VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

IBM NALAIYA THIRAN 2022

Submitted By:

Team ID: PNT2022TMID21213

Team Members:

9177119C013- Charu Sneha L R

9177119C046- Kushagra Kapoor

9177119C074 – Priyavarthan N

9177119C108 – Uma Maheswari T

1. INTRODUCTION

1.1. Project Overview

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 advanced data mining techniques are used. In this study, a Heart Disease Prediction System (HDPS) is developed using Logistic Regression, Linear Discriminant Analysis (LDA), Naive Bayes, SVM, KNN and Decision Tree algorithms for predicting the risk level of heart disease. The system uses 14 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction. The HDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge. E.g., Relationships between medical factors related to heart disease and patterns, to be established. We have employed the multi layer perceptron neural network with back propagation as the training algorithm. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases.

1.2. Purpose

The developed project predicts the likelihood of patients getting heart disease. It enables significant knowledge, e.g., relationships between medical factors related to heart disease and patterns to be established

2. LITERATURE SURVEY

2.1. Existing problem

Now-a-days, cardiovascular diseases are the main reason for huge number of deaths in the world in the last few decades and has emerged as the most life- threatening disease, not only in India but in the whole world. Approximately one person dies per minute due to heart disease. So, there is a need of reliable,accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data.

Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related.

A brief survey of that is presented here.

S.NO	Author	Title	Name of the Journal/Conference	Volume/ Issue/Year	Algorithm/Method
1.	Yash Jayesh Chauhan	Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning	*"International Journal of Science and Research (IJSR)*	*ISSN: 2319-7064* *Vol 9, issue 5,* *May2020*	they proposed a model which is based on a combination of standard machine learning algorithms such as Logistic Regression, Random Forest, K- Nearest Neighbors (KNN), support vector machine (SVM) and Decision Tree to predict whether the patient is having a Cardiovascular disease or not.
2.	Vijay Kumar T, Vinoth kanna B	Fusion based Feature Extraction Analysis of ECG Signal Interpretation – A Systematic Approach	*Journal of Artificial Intelligence and Capsule Networks (2021)*	*Vol.03/ No.01.* *March 11,2021*	Supervised and Unsupervised ML algorithms.
3.	Vamsidhar Talasila1K otakonda Madhubab u1 Meghana Chakravart hy Mahadasya m1 Naga Jyothi Atchala1 Lakshmi Sowjanya Kande1	The Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics	*International journal of Intelligent Engineering and systems.*	*January 17, 2020*	Rough Set Theory (RST) technique is used to select the most relevant features, which helps to provide the efficient classification of medical data and disease detection. The selected features are given as input to the Recurrent Neural Network (RNN) technique for disease prediction.

2.2. References

1. Yash Jayesh Chauhan Cardiovascular Disease Prediction using Classification Algorithms of *"International Journal of Science and Research (IJSR) ISSN: 2319- 7064 Vol 9, issue 5.*

2. Vijay Kumar T, Vinoth kanna B Fusion based Feature Extraction Analysis of ECG Signal Interpretation – A Systematic Approach Journal *of Artificial Intelligence and Capsule Networks (2021)* Vol.*03/ No.01. March 11,2021.*

3. Vamsidhar Talasila1Kotakonda Madhubabu1 Meghana Chakravarthy Mahadasyam1 Naga Jyothi Atchala1 Lakshmi Sowjanya Kande1 the Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics International *journal of Intelligent Engineering and systems. January 17, 2020.*

2.3. Problem Statement Definition

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it is expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, 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 be used for health diagnosis in medicinal data.

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.

- Data Exploration: To find an appropriate heart disease detection dataset and perform preprocessing.
- Exploratory Data Analysis: To perform Exploratory Data Analysis on the preprocessed dataset and getting meaningful patterns and results.
- Feature Engineering: To alter the features such that they can move forward for model building phase.
- Model Building: To create visualizations on the patterns and insights obtained from the data and creating an interactive Dash Board containing the visualizations.

