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

Heart diseases have emerged as the number one killer in the world according to the largest ever study of deaths and is considered the deadliest. About 25 per cent of deaths in the age group of 25- 69 years occur because of heart diseases. It is difficult to control most of these heart diseases as they are diagnosed at later stages at which aggravates the problem. Quick diagnosis, early treatment and constant observations is essential for patients who suffer from heart diseases. Moreover, cardiologists often diagnose cardiovascular disease based on current clinical tests and previous experience in diagnosing patients with similar symptoms. An enormous amount of data, so-called big data generated by the healthcare industries accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition based on very few data of the person available with the hospital . A set of important feature scores and rules were identified in diagnosing heart disease and cardiologists

were consulted to confirm the validity of these rules. An unsupervised learning algorithm like k mean clustering would enable us to carefully analysis the stats based on different parameters and predicts the likelihood of the occurrence of the Heart disease

2.2 References

1) Visualization and Prediction of Heart Diseases Using Data Science Framework | IEEE Conference Publication | IEEE Xplore

2) (4) (PDF) Early Prediction of Heart Diseases Using Data Mining Techniques (researchgate.net)

3) A novel approach for heart disease prediction using strength scores with significant predictors | BMC Medical Informatics and Decision Making | Full Text (biomedcentral.com)

4) (4) (PDF) Heart Disease Prediction using Exploratory Data Analysis (researchgate.net)

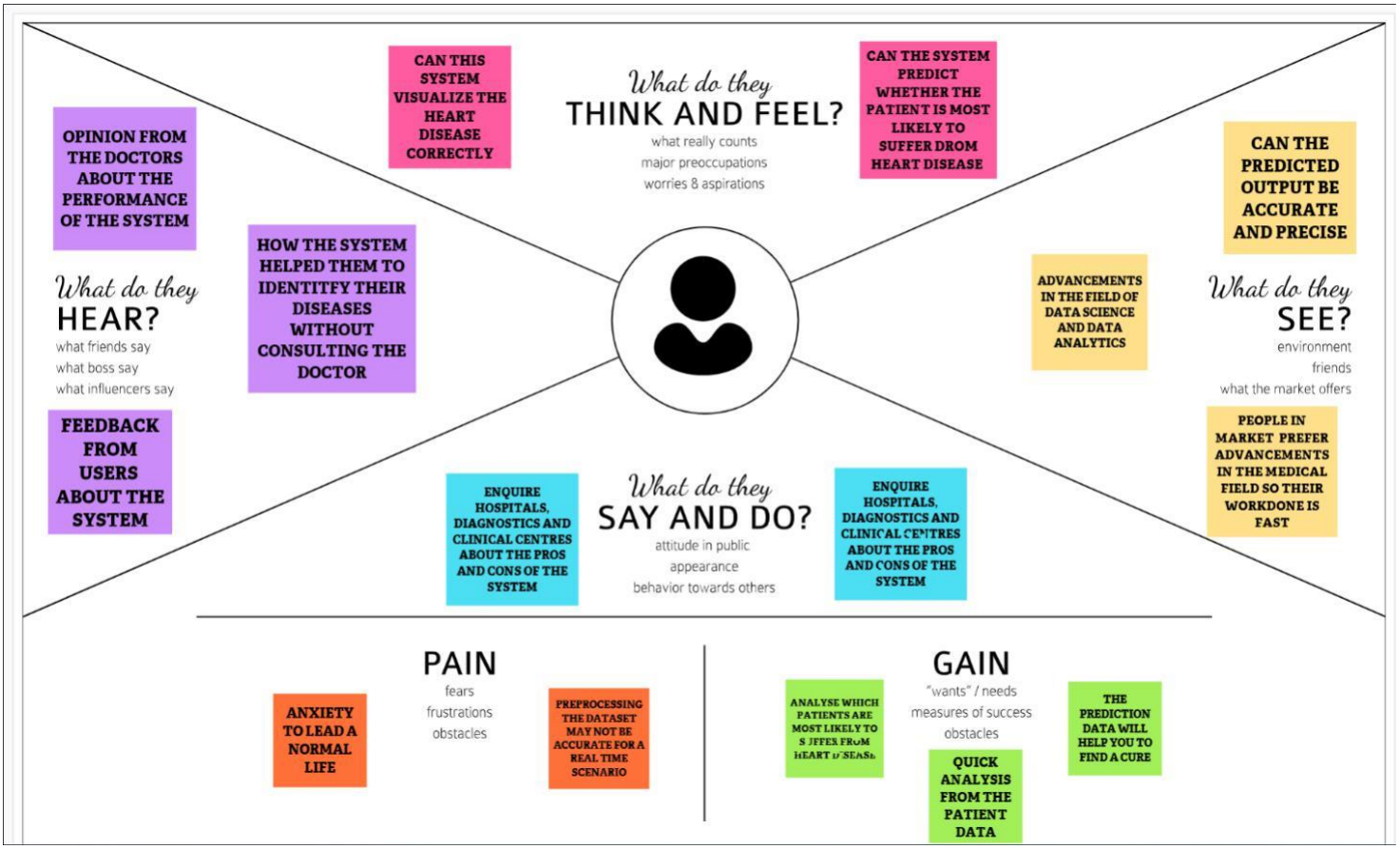
2.3 Problem Statement Definition

It is found that in India, Cardiovascular Diseases contributed to 28·1% of total deaths and 14·1% of total disability-adjusted life years. Moreover, most of those with coronary heart disease who pass away are 65 years of age or older. The 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. Although both sexes can get heart attacks in old age, women have a higher mortality rate (within a few weeks). 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. Cardiovascular Diseases happen 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 it becomes important to predict the presence/absence or the probability of getting heart disease. The user or the common people can enter the input values from their's health report and check the presence/absence of heart disease. The data is fed into the project model which predicts the probability of having heart disease. Thus the user can take appropriate steps according to the result.

