Project Report

Team ID	PNT2022TMID23536
Project Name	Visualizing and Predicting Heart Disease with an Interactive Dashboard

1. INTRODUCTION

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 organization about their disease and their medication.

Purpose

The purpose of this project is to visualize and predict the type of heart disease which is analyzed 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

Existing problem

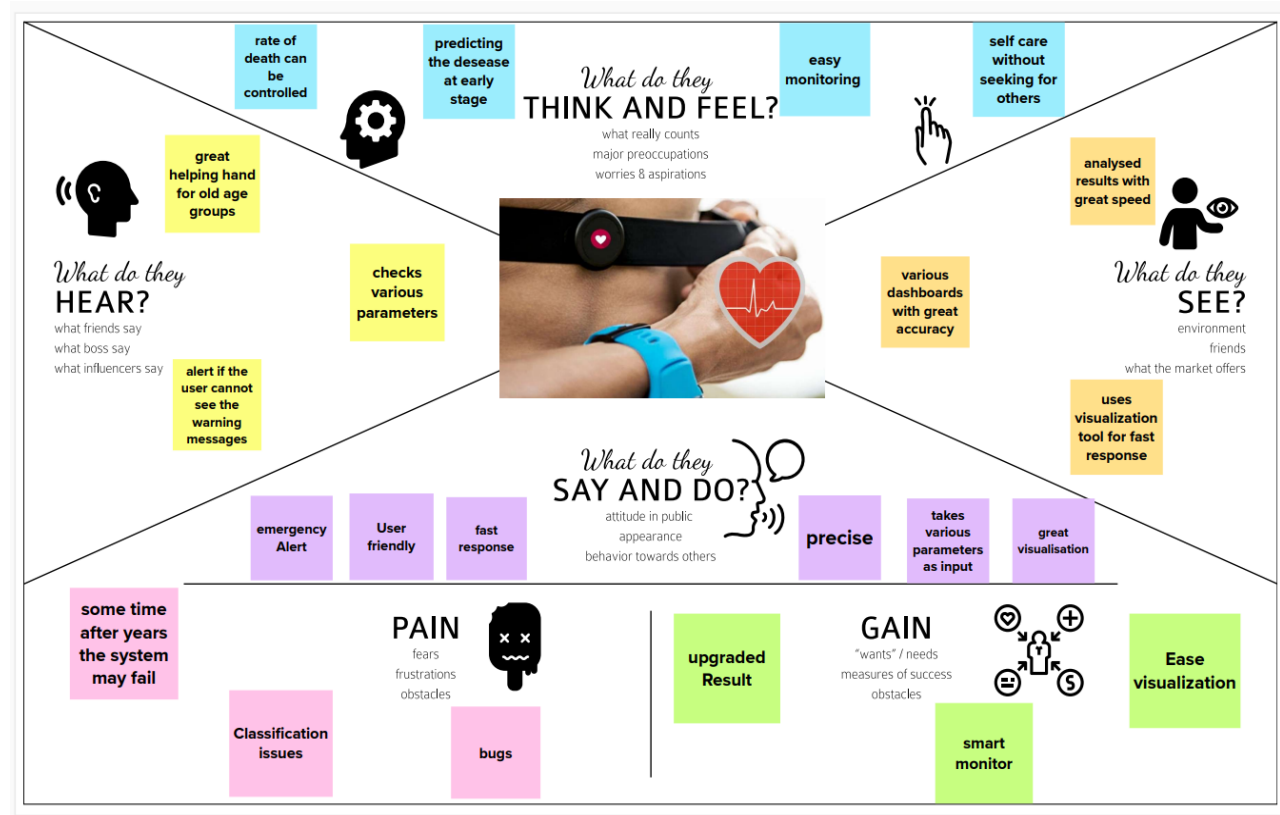Having high cholesterol increases the risk of atherosclerosis. Atherosclerosis has been linked to heart attacks and strokes.Diabetes increases the risk of heart disease. Obesity and high blood pressure increase the risk of diabetes and heart disease. Excess weight typically worsens other heart disease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

2.3 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

3.1 Empathy 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, chest pain, vomiting relating the symptoms with the dataset.

