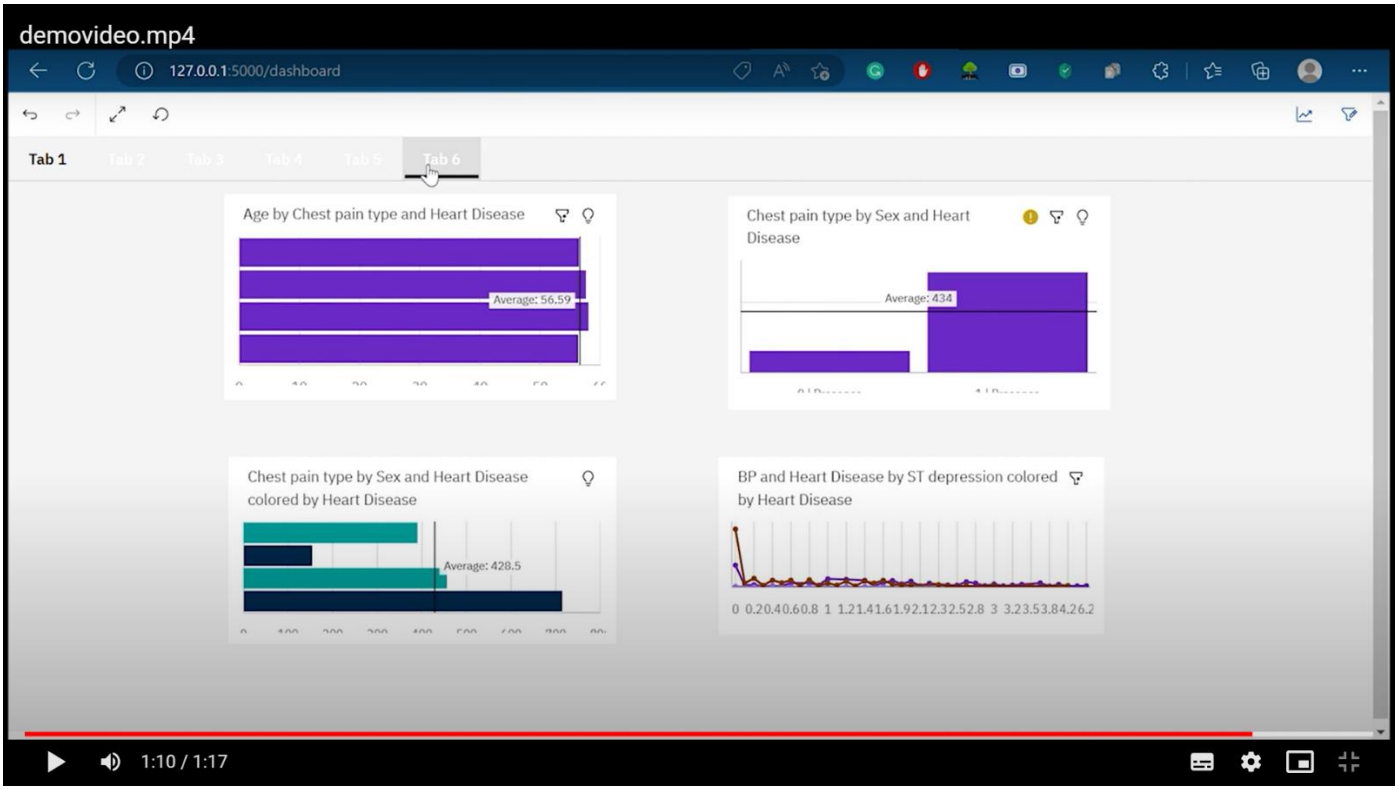
```
In [ ]: from sklearn.metrics import accuracy_score
        input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped))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





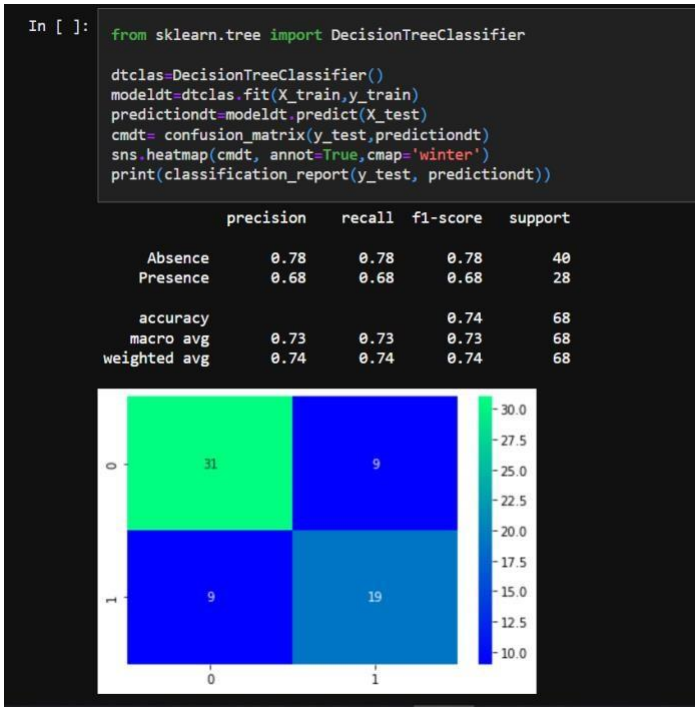

demovideo.mp4
127.0.0.1:5000/dashboard
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Chest pain type by Sex and Heart Disease colored by Heart Disease
Heart Disease
Absence
Presence
Sex • Heart Disease
0 | Absence
0 | Presence
1 | Absence
1 | Presence
Average: 428.5
0
200
400
600
800
Chest pain type (Sum)
1:05 / 1:17


demovideo.mp4
127.0.0.1:5000/dashboard
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Tab 6
Age by Chest pain type and Heart Disease
Average: 56.59
Chest pain type by Sex and Heart Disease
Average: 434
Chest pain type by Sex and Heart Disease colored by Heart Disease
Average: 428.5
BP and Heart Disease by ST depression colored by Heart Disease
0 0.20.40.60.8 1 1.21.41.61.92.12.32.52.8 3 3.23.53.84.26.2
1:10 / 1:17

9. Result

9.1 Performance Metrics

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



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-11324-1659290169

DemoLink:

https://drive.google.com/file/d/1NZEvG4zbPlpLP-DpdZyPbeip0h9zZOc9/view