VISUALIZINGAND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

NALAIYA THIRAN IBM PROJECT REPORT
Team Id: PNT2022TMID11921

Submitted by

SUBALAKSHMI P	**73771921198**
SAKTHISARAN S	**73771921184**
THILAGAVATHI A	**73771921208**
NITHISH KUMAR S R	**73771921168**

in partial fulfillment for the award of the degree

Of

BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

K.S. RANGASAMY COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

TIRUCHENGODE – 637 215

CONTENTS

1. INTRODUCTION

Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithms.

1.1 Project overview

In this fast moving world people want to live a very luxurious life so they work like a machine in order to earn lot of money and live a comfortable life therefore in this race they forget to take care of themselves, because of this there food habits change their entire lifestyle change, in this type of lifestyle they are more tensed they have blood pressure, sugar at a very young age and they don't give enough rest for themselves and eat what they get and they even don't bother about the quality of the food if sick the go for their own medication as a result of all these small negligence it leads to a major threat that is the heart disease.

1.2 Purpose

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions for providing appropriate results and making effective decisions on data, some data mining techniques are used to better the experience and conclusion that have been given.

2. LITERATURE SURVEY

The main aim of this paper is to use various classification algorithms of data science framework to somehow detect the chances of having a heart disease. Also, the main aim of this research paper is to find out the most efficient classification algorithm that can help us to detect heart diseases at early stage. This algorithm can be used on heart records of the patient or by using it on classification reports. This research was conducted and tested upon various algorithms to test its accuracy like Logistic Regression, Random Forest, Vector Support and XG-Boost. After applying these algorithms of prediction model has been developed

2.1 Existing problem

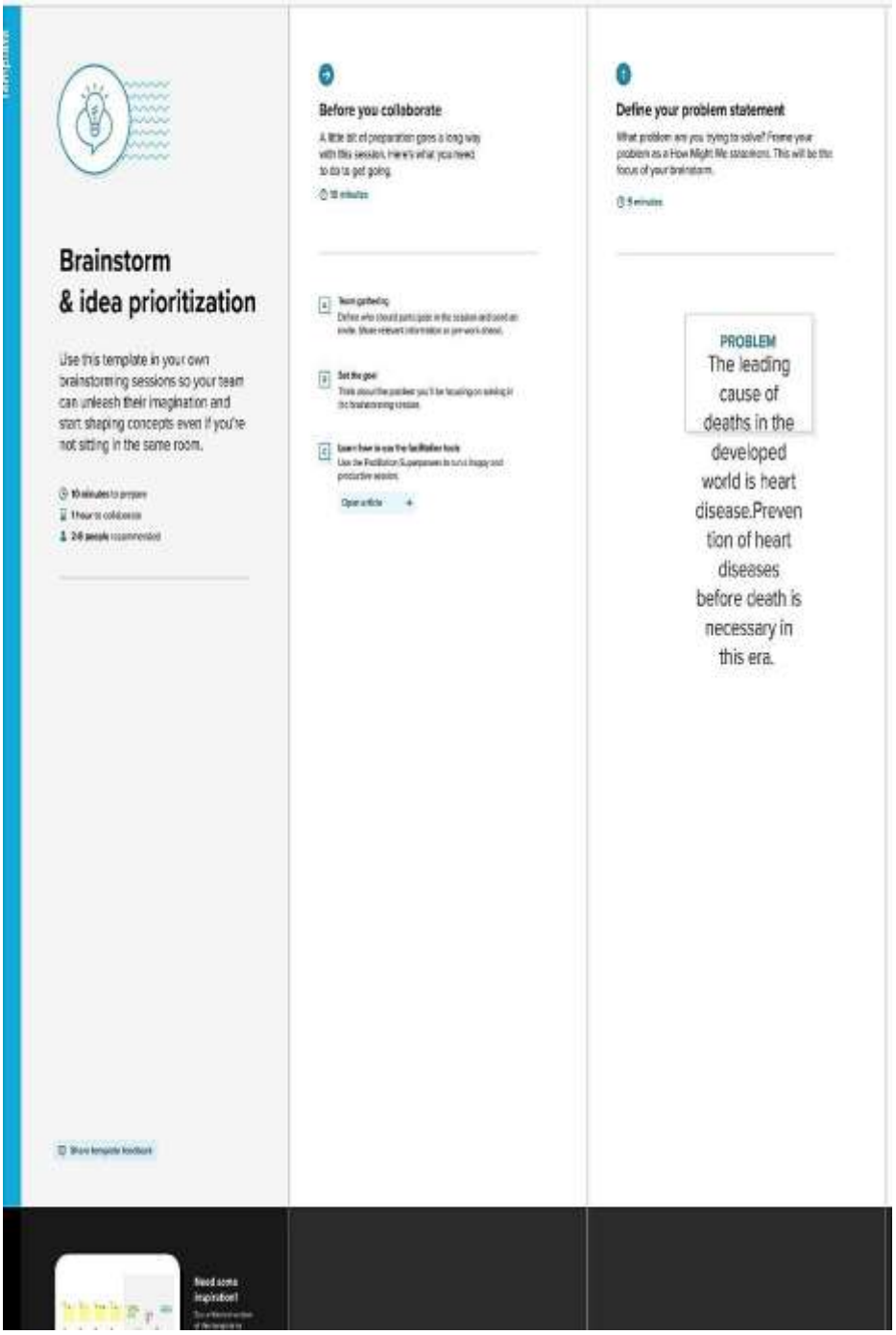
The diagnosis of heart disease is usually based on signs, symptoms and physical examination of the patient. There are several factors that increase the risk of heart disease, such as **smoking habit, body cholesterol level, family history of heartdisease, obesity, high blood pressure, and lack of physical exercise**