S.NO	Parameter	Description
1	Problem statement (problem to be solved)	Heart disease analysis is determined as cause of emerging rates to scrutinized the effect of deadly causes which can be resolved by accurate analysis what can be the appropriate analysis for effective function ?
2	Idea /Solution description	In this phase the requirements are collected and accessed the project predicts people with cardio vascular disease by extracting the patients medical history and previous analysis on trauma data. Analyzing

		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 in this research work aims to design the framework 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 analyses is effective because the treatment can be precisely and effectively made to the patients at emergency conditions .

3.4 Problem Solution fit

Project Title: Visualizing and Predicting Heart Diseases with an Interactive Dash Board   Project Design Phase-I - Solution Fit   Team ID: PNT2022TMID23536

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) `CS`
- Old age people at home - domestic level
- Private hospitals and clinics
- Lab centres who generates medical analysis reports and insights

6. CUSTOMER CONSTRAINTS `CC`
- Lack of digital knowledge for patients by themselves
- Huge variables in the parameters of analysis which is a challenging factor for customer

5. AVAILABLE SOLUTIONS `AS`
- Traditional heart disease prediction systems doesn't have interactive dynamic dashboard for user understanding
- Intensive fit bands or hardware systems are complex which are not user friendly

Explore AS, differentiate

2. JOBS-TO-BE-DONE / PROBLEMS `J&P`
- To prevent heart disease with an interactive dashboard
- To prevent heart diseases with intensive anlayis at the early stage of the cause
- Analysing patterns in heart parameters for detailed visual data

9. PROBLEM ROOT CAUSE `RC`
- Inefficient traditional systems
- No interactive visual interface
- Less accuracy in analysing heart parameters

7. BEHAVIOUR `BE`
- To provide necessary data to analyse the data pertaining to heart disease parameters and generate predictive output.
- Navigation around Dashboard features understanding the same

Identify strong TR & EM

3. TRIGGERS `TR`
- Maximum deaths due to CVD for people around the customers.
- More awareness to concentrate on heart diseases through social media, doctor advises and word of mouth about the problem

4. EMOTIONS: BEFORE / AFTER `EM`
Insecure → Confident
Unreliable → Trustworthy

10. YOUR SOLUTION `SL`
The proposed solution aims at increasing the accuracy of heart disease prediction system through interactive user friendly dashboard for analysis of heart diseases
- Heart data parameters monitoring and analysis using data analytics tools
- Visual dashboard for displaying the predicted diseases through the analysis of various parameters from patient end

8.CHANNELS of BEHAVIOUR `CH`
8.1   ONLINE
- Analysis of Data collected
- Accessing and understanding application's features.

8.2   OFFLINE
- Data collection of patients/customer
- Implementing the suggestions, insights provided by the application interface.

Identify strong TR & EM

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	Registration through Form. Registration through Gmail. Registration through Linked IN.
FR-2	User Confirmation	Confirmation via Email. Confirmation via OTP.
FR-3	User verification	Verification through CAPTCHA Verification through I'm not a robot.
FR-4	User Authentication	Recognition of correct person Resending the code in case of forgot password.
FR-5	User validation	Reconfirming the new password Sending a two digit number in (Google account) your Old devices, so that you can enter into a new device By entering the two digit number.
FR-6	User Submission	Submission through Google form Submission through Email.

4.2 Non-functional Requirements:

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

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, eg, relationships between medical factors related to heart disease and patterns, to be established.
NFR-2	**Security**	When it deals with(comes to)health factors, we should provide more security services. There shouldn't be no errors, lagging , base of data of a patient profile, while working on the software or product.
NFR-3	**Reliability**	Reliability is said to be the measure of stability or consistency of test scores shown in your product. Therefore your product will normal as a good performance one in the field of accuracy.
NFR-4	**Performance**	The performance should be fast relaying. This prediction system should be made available in cloud to ensure better accessibility and setting a milestone in providing good quality affordable healthcare.
NFR-5	**Availability**	The Availability of getting used to this software or product design is through by accessing IBM cognos Analytics and IBM cloud.

