Project Report

Team ID	PNT2022TMID25912
Project Name	visualizing and predicting heart disease with an interactive dashboard

1. INTRODUCTION

1.1 Project Overview

This project overview is mainly composed of data analytics and dataset provided it predicts the heart disease based on the symptoms and the dataset given.Early detection of cardiac diseases and continuous supervision of clinicians report reduces the rate of mortality.The data analytics gives the overview to the patient and the organisation about thier disease and thier medication.

1.2 Purpose

The purpose of this project is to visualize and predict the type of heart disease which is analysed by the dataset given and processed in platform of cognos analytics.The purpose mainly concerns on treatment on risk factors such as heart attack and stroke;prevention of repeat cardiovascular events;and reduction in deaths from cardiovascular disease.

2.LITERATURE SURVEY

2.1 Existing problem

Having high cholesterol increases the risk of atherosclerosis. Atherosclerosis has been linked to heart attacks and strokes.Diabetes increases the risk of heart disease. Obesity and high blood pressure increase the risk of diabetes and heart disease. Excess weight typically worsens other heart disease risk factors.The Existing problem mainly concerns on manual screening .

2.2. References

1.Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics—2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation. 2008;117(4):e25–146. [PubMed]

2. National Heart Lung and Blood Institute Fact Book, Fiscal Year 2006. Bethesda, Md: National Heart Lung and Blood Institute, National Institutes of Health; 2006. [12 April 2011]. Last accessed at http://www.nhlbi.nih.gov/about/factbook-06/toc.htm on.

3.Expert panel on detection evaluation and treatment and Treat high level cholestrol in adults .Third Report of National Education Program(NCEP)[PUB_MED].

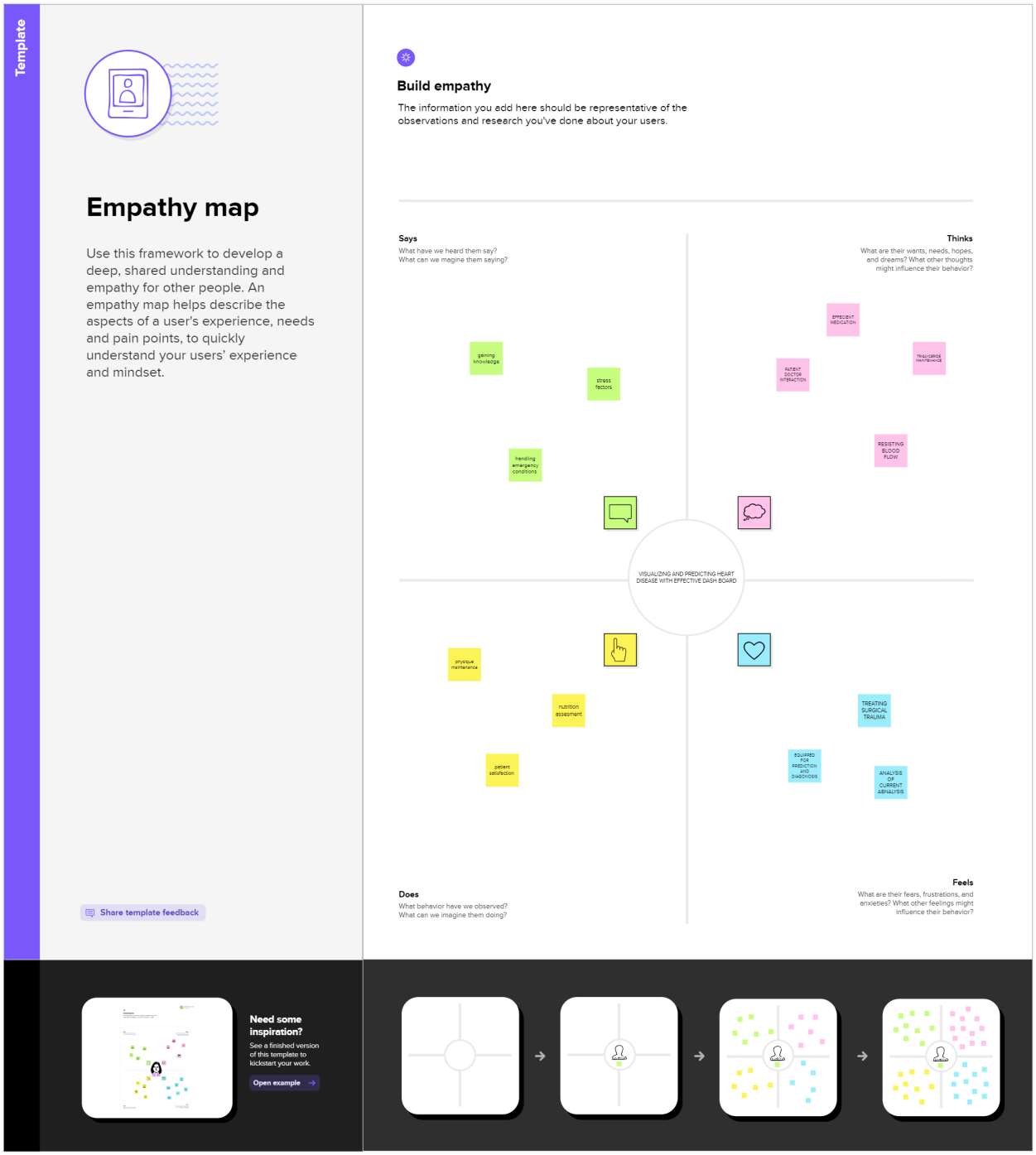
4.Tunstall pedoe.H The Dundee Coronary risk Management.BMJ;2001:323(7304):75-81

