

"VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASH BOARD"

NALAIYA THIRAN IBM PROJECT REPORT

Team Id : PNT2022TMID07543

Submitted by

*Akarsana.S
Gowripriya.R
Keerthana.M
Kowsikamani.T*

**721719106004
721719106015
721719106024
721719106028**

**UNDER THE GUIDANCE OF
Vishnu vardhan**

P.A college of engineering and technology

**Department of electronics and communication
Engineering**

CONTENT

1 INTRODUCTION

- 1.1 Project overview
- 1.2 Purpose

2 LITERATURE SURVEY

- 2.1 Existing problem
- 2.2 Problem statement

3 IDEATION PHASE

- 3.1 Brainstorm & idea prioritization
- 3.3 Empathy Map

4 REQUIREMENT ANALYTICS

- 4.1 Functional requirement
- 4.2 Non-Functional requirement

5 PROJECT DESIGN PHASE -I

- 5.1 Proposed Solution
- 5.2 Problem Solution Fit
- 5.3 Solution Architecture

6 PROJECT PLAINING

- 6.1 Sprint Planning & Estimation
- 6.2 Sprint Delivery Schedule

7. CODING & SOLUTIONING

- 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

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 their 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 they 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

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

3.1Brainstorm & idea prioritization

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

👤 2-4 people recommended

➔

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

A

Team gathering

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

B

Set the goal

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

C

Learn how to use the facilitation tools

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

Open article →

1

Define your problem statement

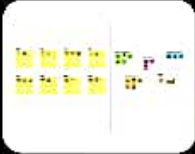
What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

🕒 5 minutes

PROBLEM

The leading cause of deaths in the developed world is heart disease.Prevention of heart diseases before death is necessary in this era.

📄 Share template feedback



Need some inspiration?

Some brainstorming of the responses to inspire your work.

Open examples →

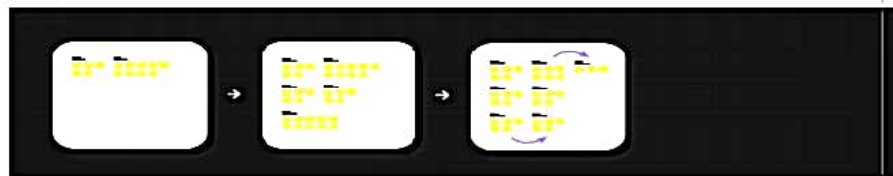
3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into another subgroup.