3.PROJECT DESIGN

Data Flow Diagrams

START

Collect the dataset

Data Module

IBM COGNOS

Story

Upload the dataset

Report

Exploration

Dashboard

PYTHON in machine learning

Apply the support vector machine algorithm to predict

Disease prediction

Yes

No

Patient has heart disease

Patient does not have heart disease
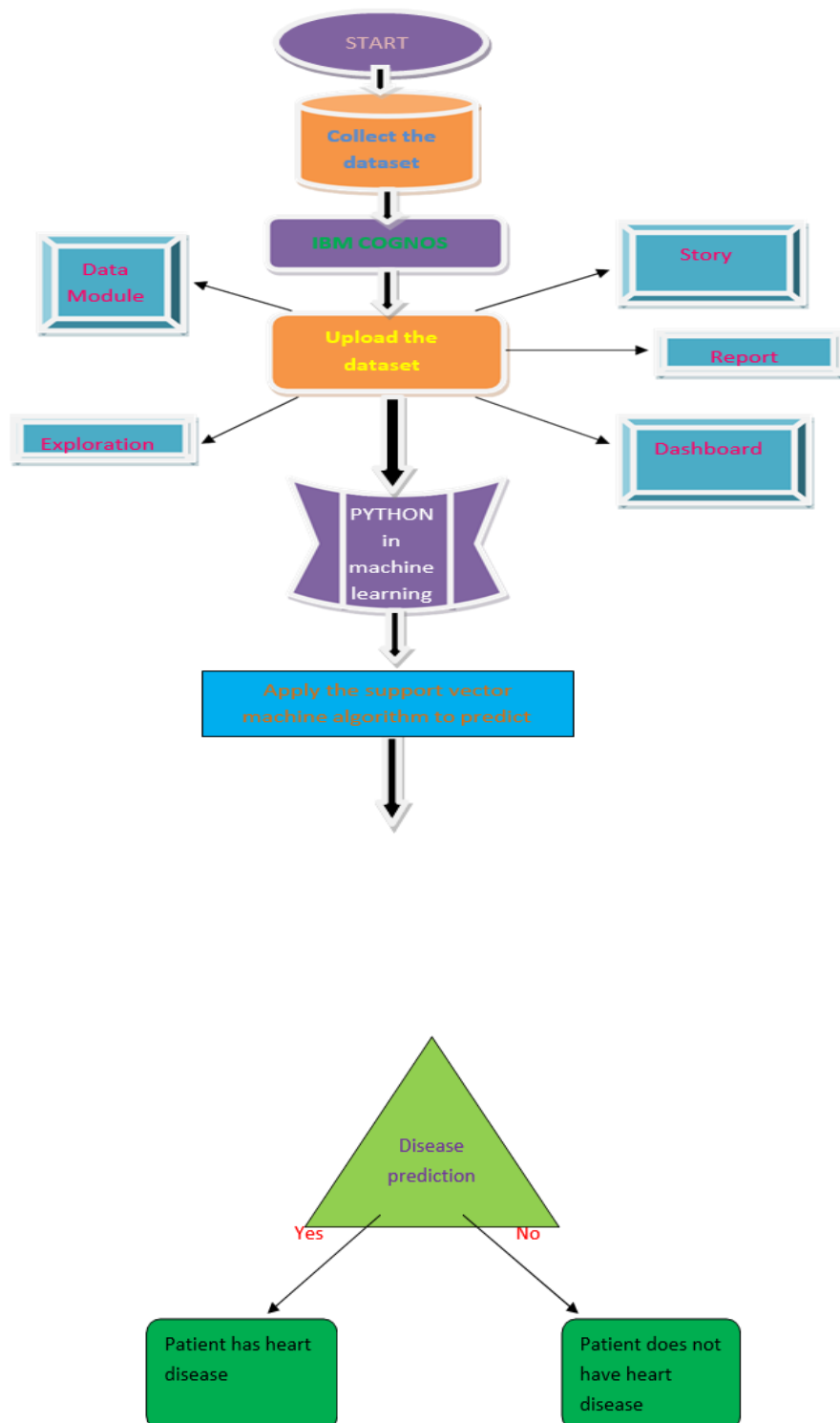
Solution & Technical Architecture
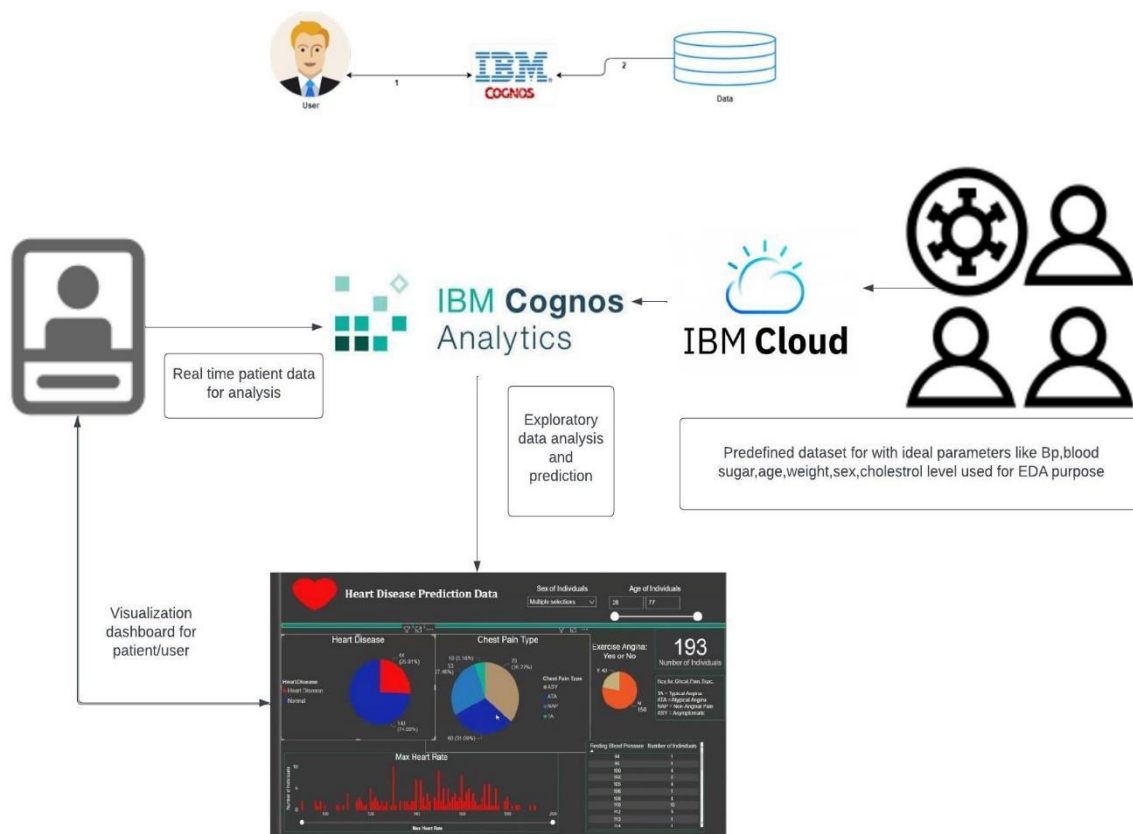
5.3 User stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application 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 application	I can receive confirmation email & click confirm.	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login.	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register & access the dashboard with Gmail Login	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can register & access the dashboard	High	Sprint-1