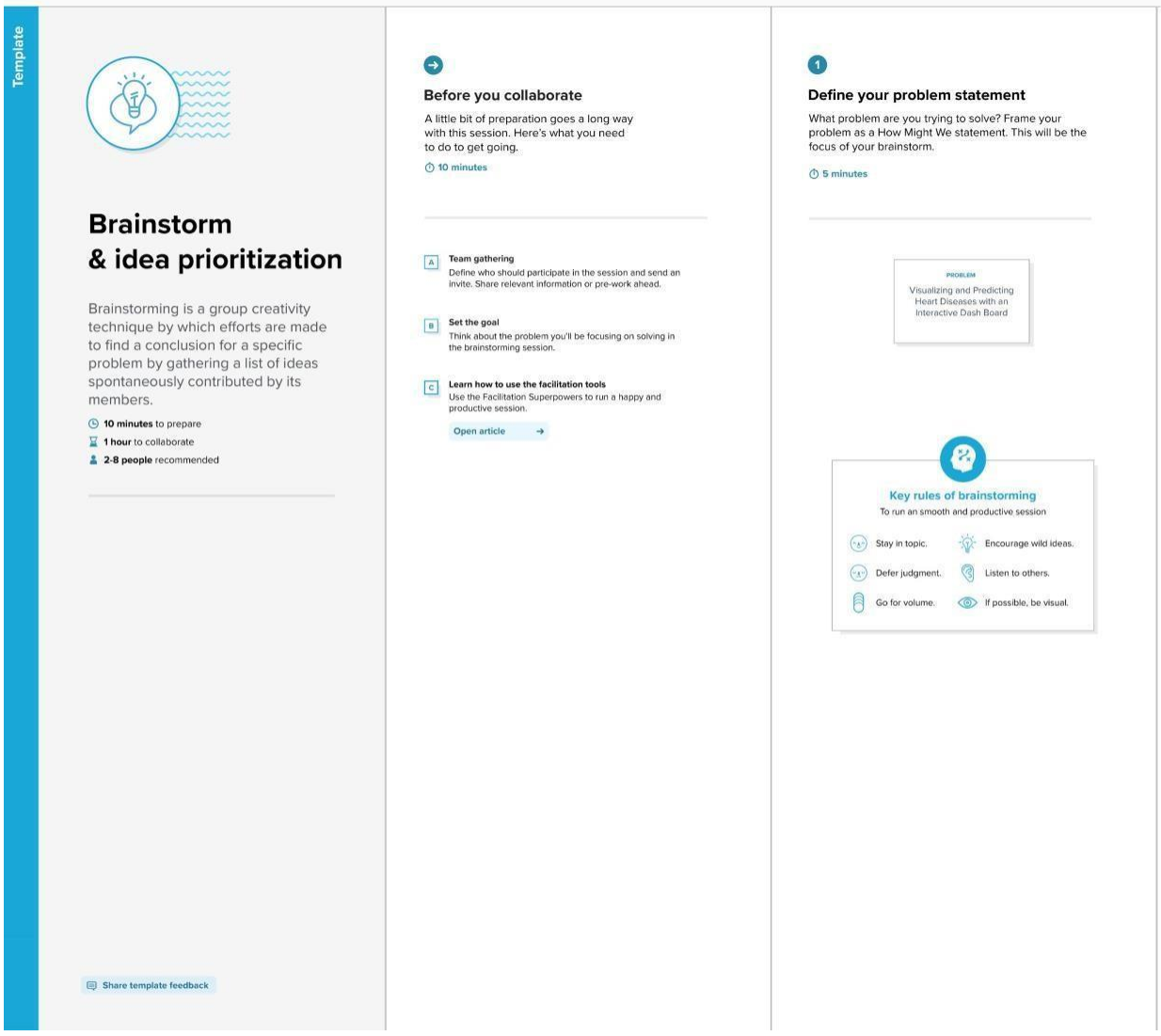
3. Ideation and Proposed Solution

3.1 Empathy Map Canvas

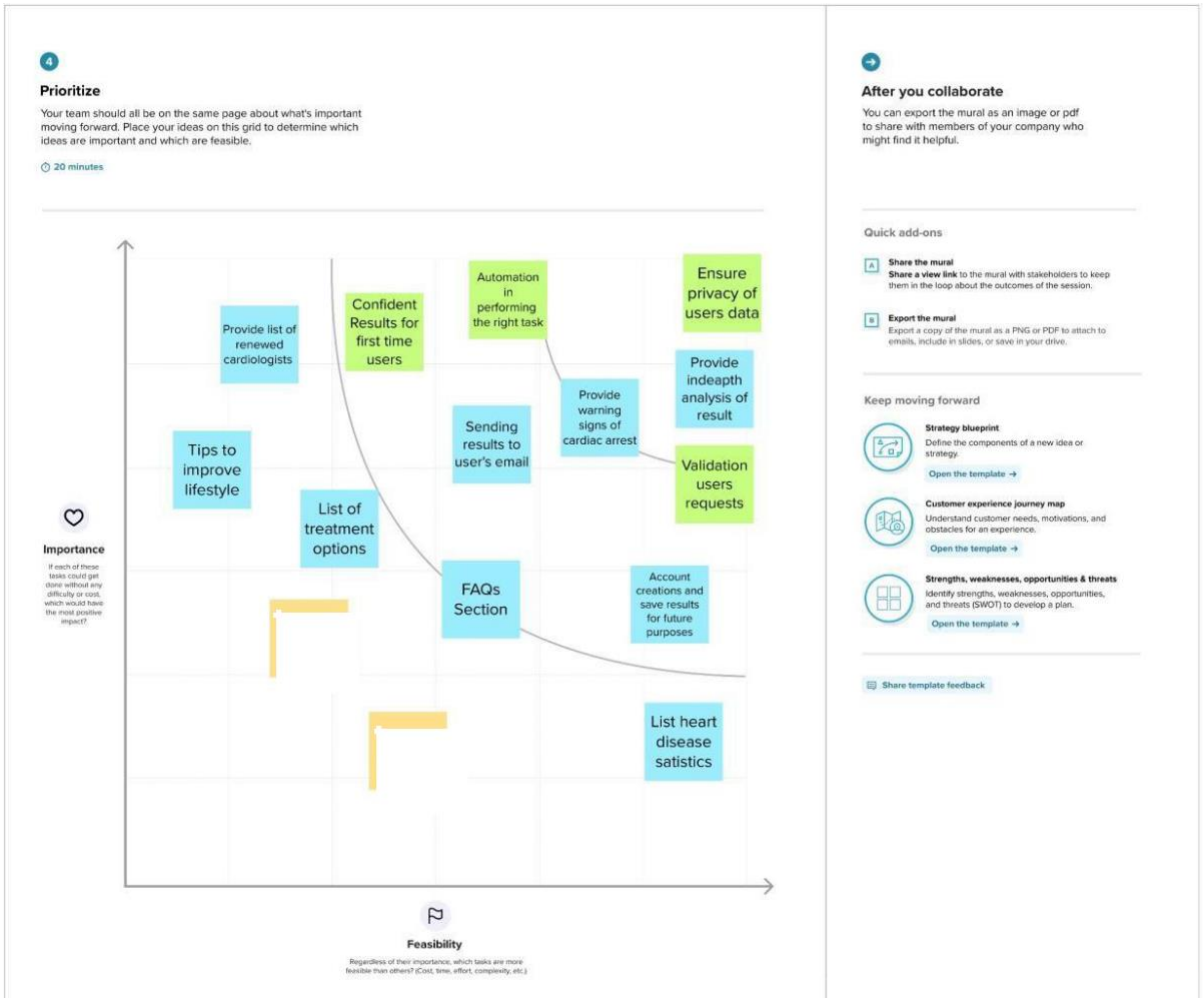


3.2 Ideation and Brainstorming

Team Gathering, Collaboration and Select the Problem Statement



Idea Prioritization



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heart disease refers to several types of abnormalities in heart conditions. The leading cause of death is heart disease. It is infeasible for a common man to frequently undergo tests for ECG and so on. Hence, there needs a replacement for this, which must be handy and reliable.
2.	Supervised Learning	The idea behind the proposed solution is to propose an interactive dashboard for visualizing and predicting heart diseases in which user can view his/her medical report analysis and the predicted final result. The dashboard will be generated using IBM Cognos. The heart disease will be predicted using Naïve Bayes Algorithm.
3.	Novelty / Uniqueness	The novelty behind the proposed system is to provide suggestions to the user based on his/her medical analysis. It will provide the preventive measures to take care of the user himself.
4.	Social Impact / Customer Satisfaction	The system helps the user as well as the doctor to make better decisions to predict heart disease. It is useful in predicting the disease in an earlier stage and makes the user alert about his current condition periodically.
5.	Business Model (Revenue Model)	This interactive dashboard for heart disease prediction can be deployed in Health care centres and Hospitals, so that it makes the analysis in a fast manner.

3.4 Problem Solution Fit

CUSTOMER END	PROBLEM CAUSES	CUSTOMER CONSTRAINTS
Who's the Customer? • Smokers • People with high blood pressure • People with high Cholestrol • Diabetic Patients • People who have Chest Pains ,Chest discomfort • People with overweight • People who lack physical excersise	What is the real reason for the problem? What is the need for this job? • Lack of excercise,obesity and smoking • Fatty plagues in the arteries • Problem in the heart affects the whole body • Thus this Visualisation is made and an interactive dasboard is made for different heart diseases.	What constraints prevent customers from taking action or limit choices for solutions? • Lack of knowledge about heart diseases. • Lack of hope in treatment • Complex symptoms • Economical background • Psychological problems • Negative thoughts of the customer
Solutions Available	**Jobs to be Done/Problems**	**BEHAVIOUR**
Solutions available for people with heart diesease are: • Quit smoking • Get Cholestrol tested regularly • Diabetic Patients • Eating healthy foods • Regular physical excersise • Along with these they have to go for regular medical checkup and tests • If heart diesease is found proper medication should be taken	Which jobs to be done is addressed to the customers? • Lives depending on medical support • Financial Insecurity • Shortness of breath • Chest pain,Chest pressure	What constraints prevent customers from taking action or limit choices for solutions? • Lack of knowledge about heart diseases. • Lack of hope in treatment • Complex symptoms • Economical background • Psychological problems • Negative thoughts of the customer miro

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	The users can register/Sign up through Gmail or through a form which gets user details(email, username, password etc ..)
FR-2	User Confirmation	The registration is confirmed via mail. From the user end it is confirmed with OTP recieved through mail.
FR-3	Loading User Data	The user will enter/give the details required to predict heart disease via the platform.
FR-4	Visualizing the Data	The user will be able to visualize the heart disease trend through Dashboard created using IBM Cognos Analytics.
FR-5	Generating Report	The user can access their report generated based on the details given.
FR-6	Recommandation	Set of recommandation to recover can be given, which should be taken only with the consent of a cardiologist.

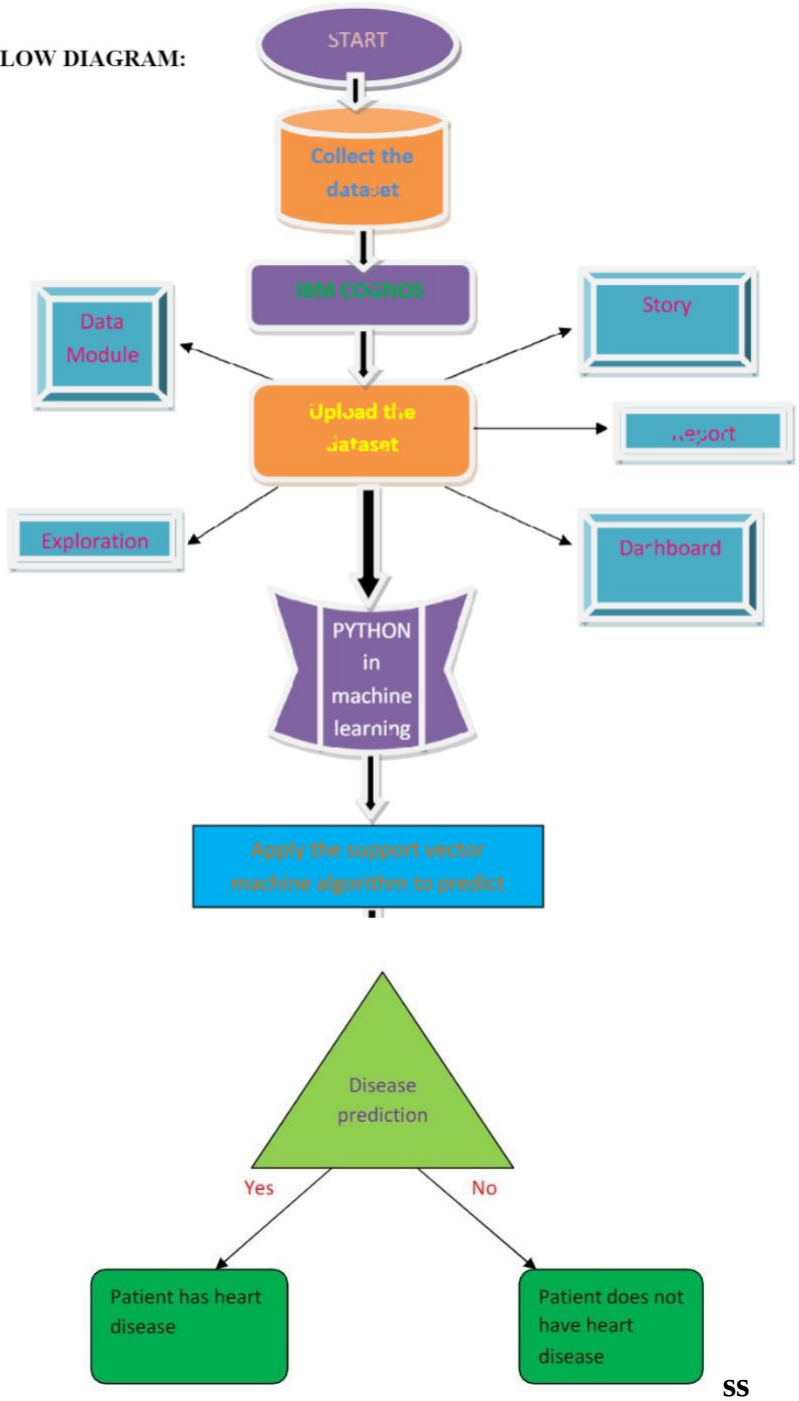
4.2 Non-Functional Requirement

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

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	The users can register/Sign up through Gmail or through a form which gets user details(email, username, password etc ..)
FR-2	User Confirmation	The registration is confirmed via mail. From the user end it is confirmed with OTP recieved through mail.
FR-3	Loading User Data	The user will enter/give the details required to predict heart disease via the platform.
FR-4	Visualizing the Data	The user will be able to visualize the heart disease trend through Dashboard created using IBM Cognos Analytics.
FR-5	Generating Report	The user can access their report generated based on the details given.
FR-6	Recommandation	Set of recommandation to recover can be given, which should be taken only with the consent of a cardiologist.

