VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

NALAIYA THIRAN IBM PROJECT REPORT

TEAM ID : PNT2022TMID09237

Submitted By

DILLI RANI R	**310619104029**
ANIRUDH K N	**310619104007**
KEERTHI SHREE S	**310619104056**
SURYA M	**310619104317**

in partial fullfillment for the award of the degree
of
BACHEOR OF ENGINEERING
in

Department Of Computer Science & Engineering
Easwari Engineering College
Chennai

CONTENT

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 inhigh-risk patients and in turn reduce the complications. This project aims to predict future HeartDisease by analyzing data of patients which classifies whether they have heart disease or not usingmachine-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 hasbeen developed

2.1 Existing problem

I am	Describe customer with 3-4 key characteristics - *who are they?*	Describe the customer and their attributes here
I'm trying to	*List their outcome or "job" the care about - what are they trying to achieve?*	List the thing they are trying to achieve here
but	Describe what problems or barriers stand in the way – *what bothers them most?*	Describe the problems or barriers that get in the way here
because	Enter the "root cause" of why the problem or barrier exists – *what needs to be solved?*	Describe the reason the problems or barriers exist
which makes me feel	Describe the emotions from the customer's point of view – *how does it impact them emotionally?*	Describe the emotions the result from experiencing the problems or barriers

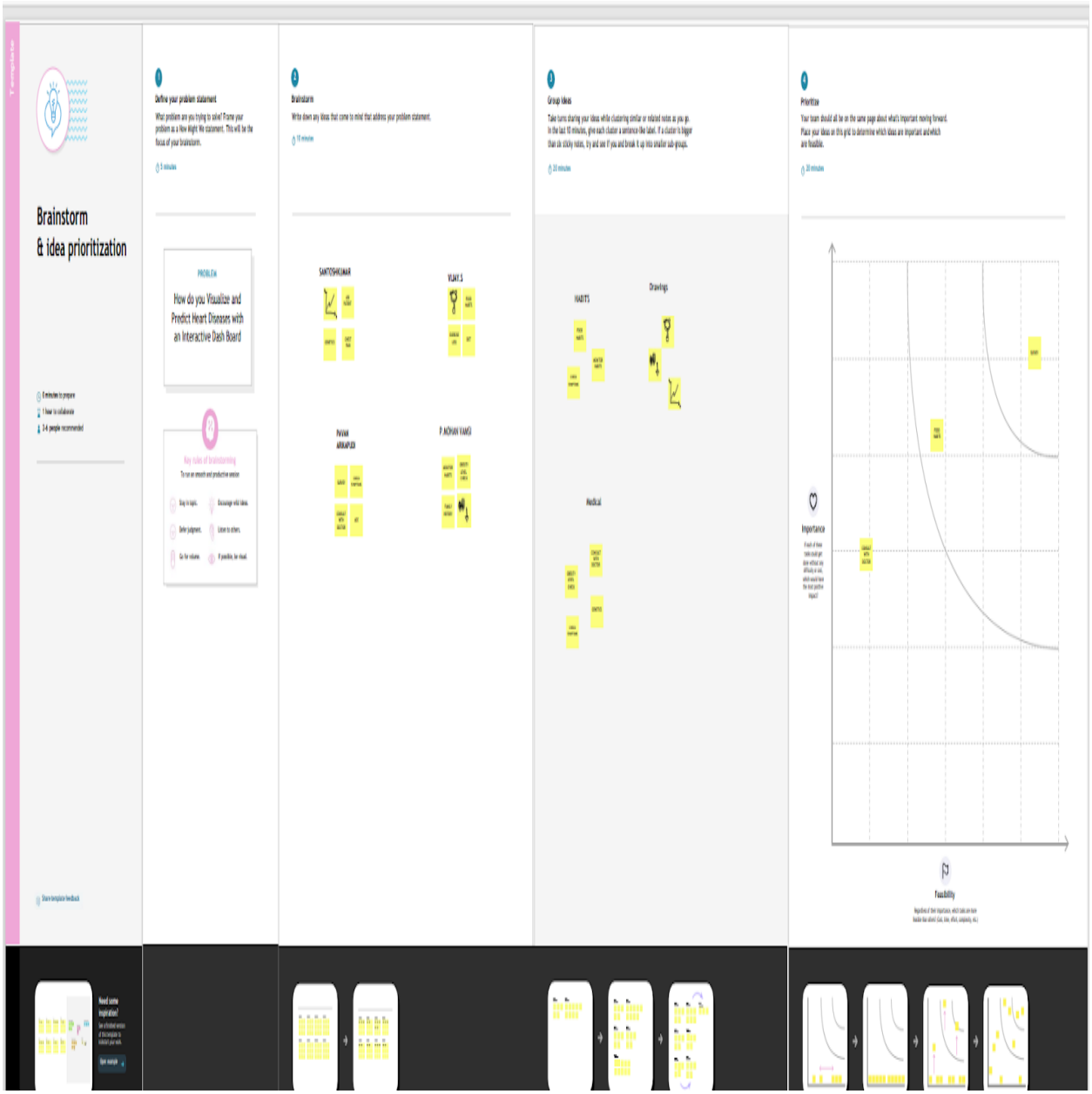
Example:

