VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

NALAIYA THIRAN PROJECT REPORT

Submitted By

TEAM LEAD

VANDHANA.V **510119205017**

TEAM MEMBERS

AARTHI.S	**510119205002**
MOHANAPRIYA.S	**510119205009**
SWETHA.S	**510119205015**

TEAM ID: PNT2022TMID39449

CONTENT

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

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 technique for 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 improves care by facilitating preventive care and EDA is a vital step while analysing data.

2.2 References

"Predicting The Risk Of Heart Failure With EHR Sequential Data Modelling" Bo Jin, Chao Che et al (2018).

Bo Jin, Chao Che et al. (2018) proposed a “Predicting the Risk of Heart Failure With EHR Sequential Data Modeling” model designed by applying neural network. This paper used the electronic health record (EHR) data from

real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the sequential nature of clinical records.

"Prediction And Diagnosis Of Heart Disease By Data Mining Techniques" Boshra Bahrami, Mirsaeid Hosseini Shirvani(2015).

Prediction and Diagnosis of Heart Disease by Data Mining Techniques" designed by Boshra Bahrami, Mirsaeid Hosseini Shirvani. This paper uses various classification methodology for diagnosing cardiovascular disease. Classifiers like KNN, SVO classifier and Decision Tree are used to divide the datasets. Once the classification and performance evaluation the Decision tree is examined as the best one for cardiovascular disease prediction from the dataset.

"An Intelligent Decision Support System For Cardiac Disease Detection" Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik(2015).

Lokanath Sarangi, Mihir Narayan Mohanty, Srikanta Pattnaik (2015) "An Intelligent Decision Support System for Cardiac Disease Detection", designed a cost efficient model by using genetic algorithm optimizer technique. The weights were optimized and fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of GA and neural networks.

"Fast Rule-Based Heart Disease Prediction Using Associative Classification Mining"

K.Prasanna Lakshmi, Dr. C.R.K.Reddy(2015).

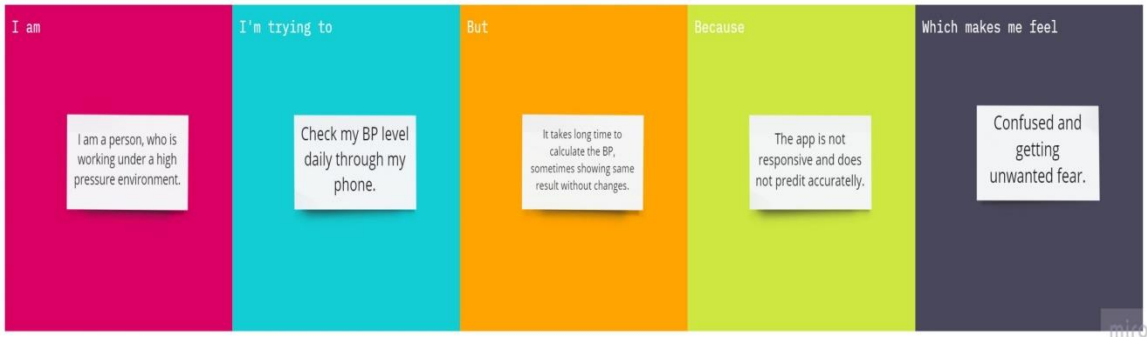
K.Prasanna Lakshmi, Dr. C.R.K.Reddy (2015) designed "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining". In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-

square testing and arranging the rules in an order to form a classifier. Using these phase to predict the heart disease easily.

"Applied Data Mining And Machine Learning Algorithms Namely Decision Tree (J48 algorithm), Naive Bayes And Artificial Neural Networks(ANN) For Heart Disease Prediction" A.Taneja(2013).

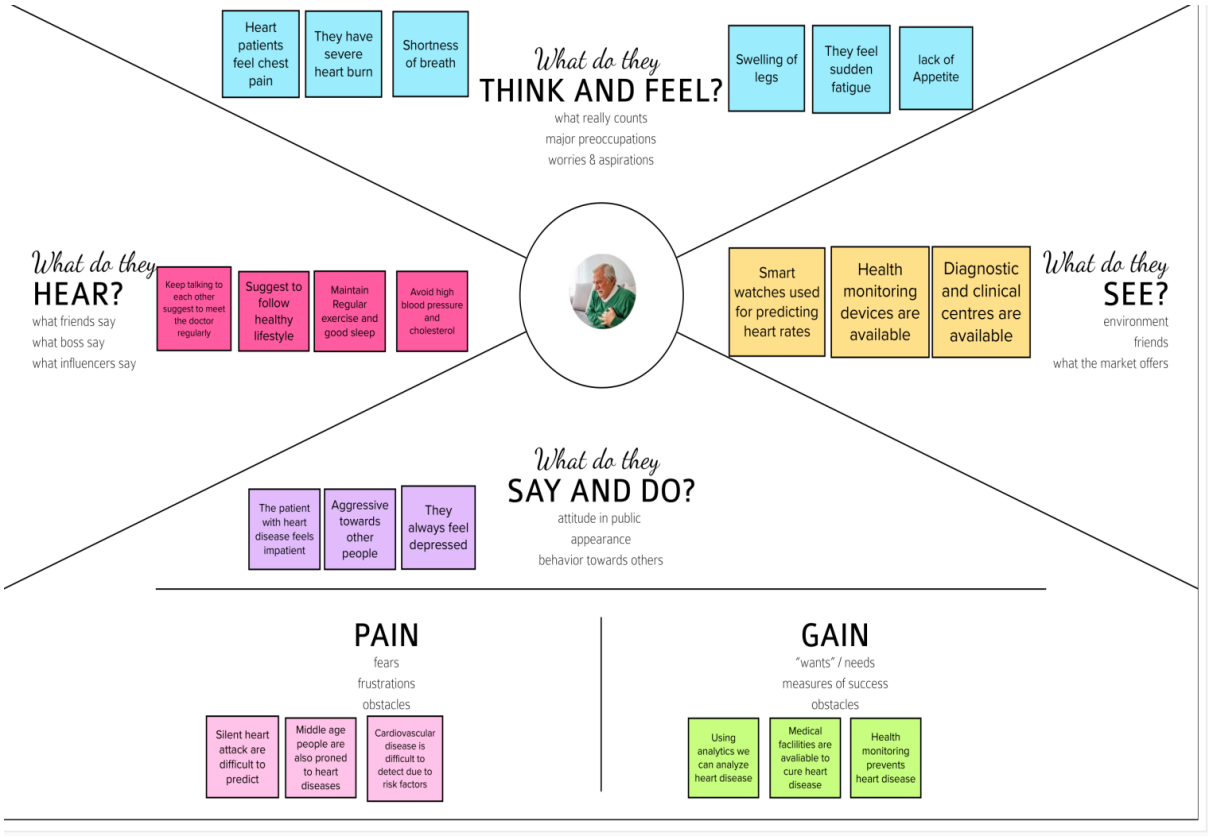
In 2013, A. Taneja, applied data mining and machine learning algorithms namely Decision Tree (J48 algorithm), Naive Bayes and Artificial Neural Networks (ANN) for heart disease prediction. A dataset of 7339 instance with 15 attributes has been taken from PGI Chandigarh. WEKA 3.6.4 tool was used for the experiment. For model training and testing 10-Fold Cross Validation techniques is used randomly. Best First Search method was used to select the best attributes from the already available 15 attributes and among them only 8 attributes has been selected. Each experiments was done on two different scenarios, first one containing all 15 attributes and the second case only 8 selected attributes. From all these experiments comparative results has been obtained and from these comparative results it has been found that J48 pruned in selected attributes case has performed best in accuracy with 95.56% and Naive Bayes with all attributes case gives less accuracy 91.96% but takes least time to build a model in the whole experiment.