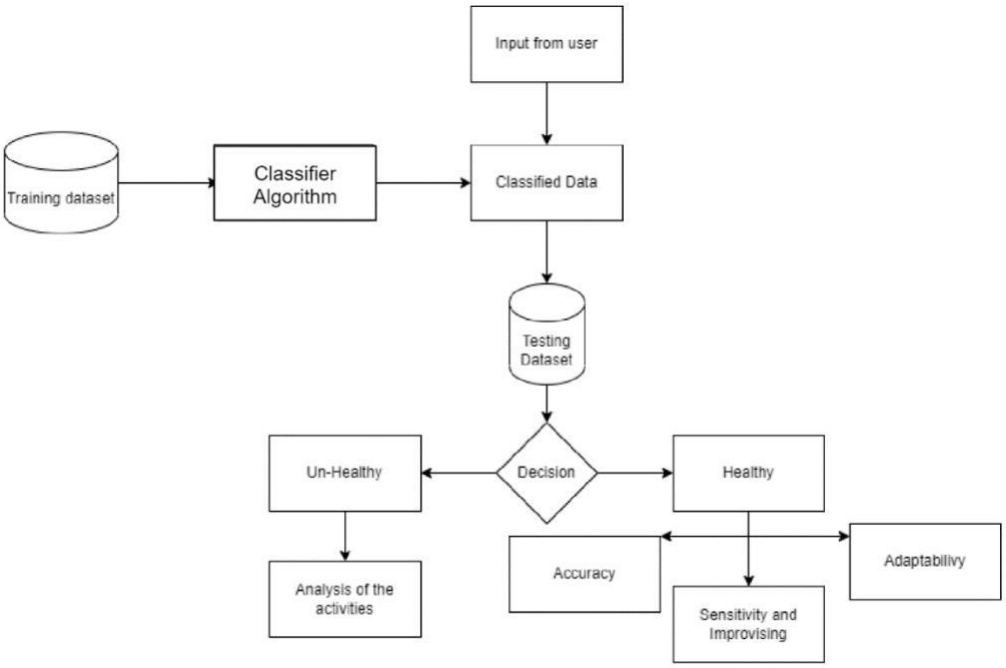
5. Project Design

DATA FLOW DIAGRAM:



SS

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

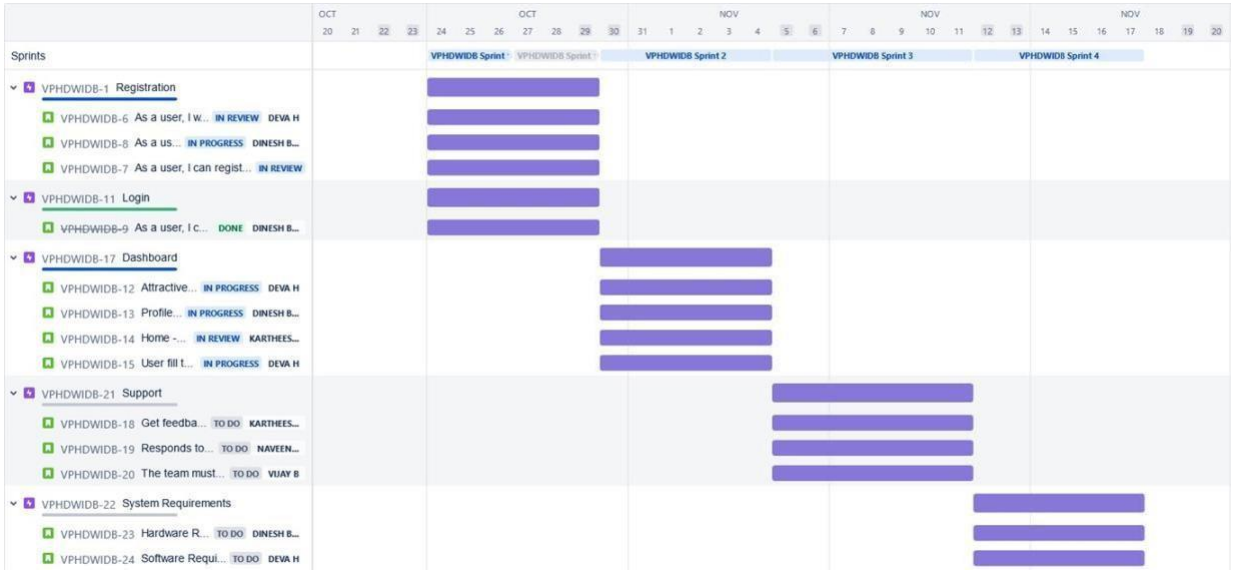
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	3	High	1
Sprint-1		USN-2	As a user, I will receive confirmation 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

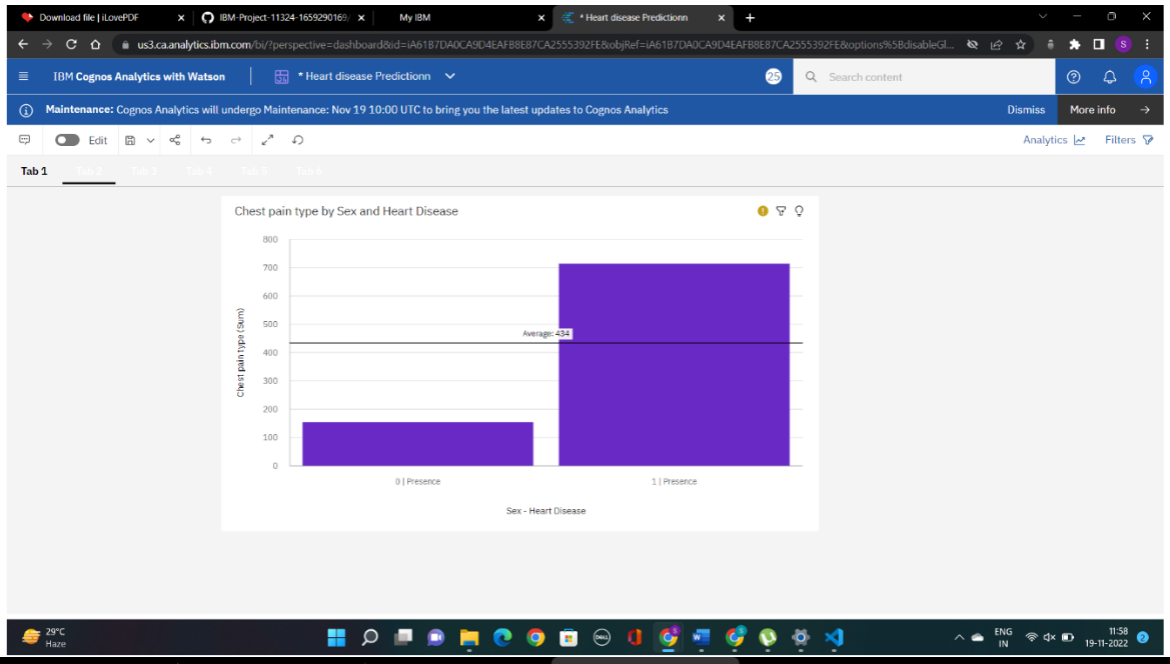
From the above result we can conclude that Logistic Regression has the hisgest accuracy for this particular dataset.

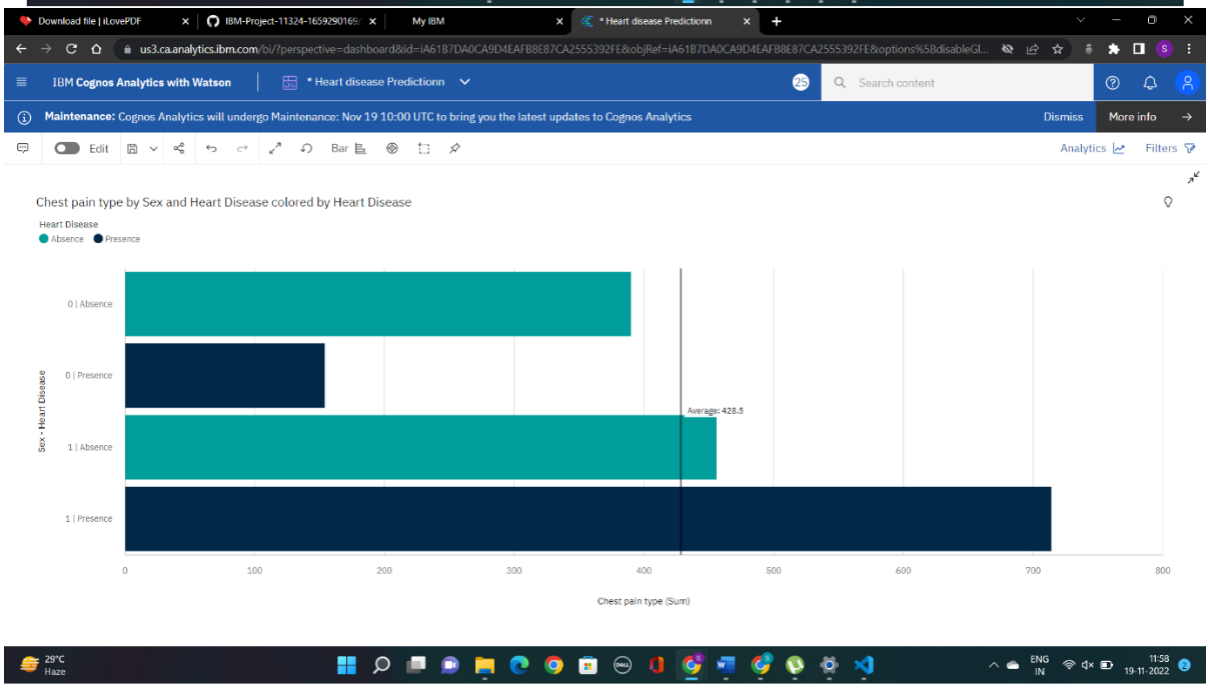
Comparing it with the accuracy gotten from Decision Tree:

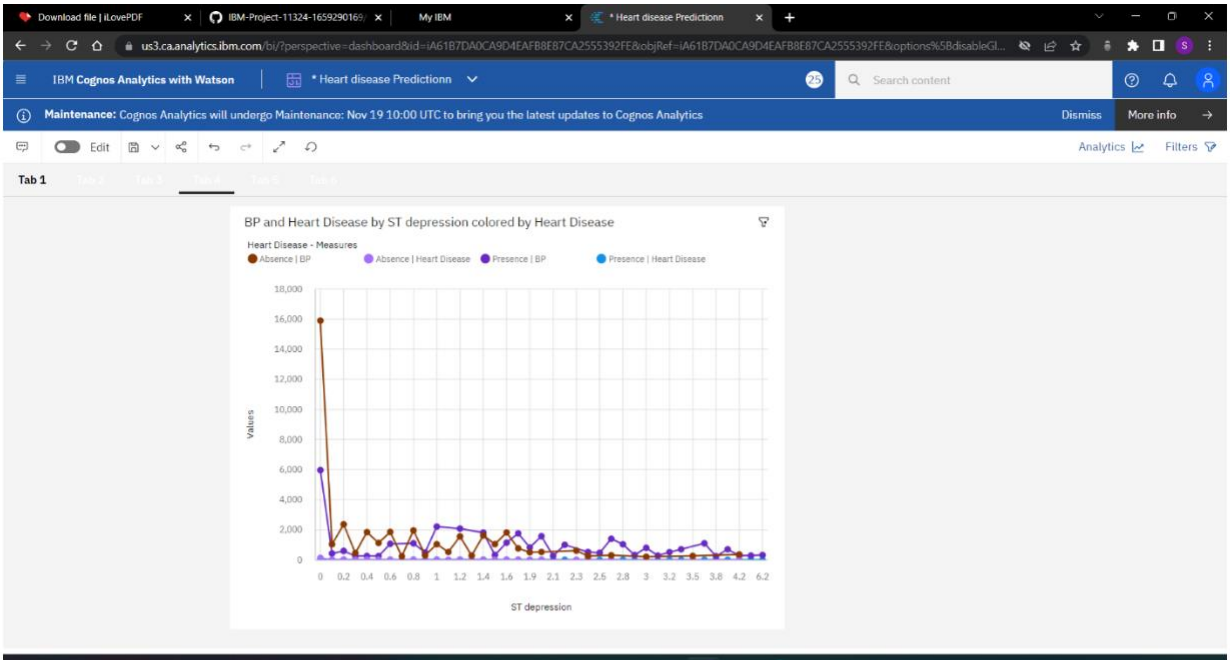
```
TP=cm[0][0] #cm=Confusion Matrix
TN=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN)))
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))
print('Testing Precision for Decision Tree:',(TP/(TP+FP)))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

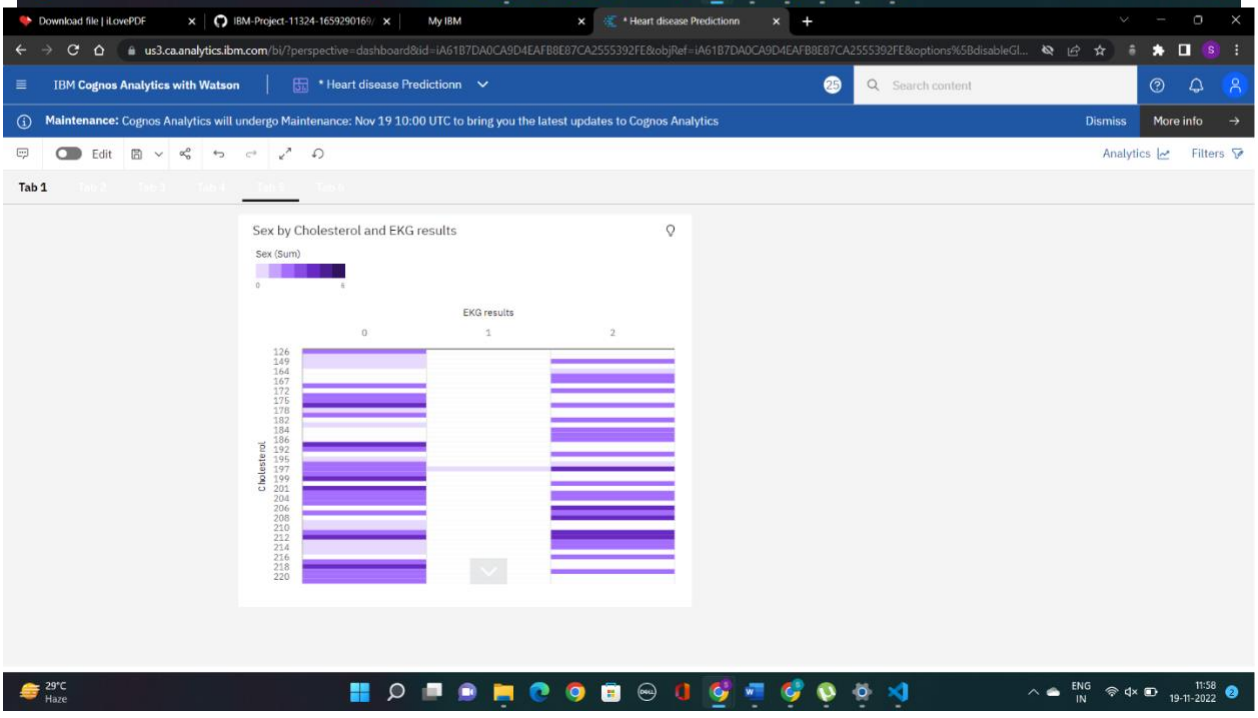
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

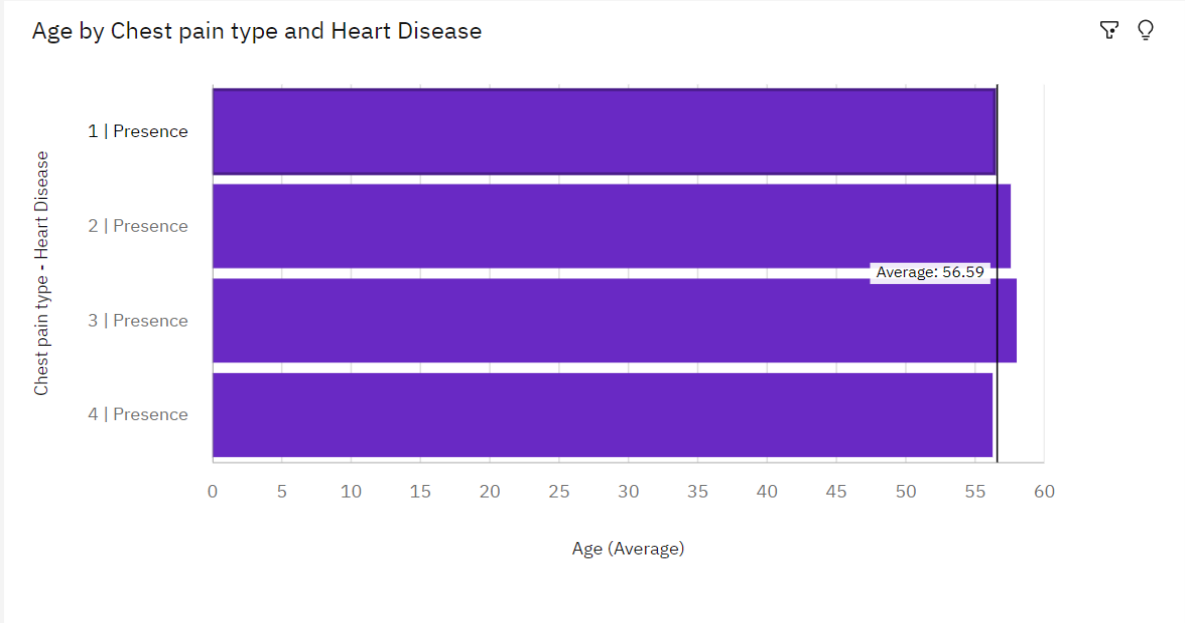
7.2 Dashboard

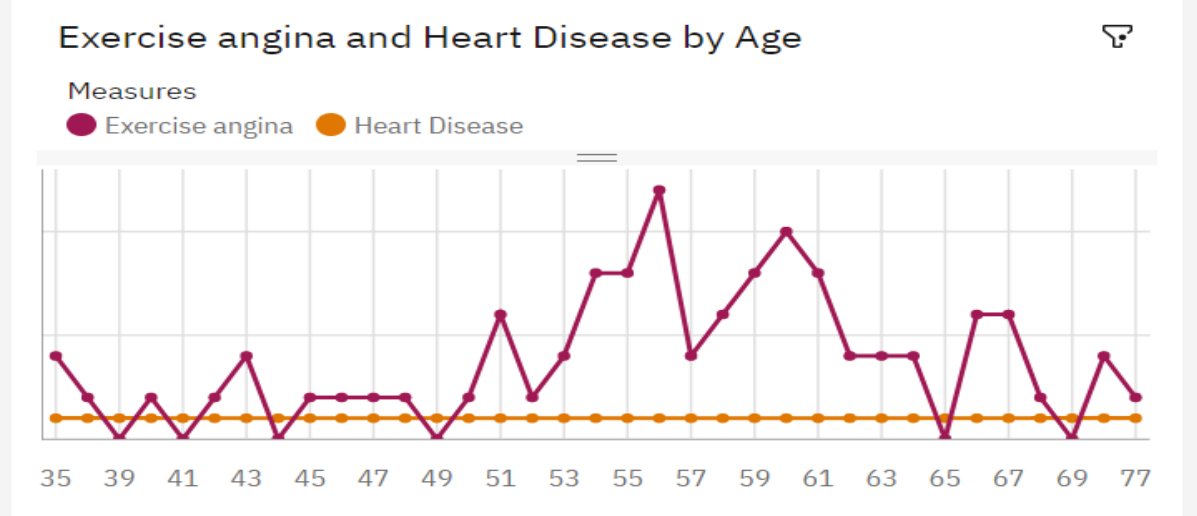


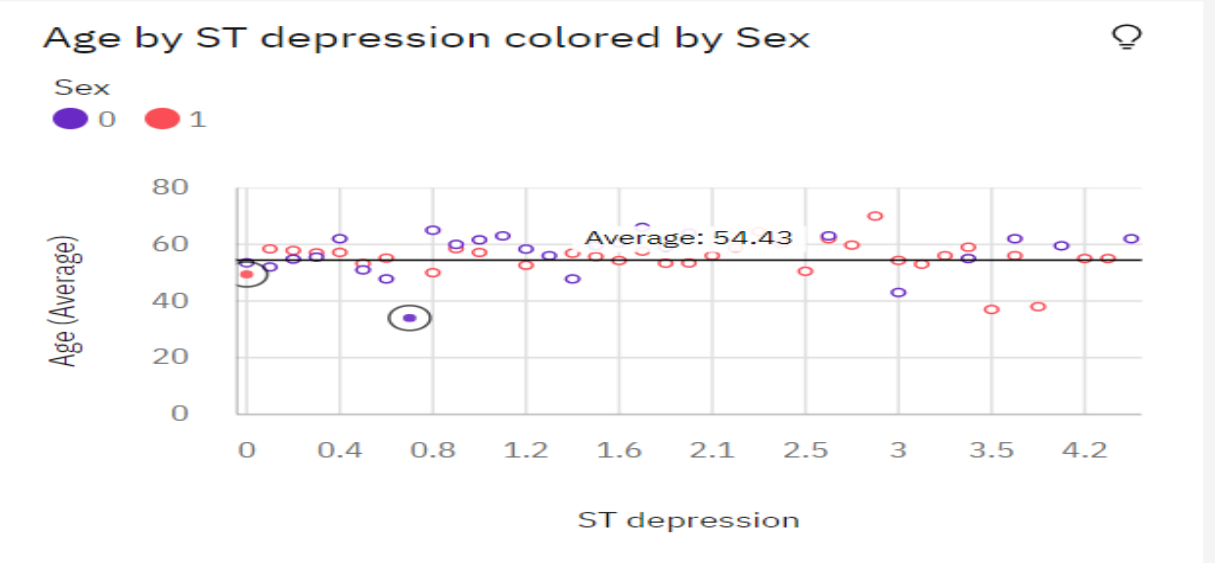



Download file | iLovePDF
IBM-Project-11324-1659290169
My IBM
* Heart disease Predictionn
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iA61B7DA0CA9D4EAFB8E87CA2555392FE&objRef=iA61B7DA0CA9D4EAFB8E87CA2555392FE&options%5BdisableGl...
IBM Cognos Analytics with Watson
* Heart disease Predictionn
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