⌚ 20 minutes

Data and Research Training and Testing Work and Deliverables

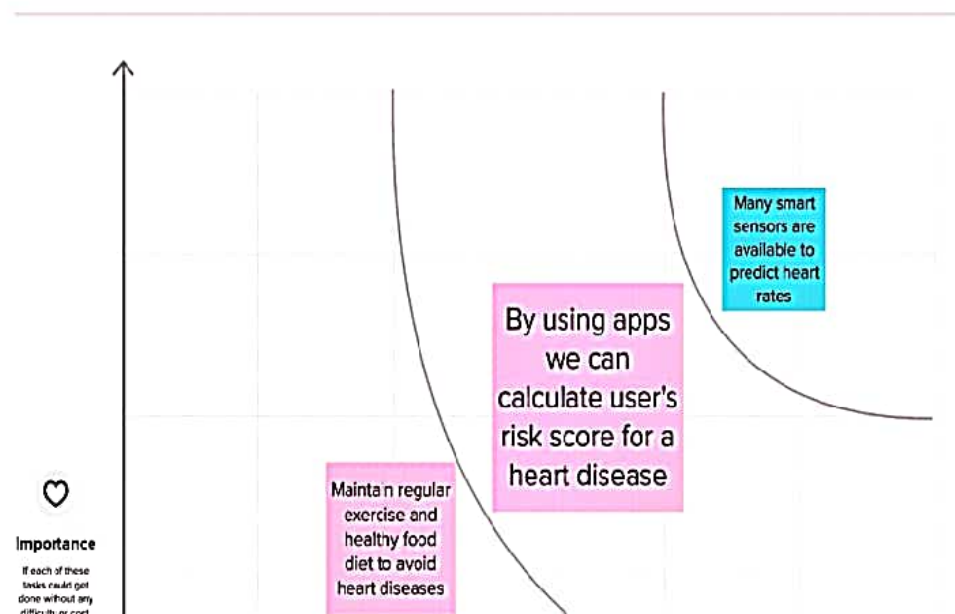


4

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



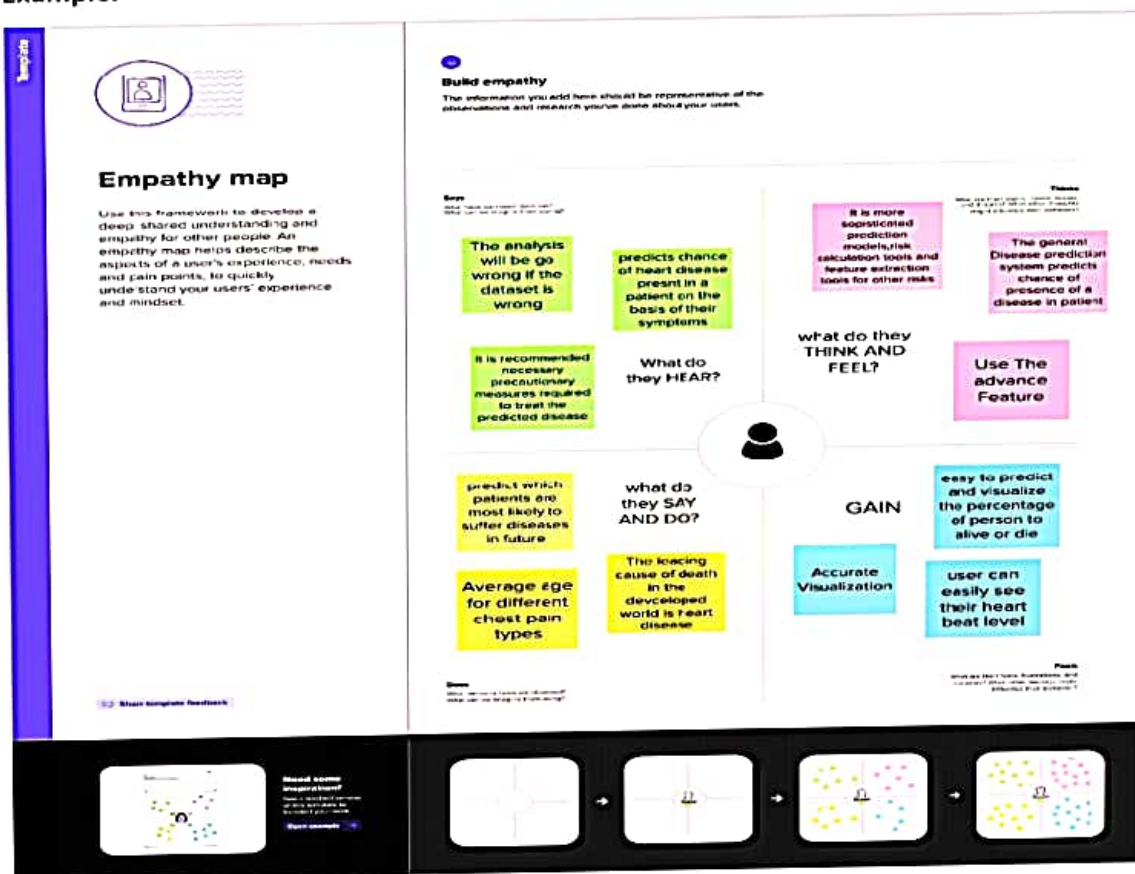
3.2 Empathy map

Empathy Map Canvas:

It is a useful tool to help teams better understand their users.

Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.

Example:



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 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. In case 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 proper 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.