2.3 Problem Statement Definition



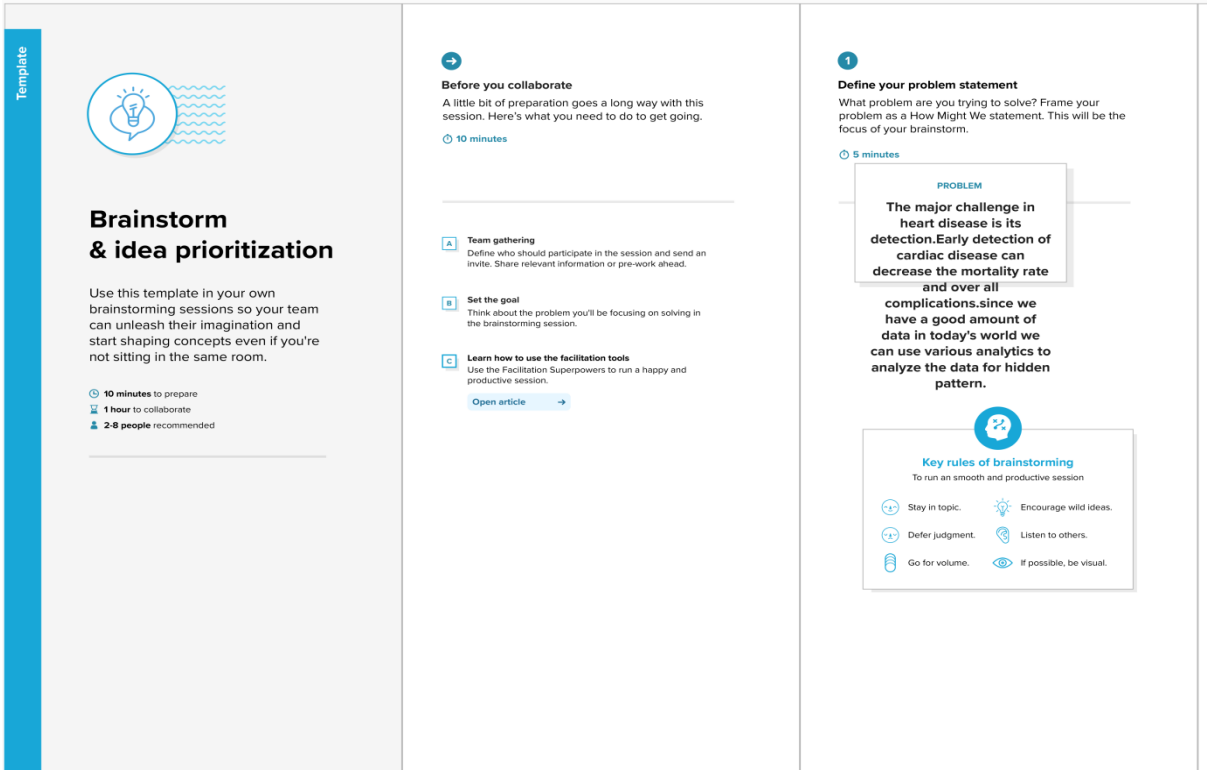
3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas

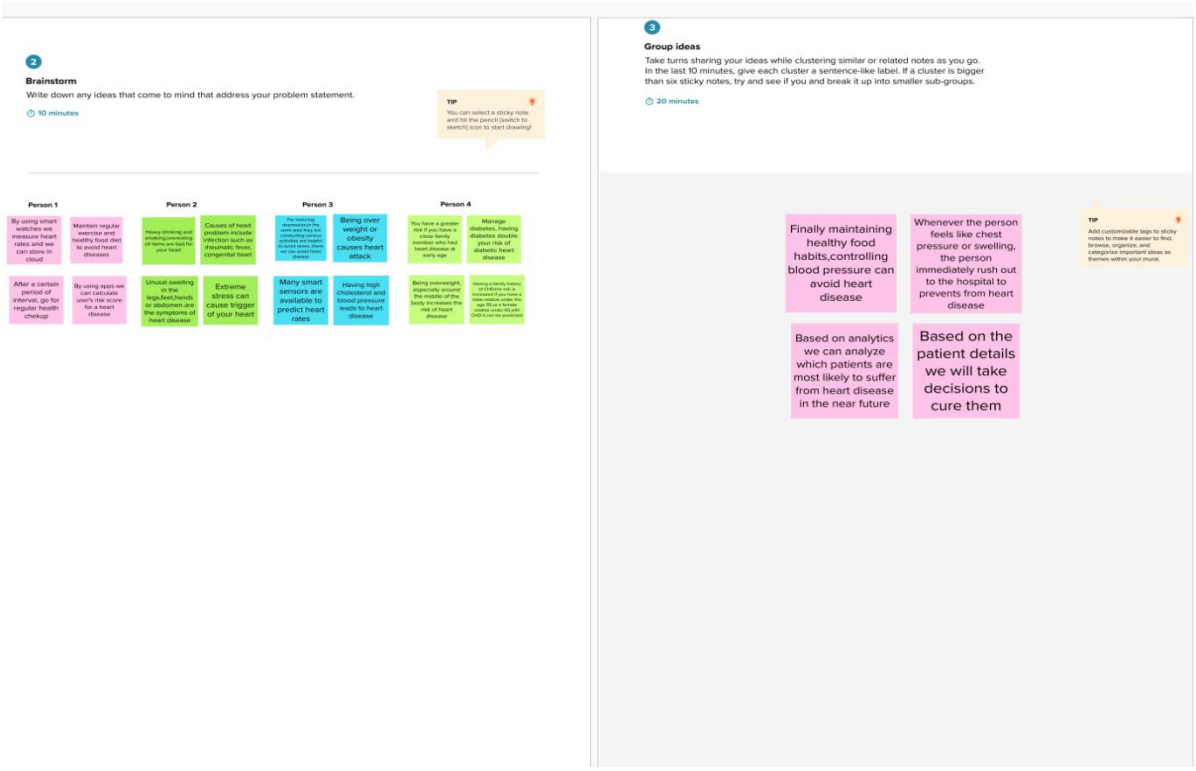


3.2 Ideation and Brainstorming

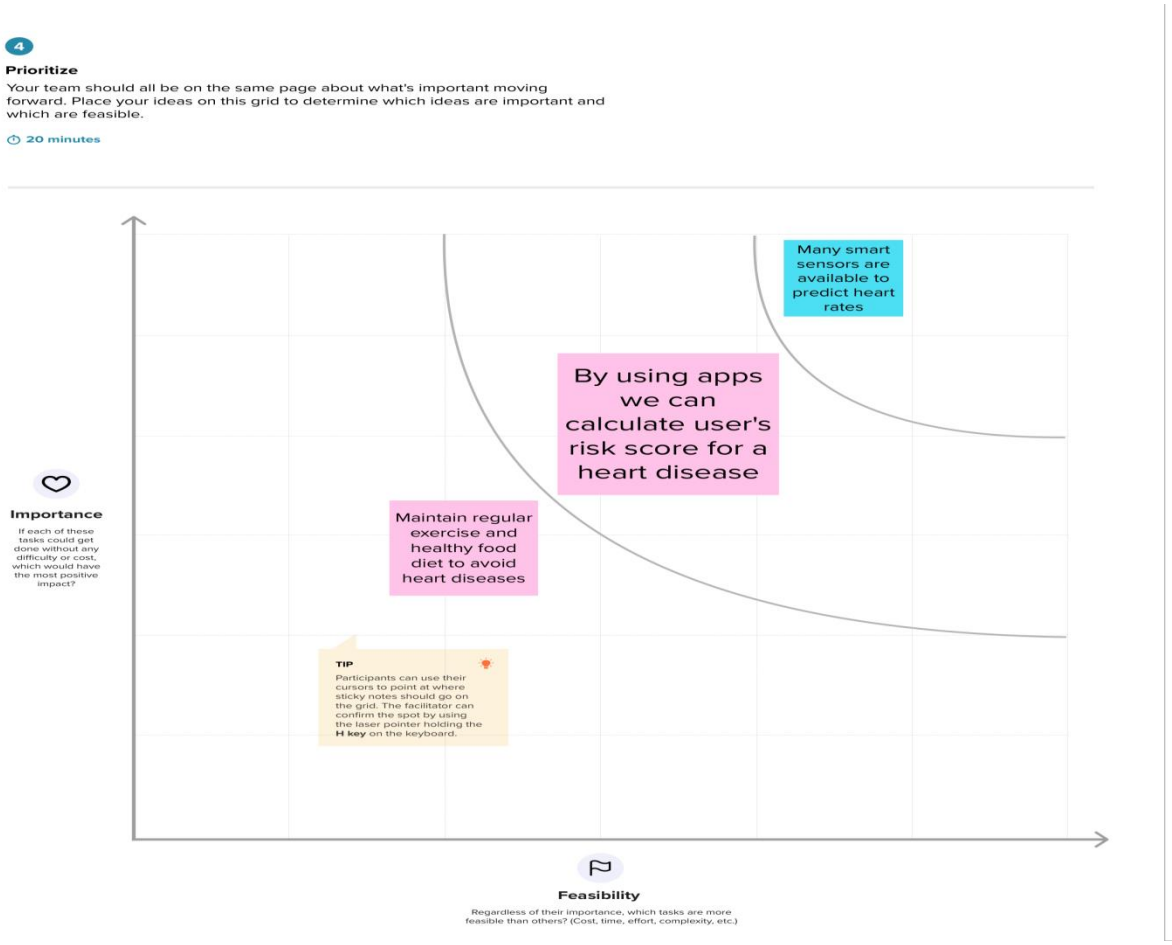
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step 2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization

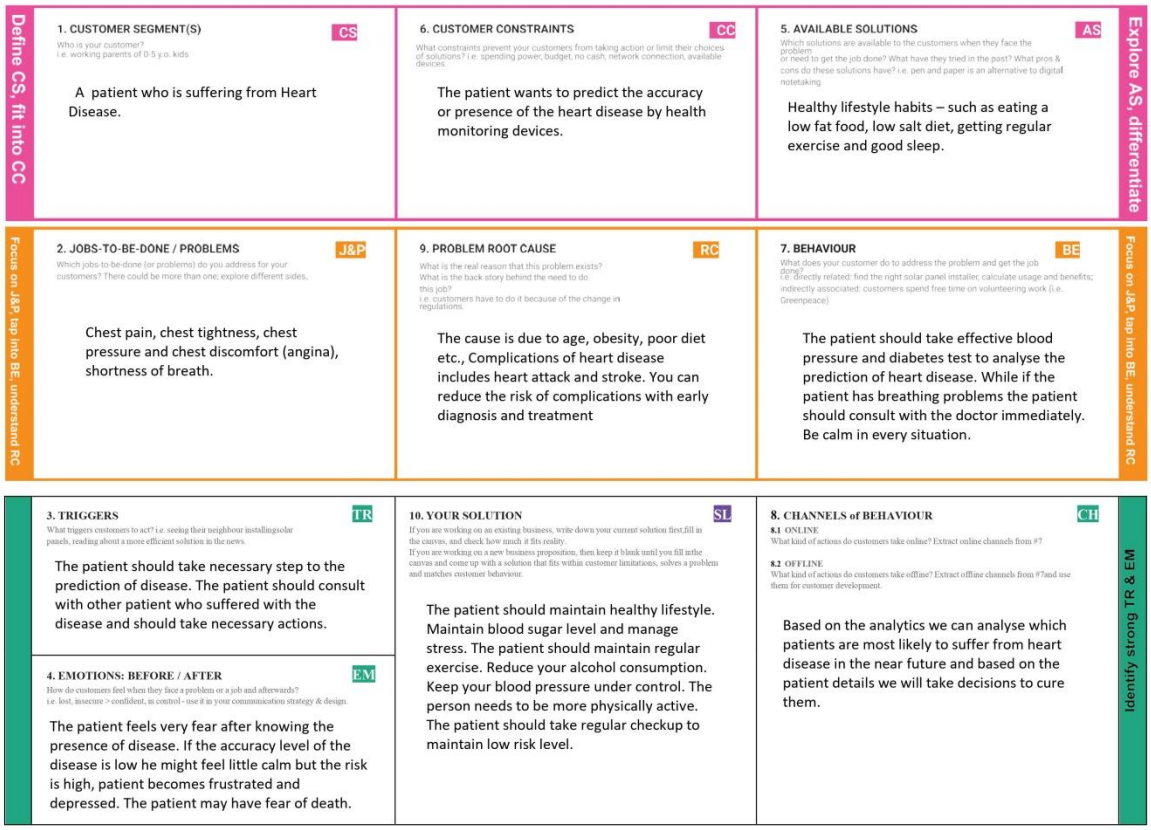


3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To find Whether a person who is working under a high pressure environment is suffered from a heart disease and also to predict effectively. If the patient suffers from heart disease or not and to determine whether the patient should be diagnosed with heart Disease or not.
2.	Idea / Solution description	The goal is 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. Based on that a

		informative and creative dashboard can be created to present the data and utilize it for future use.
3.	Novelty / Uniqueness	Treatment can be easy for the doctors on the basis of the patient heart condition. Time can be saved.
4.	Social Impact / Customer Satisfaction	By predicting, it helps the hospitals to know the health records of the heart patient. It will make the hospital to work efficiently and the patient can get immediate treatments.
5.	Business Model (Revenue Model)	Ad based revenue model – Awareness can be created among the patient through ads.
6.	Scalability of the Solution	Easy prediction of the heart disease with the patient details stored. Maintains best user experience.

3.4 Problem Solution Fit



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	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Gmail
FR-3	Data preparation	After user login they can upload a dataset and prepare the data.
FR-4	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created in IBM Cognos Analytics
FR-5	Generating Report	User can view health reports and can make decisions accordingly.

4.2 Non-Functional Requirement

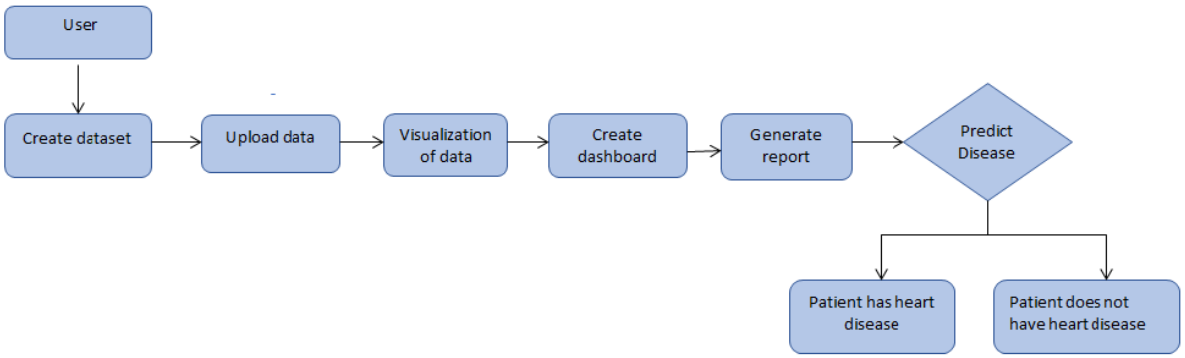
Following are the non-functional requirements of the proposed solution.

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.
NFR-2	**Security**	For Security of application, data replication technique is 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 must be reliable and strong in giving the functionalities.

NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the data submission. The application is direct and faster which depends on the efficiency of implemented algorithm.
NFR-5	**Availability**	The application will be available 24x7 for users without any interruption.
NFR-6	**Scalability**	The application can withstand increase in number of users and able to develop higher versions in future.

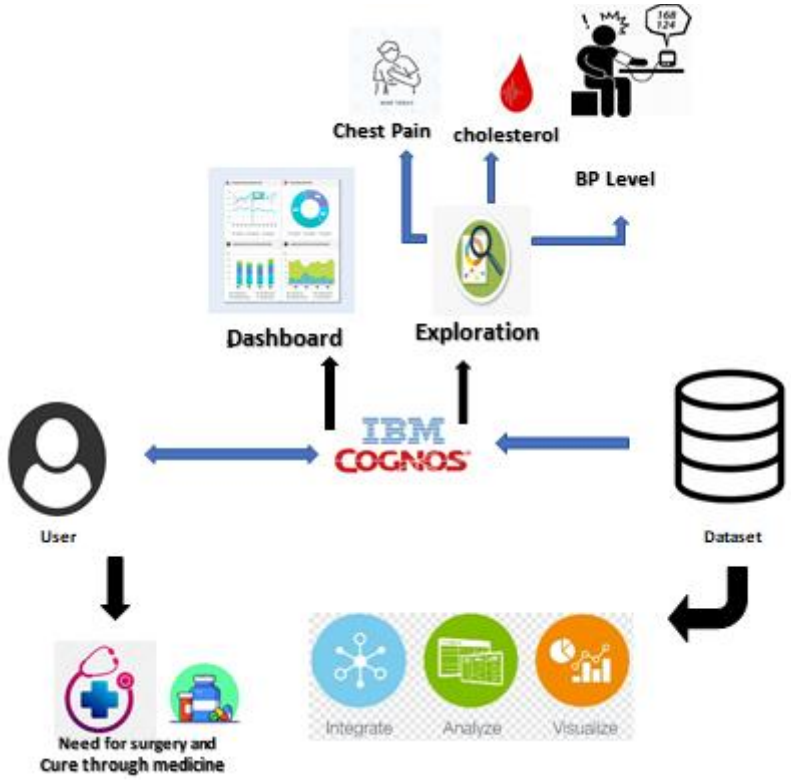
5. PROJECT DESIGN

5.1 Data Flow Diagram



5.2 SOLUTION & TECHNICAL ARCHITECTURE

Solution Architecture:



Technical Architecture:

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Registration	USN-1	As a user, I can register dashboard by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1

		USN-2	As a user, I will receive confirmation email once I have registered for the dashboard.	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a user, I can log into the dashboard by entering email & password.	I can register & access the dashboard	Low	Sprint-1
	Dashboard	USN-4	User can view his/her complete medical history.	I can view my medical history in the dashboard	Medium	Sprint-2
		USN-5	As a user, they can predict the occurrence of heart disease.	I can view the accuracy of heart disease in dashboard	High	Sprint-2
Customer Care Executive	Helpdesk	USN-6	As a customer care executive, they can view the patient problems.	I can send my problems in the dashboard	High	Sprint-3
		USN-7	As a customer care executive, they can cure the patient problems.	I can get help from the helpdesk	Medium	Sprint-3
Administrator	User Profile	USN-8	As an admin, they can add or delete the patient details.	I can access my dashboard	High	Sprint-4
		USN-9	As an admin, they can update the health issues of the patient.	I can view my updated health details	High	Sprint-4
		USN-10	As an admin, they can manage the patient details.	I can view my complete health details in dashboard	High	Sprint-4

