Visualizing and Predicting Heart Diseases with anInteractive Dashboard

NALAIYA THIRAN PROJECT REPORT

2022

Submitted by

Swetha V	**812819104045**
Boomika M	**812819104004**
Monisha R	**812819104029**
Sankari S	**812819104041**

Team ID: PNT2022TMID45611

1.

Content

a. Sprint Planning& Estimation

b. Sprint DeliverySchedule

c. Reports from JIRA

7. **CODING &SOLUTIONING (Explainthe features added in the project along with code)**
    a. Feature 1

    b. Feature 2

8. **TESTING**
    a. Test Cases

    b. UserAcceptance Testing

9. **RESULTS**
    a. Performance Metrics

10. **ADVANTAGES & DISADVANTAGES**
11. **CONCLUSION**
12. **FUTURE SCOPE**
13. **DEMO LINK**

TEAM ID : PNT2022TMID45611

VISUALIZING AND PREDICTING HEARTDISEASES WITH ANINTERACTIVE DASHBOARD

1. Introduction

a. 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 likelyto suffer from a heart disease in the near futureusing the features given.

b. 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

a. 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 techniquefor 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 improvescare by facilitating preventive care and EDAis a vital step while analysing data..

a. 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 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucoselevel, ECG in rest, heart rate and four types of chest pain. To predict the heart disease,K- means clustering algorithm is used along with data analytics and visualization tool. The paper discussesthe pre-processing methods,classifier performances and evaluation metrics.In the resultsection, the visualized data shows that the prediction is accurate.

Prediction of heart diseaseat 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 accuracylevel of each model given by different researchers. Into practice.

b. **Problem Statement Definition**

Who does the problem affect?

People with unhealthy lifestyles, stress, depression, age above 40 and when their ancestorsgot heartdisease(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 happenif we didn’t solve the problem?

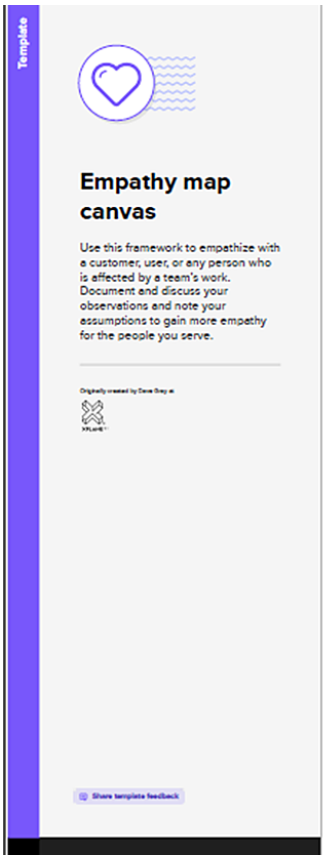
If we don’t solve the problem, many people will die at a young age. The death rate due to heart diseasewill increaserapidly.

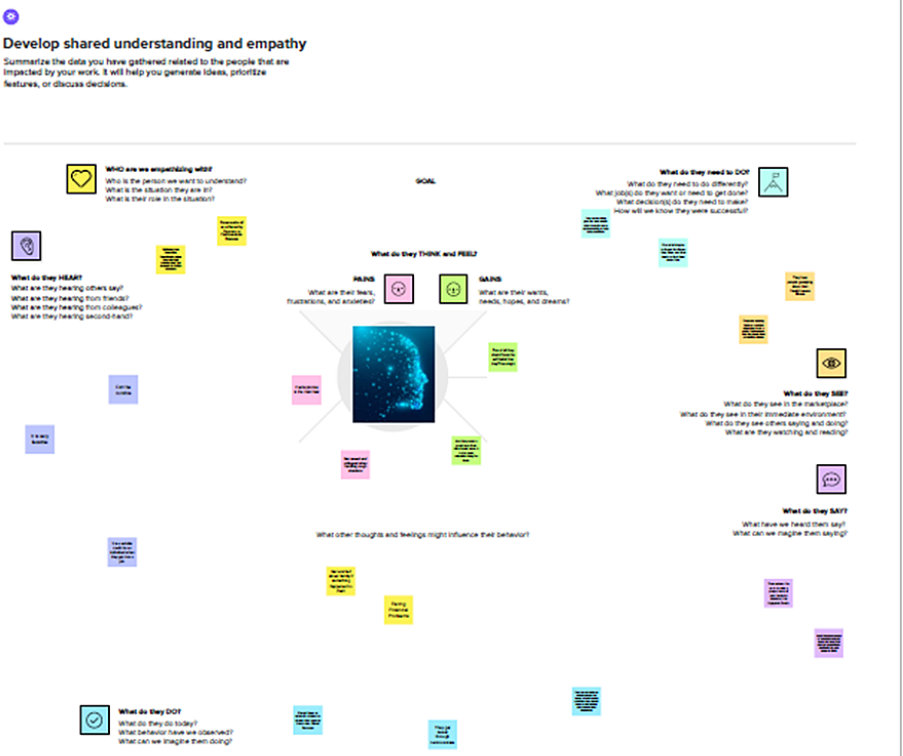
Why is it important to fix the problem?

We should predict the problem before giving treatment to the patients. As the problemis predicted early, wecan solveit easily and early.

1. Ideation and Proposed Solution

a. Empathy Map Canvas





a. Ideation and Brainstorming

Step-1: Team Gathering, Collaboration and Select the 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
- **2-8 people** recommended

Share template feedback

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

Team gathering

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

Set the goal

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

Learn how to use the facilitation tools

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

Open article

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

PROBLEM

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.