2.3 Problem Statement Definition

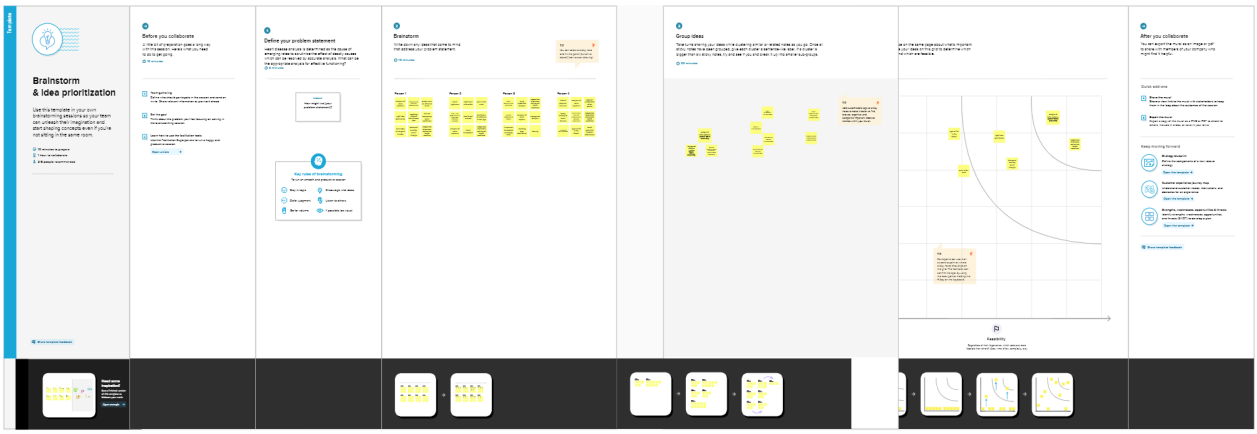
Heart disease analysis is determined as the cause of emerging rates to scrutinize the effect of deadly causes which can be resolved by accurate analysis. TWhat can be the appropriate analysis for effective functioning? The output of the algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likehood of a particular outcome.