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? i.e. working parents (30-50 yrs. kids) A patient who is suffered from Heart Disease.	6. CUSTOMER CONSTRAINTS CC What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices. The patient wants to predict the accuracy or presence of the heart disease by health monitoring devices.	5. AVAILABLE SOLUTIONS AS Which solutions are available to the customers 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 monitoring. Healthy lifestyle habit-changing food habit.	Explore AS, differentiate
Focus on J&P, tap into BE, understand	2. JOBS-TO-BE-DONE / PROBLEMS J&P What jobs to be done (or problems) do you address for your customers? There could be more than one, explore different issues. Chest pain and cardio diseases regarding chest pressure and chest discomfort (angina), breath illness.	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. Customers have to do it because of the change in regulations. The cause is due to age, obesity, poor diet etc., results in heart disease includes heart attack and stroke. with early diagnosis and treatment You can reduce the risk of complications	7. BEHAVIOUR BE What does your customer do to address the problem and get the job done? i.e. already realised, find the right doctor, given feedback, provide usage and timeline, indirectly associated, customers spend how time in waiting, waiting work i.e. symptoms. if the patient has breathing problems the patient should consult with the doctor immediately.	Focus on J&P, tap into BE, understand
Identify strong TR & EM	3. TRIGGERS TR What triggers customers to act? i.e. seeing their neighbours installing solar panels, reading about a more efficient solution in the news. Having proper awareness in health checkups. Some early symptoms of heart attack. 4. EMOTIONS: BEFORE / AFTER EM How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure > confident, at control, safe & in flow, conversational strategy & design. The patient feels panic after knowing the presence of disease.	10. YOUR SOLUTION SU If you are working on an existing business, refine down your current solution first fit 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 entire canvas and come up with a solution that fits in that customer landscape, solves a problem and matches customer behaviour. With the help of data set can be analyze the next phase of severity of illness	8. CHANNELS of BEHAVIOUR CH 8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7. 8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development. Visit 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
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	5	High	P.Divya
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High	P.Divya
Sprint-2		USN-3	As a user, I can register for the application through Facebook	2	Low	P.Divya
Sprint-1		USN-4	As a user, I can register for the application through Google	2	Medium	R.Ponmalar
Sprint-1	Login	USN-5	As a user, I can log into the application by entering email & password	3	High	R.Ponmalar

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-2	User entry	USN-7	As a User, I can enter my personal details for analysis	3	High	R.Ponmalar
Sprint-2		USN-8	As a User, I can entry my medical records & symptoms	3	High	R.Tamilarasi
Sprint-3	User profile	USN-9	As a user, I can update the health details of users.	5	High	R.Tamilarasi
Sprint-3	Helpdesk	USN-10	As a user, I can post my queries & view the frequently asked question (FAQ)	5	High	R.Tamilarasi
Sprint-3		USN-11	As an admin, I can view the user queries	3	High	R.Sowmiya

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-4		USN-14	As an admin, I can add or delete users.	3	High	R.Sowmiya
Sprint-4		USN-15	As an admin, I can manage the user details.	3	High	R.Sowmiya

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	13	6 Days	24 Oct 2022	29 Oct 2022	13	29 Oct 2022
Sprint-2	13	6 Days	31 Oct 2022	05 Nov 2022	13	05 Nov 2022
Sprint-3	13	6 Days	07 Nov 2022	12 Nov 2022	13	12 Nov 2022
Sprint-4	13	6 Days	14 Nov 2022	19 Nov 2022	13	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)))