Key rules of brainstorming

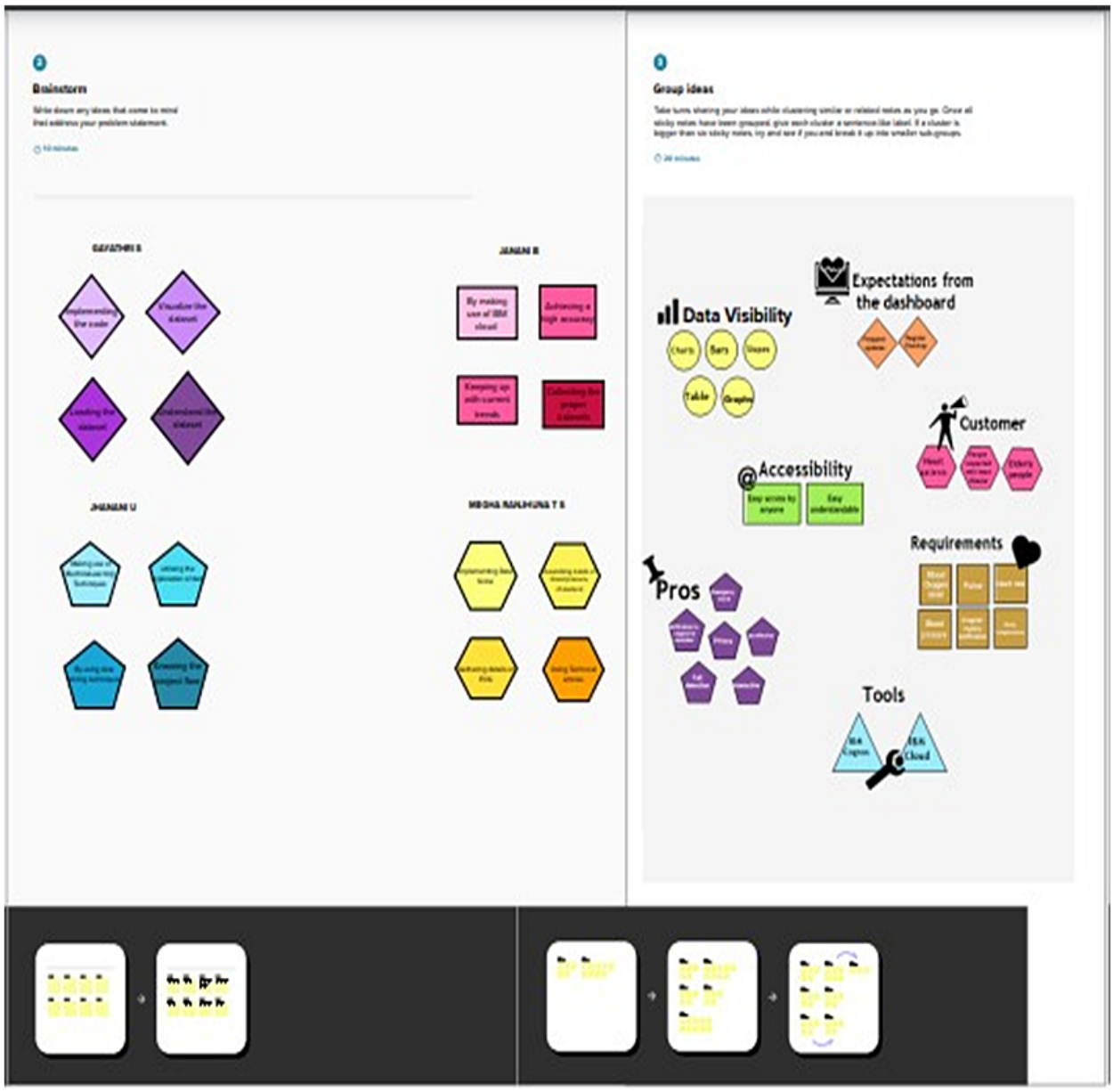
To run an smooth and productive session

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

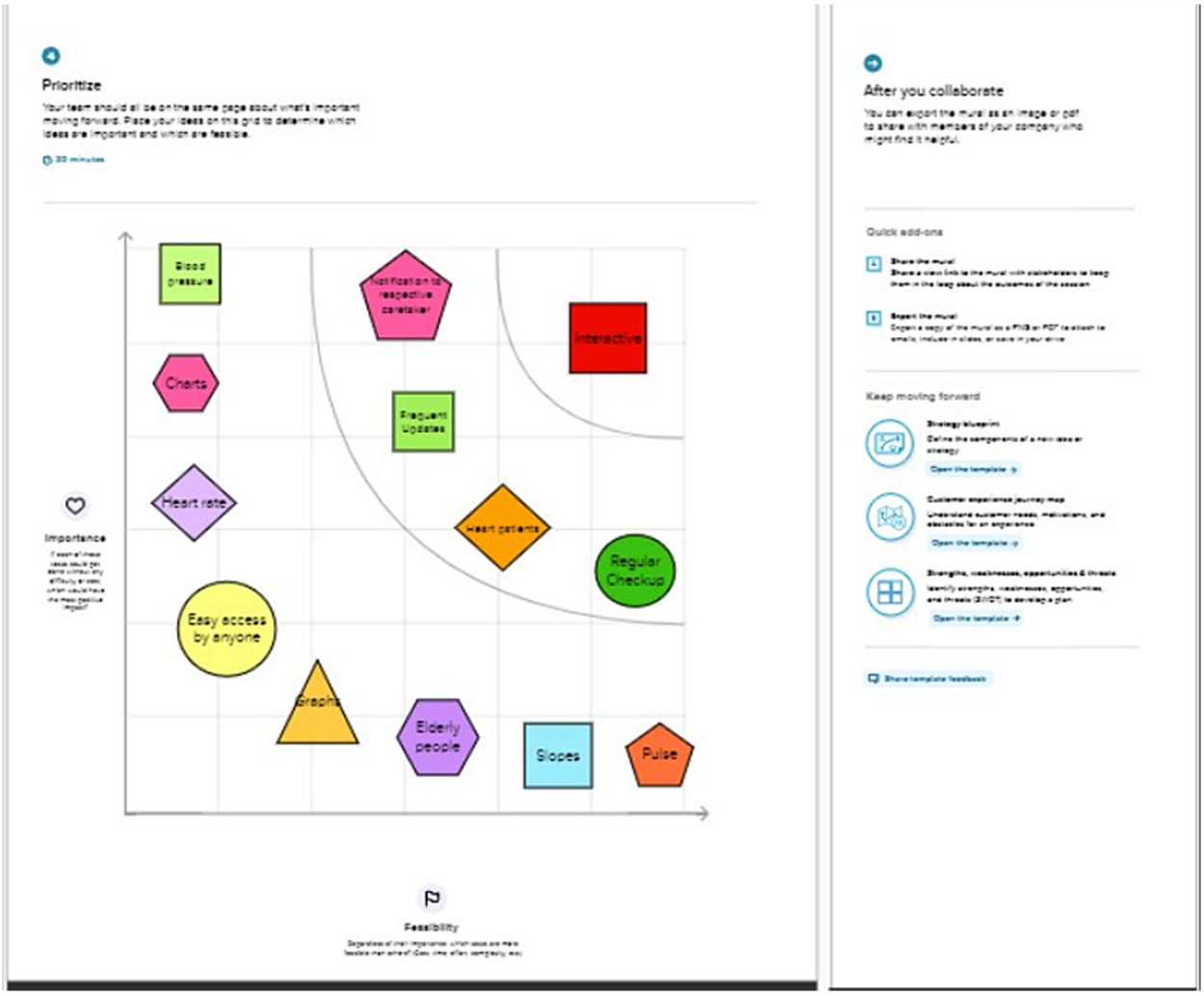
Need some inspiration?