2.2 Problem statement

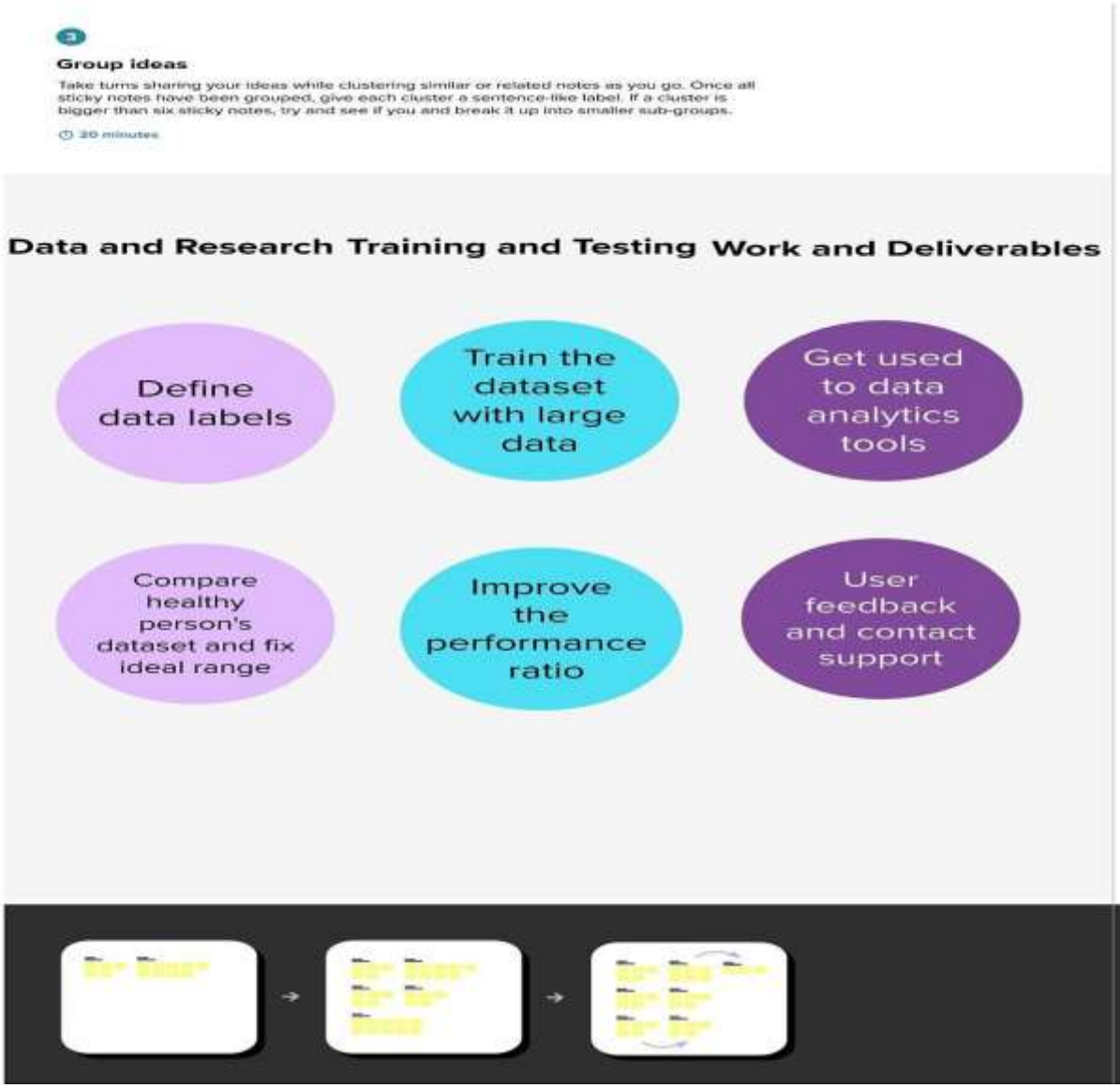
It is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. Since we have a good amount of data intoday's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