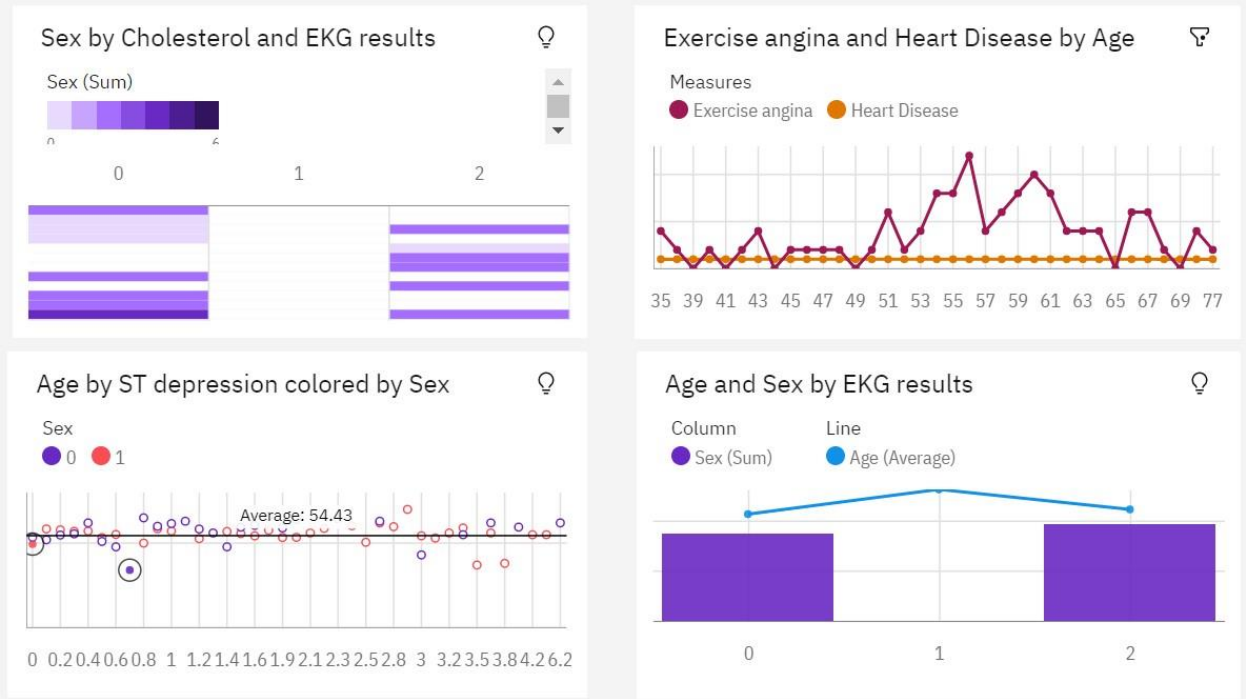

Download file | iLovePDF
IBM-Project-11324-1659290169
My IBM
* Heart disease Predictionn
us3.ca.analytics.ibm.com/bi/?perspective=dashboard&id=iA61B7DA0CA9D4EAFB8E87CA2555392FE&objRef=iA61B7DA0CA9D4EAFB8E87CA2555392FE&options%5BdisableGl...
IBM Cognos Analytics with Watson
* Heart disease Predictionn
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

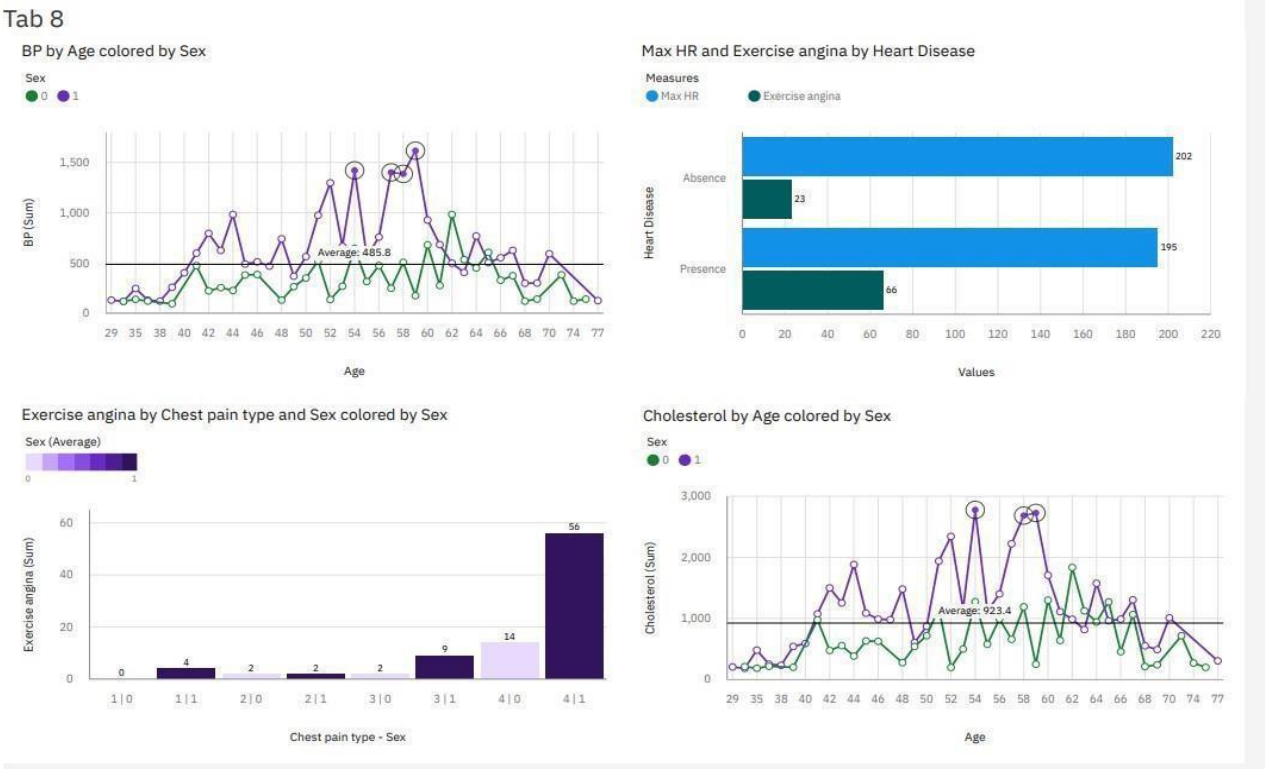

Age by Chest pain type and Heart Disease
Chest pain type - Heart Disease
1 | Presence
2 | Presence
3 | Presence
4 | Presence
Average: 56.59
0
5
10
15
20
25
30
35
40
45
50
55
60
Age (Average)


Exercise angina and Heart Disease by Age
Measures
Exercise angina
Heart Disease
35
39
41
43
45
47
49
51
53
55
57
59
61
63
65
67
69
77

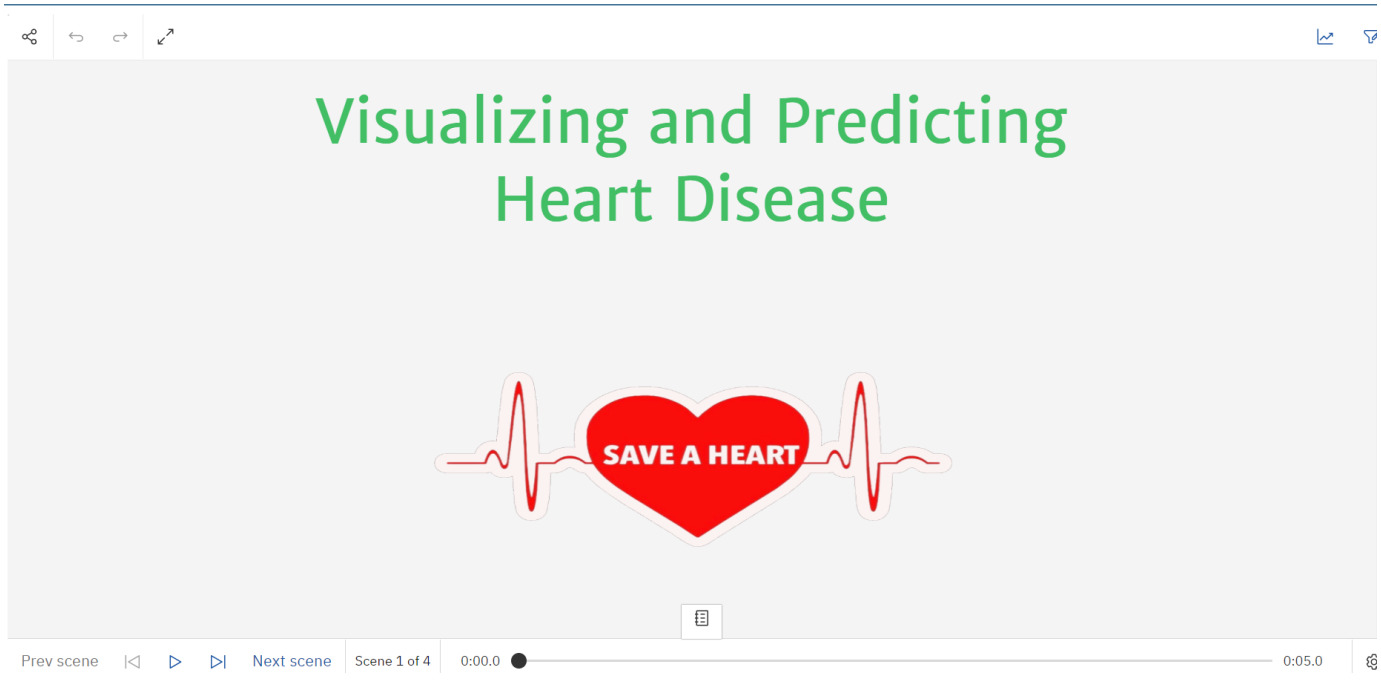

Age by ST depression colored by Sex
Sex
0
1
Age (Average)
80
60
40
20
0
Average: 54.43
0
0.4
0.8
1.2
1.6
2.1
2.5
3
3.5