Open example

Step-2: Brainstorm, Idea Listing and Grouping



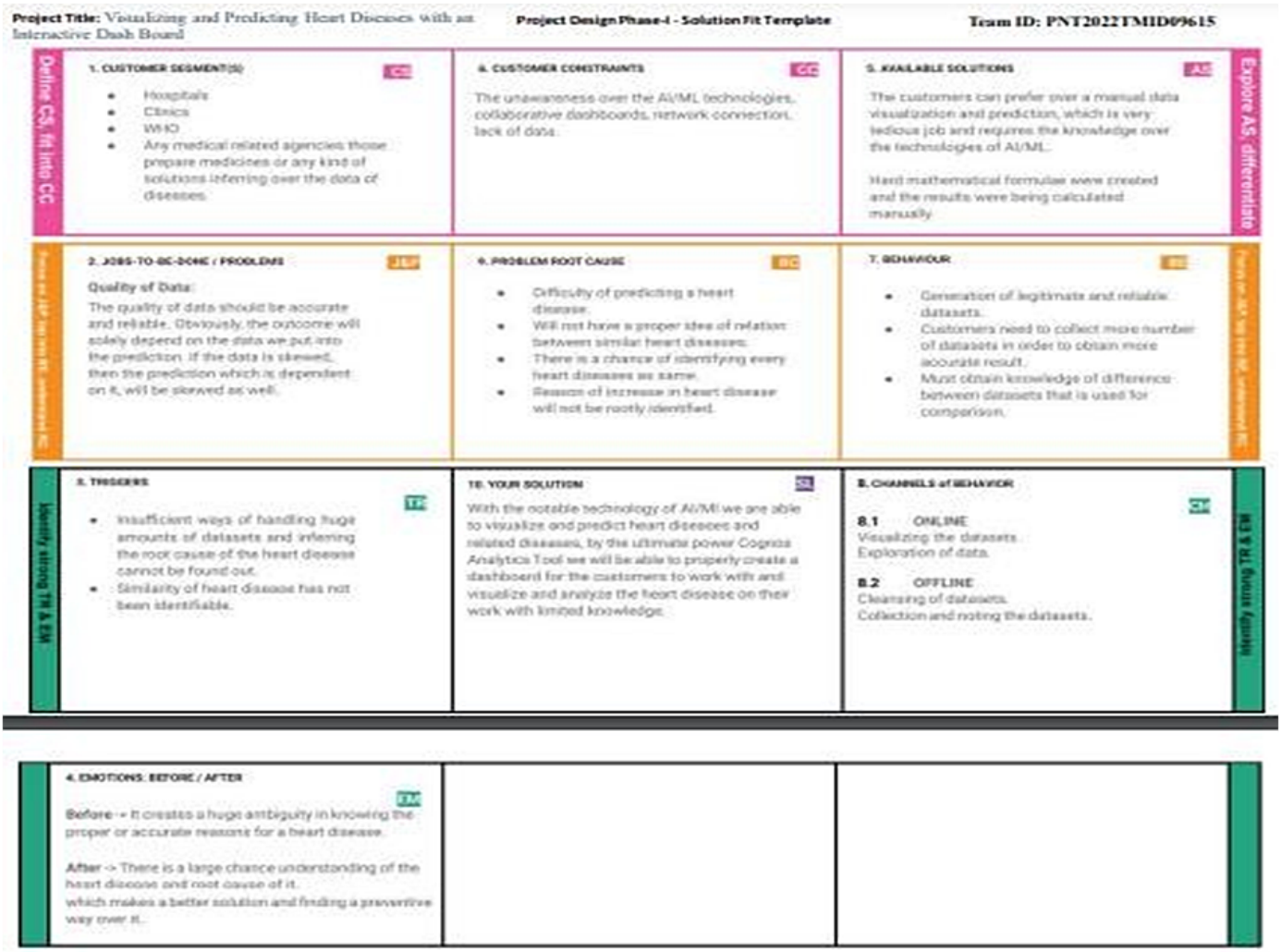
Step-3:IdeaPrioritization



3.3 ProposedSolution

S.No.	Parameter	Description
1.	Problem Statement (Problemto besolved)	The leading cause of death in the developed world isheart disease. As a result, work must be done to reducetherisksofhavingaheartattackorstroke.Itisinfeasible for a common man to frequently undergo tests for ECGandso on. Hence, itrequires a replacementthatisboth convenientanddependable.
2.	Idea/Solutiondescription	Theproposedsolutionproposesaninteractivedashboardfor visualizing and forecasting heart disorders, in whichtheuser may observehis/her medical report analysis as well as the projected endresult. IBM Cognoswill be used to create thedashboard.MachinelearningAlgorithmswillbeused toforecast cardiacdisease.
3.	Novelty/Uniqueness	Makesrecommendationstotheuserbasedonthatperson'smedical analysis.
4.	SocialImpact/CustomerSatisfaction	It helps with disease prediction at an early stage andfrequentlyalertstheusertotheircurrenthealthstatus.Both the user and the doctor can benefit from thesystem's improved decision-making regarding cardiacdisease
5.	Business Model (RevenueModel)	CanbedeployedbyHospitalsorNFOs,sothatitmakesthe analysis ina fastmanner.
6.	ScalabilityoftheSolution	The solution can work effectively on long and smalldatasets. It can also be changed to predict various otherdiseasesdepending on thedataset

3.4 ProblemSolution Fit



4 RequirementAnalysis

4.1 FunctionalRequirement

Followingarethe functionalrequirementsoftheproposed solution.