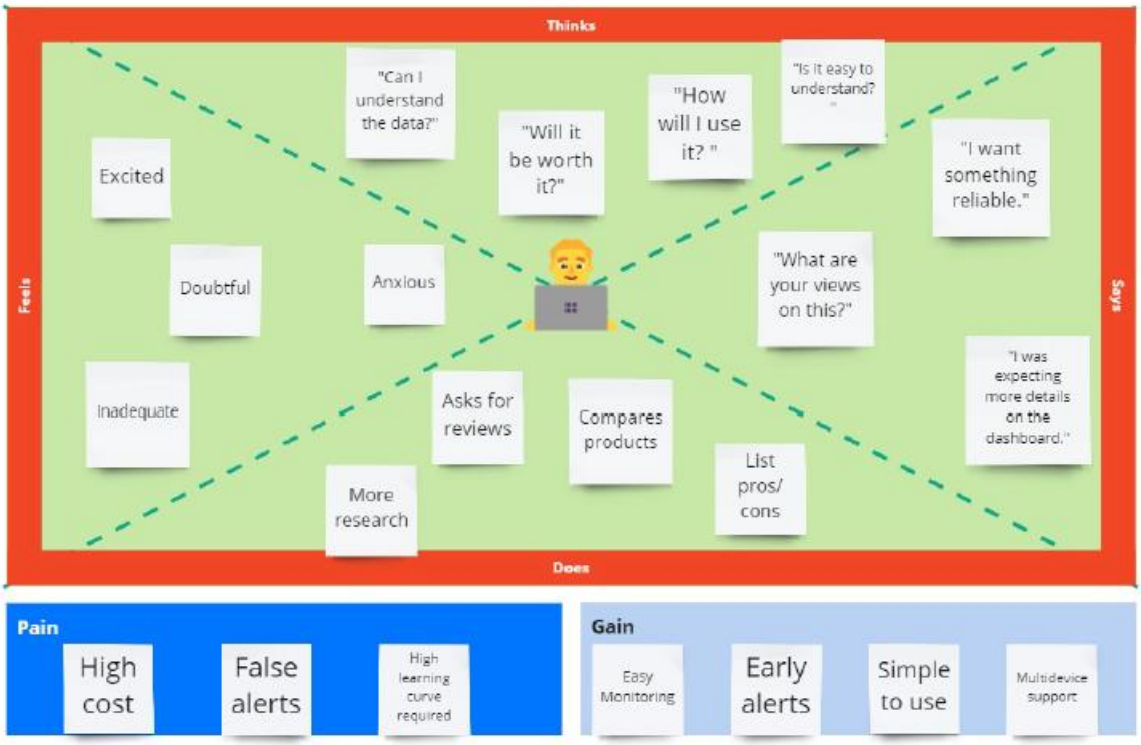
3. IDEATION & PROPOSED SOLUTION

3.1. Empathy Map Canvas

An empathy map is a collaborative visualization used to articulate what we know about a particular type of user. It externalizes knowledge about users in order to

1) create a shared understanding of user needs, and

2) aid in decision making.



The **Says** quadrant contains what the user says out loud in an interview or some other usability study. Ideally, it contains verbatim and direct quotes from research.

The **Thinks** quadrant captures what the user is thinking throughout the experience. Ask yourself (from the qualitative research gathered): what occupies the user's thoughts? What matters to the user? It is possible to have the same content in both Says and Thinks. However, pay special attention to what users think, but may not be willing to vocalize. Try to understand why they are reluctant to share — are they unsure, self-conscious, polite, or afraid to tell others something?