4.2
ST depression

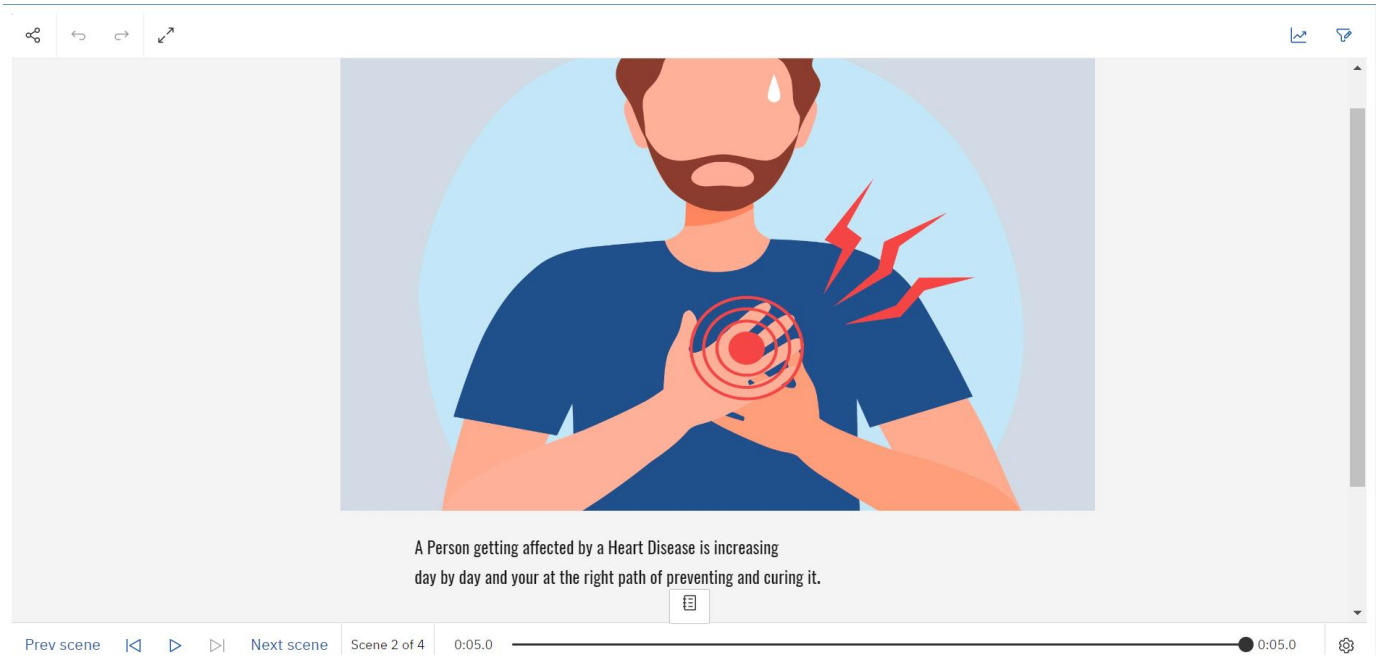
Dashboard Showing Different Types Of Visuals:

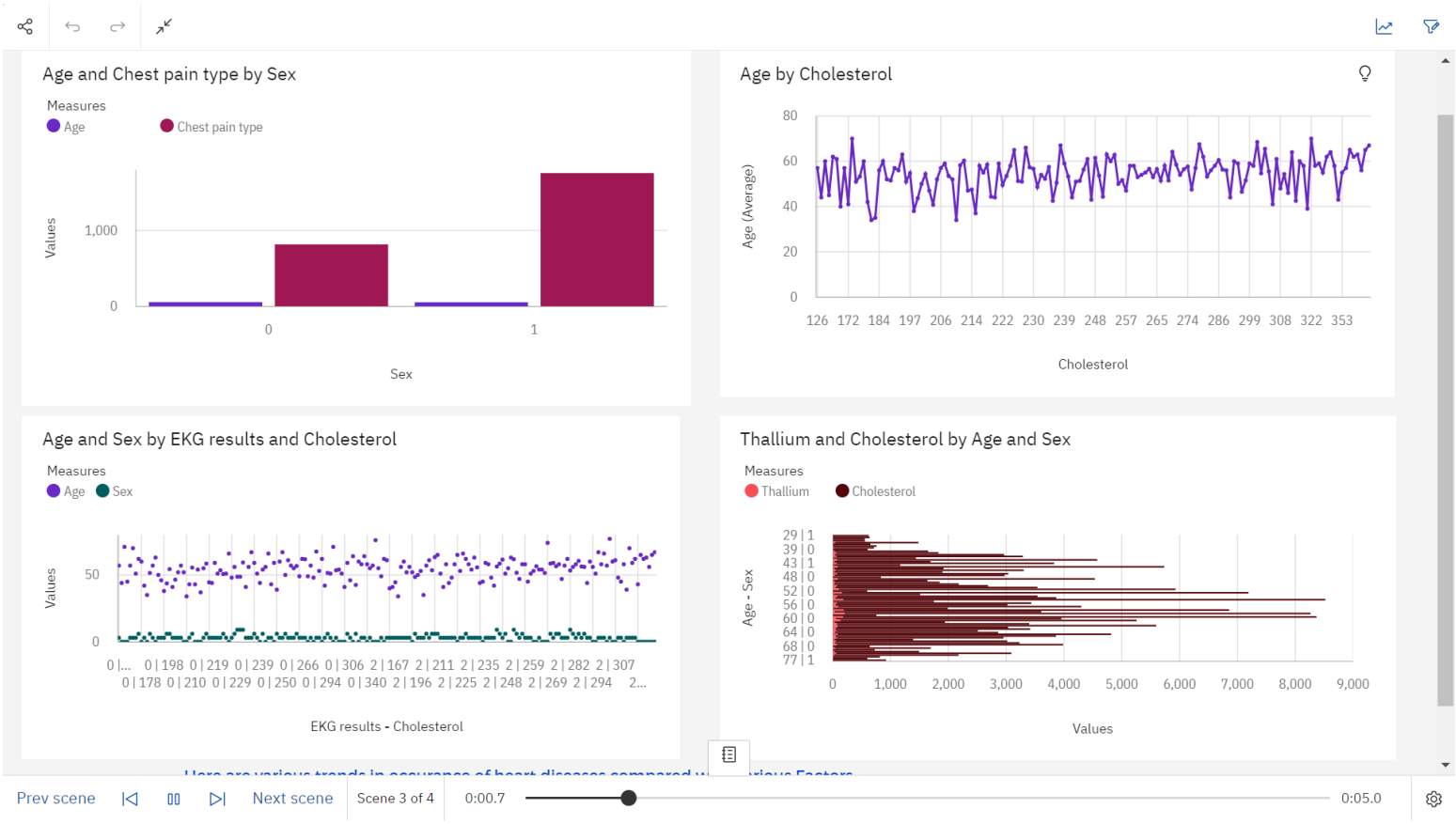




Story:

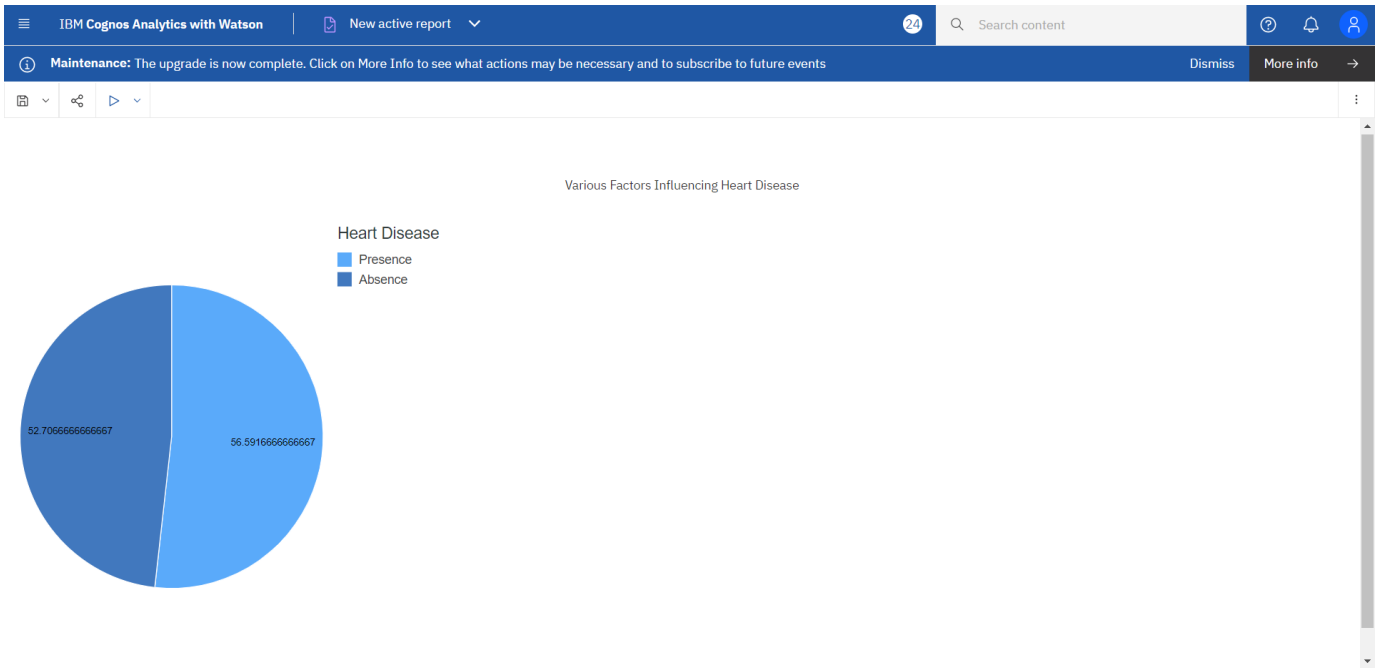

Visualizing and Predicting
Heart Disease
SAVE A HEART
Prev scene
Next scene
Scene 1 of 4
0:00.0
0:05.0


A Person getting affected by a Heart Disease is increasing
day by day and your at the right path of preventing and curing it.
Prev scene
Next scene
Scene 2 of 4
0:05.0
0:05.0


Age and Chest pain type by Sex
Measures
Age
Chest pain type
Values
1,000
0
0
1
Sex
Age by Cholesterol
Age (Average)
80
60
40
20
0
126 172 184 197 206 214 222 230 239 248 257 265 274 286 299 308 322 353
Cholesterol
Age and Sex by EKG results and Cholesterol
Measures
Age
Sex
Values
50
0
0 | ... 0 | 198 0 | 219 0 | 239 0 | 266 0 | 306 2 | 167 2 | 211 2 | 235 2 | 259 2 | 282 2 | 307
0 | 178 0 | 210 0 | 229 0 | 250 0 | 294 0 | 340 2 | 196 2 | 225 2 | 248 2 | 269 2 | 294 2...