2.2 Problem statement

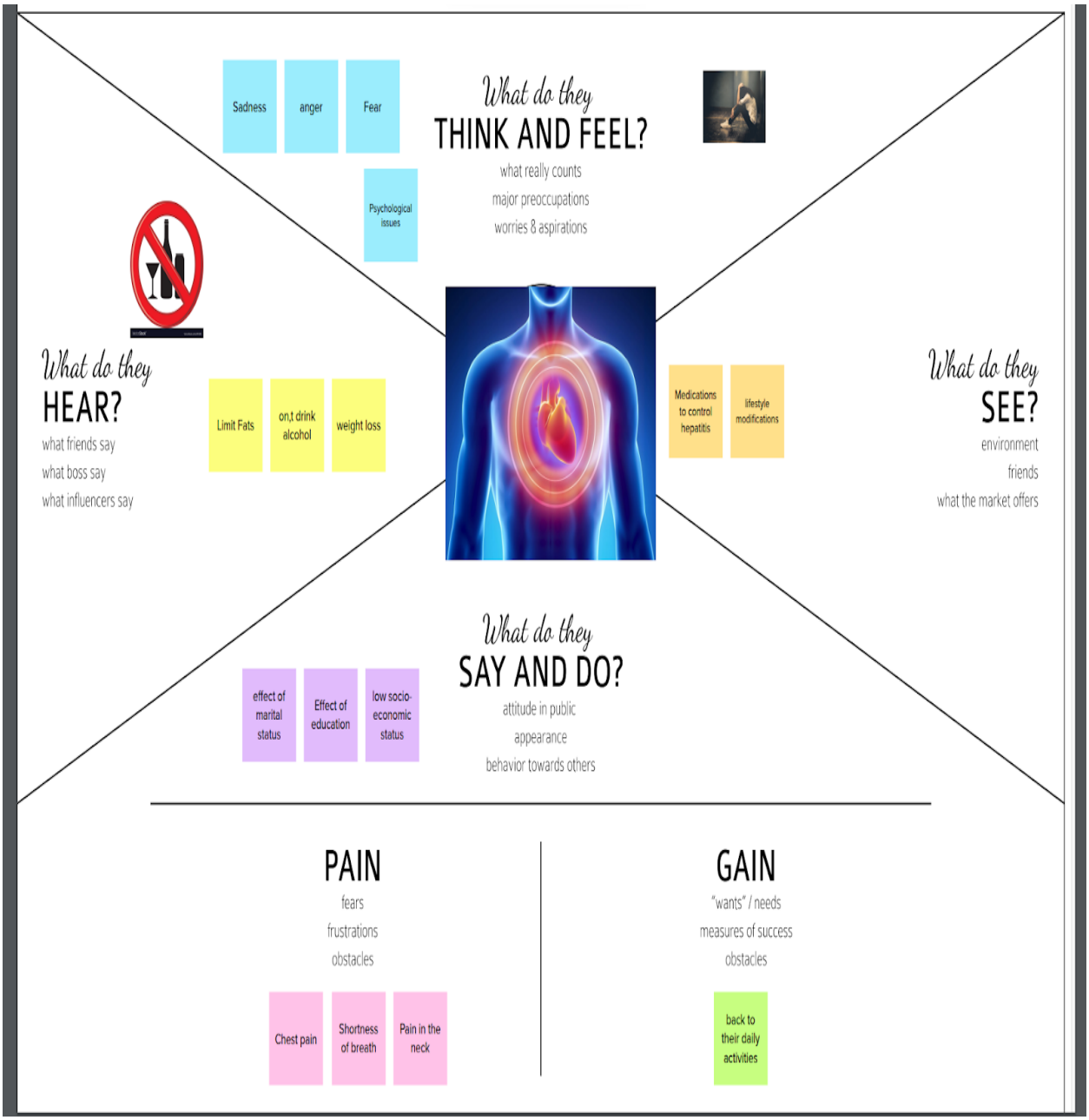
It is not possible to monitor patients every day in all cases accurately andconsultation of a patient for 24 hours by a doctor is not available since it requires more sapience,time and expertise. Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze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