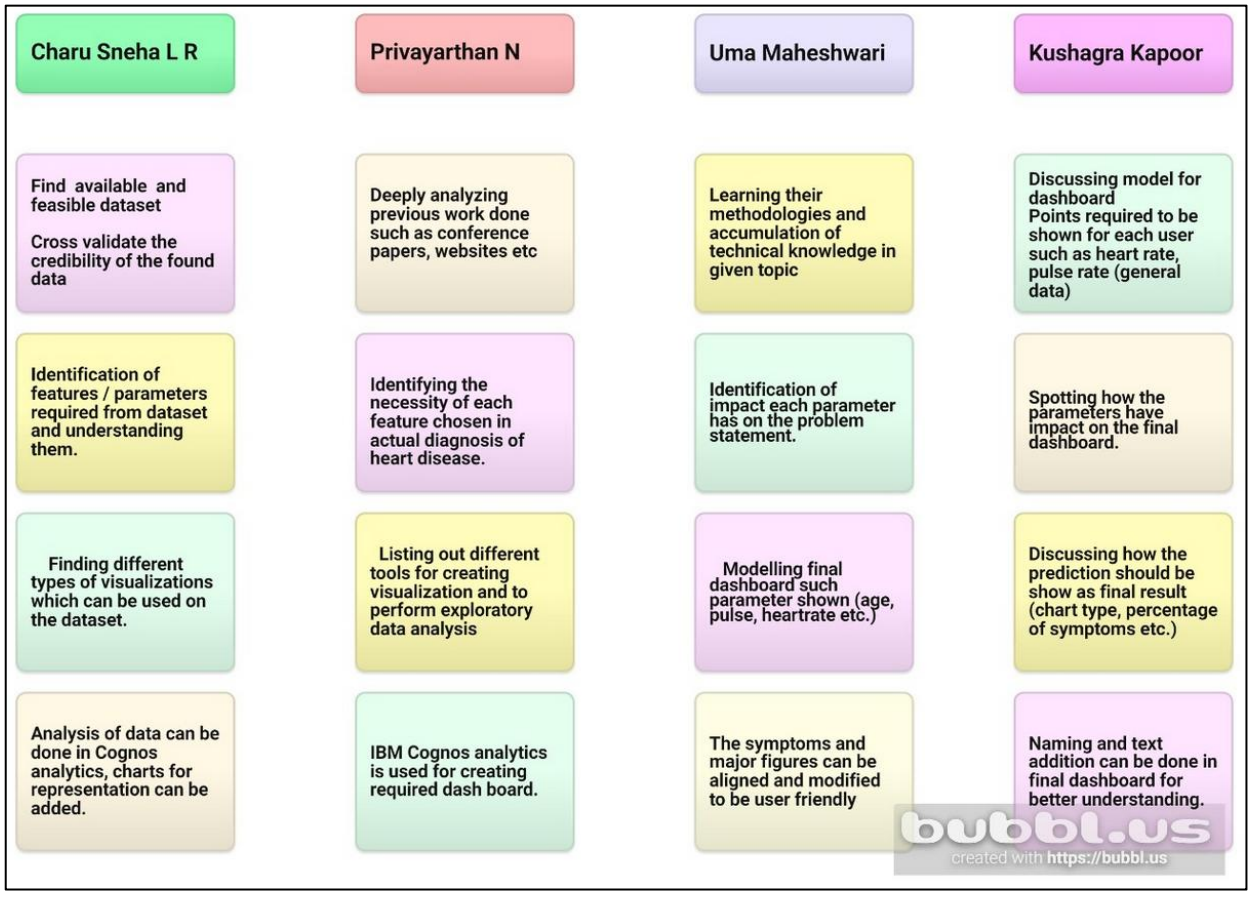
The **Does** quadrant encloses the actions the user takes. From the research, what does the user physically do? How does the user go about doing it?

The **Feels** quadrant is the user's emotional state, often represented as an adjective plus a short sentence for context. Ask yourself: what worries the user? What does the user get excited about? How does the user feel about the experience?

3.2. Ideation & Brainstorming

Ideation is the process where you generate ideas and solutions through sessions such as Sketching, Prototyping, Brainstorming, Brainwriting, Worst Possible Idea, and a wealth of other ideation techniques. Ideation is also the third stage in the Design Thinking process.

Brainstorming is a method design teams use to generate ideas to solve clearly defined design problems. In controlled conditions and a free-thinking environment, teams approach a problem by such means as "How Might We" questions. They produce a vast array of ideas and draw links between them to find potential solutions.



3.3. Proposed Solution

1. Problem Statement (Problem to be solved): Heart disease refers to several types of abnormalities in heart conditions. The leading cause of death is heart disease. It is infeasible for a common man to frequently undergo tests for ECG and so on. Hence, there needs a replacement for this, which must be handy and reliable.

2. Idea / Solution description: The idea behind the proposed solution is to propose an interactive dashboard for visualizing and predicting heart diseases in which user can view his/her medical report analysis and the predicted result. The dashboard will be generated using IBM Cognos. First the data set will be explored and pre-processed. The K-nearest neighbours, support vector machine and Decision tree classifier will be used for prediction. The best scored out of the above will be taken for final prediction and display.