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

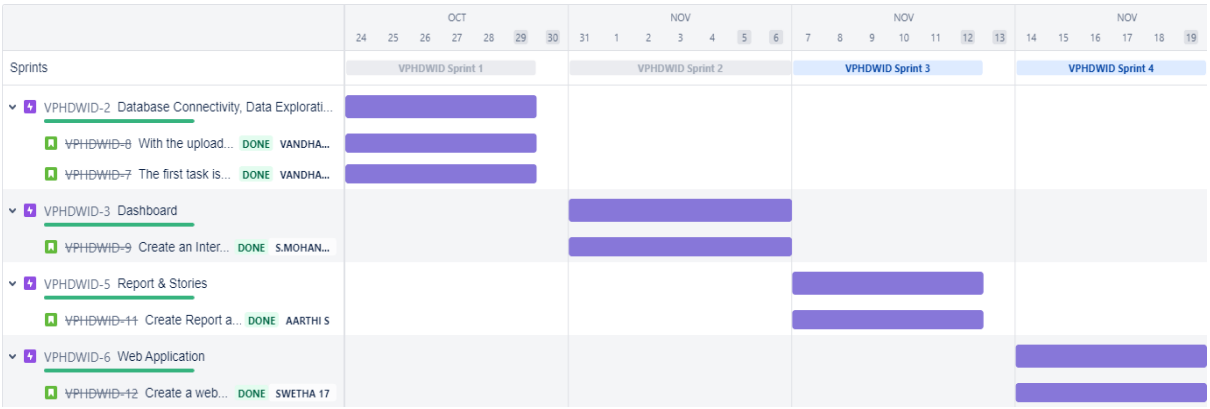
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Database Connectivity and Upload of dataset	USN-1	The first task is to collect and fetch the dataset from the external API and connect with database using IBM Cognos and IBM Cloud, then upload the dataset.	2	High	V.VANDHANA S.AARTHI S.MOHANAPRIYA S.SWETHA

Sprint-1	Data Modules & Data Exploration	USN-2	With the uploaded dataset we create a data module and perform data exploration by creating visualizations.	1	High	V.VANDHANA S.AARTHI S.MOHANAPRIYA S.SWETHA
Sprint-2	Dashboard	USN-3	Create an Interactive Dashboard after preparing the data exploration.	2	Low	V.VANDHANA S.AARTHI S.MOHANAPRIYA S.SWETHA
Sprint-3	Report and Stories	USN-4	Create Report and User stories based on the dashboard.	2	Medium	V.VANDHANA S.AARTHI S.MOHANAPRIYA S.SWETHA
Sprint-4	Web Application UI	USN-5	Create a web application for dashboard, report and user stories.	1	High	V.VANDHANA S.AARTHI S.MOHANAPRIYA S.SWETHA

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	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 Reports from JIRA