FRNo.	Functional Requirement (Epic)	SubRequirement(Story/Sub-Task)
FR-1	UserRegistration	Enablesusertomakeregistrationfortheapplication throughGmail
FR-2	UserConfirmation	Onceafterregistration,theuserwillgetconfirmationviaEmail
FR-3	VisualizingData	Usercan visualizethetrendsontheheart diseasethroughDashboardcreatedusing IBMCognos Analytics
FR-4	GenerationReport	User can view his/her health report and can makedecisionsaccordingly

a. Non-FunctionalRequirement

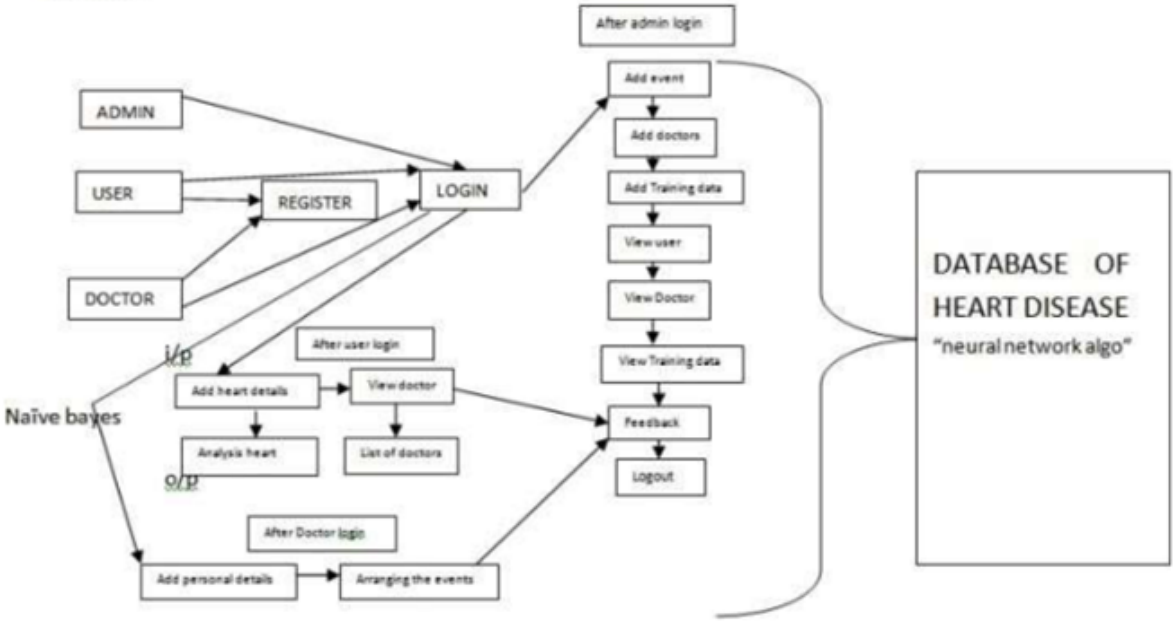
Followingarethenon-functionalrequirementsoftheproposedsolution.

FRNo.	Non-FunctionalRequirement	Description
NFR-1	Usability	The application will have a simple and userfriendlygraphical interface. Users will be ableto understandandusealltheFeaturesoftheapplicationeasily.Any actionhastobeperformedwithjust afewclicks
NFR-2	Security	For security of the application the technique known as databasereplication should be used so that all the important data should bekept safe. Incase of crash, the system should be able to backupand recoverthe data
NFR-3	Reliability	The application has to be consistent at everyscenarioandhastoworkwithoutfailureinany environment
NFR-4	Performance	Performance of the application depends on the response time andthe speed of the data submission. The response time of theapplicationisdirect and faster whichdependsontheefficiency ofimplementedalgorithm
NFR-5	Availability	Theapplicationhasto beavailable24x 7forusers withoutanyinterruption
NFR-6	Scalability	The application can withstand the increase in the no. of usersandhas tobe ableto develop Higher versions

1. ProjectDesign

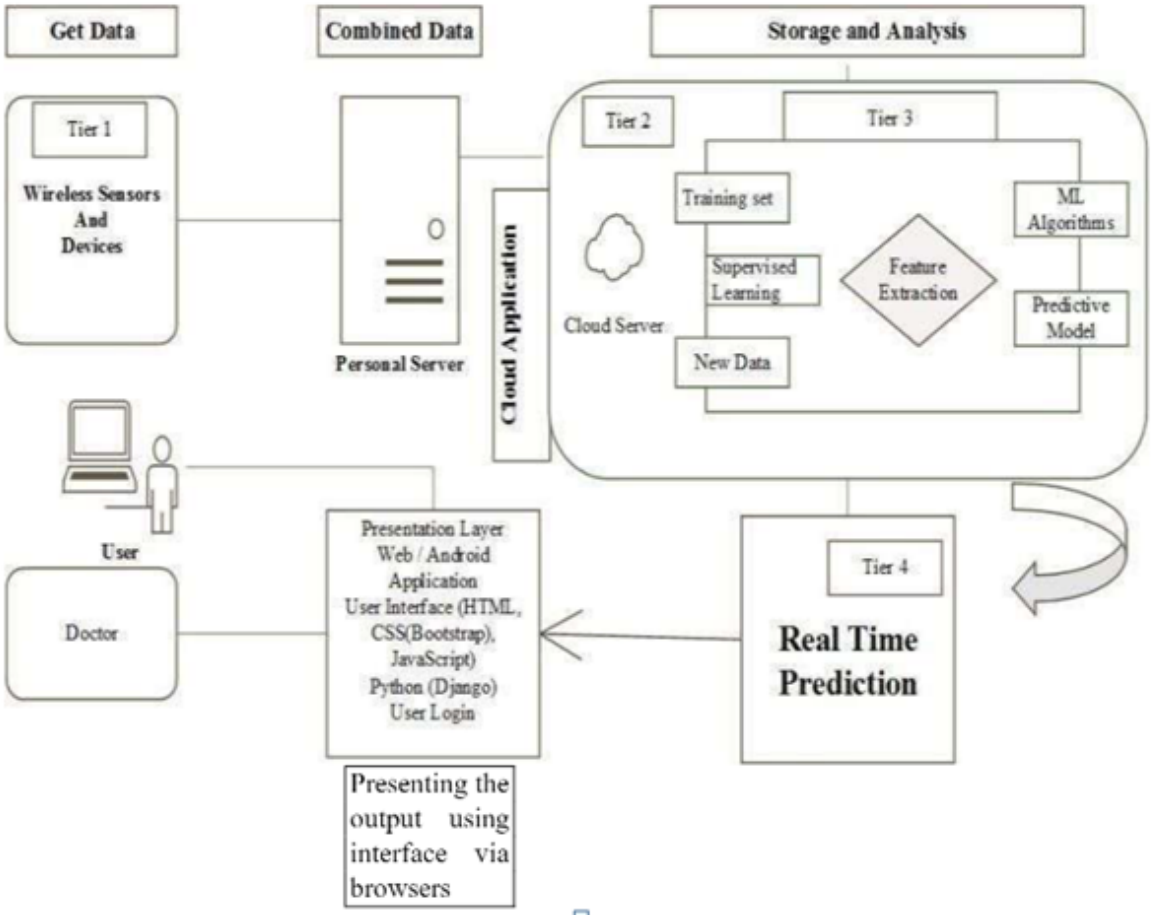
a. DataFlowDiagram

DIAGRAM:



SolutionandTechnicalArchitecture5

Architecture:



1. **ProjectPlanning andScheduling**

ScriptPlanningandExecution

Sprint	Functional Requirement (Epic)	User StoryNumber	UserStory/Task	Story Points	Priority	TeamMembers
Sprint-1	Registration	USN-1	As a user, I can register for theapplication by entering my email,password,and confirming mypassword	3	High	1
Sprint-1		USN-2	Asauser,Iwillreceiveconfirmationemailc nceIhaveregisteredforthe application	3	High	3
Sprint-1		USN-3	Asauser,Icanregisterfortheapplicatio nthrough Gmail	3	Medium	1
Sprint-1	Login	USN-4	Asauser,Icanlogintotheapplicationby enteringemail&password	6	High	5
Sprint-2	Dashboard	USN-5	Attractive dashboard For theApplication	3	Medium	3
Sprint-2		USN-6	Profile-view& updateyour profile	5	Low	2
Sprint-2		USN-7	Home-AnalyzeyourHeartproblem	2	High	4
Sprint-2		USN-8	Theuserwillhavetofillinthebelow13 fields for the system to predict adisease -Ageinyear-Gender	7	High	2
Sprint-3	Support	USN-9	Getfeedback fromusers	10	Medium	3
Sprint-3		USN-10	Responds to user queries viatelephone,emailetc.	3	Medium	2
Sprint-3		USN-11	Theteammustrespondimmediatelyto thequeriesbased onthepriority	5	High	5
Sprint-4	SystemRequirements	USN-12	HardwareRequirement 3. LaptoporPC • i5processorsystemor higher • 4GB RAM orhigher • 128GB ROM orhigher 4. Mobile • (12.0and above)	5	Low	2
Sprint-4		USN-13	SoftwareRequirement 2. LaptoporPC	8	Medium	4

			• Windows10 orhigher • AndroidStudio			

6.2SprintDeliverySchedule

Sprint	Total Story Points	Duration	SprintStartDate	Sprint End Date(Planned)	Story PointsCompleted(ason Planned End Date)	Sprint Releas Date(Actual)	e
Sprint-1	20	6Days	24Oct2022	29Oct2022	20	29Oct2022	
Sprint-2	20	6Days	30Oct2022	04 Nov 2022	17	04 Nov 2022	
Sprint-3	20	6Days	05 Nov 2022	11 Nov 2022	18	11 Nov 2022	
Sprint-4	20	6Days	12 Nov2022	17 Nov 2022	19	17 Nov 2022	

1. CodingAndSolutioning

a. MachineLearning

Learningwhichmodelisbest forthegivenDataset

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.

Comparingitwiththeaccuracygotten fromDecisionTree:

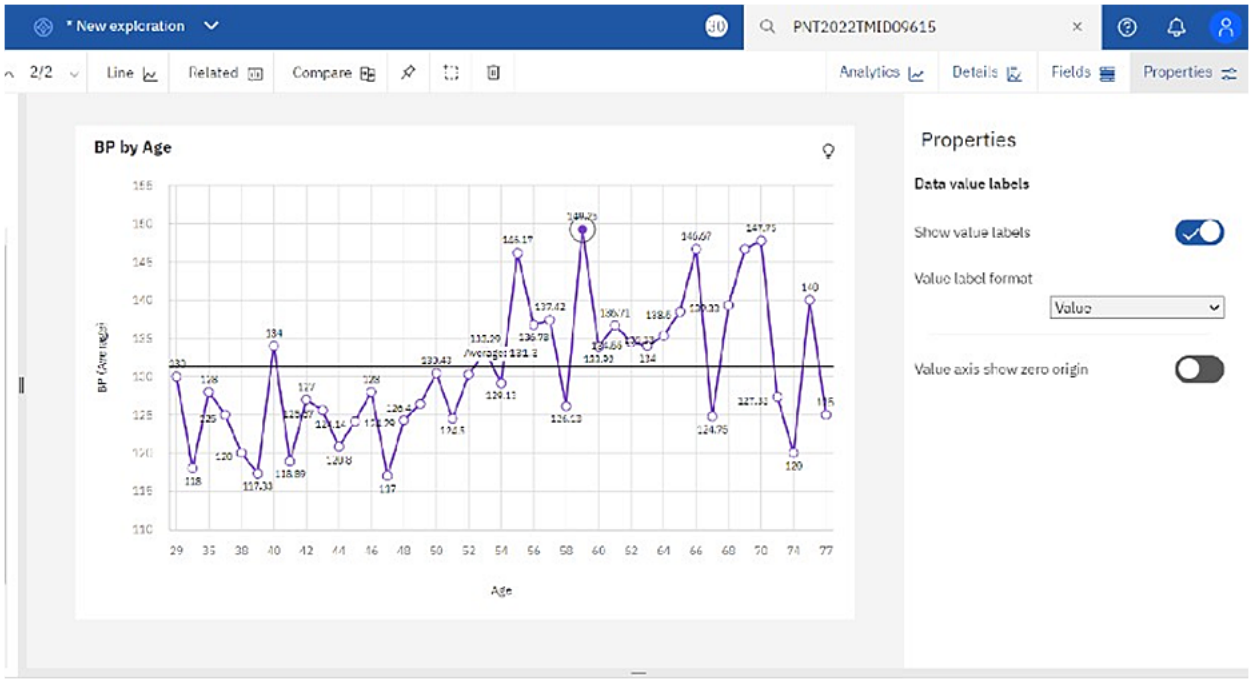
TP=cm[0][0]
#cm=Confusi
onMatrixTN=
cm[1][1]
FN=cm[1][0]
FP=cm[0][1]

print('TestingAccuracyforDecisionTree:',(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('TestingPrecision forDecision Tree:',(TP/(TP+FP)))