3 IDEATION PHASE

3.1 Brainstorm & idea prioritization



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 user friendly 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. 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

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To find Whether a person who is having heart attack and possibilities of major and minor attacks and its proprer medication.
2.	Idea / Solution description	To accurately create a data set about the Heart Patients and to store it in cloud, so the hospitals can use this information to easily analyse and predict the patient details.
3.	Novelty / Uniqueness	Treatment can be effective and accuracy on the basis of the patient heart condition. Time and life can be saved.
4.	Social Impact / Customer Satisfaction	It will make the hospital to work efficiently and the patient can get immediate treatments.
5.	Business Model (Revenue Model)	Application can be built using low cost and minimum effort.
6.	Scalability of the Solution	Accurate prediction of the heart disease with the patient details stored.

5.2 Problem SolutionFit

Project Title: Visualizing and Predicting Heart Diseases with an Interactive Dashboard
Project Design Phase-I - Solution Fit Template
Team ID: PNT2022TMID09237

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

Who is your customer?

Doctois in hospitals

E.g.: Doctois can use this along with the patients' medical data to analyze the risk ofheait disease.

6. CUSTOMER CONSTRAINTS CC

What constraints prevent your customeis from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

Since we are dealing with sensitive medical data, it is not recommended for customeis to self-diagnose as it is very risky. It can however be used as a tool to increase awareness regarding this issue.

5. AVAILABLE SOLUTIONS AS

Which solutions are available to the customeis when they face the problem

or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an alternative to digital notetaking

Customeis can go to the doctor for a medical checkup. Based on the test results, doctois will advise them.

Explore AS, differentiate

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

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

Which jobs-to-be-done (or problems) do you address for your customeis? There could be more than one; explore different sides.

Visualizations give doctois very good insights on the potential chances for a patient to get heart disease. It is also very useful to explain to patients so that they can easily understand the risk factor and take care of themselves to reduce the likelihood of getting heart disease.

9. PROBLEM ROOT CAUSE RC

What is the real reason that this problem exists? What is the back story behind the need to do this job?

i.e. customeis have to do it because of the change in regulations.

Not storing and analyzing data properly to help doctois make informed decisions

7. BEHAVIOUR BE

What does your customer do to address the problem and get the job done?

i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customeis spend free time on volunteering work (i.e. Greenpeace)

Ensure data is stored in an organized and sequential order like an excel sheet for example right from the start so that is ready to be used for analysis.

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

3. TRIGGERS TR

What triggers customers to act? i.e., seeing their neighbor installing solar panels, reading about a more efficient solution in the news.

Patients who have a history with heart disease or those patients who are currently experiencing similar symptoms to those who have heart disease.