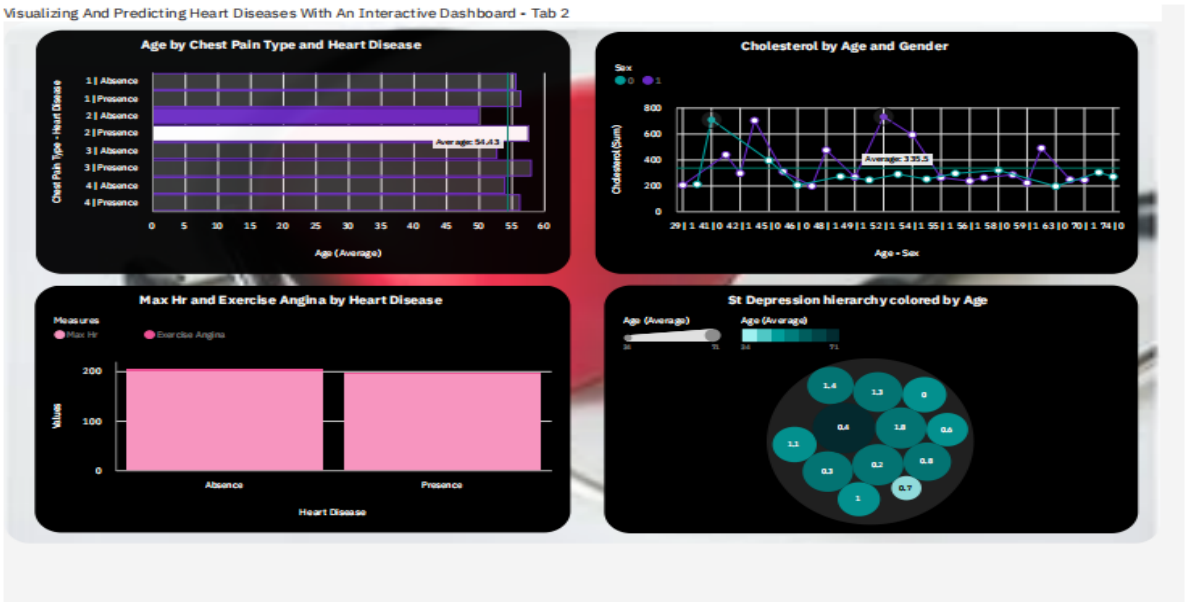


7. CODING & SOLUTIONING

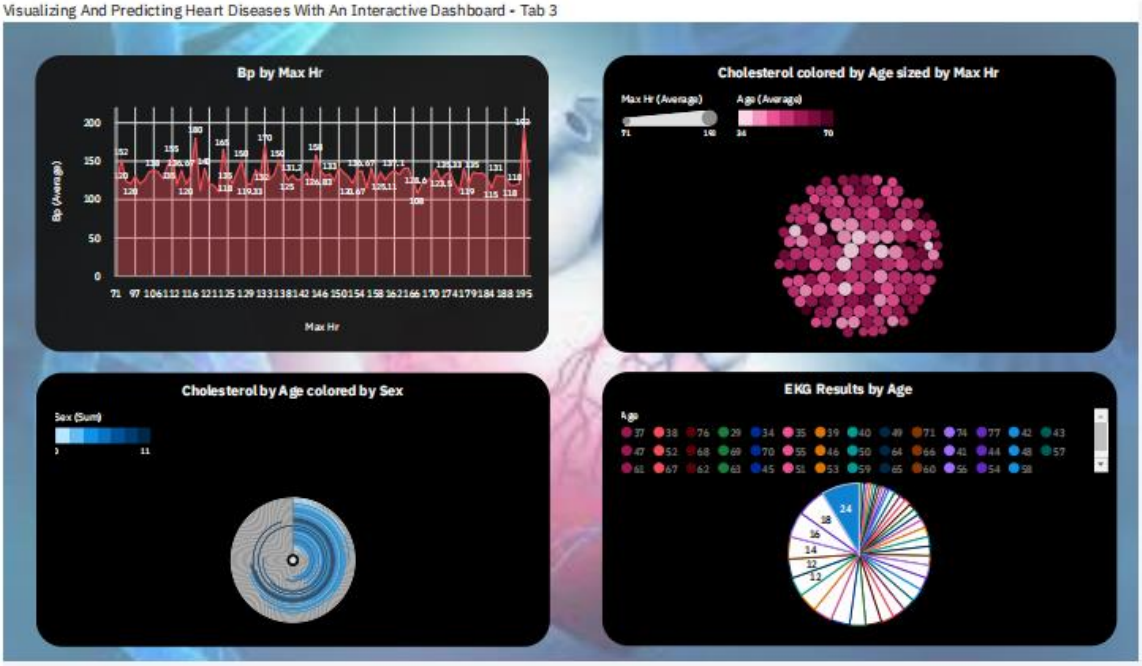
7.1 Feature 1

Dashboard:

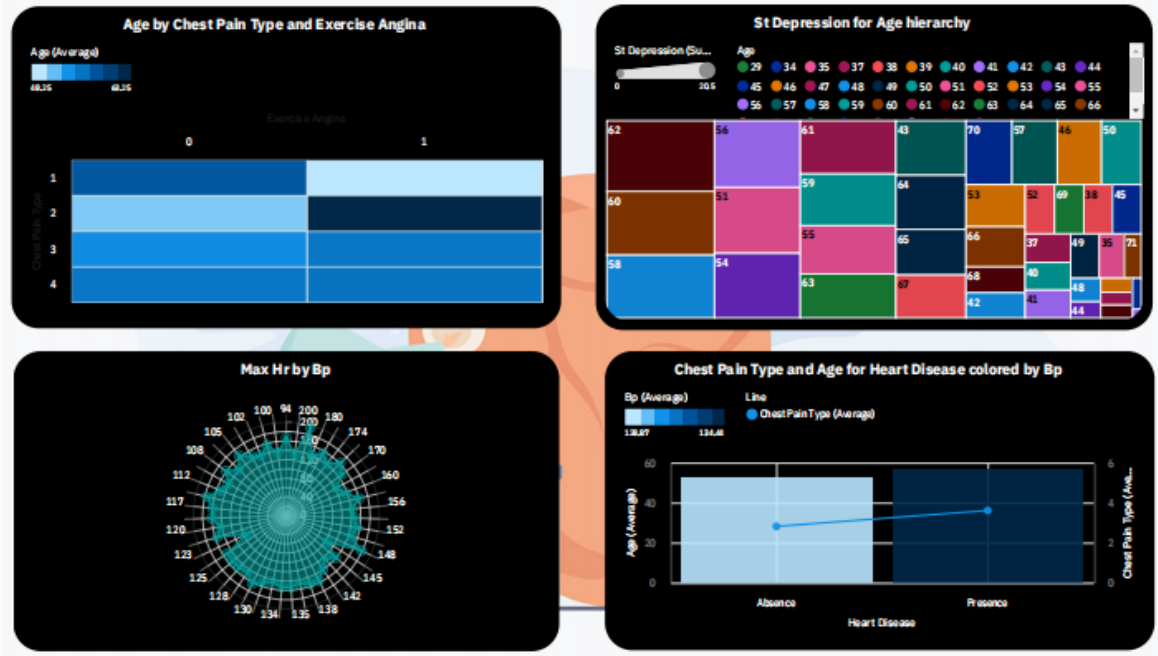
A dashboard for data analytics is **a tool used to multi-task, organize, visualize, analyze, and track data**. The overall purpose of a data analytics dashboard is to make it easier for data analysts, decision makers, and average users to understand their data, gain deeper insights, and make better data-driven decisions.