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
				with Gmail Login		
	Dashboard	USN-6	Profile - view & update your profile	I can see the profile.	Medium	Sprint-2
		USN-7	Change Password - user can change the password	I can able to change the password.	High	Sprint-1
		USN-8	Home - Analyze your Heart	I can detect the health condition from where ever I want.	High	Sprint-1
		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)	These are the categories available in that application.	High	Sprint-2
		USN-10	View Doctors - view doctor detail by searching by names or filter by specialty	Using this application, people can known that the speciality doctors.	Medium	Sprint-1

Customer (Web user)	System Requirement	USN-11	I. Hardware Requirement
			
			i. Laptop or PC
			    • I5 processor system or higher
			    • 4 GB RAM or higher
				These are all the specification available in your PC.	High	Sprint-2

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
				with Gmail Login		
	Dashboard	USN-6	Profile - view & update your profile	I can see the profile.	Medium	Sprint-2
		USN-7	Change Password - user can change the password	I can able to change the password.	High	Sprint-1
		USN-8	Home - Analyze your Heart	I can detect the health condition from where ever I want.	High	Sprint-1
		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)	These are the categories available in that application.	High	Sprint-2
		USN-10	View Doctors - view doctor detail by searching by names or filter by specialty	Using this application, people can known that the speciality doctors.	Medium	Sprint-1
Customer (Web user)	System Requirement	USN-11	I. Hardware Requirement i. Laptop or PC • I5 processor system or higher • 4 GB RAM or higher	These are all the specification available in your PC.	High	Sprint-2

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
			• 128 GB ROM or higher ii. Android Phone (12.0 and above)			
		USN-12	II. Software Requirement iii. Laptop or PC • Windows 10 or higher • Android Studio	Install your application. This system can be used to predict the presence of heart disease.	Medium	Sprint-2
		USN-13	*Reference-* https://ieeexplore.ieee.org/document/9619208/	Go and Check our Reference link.	Medium	Sprint-1
Customer Care Executive	Dashboard	USN-14	Query	You can post your queries in the text box available in that application.	High	Sprint-1
		USN-15	Toll Free	Ask your doubts in given number(8365492107).	High	Sprint-1
		USN-16	Ratings	Give your ratings as your wish.	Medium	Sprint-1
Administrator	Dashboard	USN-17	Verification	Verification through CAPTCHA Verification through I'm not a robot	High	Sprint-1
		USN-18	validation	Reconfirming the new password Sending a two digit number in (Google account) your Old devices, so that you can enter into	High	Sprint-2

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
				a new device By entering the two digit number.		
		USN-19	Feedback - send feedback to the Admin.	Please send your feedback to host.	Medium	Sprint-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

REPORTS FROM JIRA

Chart Title

PLANNED   ACTUAL
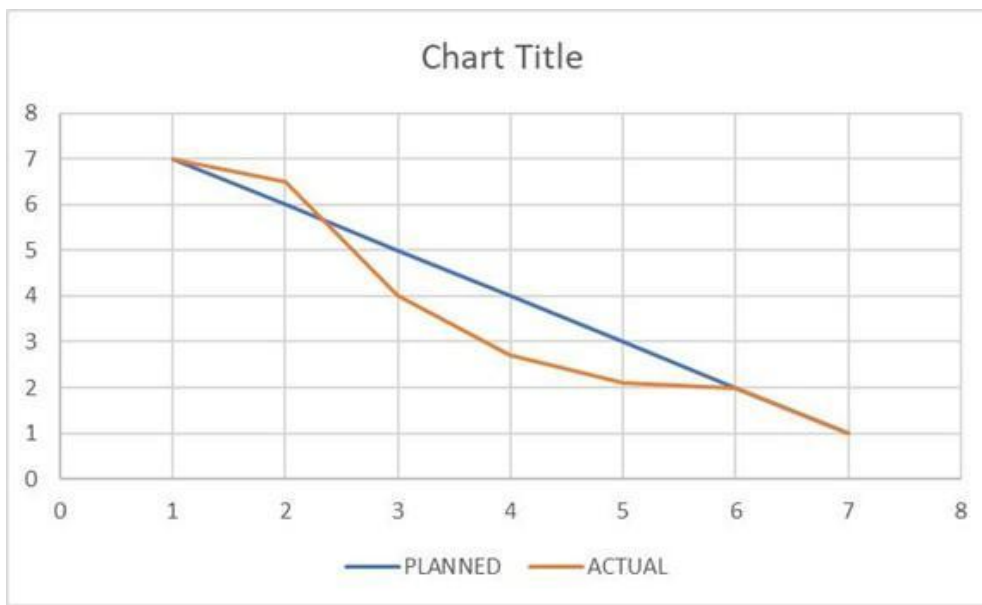
5. CODING & SOLUTIONIN

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:

```html
<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>
```

```html
        <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</opt
```

```html
</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>
```

```html
</center>

</b>

</table>

  </form>

</center>

</div>
```