4. EMOTIONS: BEFORE / AFTER EM

How do customeis feel when they face a problem or a job and afterwards?

i.e. lost, insecure > confident, in control - use it in your communication strategy & design.

Feeling afraid and depressed.
Develop a feeling of awareness which mean people

10. YOUR SOLUTION SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.

If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior.

To clean data and provide visualizations to help doctois in their diagnosis of patient as well as make customeis more aware of this issue.

8. CHANNELS of BEHAVIOUR CH

8.1 **ONLINE**

What kind of actions do customeis take online? Extract online channels from #

8.2 **OFFLINE**

What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

ONLINE: Useis look at the data and compare it with their test results

OFFLINE: Doctois use it as a tool to diagnose patients and make accurate predictions.

6. PROJECT PLAINING

6.1 Sprint Planning & Estimation

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

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
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
		USN-10	View Doctors - view doctor detail by searching by names or filter by specialty	1	Medium	4
Sprint-3	System Requirment	USN-11	I. Hardware Requirement i. Laptop or PC • I5 processor system or higher	2	High	2

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

6.2. Sprint Delivery Schedule

Project Tracker, Velocity

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

Velocity:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

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]

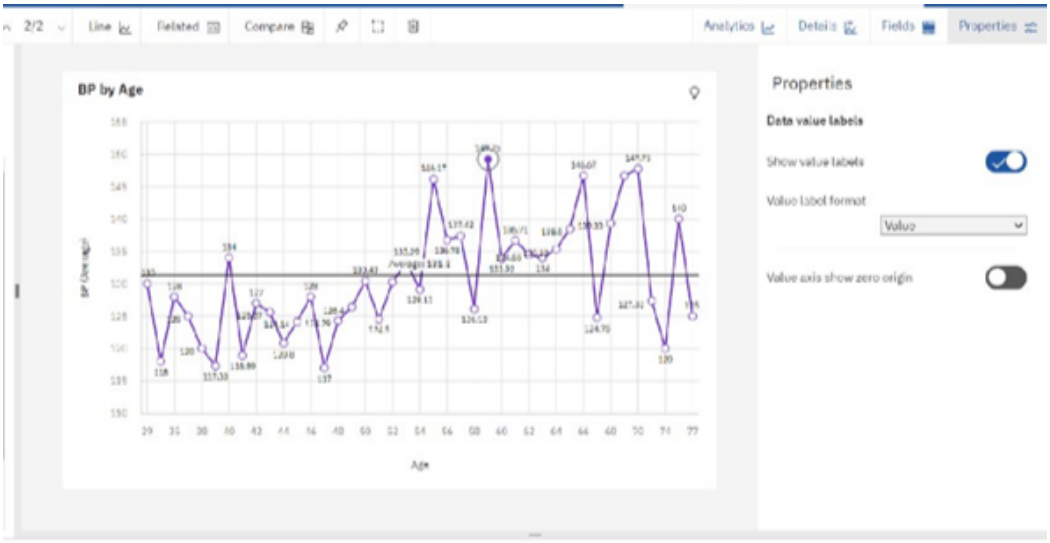
FP=cm[0][1]

```
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)))
```