3 IDEATION PHASE

3.1 Brainstorm & idea prioritization



Step-3: Idea prioritazion

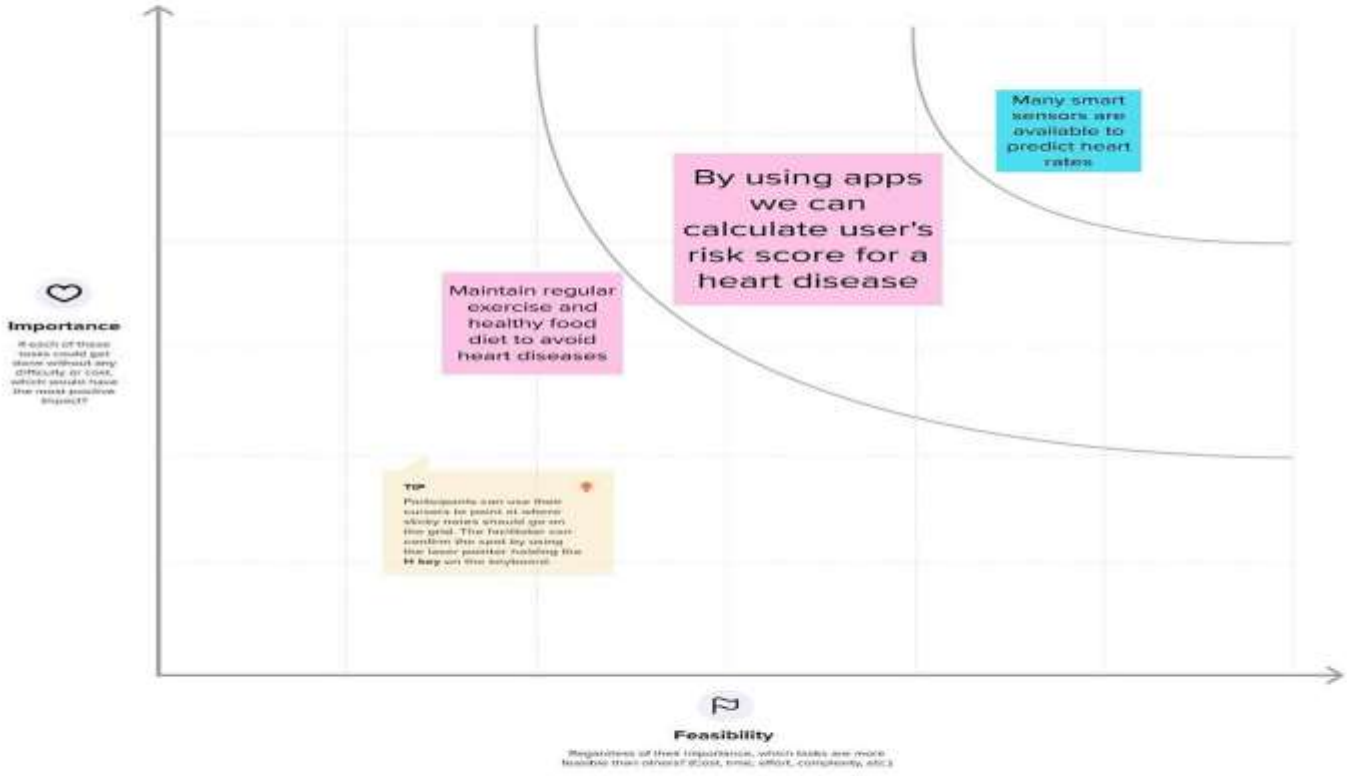


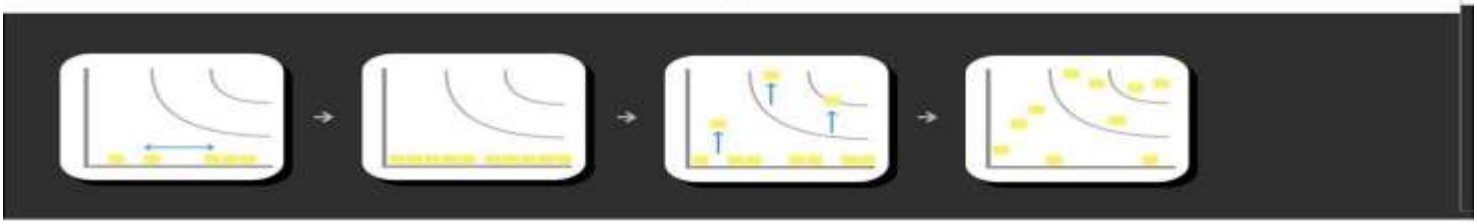
4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes





3.2 Empathy map

4 REQUIREMENT ANALYTICS

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

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and userfriendly graphical interface. Users will be ableto 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. Incase of crash, the system should be able to backup and recover the data
NFR-3	Reliability	The application has to be consistent at every scenario and has to work without failure in any environment
NFR-4	Performance	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
NFR-5	Availability	The application has to be available 24 x 7 for users without any interruption
NFR-6	Scalability	The application can withstand the increase in the no. of users and has to be able to develop Higher versions

5 PROJECT DESIGN

5.1 Proposed Solution

Project Title: Visualizing and Predicting Heart Diseases With An Interactive Dash Board

Project Design Phase-I - Solution Fit Template

Team ID: PNT2022TMID30513

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

A patient who is suffered from Heart Disease.

6. CUSTOMER CONSTRAINTS CC

The patient wants to predict the accuracy or presence of the heart disease by health monitoring devices.

5. AVAILABLE SOLUTIONS AS

Healthy lifestyle habit-changing food habit.

Explore AS, differentiate

Focus on J&A top into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

Chest pain and cardio diseases regarding chest pressure and chest discomfort (angina), breath illness.

9. PROBLEM ROOT CAUSE RC

The cause is due to age, obesity, poor diet etc.,results in heart disease includes heart attack and stroke. with earlydiagnosis and treatment You can reduce the risk of complications

7. BEHAVIOUR BE

If the patient has breathing problems the patient should consult with the doctor immediately.