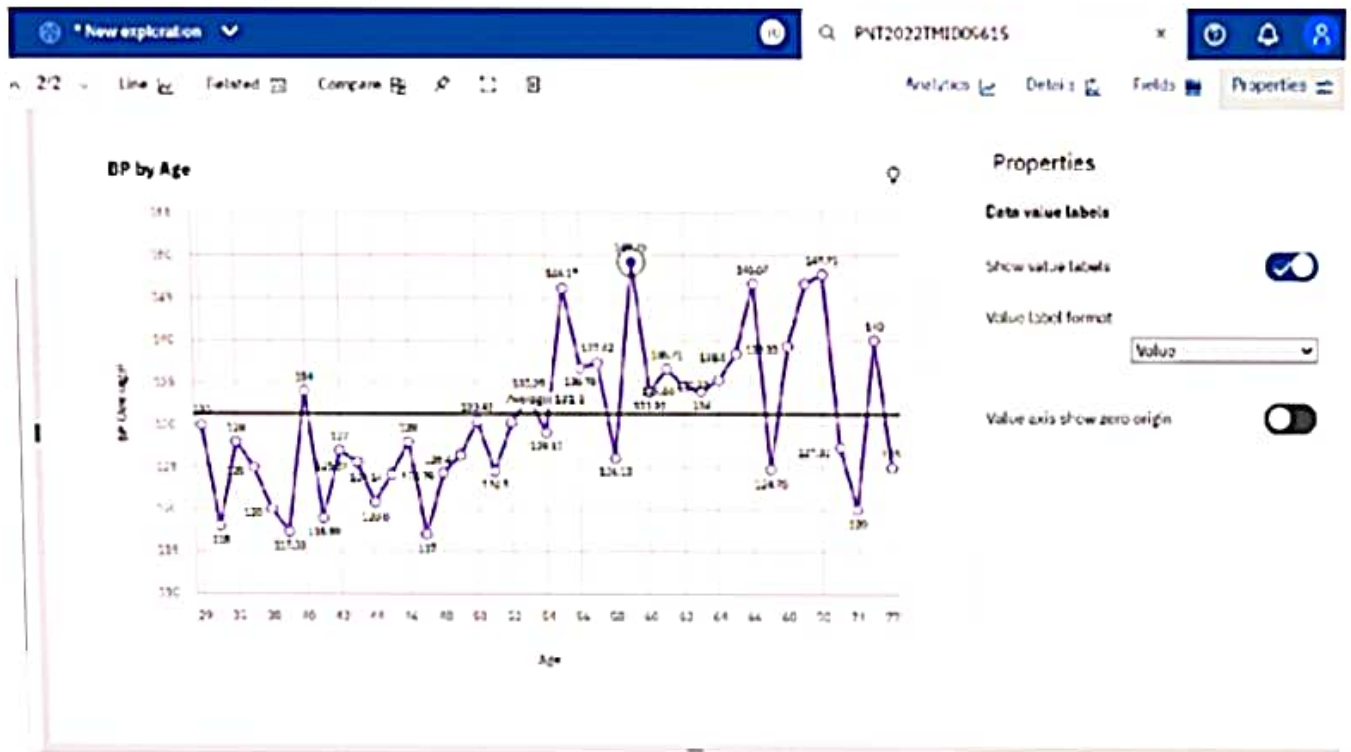
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))