3. Novelty / Uniqueness: The novelty behind the proposed system is to provide suggestions to the user in non-medical way and to help them understand their body level. This will help them take required precautions and tune to their body requirement.

4. Social Impact / Customer Satisfaction: This helps the user, doctor or concerned personal understand the patient and to take better decisions according to situation. It is useful in predicting the disease in an earlier stage and makes the user alert about his current condition periodically.

5. Business Model (Revenue Model) This interactive dashboard for heart disease prediction can be deployed in Health care centres and Hospitals, so that it makes the analysis in a fast manner.

6.Scalability of the Solution The proposed solution will work efficiently in both smaller and larger datasets in a similar manner. In future, this can be scaled to other diseases too.

3.4. Problem Solution fit

The Problem-Solution Fit canvas is based on the principles of Lean Startup, LUM (Lazy User Model) and User Experience design. It helps entrepreneurs, marketers and corporate innovators identify behavioral patterns and recognize what would work and why. It is a template to help identify solutions with higher chances of solution adoption, reduce time spent on testing and get a better overview of the current situation.

It helps you to:

- Solve complex problems in a way that fits the state of your customers.
- Succeed faster and increase your solution adoption by tapping into existing mediums and channels of behaviour.
- Sharpen your communication and marketing strategy with the right triggers and messaging.
- Increase touch-points with your company by finding the right problem-behaviour fit and building trust by solving frequent annoyances, or urgent or costly problems.
- Understand the existing situation in order to improve it for your target group.

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

- Doctors in hospitals
- Clinics
- Health Centers

E.g.: Doctors can use this along with the patients' medical data to analyze the risk of heart disease.

6. CUSTOMER CONSTRAINTS CC

- Budget
- No accuracy in prediction
- Interactive Dashboards
- Network Connection
- Need of dataset
- There is no awareness about the

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 notetaking

- Customers can go to the doctor for a medical checkup.
- Based on the test results, doctors will advise them.
- The patient can do manual prediction

Explore AS, differentiate

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

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

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

- Visualizations give doctors very good insights on the potential chances for a patient to get heart disease.
- It is also very useful to explain to patients so that they can easily understand the risk factor and take care of themselves to reduce the likelihood of getting heart disease.
- Standard of Data: The outcome is fully depends on the accurate and relative dataset
- Lives depending on medical support

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 inregulations.

Buildup of fatty plaques in the arteries is the most common cause of coronary artery disease.

Not storing and analyzing data properly to help doctors makeinformed decisions

Increasing in heart disease will not be identified firstly is major reason.

There is a possibility of consideringevery heart disease as same

There is no idea about relationbetween similar heart disease

7. BEHAVIOUR BE

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

i.e. directlyrelated: findtherightsolarpanel installer, calculateusageandbenefits; indirectly associated: customers spend free time on volunteering work (i.e.Greenpeace)

Regular, daily physical activity can lower the risk of heart disease. Physical activity helps control your weight.
A healthy diet can help protect the heart, improve blood pressure and cholesterol, and reduce the risk of type 2 diabetes.
One of the best things you can do for your heart is to stop smoking or using smokeless.tobacco.
Even if you're not a smoker, be sure to avoid secondhand smoke.
Maintain a healthy weight
Get good quality sleep
Manage stress

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

3. TRIGGERS TR

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

Lifestyle changes ,Lives depending on medical support,need to search for heart specialist with manageable price , need to apply for health insurance , Financial insecurity ,Anxiety ,shortness of breath ,may feel emotional stress ,may feel chest pain,chest tightness,chest pressure and feel for fatigue

4. EMOTIONS: BEFORE / AFTER EM

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

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

Feeling afraid and depressed.
Develop a feeling of awareness which mean people
There is huge uncertainity in knowing the accurate and correct Reason for a disease and predicting it.

10. YOUR SOLUTION SL

If you are working on an existing business, write down your currentsolution first, fill inthecanvas, andcheckhowmuchit fits reality.

Ifyou areworkingon anew businessproposition, thenkeepit blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior.

To clean data and provide visualizations tohelp doctors in their diagnosis of patient aswell as make customers more aware of thisissue.
This can help to prevent casualties and to take action either from immediate medical help or by self remedies.