Focus on J&P, top into BE, understand RC

5.2 Problem Solution Fit

3. TRIGGERS TR

Having proper awareness in health checkups. Some early symptoms of heart attack.

4. EMOTIONS: BEFORE / AFTER EM

The patient feels panic after knowing the presence of disease.

10. YOUR SOLUTION SL

With the help of data set can be analyze the next phase of severity of illness

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE

8.2 OFFLINE

Vist the doctor in proper time even after any minor attacks .

Identify strong TR & EM

6 PROJECT PLAINING

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
			• 4 GB RAM or higher • 128 GB ROM or higher ii. Android Phone (12.0 and above)			
Sprint-3		USN-12	II. Software Requirement iii. Laptop or PC • Windows 10 or higher • Android Studio	2	Medium	2
Sprint-4	Dashboard	USN-13	Query	1	High	1
		USN-14	Toll Free	1	High	1
		USN-15	Ratings	2	Medium	2
		USN-16	Verification	2	High	2
		USN-17	Validation	1	High	2
		USN-18	Feedback – send feedback to the Admin	2	Medium	3

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Member
Sprint-3		USN-9	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(Restecg) -Exercise Induced Angina(Exang) -The slope of the peak exercise ST segment -CA – Number of major vessels colored by fluoroscopy -Thal -Trest Blood Pressure -Serum Cholesterol -Maximum heart rate achieved(Thalach) -ST depression induced by exercise(Oldpeak)	2	High	5

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	2	High	1
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	2
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	4
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	3
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	2
Sprint-2	Dashboard	USN-6	Profile - view & update your profile	2	High	5
Sprint-1		USN-7	Change Password - user can change the password	1	High	2
Sprint-1		USN-8	Home - Analyze your Heart	2	High	5

6.2 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	18	06 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	11 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	19	19 Nov 2022

7. CODING & 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:

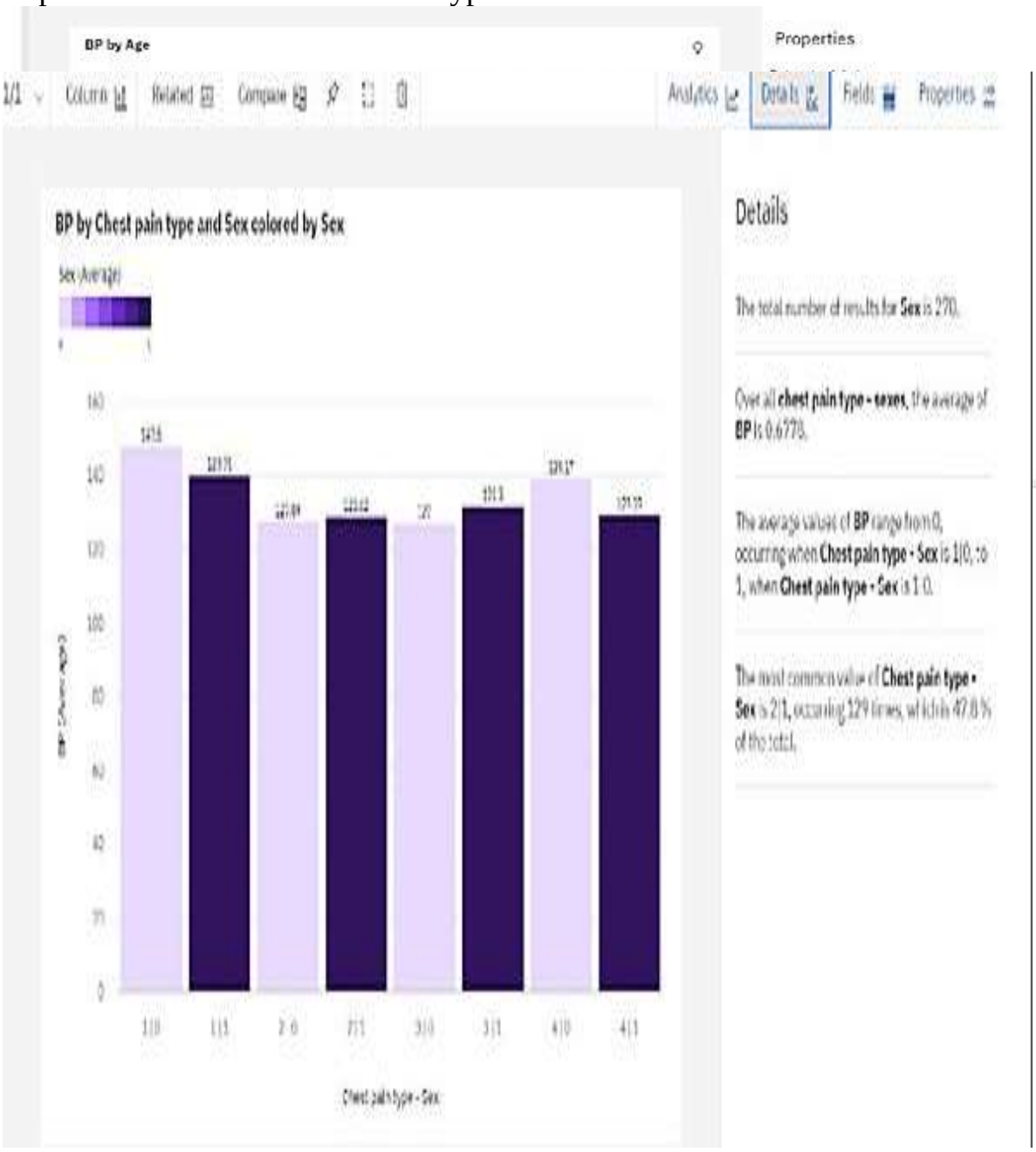
```
TP=cm[0][0] #cm=Confusion Matrix
TN=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN)))
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))
print('Testing Precision for Decision Tree:',(TP/(TP+FP)))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