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



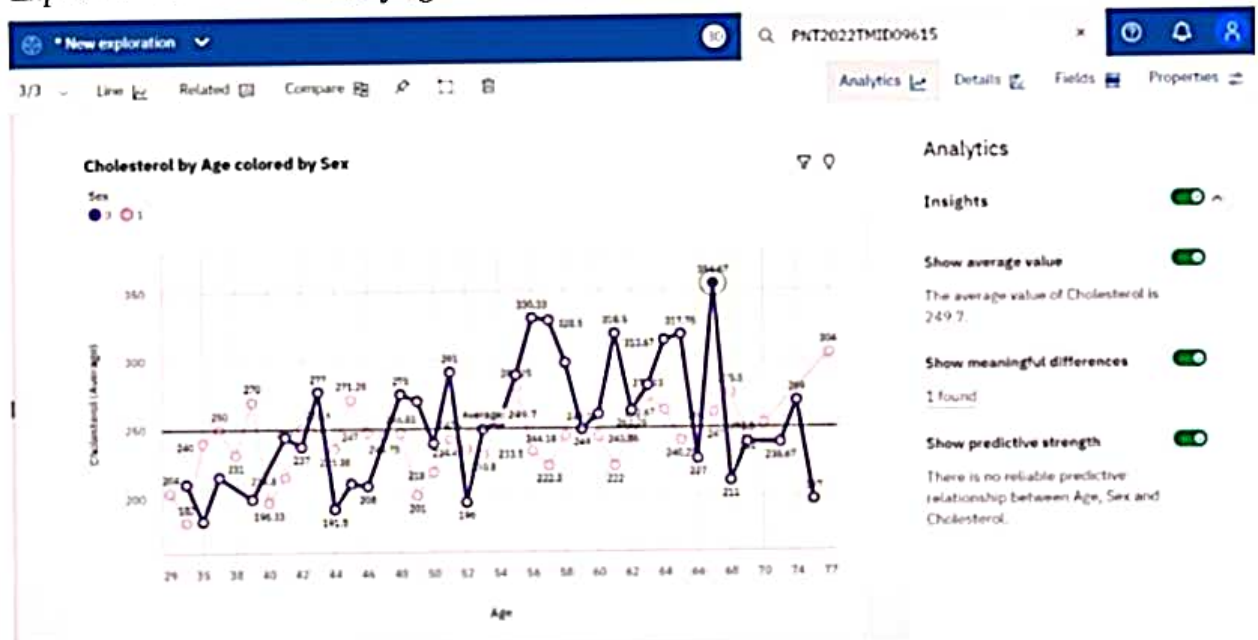
Exploration Of BPvsChestPainType And Gender:



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)
input_as_numpy=np.asarray(input)
input_resshaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resshaped)
print(pre1)
a1 = accuracy_score(pre1,model1.predict(input_resshaped)) * 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_resshaped=input_as_numpy.reshape(1,-1)
pre1=tree_model.predict(input_resshaped)
print(pre1)
a1 = accuracy_score(pre1,model1.predict(input_resshaped)) * 100
print(a1)

['Presence']
100.0
```

8.2 User acceptance Testing

The screenshot shows a web application interface for heart disease risk assessment. It features several input fields for user data, a 'Submit' button, and a modal dialog box.

Form Fields:

- Exercise angina (exercise induced angina (1 = yes, 0 = no)): 0
- ST depression: 2.4
- Slope of ST: 2
- Number of vessels fluo (number of major vessels (0-3) colored by fluoroscopy): 3
- Thallium: 3 (normal, 6 = fixed defect, 7 = reversible defect)

Submit Button: A button labeled 'Submit' is located at the bottom of the form.

Modal Dialog: A modal dialog box is open, displaying the text: "localhost:4200 says: The patient has increased risk of heart disease." with an 'OK' button.

Testing a case where user does not have heart disease

localhost:4200 says
The patient has no risk of heart disease

Max HR (maximum heart rate achieved)
160

Exercise angina (exercise induced angina (1 = yes, 0 = no))
0

ST depression
1.6

Slope of ST
2

Number of vessels fluro (number of major vessels (0-3) colored by flourosopy)
0

Thallium: 3 = normal, 6 = fixed defect, 7 = reversible defect
7

Submit

9. Result

9.1 Performance Metrics

The confusion matrix below shows the performance metrics

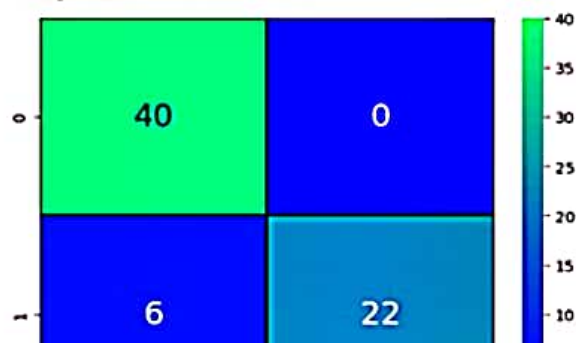
```
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

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:

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 informations

Disadvantages:

7. Needs work
8. Users need to know all the fields
9. Does Not take null value as input
10. 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-2202-1658465973/tree/main/Final%20deliverable>