3. IDEATION AND PROPOSED SOLUTION

3.1Empathy Map Canvas



3.2 Ideation & Brainstorming

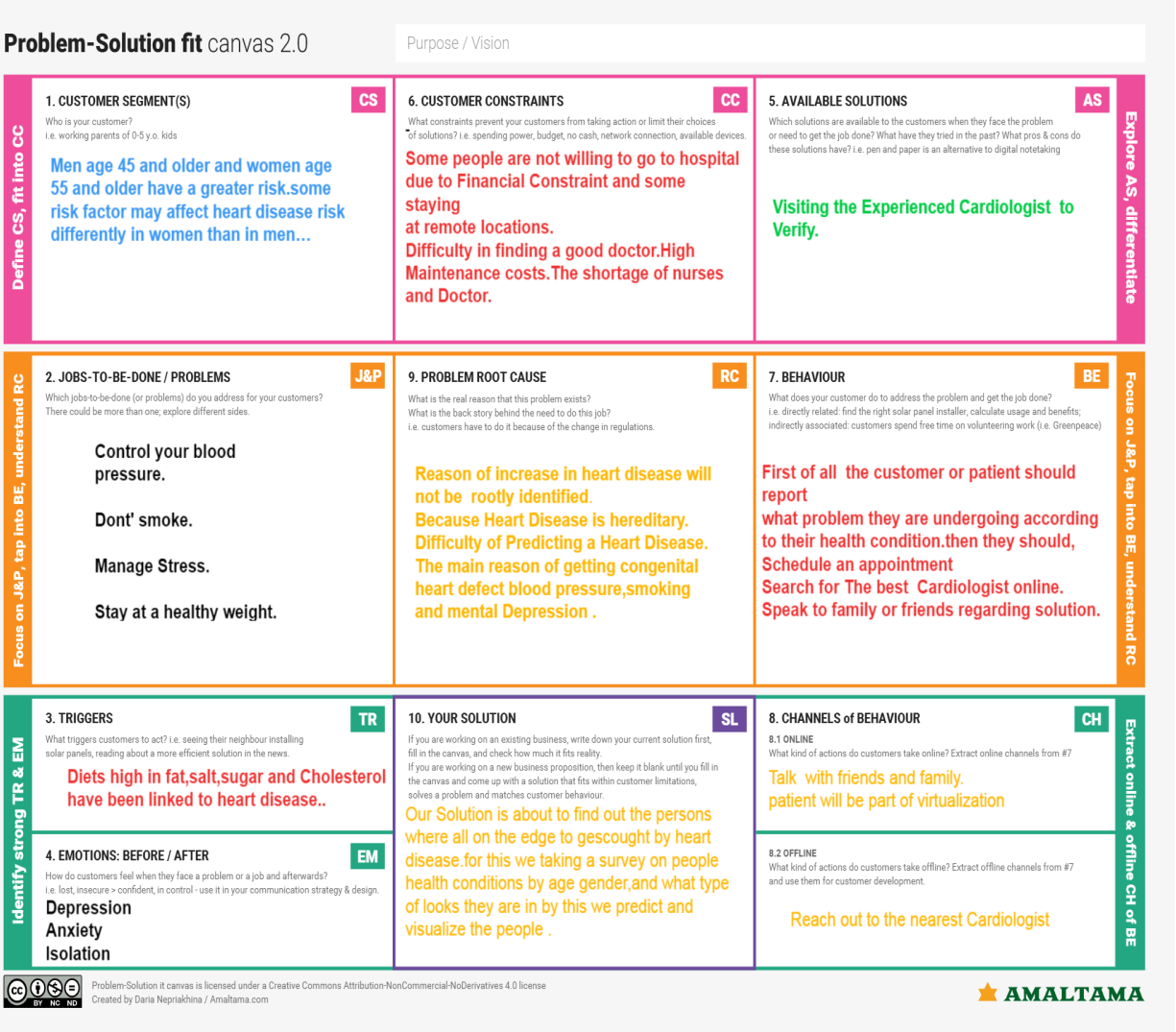


3.3 Proposed Solution

In this phase the requirement are collected and accessed .this project predicts people with cardiovascular disease by extracting the patient medical history and previous analysis on trauma data . analysis in medicine is becoming more and more frequent to clarify analyses and optimal prediction of results .The proposed solution customizes on the prediction with the symptoms such as lightheadness,chestpain,vomiting relating the symptoms with the dataset.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heart disease analysis is determined as the cause of emerging rates to scrutinize the effect of deadly causes which can be resolved by accurate analysis what can be the appropriate analysis for effective functioning?
2.	Idea / Solution description	In this phase the requirement are collected and accessed .this project predicts people with cardiovascular disease by extracting the patient medical history and previous analysis on trauma data . analysis in medicine is becoming more and more frequent to clarify analyses and optimal prediction of results
3.	Novelty / Uniqueness	This work is particularly interested in the category of data obtained this research work aims to design a frame work for heart disease prediction by using major risk factors based on different algorithms and dashboard as an optimal technique
4.	Social Impact / Customer Satisfaction	The customer satisfaction on this data analysis is effective because the treatment can be precisely and effectively made to the patients at emergency condition .it can be even more cost effective as it rely on the analysis and immediate cure instead of evaluating with all testing procedures.

3.4 Problem Solution fit



4.REQUIREMENT ANALYSIS

4.1 Functional requirements

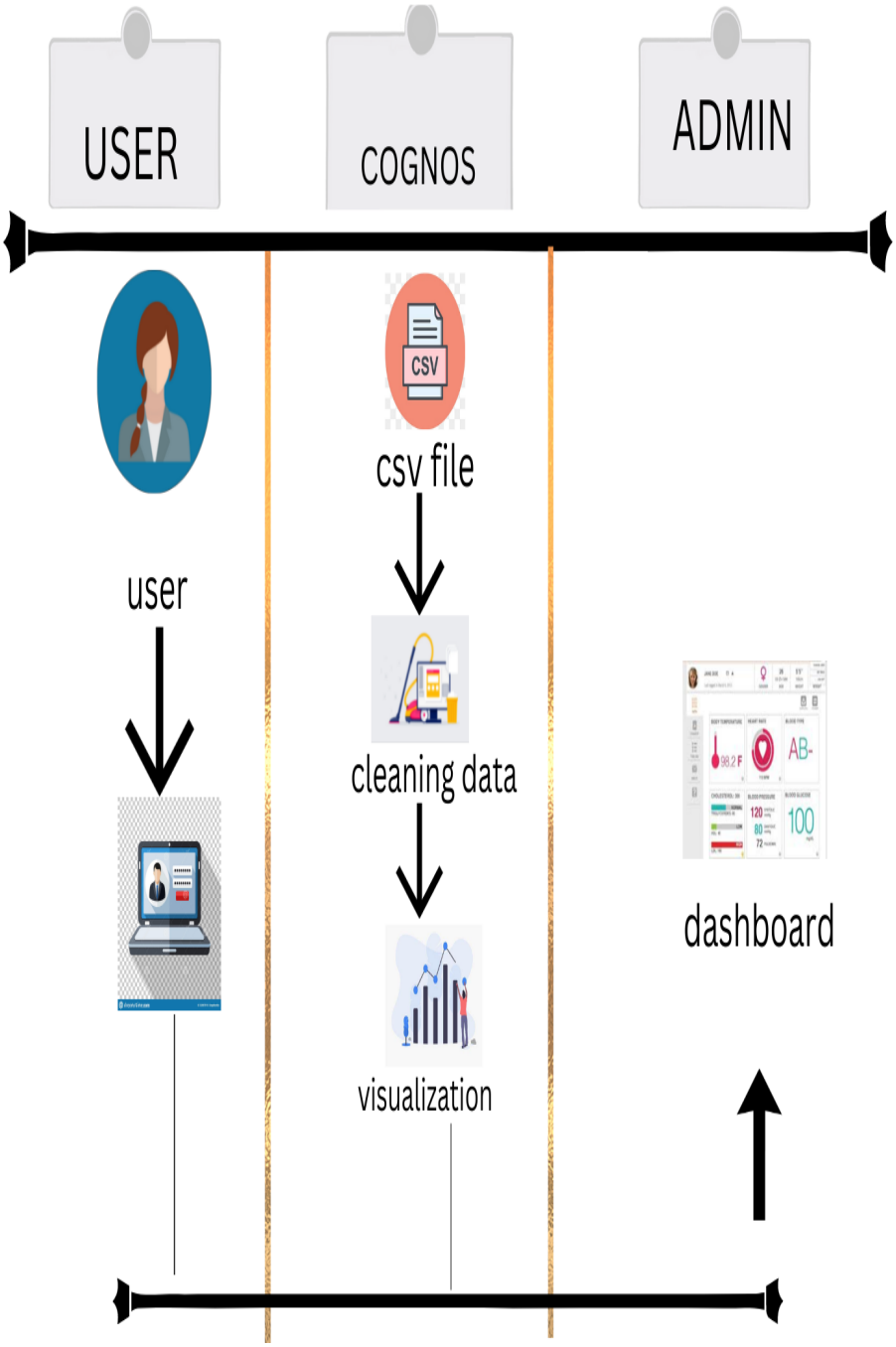
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Login	Login using Username and Password Login using Mobile No or Email Id