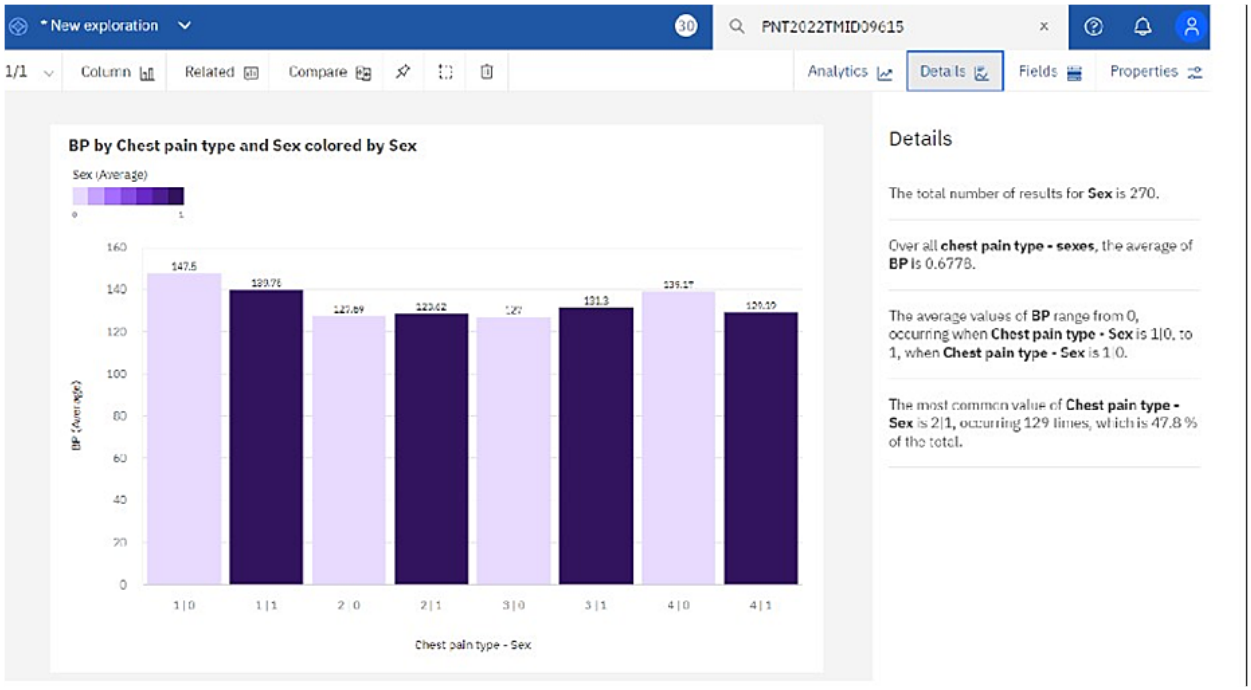
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