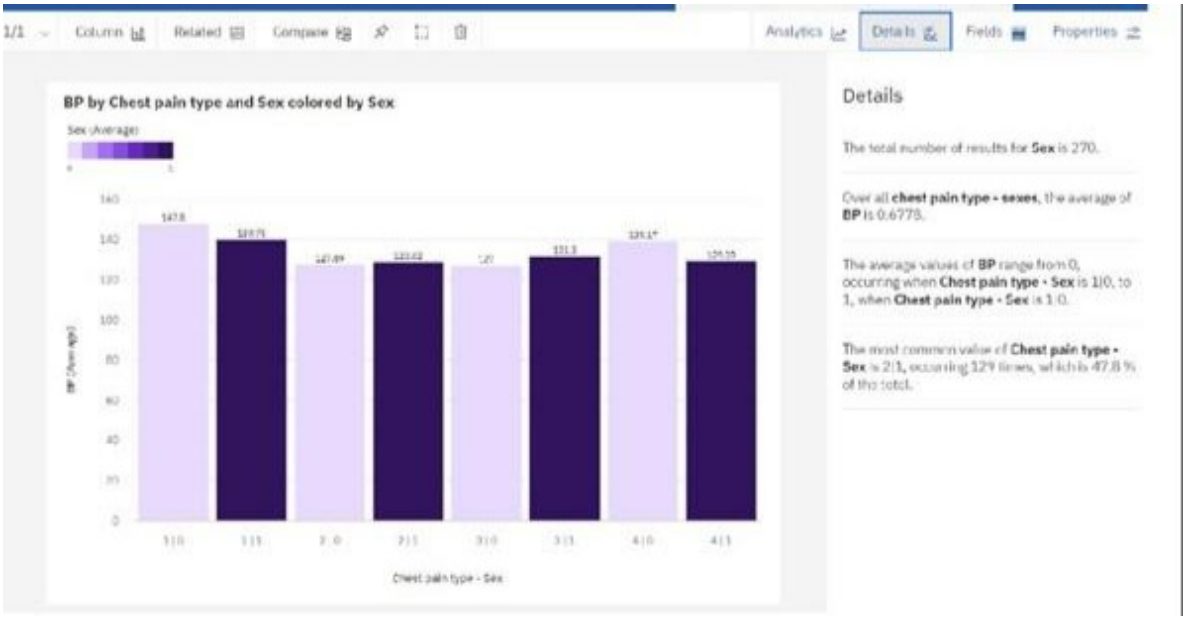
```
Testing Accuracy for Decision Tree: 0.9264705882352942
Testing Sensitivity for Decision Tree: 0.8888888888888888
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```

7.2 DashBoard

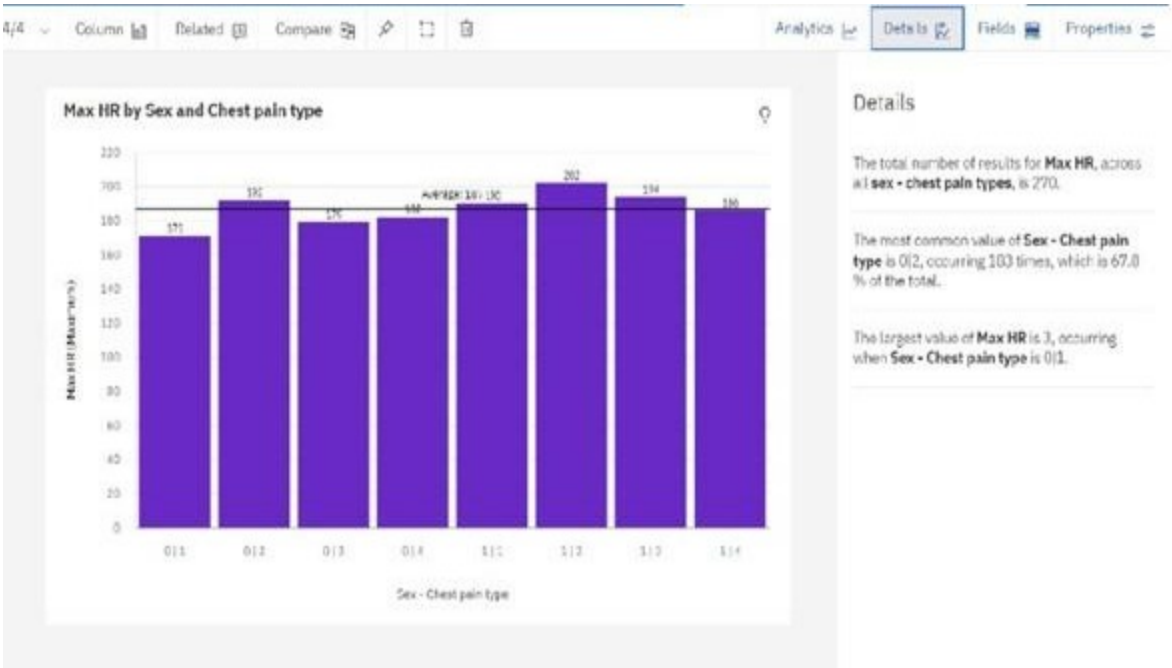