4.2 Non Functional Requirements

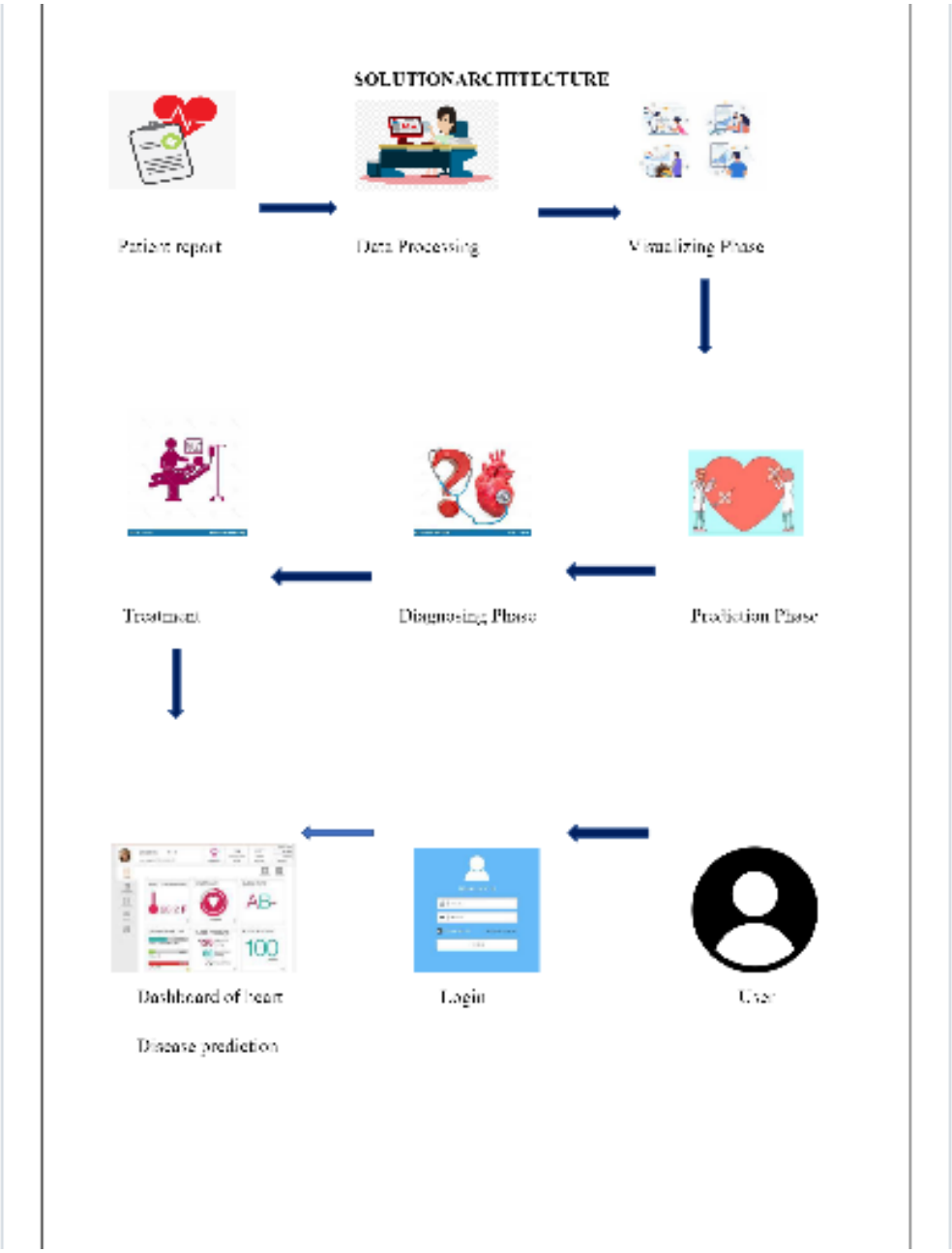
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	In order to overcome this issue, this work proposes a new architecture for real time health status prediction and It involves watching a group of users interacting with your dashboard.
NFR-2	**Security**	Providing a secure platform to users is crucial as it involves user upload their report details on the dashboard.
NFR-3	**Reliability**	Specification can be depended on to be accurate and the quality of being trustworthy or of performing consistently well.
NFR-4	**Availability**	To make sure that the dashboard is easily accessible and the quality of being able to be used or obtained.
NFR-5	**Scalability**	The ability of a computing process to be used or the measure of the dashboard's ability to increase or decrease in performance and produced in a range of capabilities.

5.PROJECT DESIGN

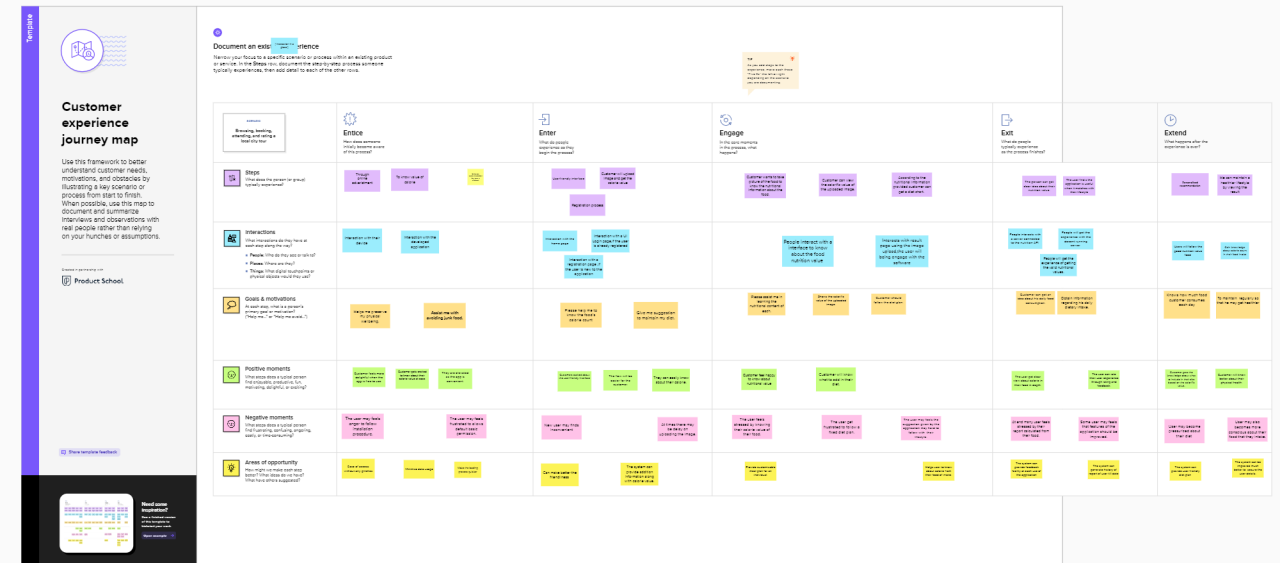
5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