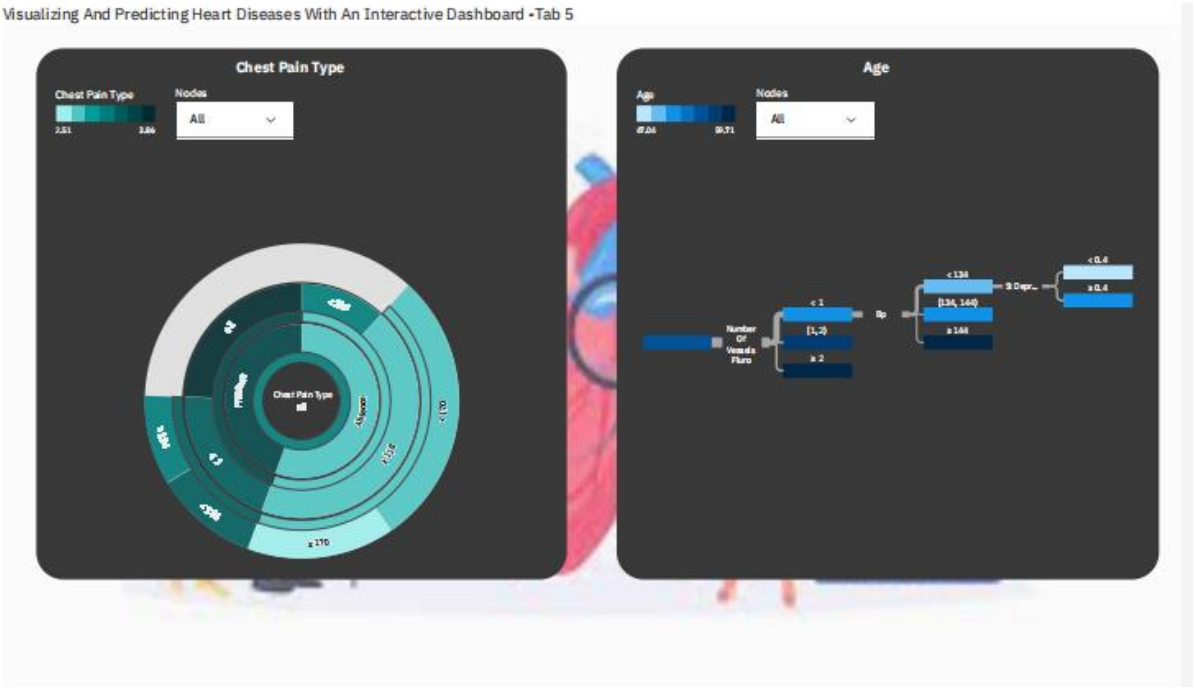


Visualizing And Predicting Heart Diseases With An Interactive Dashboard - Tab 3



Visualizing And Predicting Heart Diseases With An Interactive Dashboard - Tab 4





7.2 Feature 2

Working With Colab

```
import numpy as np

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
data=pd.read_csv("/content/Heart_Disease_Prediction.csv")
data
data.info
data.isnull().sum()
sns.countplot(x=data['Heart Disease'],hue='Sex',data=data)
sns.countplot(x=data['Heart Disease'],hue='Chest pain type',data=data)
sns.countplot(x=data['Sex'],hue='Chest pain type',data=data)
sns.barplot(x=data['Sex'],y=data['Cholesterol'],data=data)
sns.barplot(x=data['Sex'],y=data['BP'],data=data)
sns.barplot(x=data['Heart Disease'],y=data['Cholesterol'],data=data)
sns.barplot(x=data['Heart Disease'],y=data['BP'],data=data)
sns.lineplot(x=data['Age'],y=data['BP'],data=data)
sns.lineplot(x=data['Age'],y=data['Cholesterol'],data=data)
sns.lineplot(x=data['Age'],y=data['ST depression'],data=data)
sns.barplot(x=data['Heart Disease'],y=data['Exercise angina'],data=data)
sns.barplot(x=data['Heart Disease'],y=data['Number of vessels fluro'],data=data)
sns.barplot(x=data['Heart Disease'],y=data['Thallium'],data=data)
```

```
sns.barplot(x=data['Sex'],y=data['FBS over 120'],data=data)



sns.heatmap(data.corr())

from sklearn.preprocessing import LabelEncoder,StandardScaler
le=LabelEncoder()
data['Heart Disease']=le.fit_transform(data['Heart Disease'])

y=data['Heart Disease']
x=data.drop(['Heart Disease'],axis=1)

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=0,test_size=0.2
)

from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
lr=LogisticRegression(max_iter=10000)
lr.fit(x_train,y_train)
pred_1=lr.predict(x_test)
score_1=accuracy_score(y_test,pred_1)

score_1

from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rfc.fit(x_train,y_train)
pred_2=rfc.predict(x_test)
score_2=accuracy_score(y_test,pred_2)

score_2

from xgboost import XGBClassifier
xgb=XGBClassifier()
xgb.fit(x_train,y_train)
pred_3=xgb.predict(x_test)
score_3=accuracy_score(y_test,pred_3)

score_3
```

```
from sklearn.neighbors import KNeighborsClassifier
list_1=[]
for i in range(1,21):
  knn=KNeighborsClassifier(n_neighbors=1)
    knn.fit(x_train,y_train)
    preds=knn.predict(x_test)
    scores=accuracy_score(y_test,preds)
    list_1.append(scores)

max(list_1)
```

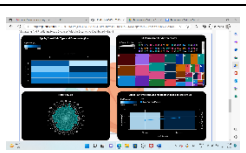
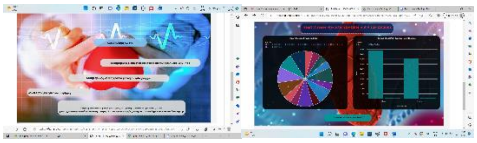
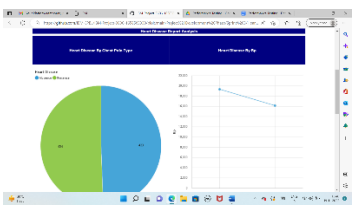
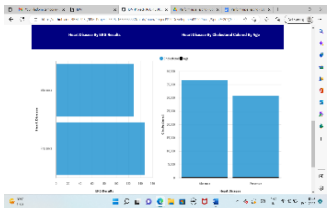
8. TESTING

8.1 Test Cases

```
x_train_prediction = model.predict(x_train)
training_data_accuracy = accuracy_score(x_train_prediction,y_train)

print("Accuracy of x train data:",training_data_accuracy)

Accuracy of x train data: 0.875

x_test_prediction = model.predict(x_test)
test_data_accuracy = accuracy_score(x_test_prediction,y_test)

print(test_data_accuracy)

0.8333333333333334
```

8.2 Performance Testing

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations / Graphs – Eighteen Visulizations in Five Tabs 

		 
2.	Data Responsiveness	Data is Easily Identified, Change and Deletion of Data is Secured, Easily Identified and Adapted, Reliable and Scalable.
3.	Amount Data to Rendered (DB2 Metrics)	All the data sets provided are rendered into visualization, graphs and dashboards. The dataset is trained and visualized using cognos and it connected with IBM Cloud.
4.	Utilization of Data Filters	The datasets are examined and excluded, rearranged, data's are replaced easily, count of number data sets are recorded, dataset is evaluated.Visualizations are utilized in order to filter data.
5.	Effective User Story	No of Scene Added – 2 Scene 
6.	Descriptive Reports	No of Visualizations / Graphs – Visualizations=4/graphs=  

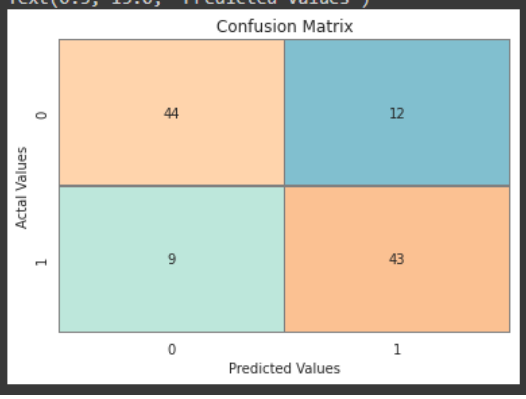
9. RESULTS

9.1 Performance Metrics

```
[23] cm = confusion_matrix(y_test,pred)
     print(classification_report(y_test,pred))
```

	precision	recall	f1-score	support
0	0.83	0.79	0.81	56
1	0.78	0.83	0.80	52
accuracy			0.81	108
macro avg	0.81	0.81	0.81	108
weighted avg	0.81	0.81	0.81	108

```
[24] sns.heatmap(cm, annot = True, fmt = 'g', cbar = False, cmap = 'icefire', linewidths= 0.5, linecolor= 'grey')
     plt.title('Confusion Matrix')
     plt.ylabel('Actal Values')
     plt.xlabel('Predicted Values')

     Text(0.5, 15.0, 'Predicted Values')
```




```
print("Accuracy Score = {}".format(round(accuracy_score(y_test,pred),5)))

Accuracy Score = 0.80556
```

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
- Very useful in case of emergency.
- User Friendly
- Easy to understand
- Secure
- Dashboard provides insightful information

Disadvantages:

- Accuracy Issues: A Computerized system alone does not ensure accuracy, and warehouse data is only as good as the data entry that created it.
- The system is not fully automated, it needs data from user for full diagnosis.