```
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
```

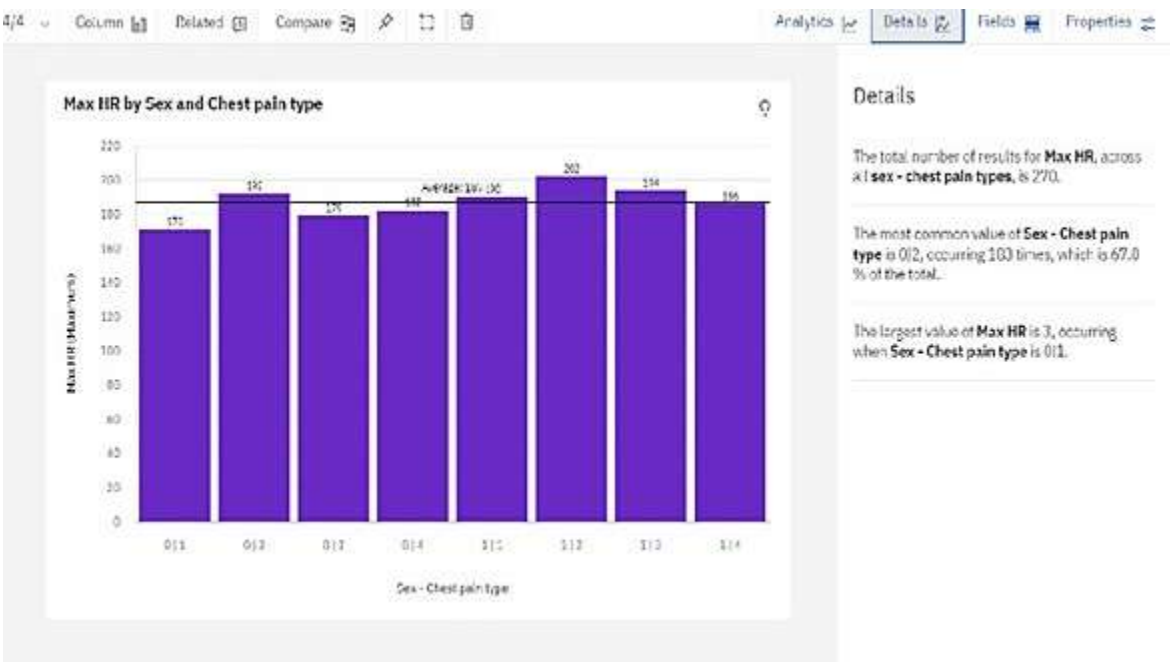
7.2 Dash Board

Average BP during chest pain

Exploration Of BPvs Chest Pain Type And Gender:



Exploration Of Max Heart Rate During The Chest Pain:



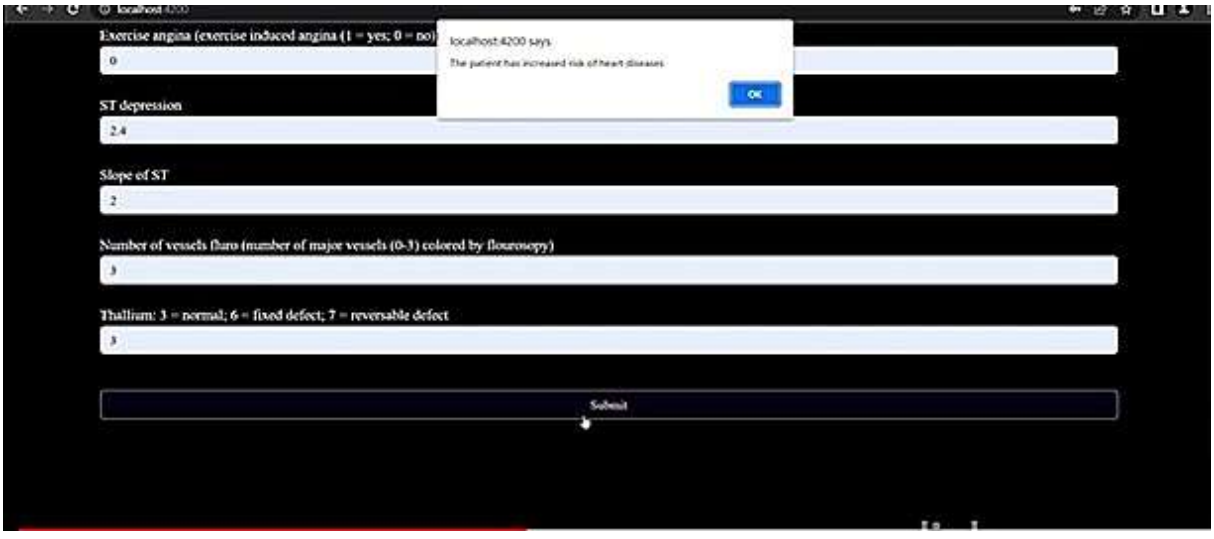
Exploration Of Cholesterol by age and Gender:

8. Testing

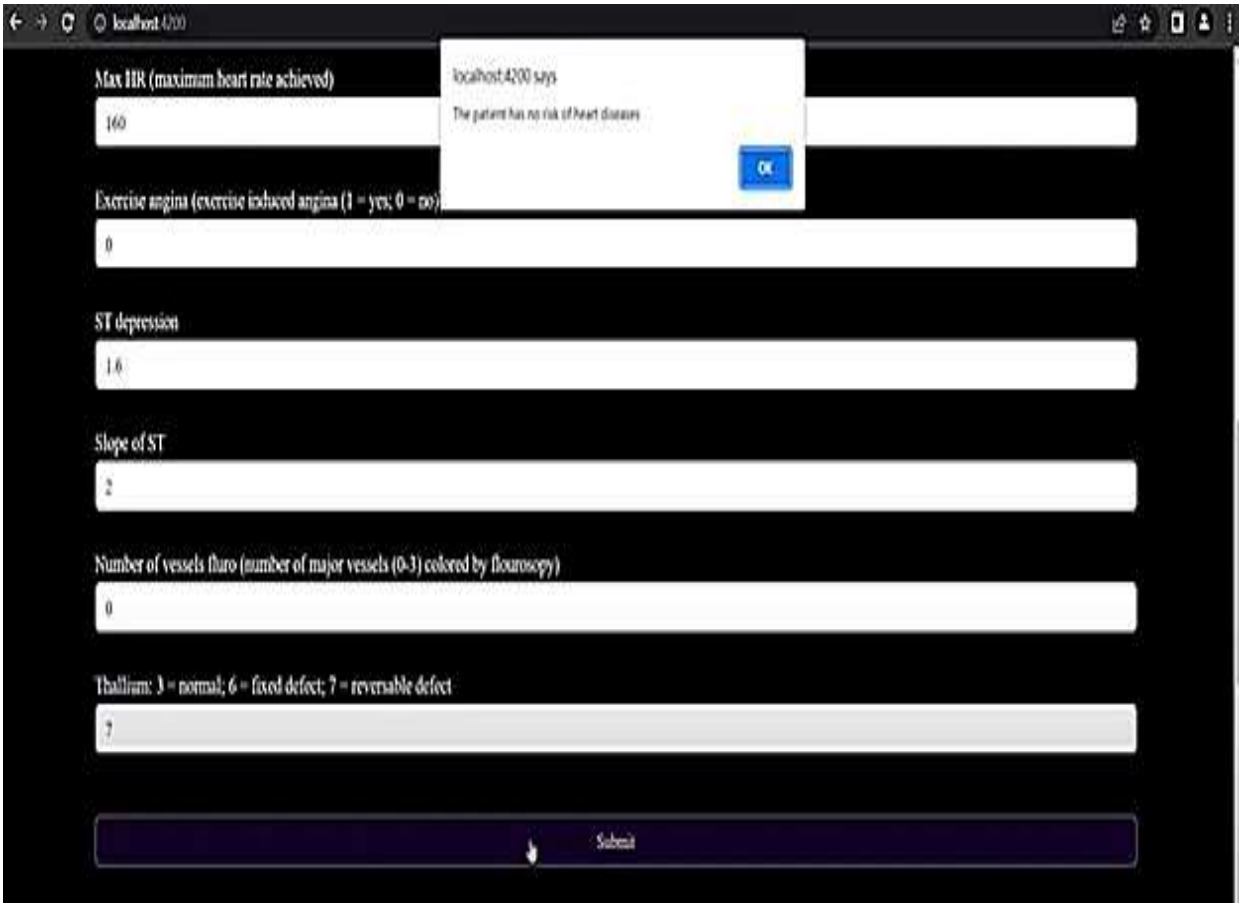
8.1 Test Cases

```
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



Testing a case where user does not have heart disease



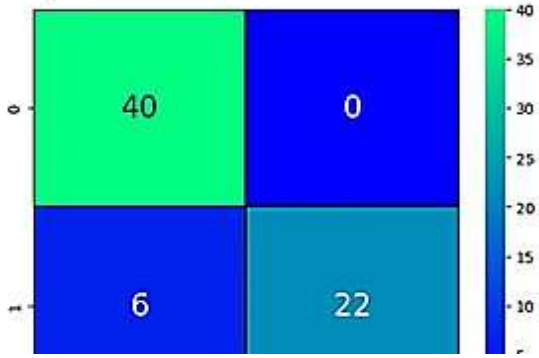
9. RESULT

9.1 Performance Metrics

The confusion matrix below shows the performance metrics

```
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
```



10. ADVANTAGES AND DISADVANTAGES

Advantages:

1. This is one of the fastest ways to determine if a person is likely to suffer from a heart disease or not.
2. Useful for medical practitioners to easily classify their patients.
3. User Friendly
4. Easy to understand
5. Secure
6. Dashboard provides insightful information

Disadvantages:

1. Needs work
2. Users need to know all the fields
3. Does Not take null value as input
4. 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 bebetter if we require less attributes and still give an effective result