5.3 User Stories



6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

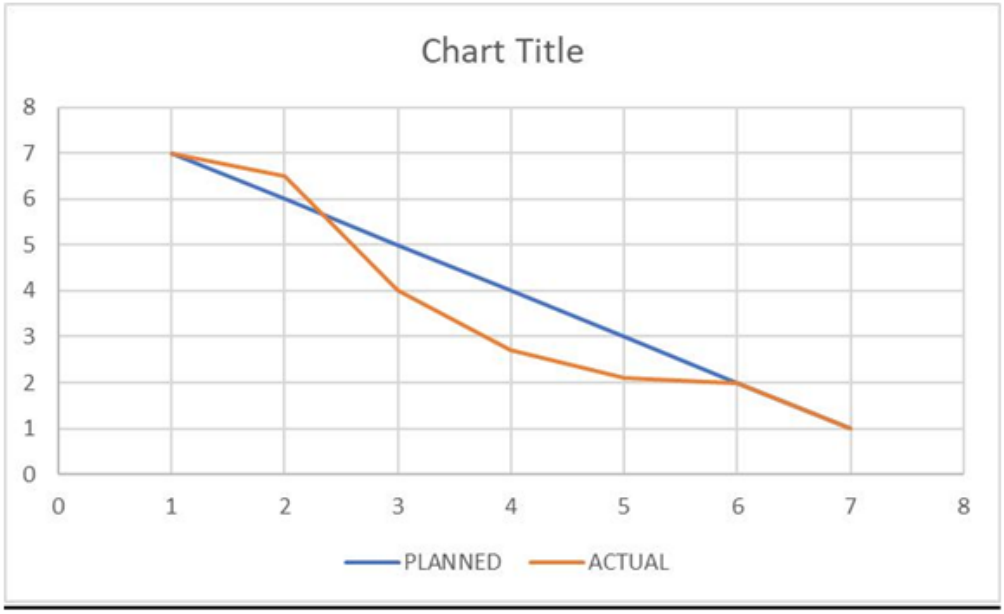
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	9	High	Sowmiya
			As a user, I will receive confirmation email once Ihave registered for the application.	4	Low	Vaishnavi

		USN-2	As a user, I can register for the applicationthrough Gmail.	7	Medium	Nishitha
	Login	USN-3	As a user, I can log into the application by entering email & password.	9	High	Sanju
Sprint-2	Working with the Dataset	USN-4	Importing the dataset on cognos platform and understand, clean and prepare the dataset.	9	High	Sowmiya, Sanju
	Data Visualization chart	USN-5	After importing the dataset, we create some visualizations to understand more about the predicting heart diseases.	7	Medium	Vaishnavi, Nishitha
Sprint-3	Creating the Dashboard	USN-6	Creating the dashboard to display the visualizations which gives insights of predicting the Heart diseases.	9	High	Nishitha, Sowmiya, Vaishnavi
Sprint-4	Export the Analytics	USN-7	Exporting the created dashboard to showcase the work to others.	9	High	Sanju, Sowmiya, Vaishnavi

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

REGISTRATION:

The new user sholud register with user name, password and the email id to access the website. After registering, user's login credentials will be stored in the database.

Coding:

```
<div class="container">
<body background="hospital_loop.gif"/>
<center>
  <h1>REGISTRATION FORM</h1>
  <form name="registration" class="registration-form" onsubmit="return formValidation()">
    <table>
      <tr>
        <td><b><label for="name">Name:</label></b></td>
        <td><input type="text" name="name" id="name" placeholder="your name"></td>
      </tr>
      <tr>
        <td><b><label for="email">Email:</label></b></td>
        <td><input type="text" name="email" id="email" placeholder="your email"></td>
```

```
</tr>
<tr>
  <td><b><label for="password">Password:</label></b></td>
  <td><input type="password" name="password" id="password"></td>
</tr>
<tr>
  <td><b><label for="phoneNumber">Phone Number:</label></b></td>
 <td><input type="number" name="phoneNumber" id="phoneNumber"></td>
</tr>
<tr>
  <td><b><label for="gender">Gender:</label></b></td>
  <td>Male: <input type="radio" name="gender" value="male">
    Female: <input type="radio" name="gender" value="female">
    Other: <input type="radio" name="gender" value="other"></td>
</tr>
<tr>
  <td><b><label for="language">language</label></b></td>
  <td>
    <select name="language" id="language">
      <option value="">Select language</option>
      <option value="English">English</option>
      <option value="Spanish">Spanish</option>
      <option value="Hindi">Hindi</option>
      <option value="Arabic">Arabic</option>
      <option value="Russian">Russian</option>
```

```
        </select>
      </td>
    </tr>
<tr>
      <td><b><label for="address">Address:</label></b></td>
      <td><textarea name="address" id="address" placeholder="Write your address..."></textarea></td>
    </tr>
    <tr>
      <td><b><label for="zipcode">Zip Code:</label></b></td>
      <td><input type="number" name="zipcode" id="zipcode"></td>
    </tr>
 </table>
<p>By creating an account you agree to our Terms & Privacy.</p>
<button type="submit"
 class="registerbtn"><strong><a href="https://colab.research.google.com/drive/1K4AonvF-l4oozaEbZEr9OVLIW5H0yqGK#scrollTo=54mnPMaupFME">Register</a></strong></button>

</div>
<div class="container signin">
<center>
<p>Already have an account? <a href="loo.html">Sign in</a>.</p>
</center>
```

```
</b>
</table>
  </form>
</center>
</div>
```