Average BP during chest pain



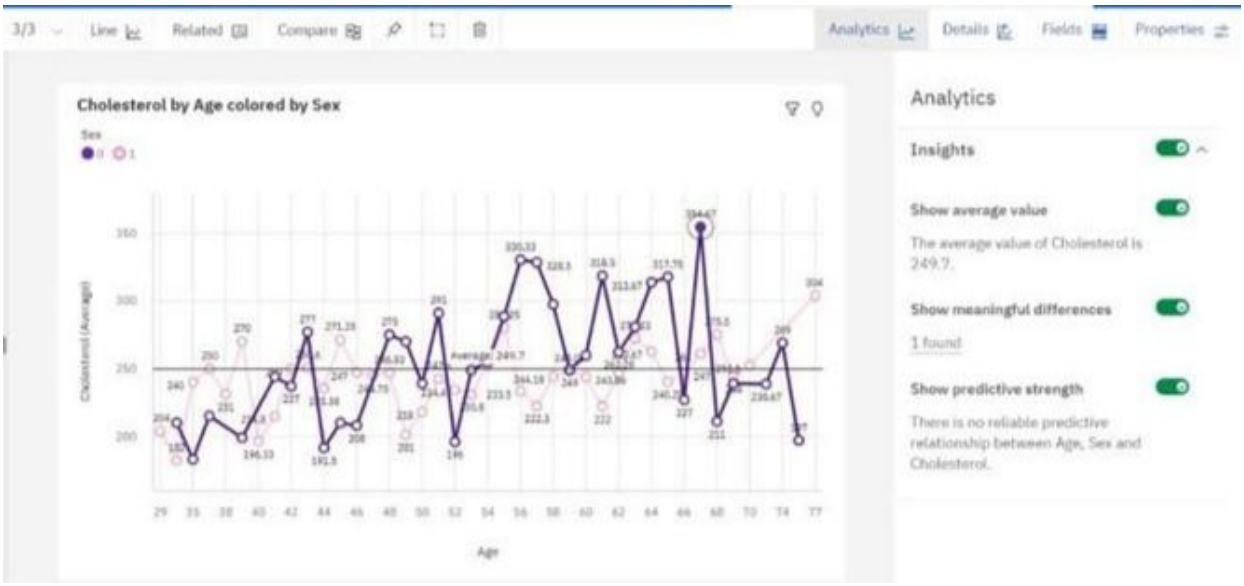
Exploration Of BPvsChestPainType And Gender:



Exploration Of Max Heart Rate During The Chest Pain:



Exploration Of Cholesterol by age and Gender:

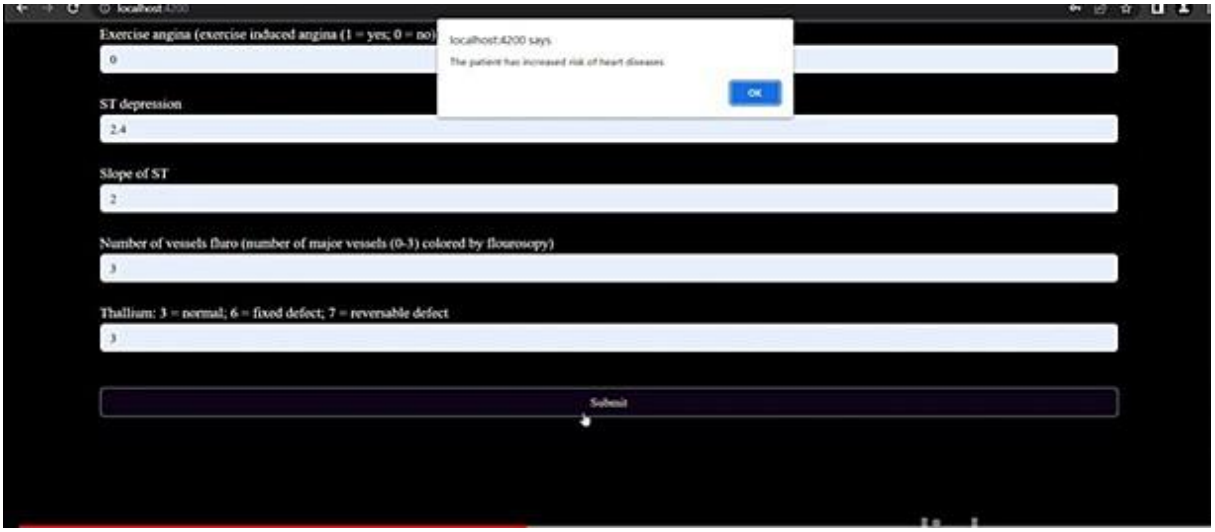


8. Testing