11. CONCLUSION

Heart diseases are a major killer in India and throughout the world, application of promising technology like dashboard in data analytics and machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients.

12.FUTURE SCOPE

For the Future Scope more dashboard in data analytics and machine learning approach will be used for best analysis of the heart diseases and for earlier prediction of diseases so that the rate of the death cases can be minimized by the awareness about the diseases.

13. APPENDIX

Source Code:

Web Application:

Index.html

```
<html>

<head>

<title> HEART DISEASE PREDICTION </title>

 <link rel="stylesheet" href="Style.css">

 <style>

 a:hover {

 background-color: crimson;

 }
```

```
</style>

</head>

<body>

<div class="menu-bar">

<ul>

<li class="active"><a href="#">Home</a></li>

<li><a
href="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;
pathRef=.my_folders%2FHeart%2BDisease%2BPrediction%2BDashboard&a
mp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=
false&amp;shareMode=embedded&amp;action=view&amp;mode=dashbo
ard&amp;subView=model00000184655ad0d6_00000000" width="320"
height="200" frameborder="0" gesture="media" allow="encrypted-media"
allowfullscreen="">Dashboard</a></li>

<li><a
href="https://us3.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2FHeart
%2BDisease%2BPrediction%2BReport&amp;closeWindowOnLastView=true
&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedde
d&amp;action=edit" width="320" height="200" frameborder="0"
gesture="media" allow="encrypted-media"
allowfullscreen="">Report</a></li>

<li><a
href=https://us3.ca.analytics.ibm.com/bi/?perspective=story&amp;pathRe
f=.my_folders%2FHeart%2BDisease%2BPrediction%2BStory&amp;closeWi
ndowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;
shareMode=embedded&amp;action=view&amp;sceneId=model000001847
6737829_00000000&amp;sceneTime=0" width="320" height="200"
frameborder="0" gesture="media" allow="encrypted-media"
allowfullscreen="">Story</a></li>

</ul>

</div>

<div class="content">
```

```
<h1>HEART DISEASE PREDICTION</h1>

<p>Let us have a look on Dashboard,Report and Story that is used for predicting the heart disease</p>

</div>

</body>

</html>
```

Style.css:

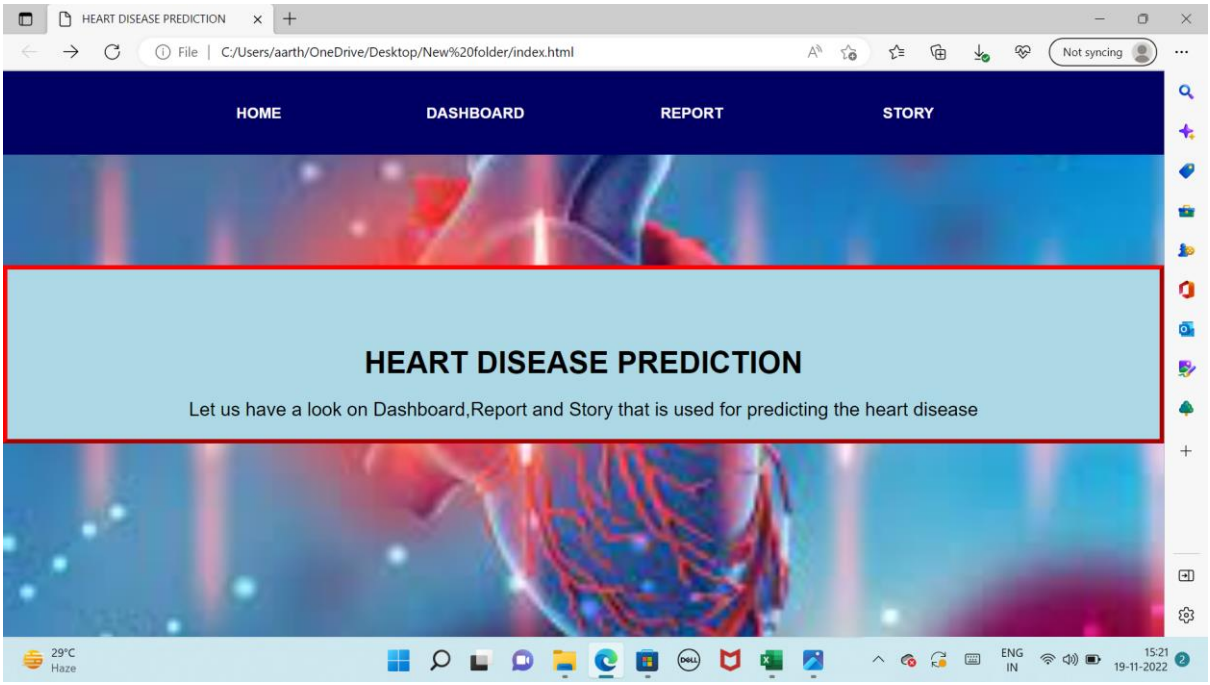
```
*{

margin: 0;

padding: 0;

box-sizing: border-box;

}

body

{

background-image: url(Heart.jpg);

background-size: cover;

background-position: center;

font-family: sans-serif;

}

.menu-bar

{

background: rgba(0,0,100);

text-align: center;

}
```

```
.menu-bar ul
{
display: inline-flex;
list-style: none;
color: #fff;
}
.menu-bar ul li
{
width: 200px;
margin: 15px;
padding: 20px;
}
.menu-bar ul li a
{
text-decoration: none;
color: #fff;
text-transform:uppercase;
font-family:Arial;
font-weight:bold;
}
.menu-bar ul li a: hover;
{
background-color: crimson;
```

```
}
.content
{
width:100%;
position:absolute;
top:50%;
transform:translateY(-50%);
text-align:center;
color:black;
border: 5px outset red;
background-color: lightblue;
text-align: center;
}
.content h1
{
font-size:7-px;
margin-top:80px;
}
.content p
{
margin:20px auto;
font-weight:100;
line-weight:25px;
```

font-size:20px;

]



GitHub & Project Demo Link

GitHub Link: https://github.com/IBM-EPBL/IBM-Project-3606-1658583330

Demo Link:
https://drive.google.com/file/d/1KNWu411zrvLSpM54DJR1UGGb2juoJInP/view?usp=share_link