7.2 Feature 2

Working with colab:

we are visualzing and predicting heart disease with an interactive dashboard by using the colab as well as with ibm cognos and cloud.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
df = pd.read_csv('dataset.csv')
df.info()
df.describe()
import seaborn as sns
```

```
#get correlations of each features in dataset
corrmat = df.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))
#plot heat map
g=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap="RdYlGn")
df.hist()
sns.set_style('whitegrid')
sns.countplot(x='target',data=df,palette='RdBu_r')
dataset = pd.get_dummies(df, columns = ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope',
'ca', 'thal'])
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
standardScaler = StandardScaler()
columns_to_scale = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
dataset[columns_to_scale] =
standardScaler.fit_transform(dataset[columns_to_scale])
dataset.head()
y = dataset['target']
X = dataset.drop(['target'], axis = 1)
from sklearn.model_selection import cross_val_score
knn_scores = []
for k in range(1,21):
    knn_classifier = KNeighborsClassifier(n_neighbors = k)
    score=cross_val_score(knn_classifier,X,y,cv=10)
    knn_scores.append(score.mean())
plt.plot([k for k in range(1, 21)], knn_scores, color = 'red')
```

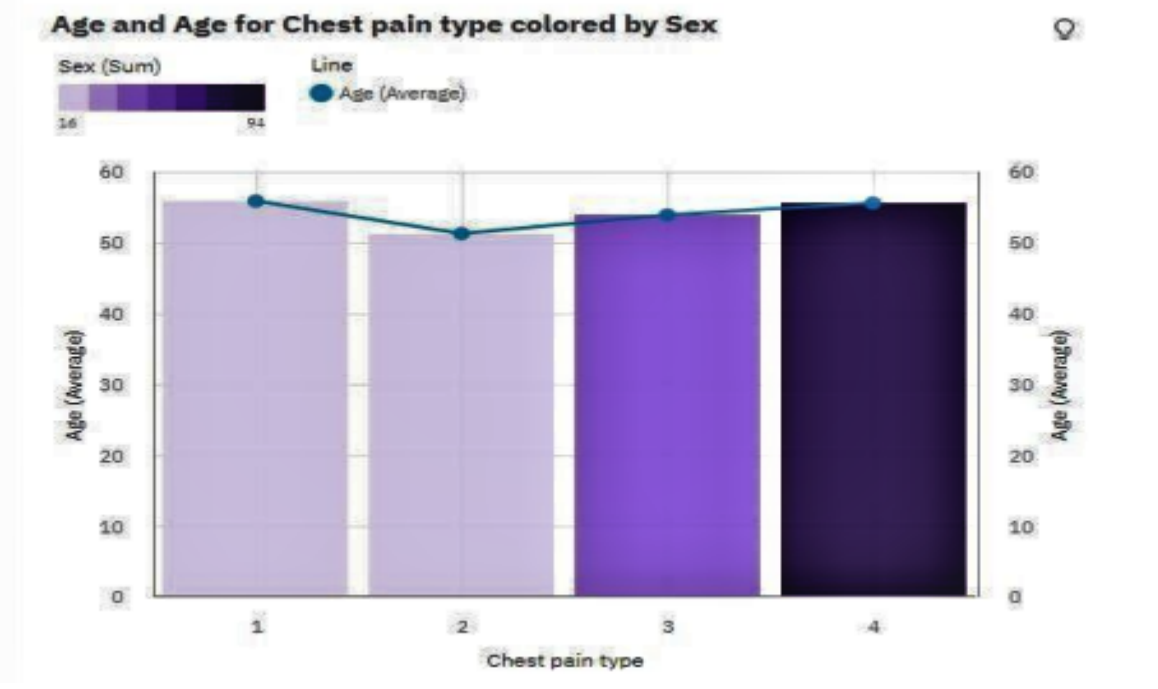
```
for i in range(1,21):
    plt.text(i, knn_scores[i-1], (i, knn_scores[i-1]))
plt.xticks([i for i in range(1, 21)])
plt.xlabel('Number of Neighbors (K)')
plt.ylabel('Scores')
plt.title('K Neighbors Classifier scores for different K values')
knn_classifier = KNeighborsClassifier(n_neighbors = 12)
score=cross_val_score(knn_classifier,X,y,cv=10)
score.mean()
from sklearn.ensemble import RandomForestClassifier
randomforest_classifier= RandomForestClassifier(n_estimators=10)
score.mean()
```

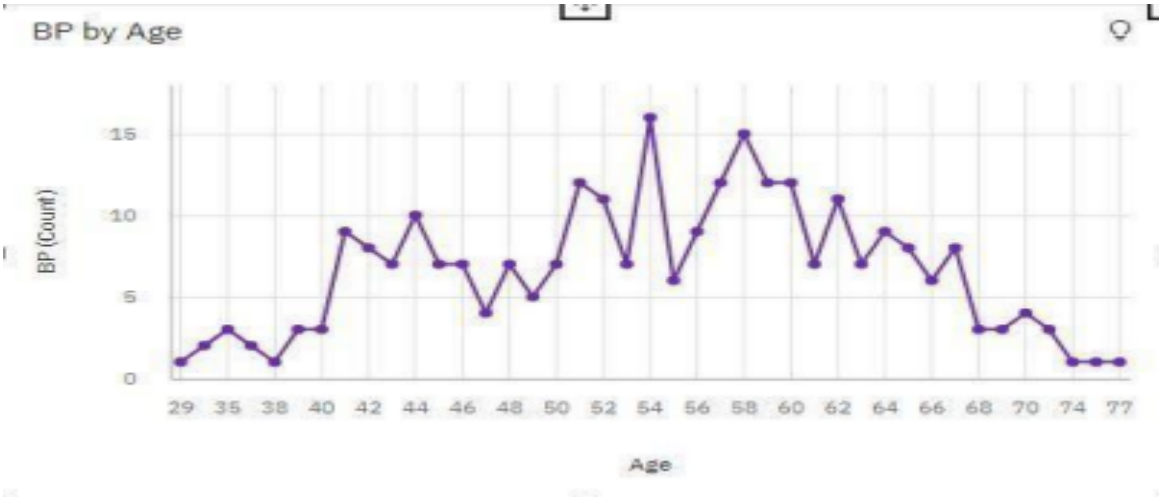
RANDOM FOREST ALGORITHM:

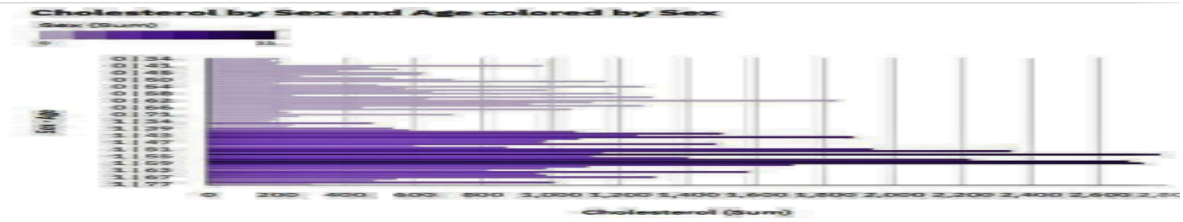
Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems.

K-NEAREST NEIGHBOR ALGORITHM:

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point


Age and Age for Chest pain type colored by Sex
Sex (Sum)
16
94
Line
Age (Average)
60
50
40
30
20
10
0
Age (Average)
1
2
3
4
Chest pain type


BP by Age
BP (Count)
15
10
5
0
29 35 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 74 77
Age


Cholesterol by Sex and Age colored by Sex

8.TESTING

8.1 Test Cases

Test caseID	Feature Type	Component	Test Scenario
LoginPage_TC_OO1	Functional	Home Page	Verify user is able to see the Login/Signup popupwhen userclicked on My accountbutton
LoginPage_TC_OO2	UI	Home Page	Verify the UI elements in Login/Signup popup
LoginPage_TC_OO3	Functional	Home page	Verify user is able to log into application with Valid credentials