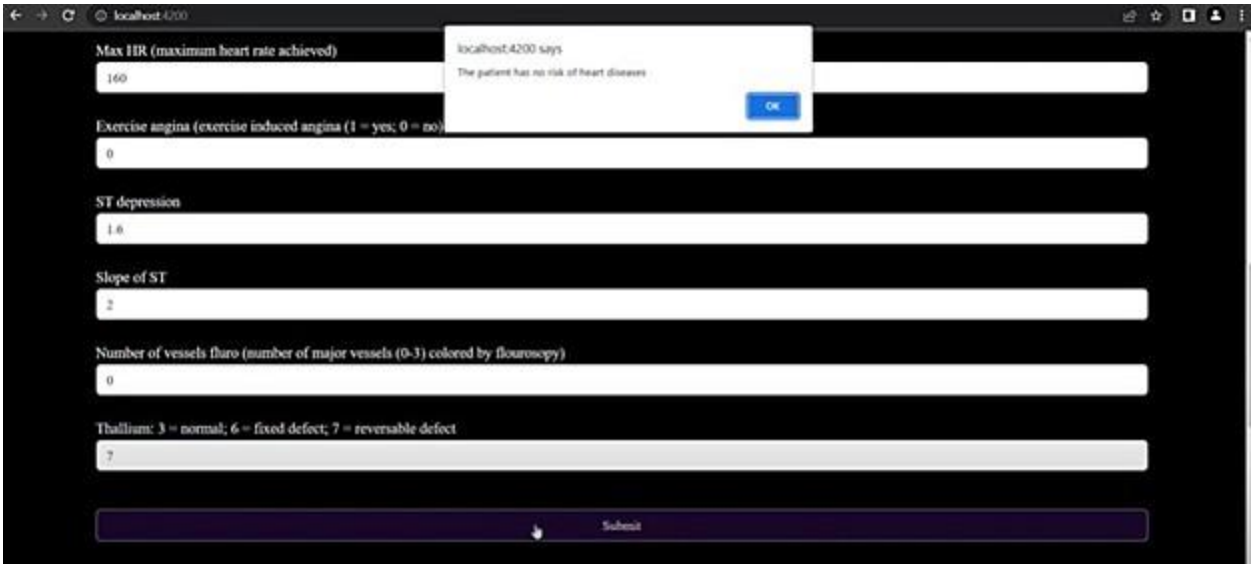
8.1 Test Cases

```
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



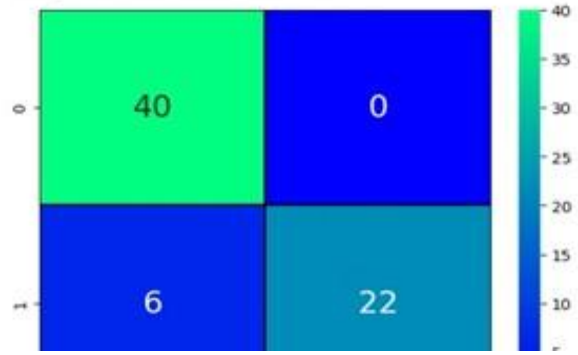
9. Result

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 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-35162-1660282022