EKG results - Cholesterol
Thallium and Cholesterol by Age and Sex
Measures
Thallium
Cholesterol
Age - Sex
29 | 1
39 | 0
43 | 1
48 | 0
52 | 0
56 | 0
60 | 0
64 | 0
68 | 0
77 | 1
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
Values
Prev scene
Next scene
Scene 3 of 4
0:00.7
0:05.0


HEART ATTACK
PREVENT IT BEFORE IT'S TOO LATE
ODETA STUIKYS ROSE
Prev scene
Next scene
Scene 4 of 4
0:05.0
0:05.0

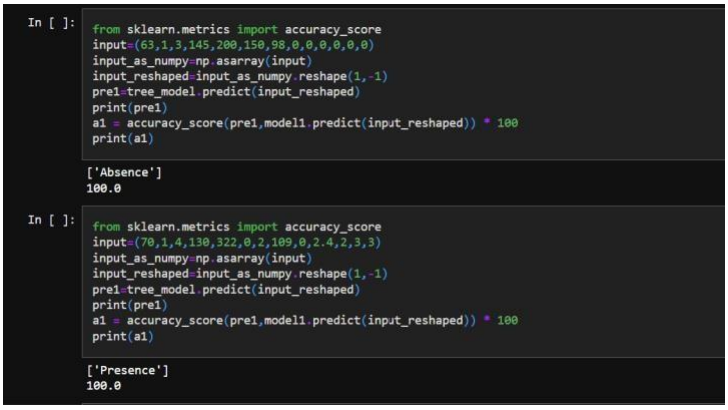
Report:



8. Testing

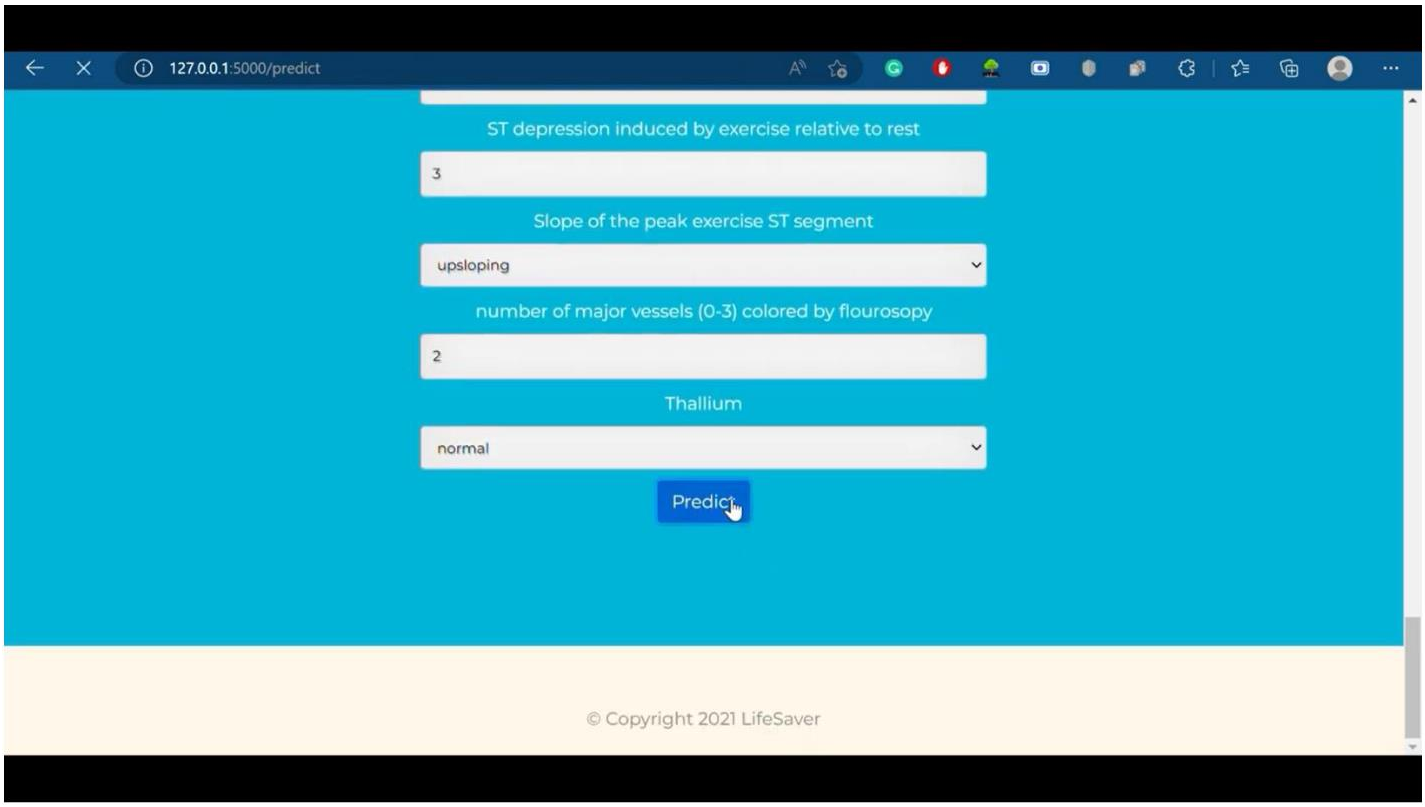
8.1 Test Cases

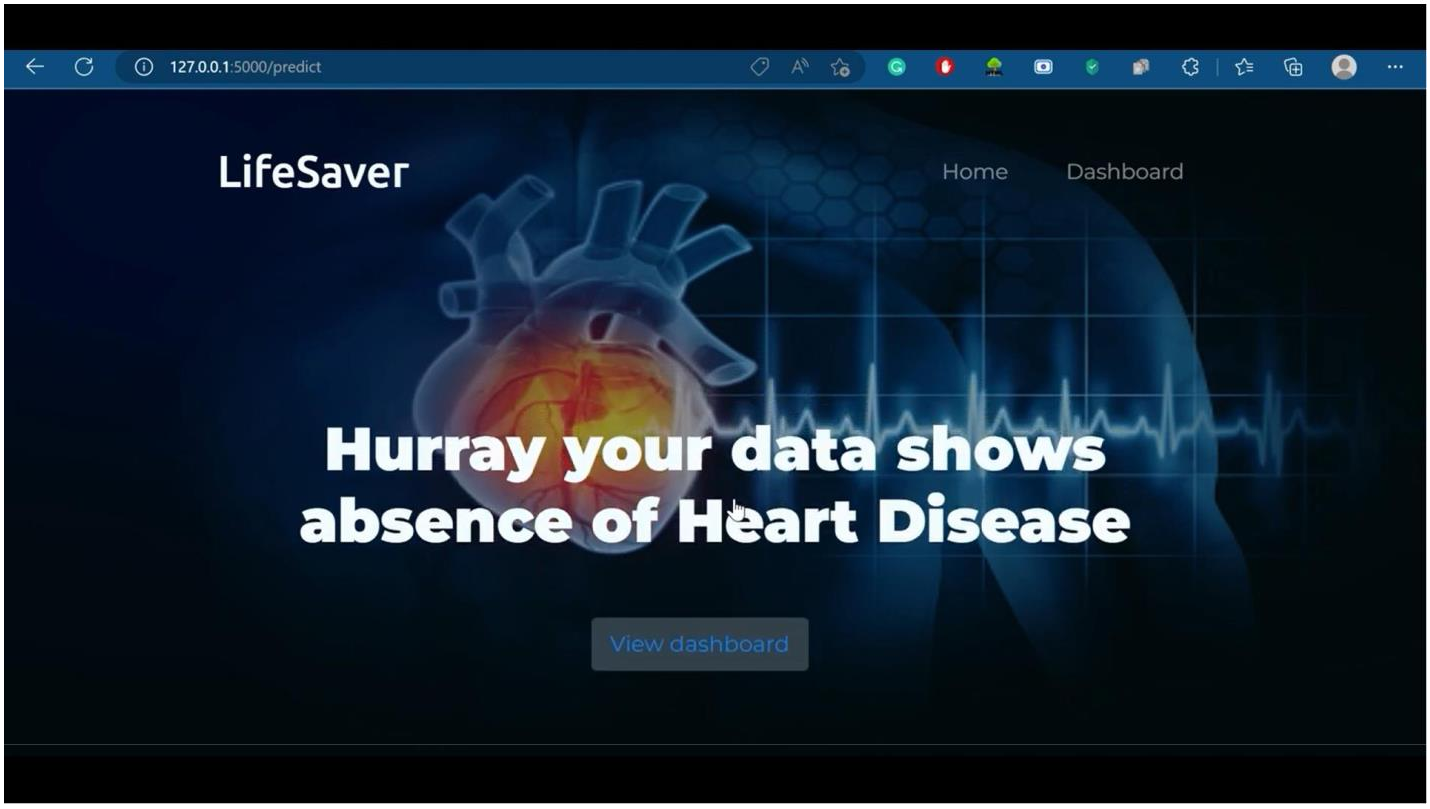
Testing the data model for various input values

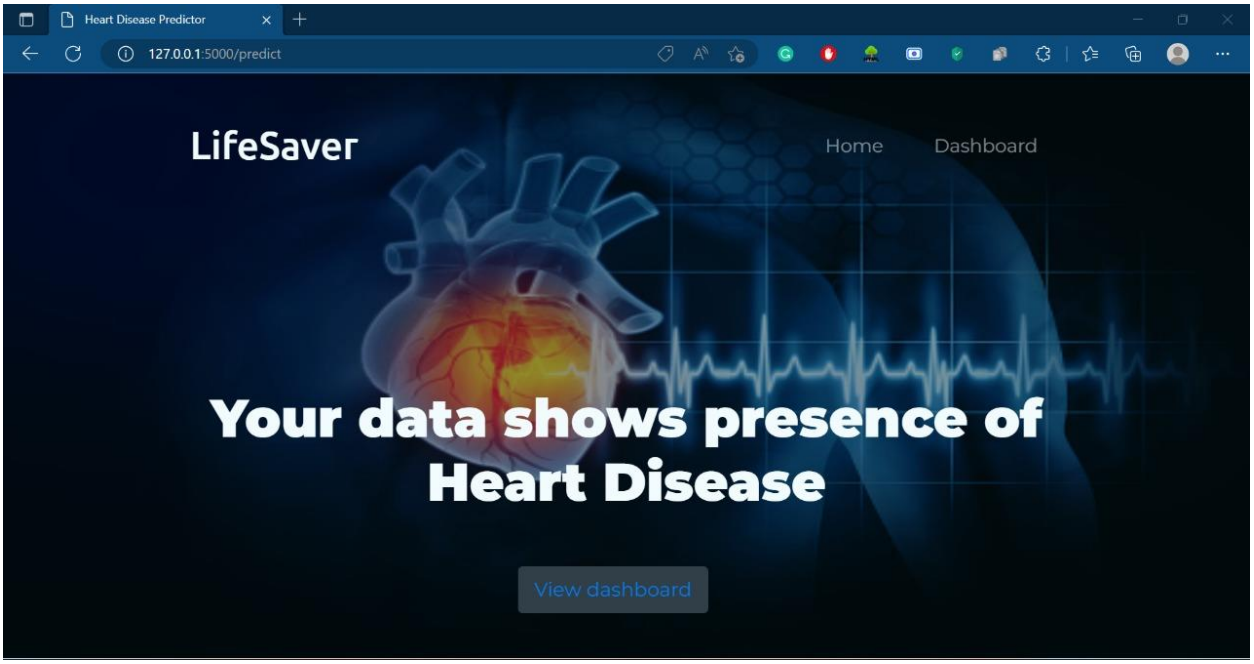


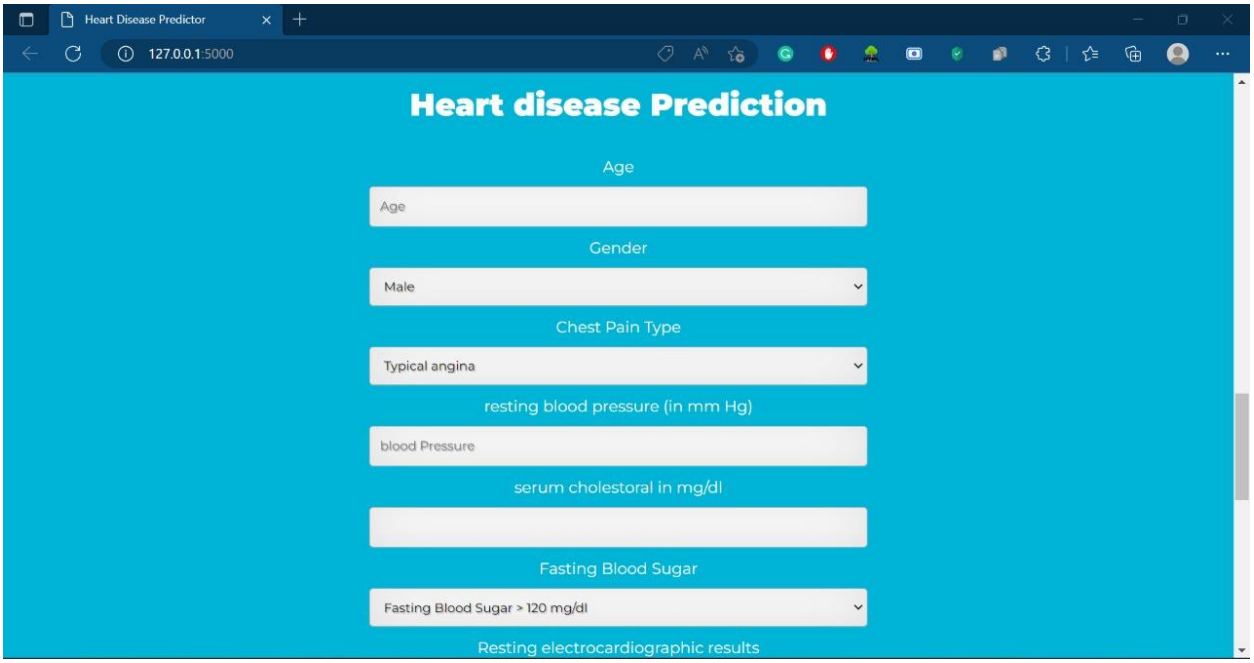
```
In [ ]: from sklearn.metrics import accuracy_score
        input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped))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