Feature 2

Working :

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

```python
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
```

```python
#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)
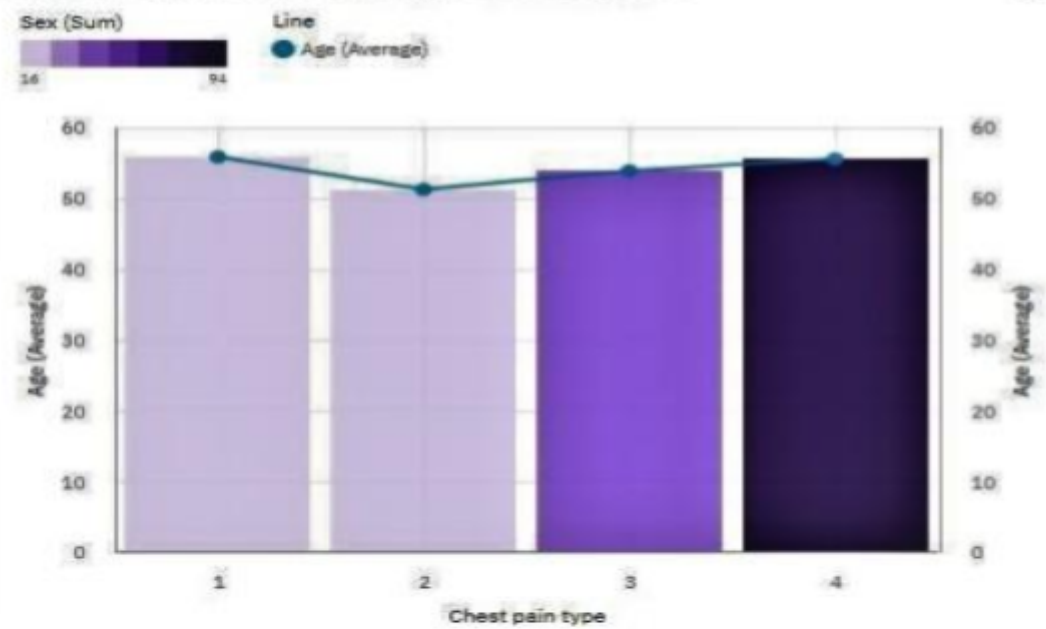
BP by Age
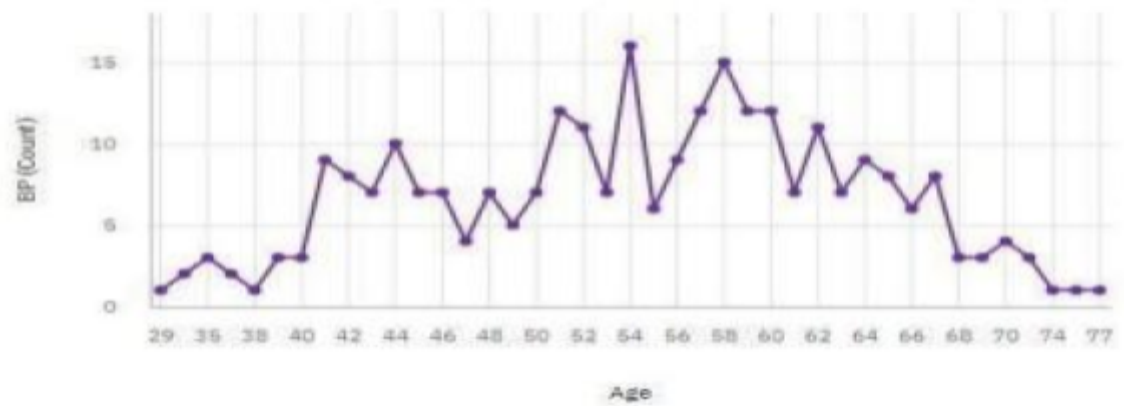
Cholesterol by Sex and Age colored by Sex

Sex (Sum)
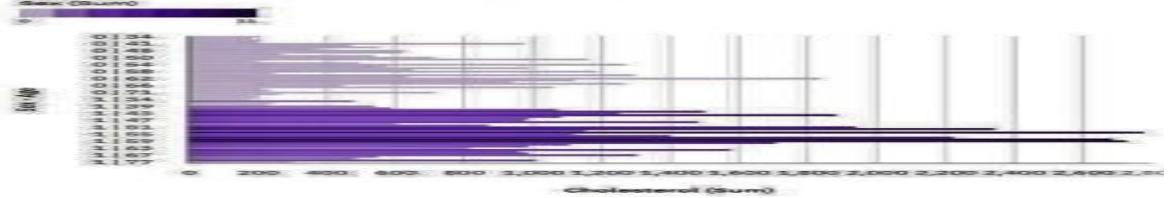
4.TESTING

Test Cases

Test caseID	Feature Type	Component	Test Scenario
LoginPage_TC_OO 1	Functional	Home Page	Verify user is able to see the Login/Signup popup when user clicked on My account button
LoginPage_TC_OO 2	UI	Home Page	Verify the UI elements in Login/Signup popup
LoginPage_TC_OO 3	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 into application with InValid credentials

Testcase ID	Feature Type	Component	TestScenario
LoginPage_TC_OO5	Functional	Login page	Verify user is able to log into application 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.

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

5.RESULTS

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

Locust Test Report
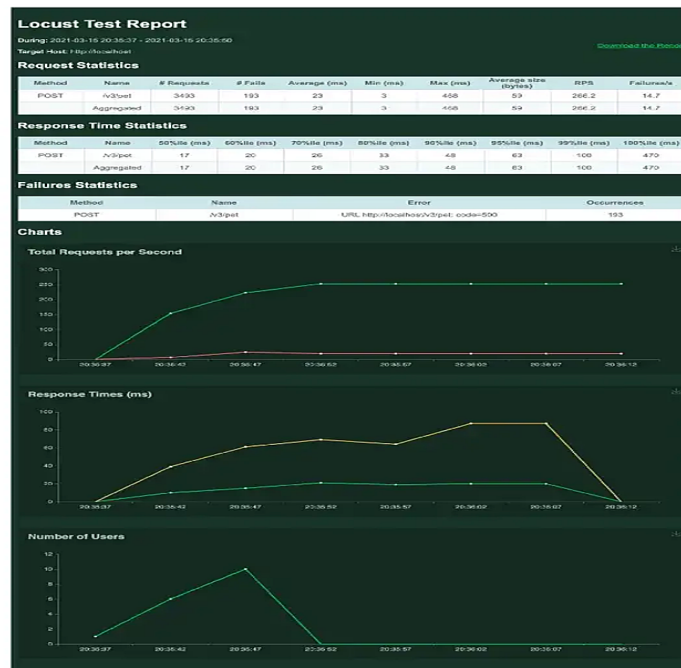
6. ADVANTAGES AND DISADVANTAGES

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.

7. 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%.

8. 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

9. 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>
```

```html
<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>
```

```html
        <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>
```

```html
<p>By creating an account you agree to our Terms & Privacy.</p>

<button                                type="submit"
 class="registerbtn"><strong><a
href="">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:
```html
<head>
<link rel="stylesheet" type="text/css" href="css/style.css"/>
</head>
<!DOCTYPE html>
```

```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="">login</a></strong></button>


   </form>
</div>
</center>
</body>
```

</html>

VISUALIZING AND PREDICTING HEART DISEASE

```python
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')
```

```python
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()
```

```python
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=cross_val_score(randomforest_classifier,X,y,cv=10)

score.mean()

PROJECT DEMO LINK:

https://drive.google.com/file/d/1qedYKAZxNPEgcO7OT_k8w3um9FdD5kuG/view?usp=sharing

GITHUB :

https://github.com/IBM-EPBL/IBM-Project-22961-1659862481