LoginPage_TC_OO4	Functional	Login page	Verify user is able to log intoapplication with InValid credentials

LoginPage_TC_OO4	Functional	Login page	Verify user is able to log intoapplication with InValid credentials

Testcase ID	Feature Type	Component	TestScenario
LoginPage_TC_OO5	Functional	Login page	Verify user is able to log intoapplication with InValid credentials
LoginPage_TC_OO6	Functional	Login page	Verify user is able to the log into application with valid credentials.
LoginPage_TC_OO7	Functional	Home page	User must be able to navigate between different sections.
LoginPage_TC_OO8	Functional	Home page	User must be able to change the visualization according to their requirements.
LoginPage_TC_OO9	Functional	Home page	User must be able to navigate to different tabs using the given link.

8.2 USER ACCEPTANCE TESTING

PURPOSE :

The purpose is to brefly explain the test coverage and open issues of the project at the time of the release to User Acceptance Testing.[UAT]

DEFECT ANALYSIS:

This report shows the number of resolved or closed bugs at each severity level, and show they were resolved.

Resolution	Severity1	Severity2	Severity3	Severity4	Subtotal
By Design	8	2	2	3	17
Duplicate	1	0	3	0	4
External	3	2	0	1	6
Fixed	9	2	4	16	29
Not Reproduced	0	0	1	0	1
Skipped	0	0	0	0	0
Won'tFix	0	5	2	1	8
Totals	21	11	12	22	65

TEST CASE ANALYSIS:

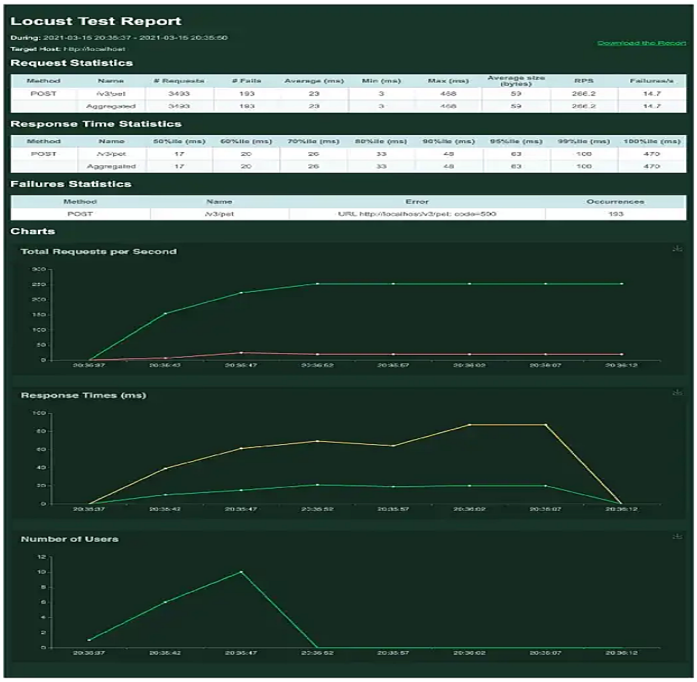
This report shows the number of test cases that have passed, failed and untested.

Section	TotalCases	Not Tested	Fail	Pass
PrintEngine	7	0	0	7
ClientApplication	51	0	3	48
Security	2	0	1	1
OutsourceShipping	3	0	2	1
ExceptionReporting	9	0	0	9
FinalReportOutput	4	0	0	4
VersionControl	2	0	1	1

9.RESULTS

9.1 Performance Metrics

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visualizations / Graphs - 6
2.	Data Responsiveness	People experiencing the chest pain • Average BP during the chest pain • Maximum heart rate during the chest pain • BP along with age • Cholesterol by age and gender
3.	Amount Data to Rendered (DB2 Metrics)	The dataset is trained and visualized using Cognos and it is connected to the IBM cloud.
4.	Utilization of Data Filters	Visualizations are utilized in order to filter the data.
5.	Effective User Story	No of Scene Added - 6
6.	Descriptive Reports	No of Visualizations / Graphs - 6



10.ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

- The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction.

DISADVANTAGES:

- Data mining techniques does not help to. provide effective decision making.
- Cannot handle enormous datasets for.

11.CONCLUSION

This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in building the given model are Logistic regression, Random Forest Classifier and KNN [22]. The accuracy of our model is 87.5%.

12.FUTURE SCOPE

The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc.

13.APPENDIX

Source Coding:

```
<div class="container">
<body background="hospital_loop.gif"/>
<center>
  <h1>REGISTRATION FORM</h1>
  <form name="registration" class="registration-form" onsubmit="return
formValidation()">
    <table>
      <tr>
        <td><b><label for="name">Name:</label></b></td>
        <td><input type="text" name="name" id="name" placeholder="your
name"></td>
      </tr>
      <tr>
        <td><b><label for="email">Email:</label></b></td>
        <td><input type="text" name="email" id="email" placeholder="your
email"></td>
      </tr>
      <tr>
        <td><b><label for="password">Password:</label></b></td>
        <td><input type="password" name="password" id="password"></td>
      </tr>
      <tr>
```

```
      <td><b><label for="phoneNumber">Phone Number:</label></b></td>
     <td><input type="number" name="phoneNumber" id="phoneNumber"></td>
</tr>
    <tr>
      <td><b><label for="gender">Gender:</label></b></td>
      <td>Male: <input type="radio" name="gender" value="male">
        Female: <input type="radio" name="gender" value="female">
        Other: <input type="radio" name="gender" value="other"></td>
    </tr>
    <tr>
      <td><b><label for="language">language</label></b></td>
      <td>
        <select name="language" id="language">
          <option value="">Select language</option>
          <option value="English">English</option>
          <option value="Spanish">Spanish</option>
          <option value="Hindi">Hindi</option>
          <option value="Arabic">Arabic</option>
<option value="Russian">Russian</option>
        </select>
      </td>
    </tr>
<tr>
      <td><b><label for="address">Address:</label></b></td>
      <td><textarea name="address" id="address" placeholder="Write your
```

```
address..."></textarea></td>
      </tr>
      <tr>
        <td><b><label for="zipcode">Zip Code:</label></b></td>
        <td><input type="number" name="zipcode" id="zipcode"></td>
      </tr>
 </table>
<p>By creating an account you agree to our Terms & Privacy.</p>
<button type="submit"
 class="registerbtn"><strong><a
href="https://colab.research.google.com/drive/1K4AonvF-
l4oozaEbZEr9OVLIW5H0yqGK#scrollTo=54mnPMaupFME">Register</a></stro
ng></button>

</div>
<div class="container signin">
<center>
<p>Already have an account? <a href="loo.html">Sign in</a>.</p>
</center>
</b>


   </table>
  </form>
</center>
```

```
</div>
```

LOGIN FORM:

```
 <head>
   <link rel="stylesheet" type="text/css" href="css/style.css"/>
</head>
<!DOCTYPE html>
<html>
<head>
<center>
    <title>Login Form</title>
    <link rel="stylesheet" type="text/css" href="css/style.css">
</center>
</head>
<body background="login gif.gif"/>
<center>
    <h2>Login Page</h2><br>
    <div class="login">
    <form id="login" method="get" action="login.php">
        <label><b>User Name
        </b>
        </label>
        <input type="text" name="Uname" id="Uname" placeholder="Username">
        <br><br>
        <label><b>Password
        </b>
        </label>
```

```
        <input type="Password" name="Pass" id="Pass" placeholder="Password">
        <br><br>
        <button type="submit"
 class="loginbtn"><strong><a
href="https://colab.research.google.com/drive/1K4AonvF-
l4oozaEbZEr9OVLIW5H0yqGK?authuser=3">login</a></strong></button>
 </form>
</div>
</center>
</body>
</html>
```

VISUALIZING AND PREDICTING HEART DISEASE

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import rcParams
from matplotlib.cm import rainbow
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
df = pd.read_csv('dataset.csv')
df.info()
df.describe()
import seaborn as sns
```

```
#get correlations of each features in dataset
corrmat = df.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))
#plot heat map
g=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap="RdYlGn")
df.hist()
sns.set_style('whitegrid')
sns.countplot(x='target',data=df,palette='RdBu_r')
dataset = pd.get_dummies(df, columns = ['sex', 'cp', 'fbs', 'restecg', 'exang', 'slope',
'ca', 'thal'])
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
standardScaler = StandardScaler()
columns_to_scale = ['age', 'trestbps', 'chol', 'thalach', 'oldpeak']
dataset[columns_to_scale] =
standardScaler.fit_transform(dataset[columns_to_scale])
dataset.head()
y = dataset['target']
X = dataset.drop(['target'], axis = 1)
from sklearn.model_selection import cross_val_score
knn_scores = []
for k in range(1,21):
    knn_classifier = KNeighborsClassifier(n_neighbors = k)
    score=cross_val_score(knn_classifier,X,y,cv=10)
    knn_scores.append(score.mean())
plt.plot([k for k in range(1, 21)], knn_scores, color = 'red')
```

```
for i in range(1,21):
    plt.text(i, knn_scores[i-1], (i, knn_scores[i-1]))
plt.xticks([i for i in range(1, 21)])
plt.xlabel('Number of Neighbors (K)')
plt.ylabel('Scores')
plt.title('K Neighbors Classifier scores for different K values')
knn_classifier = KNeighborsClassifier(n_neighbors = 12)
score=cross_val_score(knn_classifier,X,y,cv=10)
score.mean()
from sklearn.ensemble import RandomForestClassifier
randomforest_classifier= RandomForestClassifier(n_estimators=10)
score.mean()
```

PROJECT DEMO LINK:

https://youtu.be/235yGt0F2Mg

GITHUB :

https://github.com/IBM-EPBL/IBM-Project-1261-1658382104