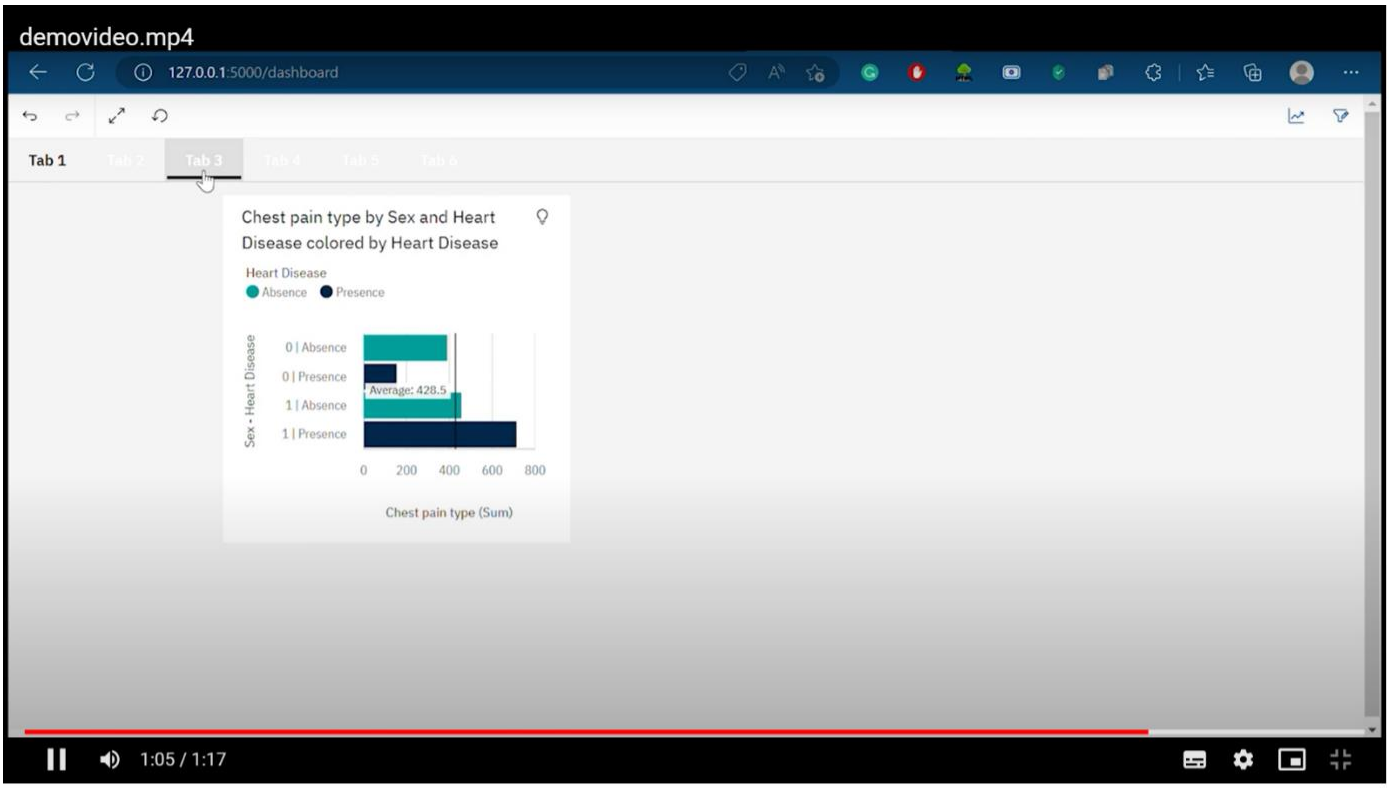
8.2 User acceptance Testing

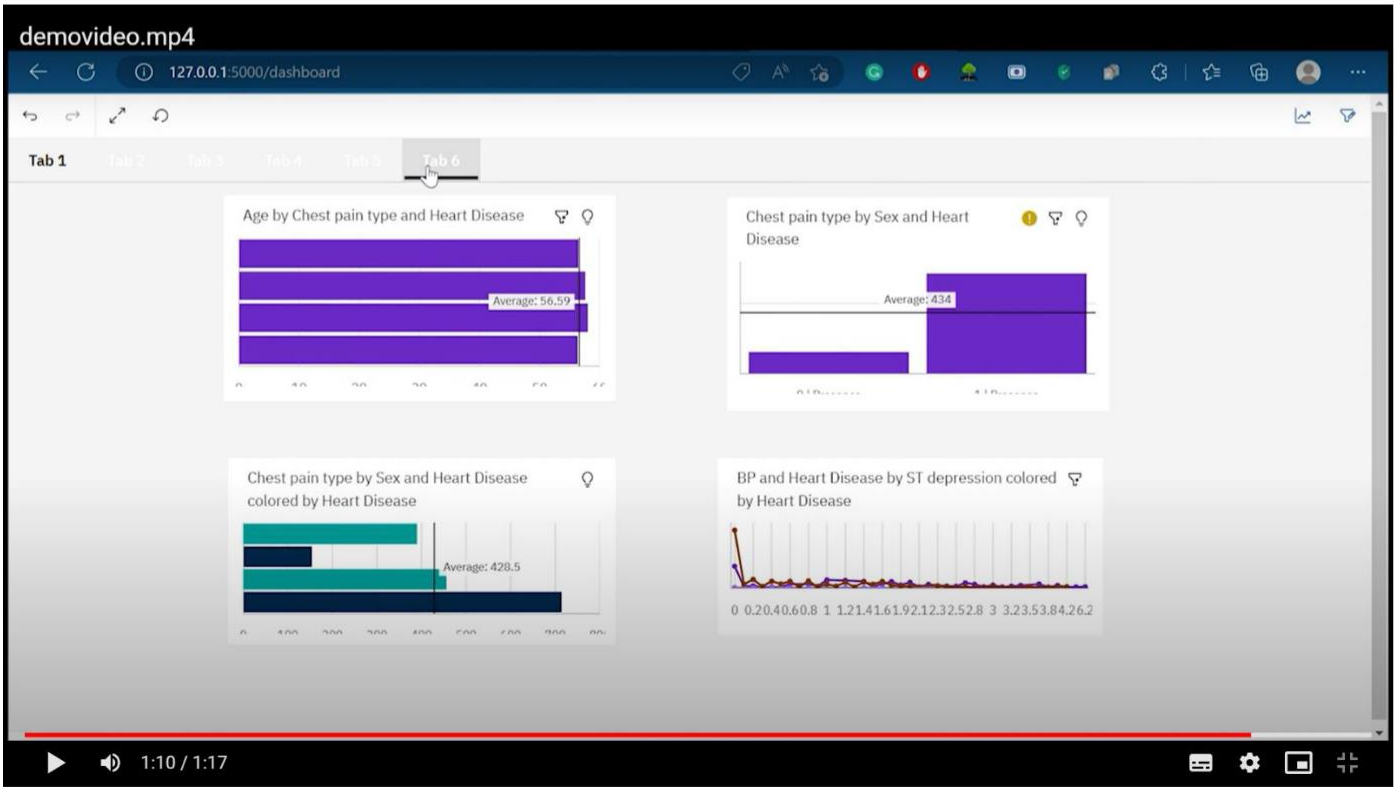





Heart Disease Predictor
127.0.0.1:5000/predict
LifeSaver
Home
Dashboard
Your data shows presence of Heart Disease
View dashboard


Heart Disease Predictor
127.0.0.1:5000
Heart disease Prediction
Age
Age
Gender
Male
Chest Pain Type
Typical angina
resting blood pressure (in mm Hg)
blood Pressure
serum cholestoral in mg/dl
Fasting Blood Sugar
Fasting Blood Sugar > 120 mg/dl
Resting electrocardiographic results


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


demovideo.mp4
127.0.0.1:5000/dashboard
Tab 1
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

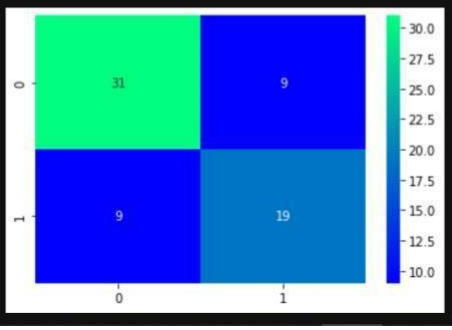
9.1 Performance Metrics

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

In [ ]:

```
from sklearn.tree import DecisionTreeClassifier

dtclas=DecisionTreeClassifier()
modeldt=dtclas.fit(X_train,y_train)
predictiondt=modeldt.predict(X_test)
cmdt= confusion_matrix(y_test,predictiondt)
sns.heatmap(cmdt, annot=True,cmap='winter')
print(classification_report(y_test, predictiondt))
```

```
              precision    recall  f1-score   support

     Absence       0.78      0.78      0.78        40
    Presence       0.68      0.68      0.68        28

    accuracy                           0.74        68
   macro avg       0.73      0.73      0.73        68
weighted avg       0.74      0.74      0.74        68
```



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/1pnlT9l5G3j0TMmJ8U3Oh8VwhC9fjux3C/view?usp=sharing