a. Dashboard

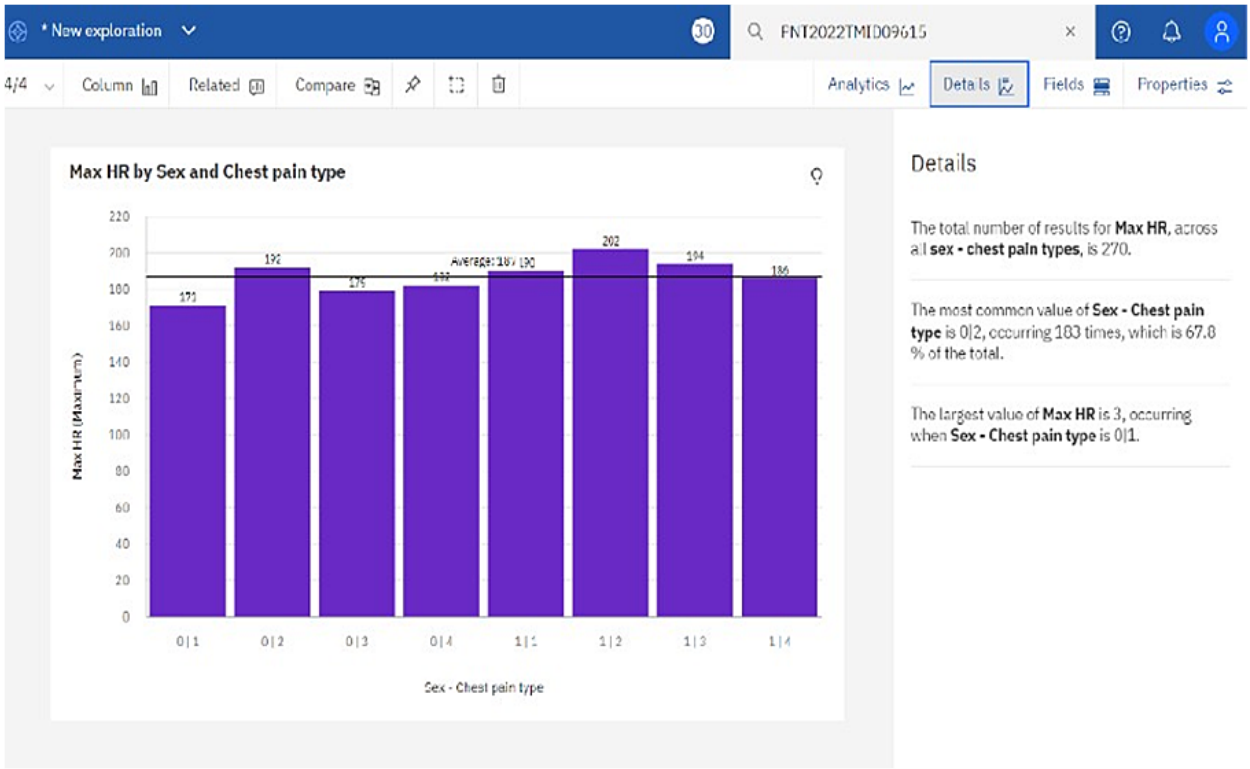
AverageBPduringchestpain



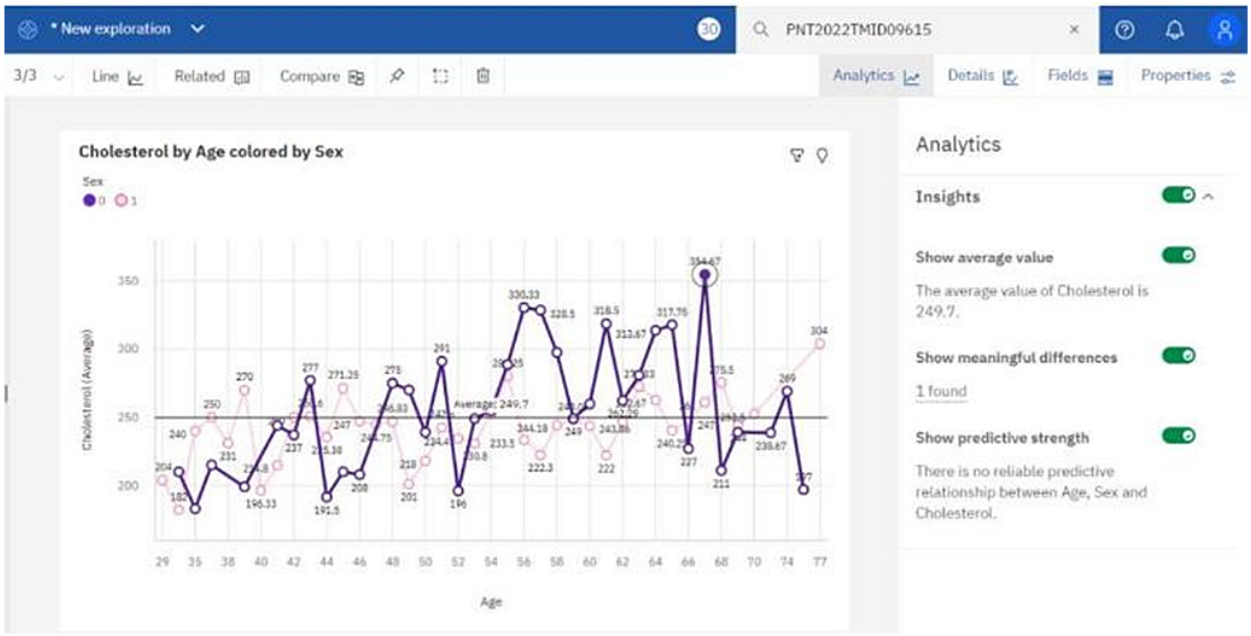
ExplorationOfBPvsChestPainTypeAndGender:



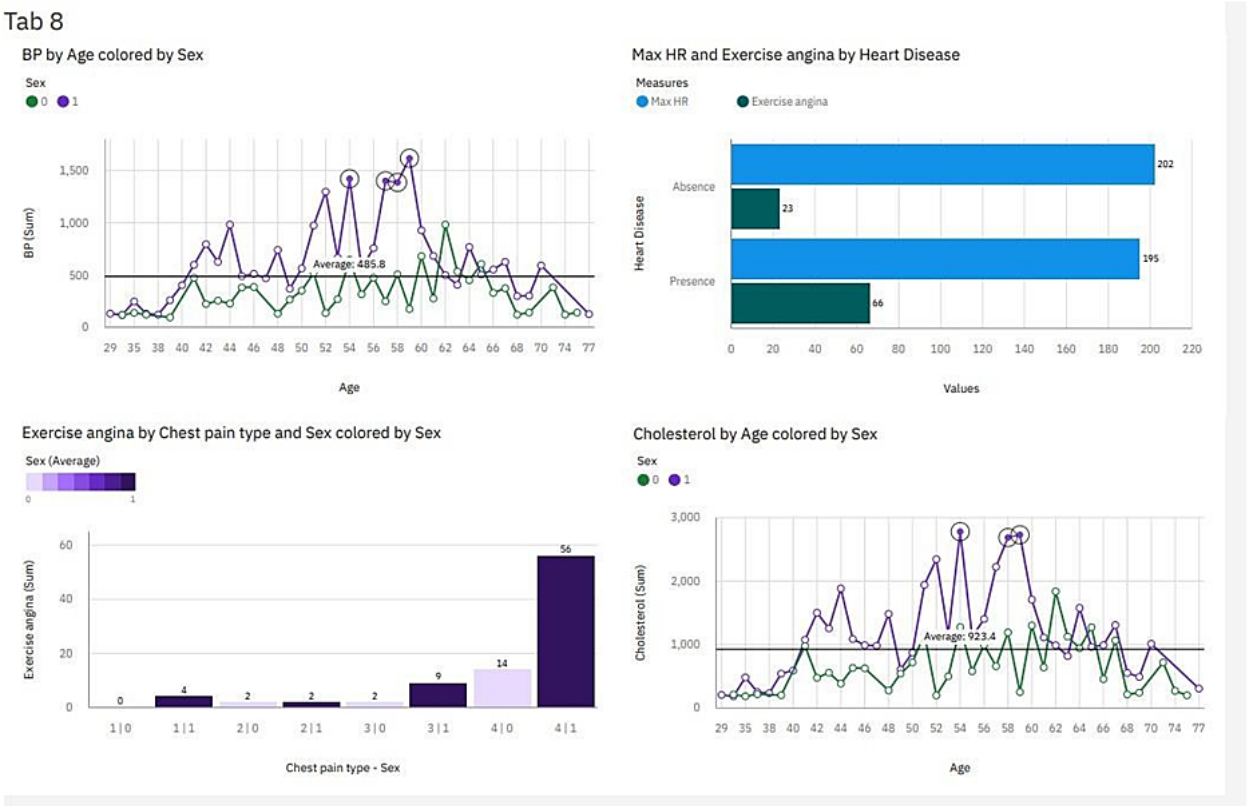
ExplorationOfMax HeartRateDuring TheChestPain:



ExplorationOfCholesterol byageandGender:



DashboardShowingDifferentTypesOfVisuals:



1. Testing

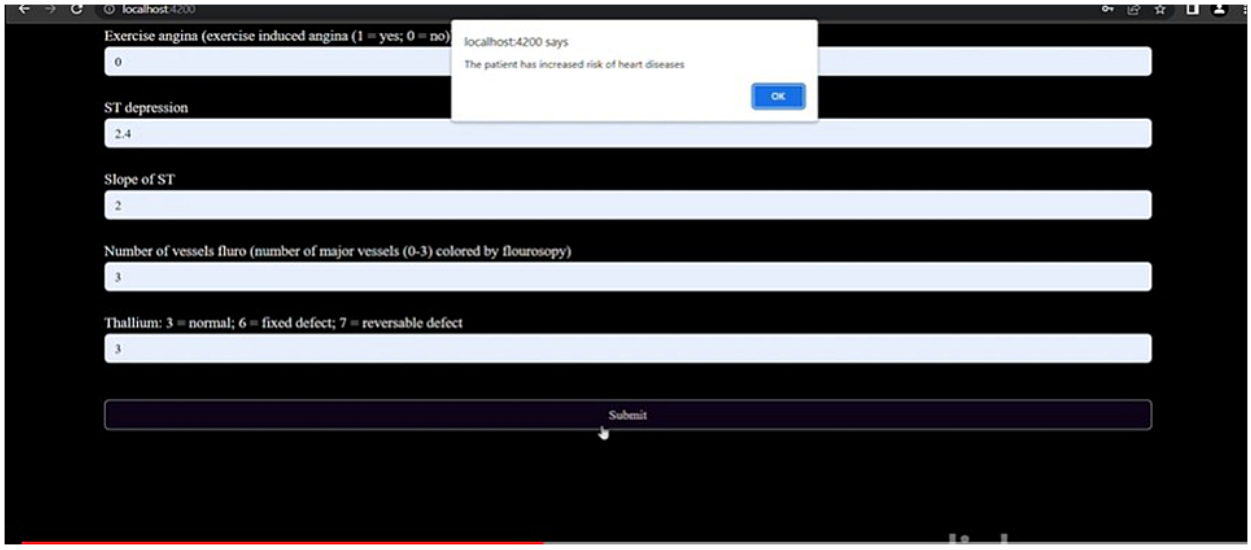
a. TestCases

Testingthedatamodel forvariousinputvalues.

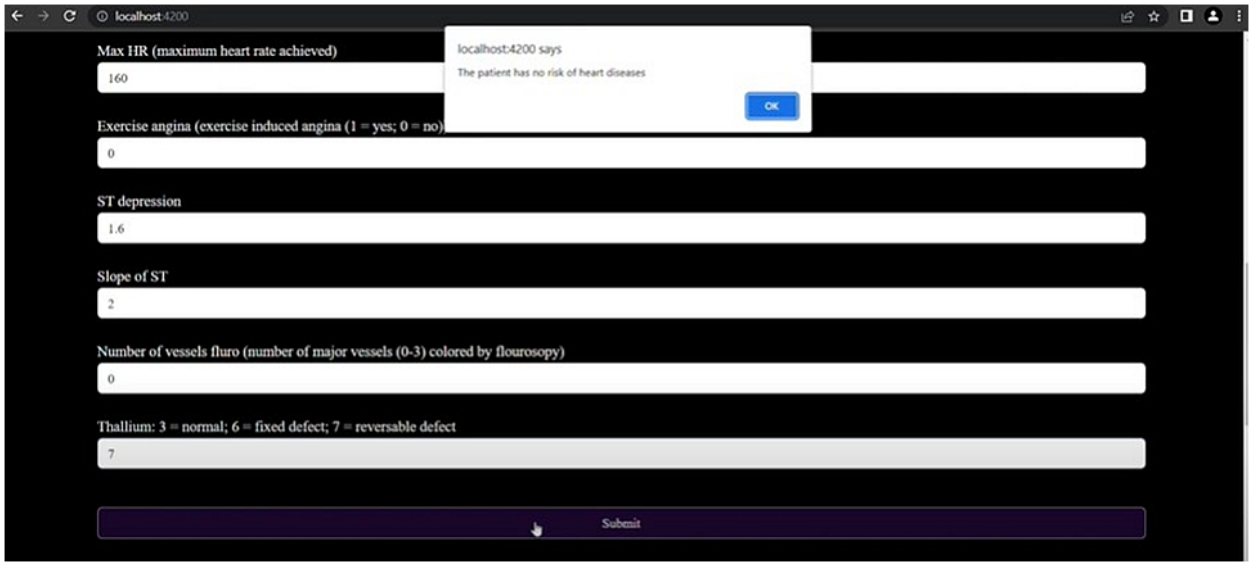
```
In [ ]: from sklearn.metrics import accuracy_score
        input=(63,1,3,145,200,150,98,0,0,0,0,0,0)
        input_as_numpy=np.asarray(input)
        input_reshaped=input_as_numpy.reshape(1,-1)
        pre1=tree_model.predict(input_reshaped)
        print(pre1)
        a1 = accuracy_score(pre1,model1.predict(input_reshaped))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

a. UseracceptanceTesting

Testingacase whereuserhasheartdisease



Testingacase whereuserdoesnot haveheartdisease



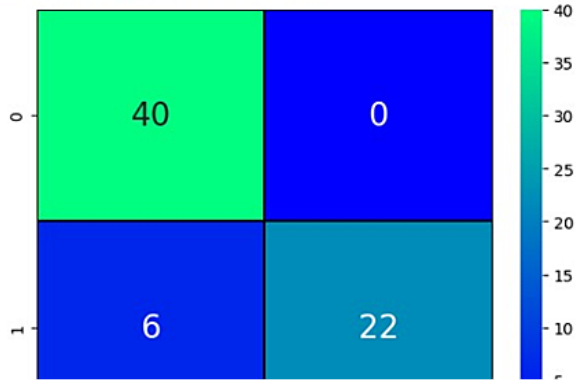
1. Result

a. PerformanceMetrics

Theconfusion matrixbelow shows theperformancemetricsofthemachinelearning 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))

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
```



1. AdvantagesDisadvantages

Advantages:

1. Thisisoneofthefastest waystodetermineif apersonislikelytosufferfromaheartdiseaseor not.
2. Usefulformedicalpractitionerstoeasilyclassifytheirpatients.
3. UserFriendly
4. Easytounderstand
5. Secure
6. Dashboardprovidesinsightfulinformations

Disadvantages:

7. Needswork
8. Usersneedtoknowallthe fields
9. DoesNottakenull value asinput
10. Doesnotprovide suggestionstouser

1. Conclusion

Complications of heart disease include heart attack andstroke. You can reduce the risk ofcomplications with early diagnosis and treatment. So the suggestion that we get from the websitemighthelp savepatients. Itis alwaysto get treatedin theearly stagesof heartdisease.

2. FutureScope

Like the saying goes “Prevention is better than cure”. We have to look into methods to preventheartdiseases altogether otherthan justpredictingit in early stages.
To use this website we need to take a lot of tests beforehand. So it would be better if we requirelessattributes and still givean effectiveresult

DEMO LINK:

https://drive.google.com/file/d/1uxp39V3Sp4X_oyZvG1kZb26kSLKQ8Hcv/view?usp=drivesdk

https://github.com/IBM-EPBL/IBM-Project-20154-1659713526.git