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE

What kind of actions do customers take online? Extract online channels from #7

8.2 OFFLINE

Whatkindofactionsdocustomerstakeoffline? Extractoffline channels from #7 and use them for customer development.

ONLINE:
Users look at the data and compare it withtheir testresults Upload data.Prepare data,Exploration of data.
OFFLINE: Doctors use it as a tool to diagnosepatients and make accurate predictions.

4. REQUIREMENT ANALYSIS

4.1. Functional requirement

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in use cases.

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 LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics and understand the insights.
FR-4	Generating Report	User can view their health reports and share them with their family doctor.

4.2. Non-Functional requirements

Nonfunctional Requirements (NFRs) define system attributes such as security, reliability, performance, maintainability, scalability, and usability. They serve as constraints or restrictions on the design of the system across the different backlogs.

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

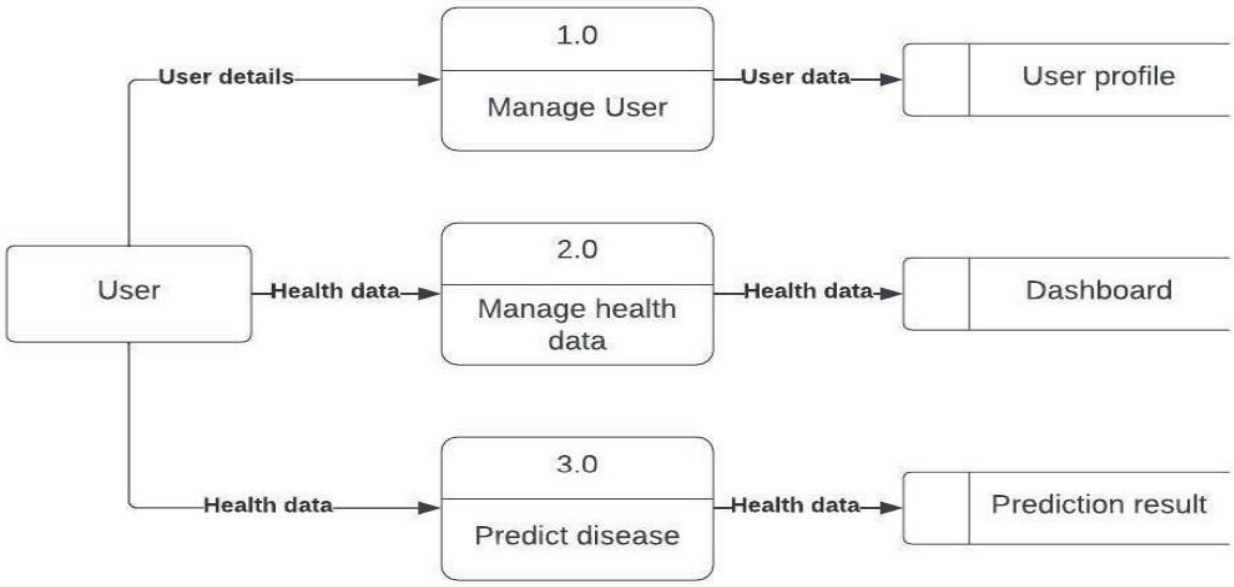
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user- friendly graphical interface. Users will be able to understand and use all the features of the application easily. The design

		will be improved regularly and made sure that it is interactive.
NFR-2	**Security**	Encryptions can be utilized to ensure that the data and records about the users are kept safe. Passwords must be used everywhere wherever an access to critical data is required.
NFR-3	**Reliability**	Storage infrastructure can be made reliable by making use of backup systems. Also, the systems made should be fault tolerant.
NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the data submission. Algorithms designed should be very efficient and make use of minimum resources and yield maximum output.
NFR-5	**Availability**	The application must be available 24 x 7 for users without any interruption. Proper use of backup servers can help in achieving such targets.
NFR-6	**Scalability**	The application should support large number of users and must be able to grow as per the demand.

5. PROJECT DESIGN

5.1. DATA FLOW DIAGRAMS:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



Flow:

1) User creates an account in the application.
2) User enters the medical records in the dashboard.
3) User can view the visualizations of trends in the form of graphs and charts for his/her medical records with the trained dataset.
4) User can view the accuracy of probability of occurrence of heart disease inthe dashboard.

5.2. Solution and Technical Architecture

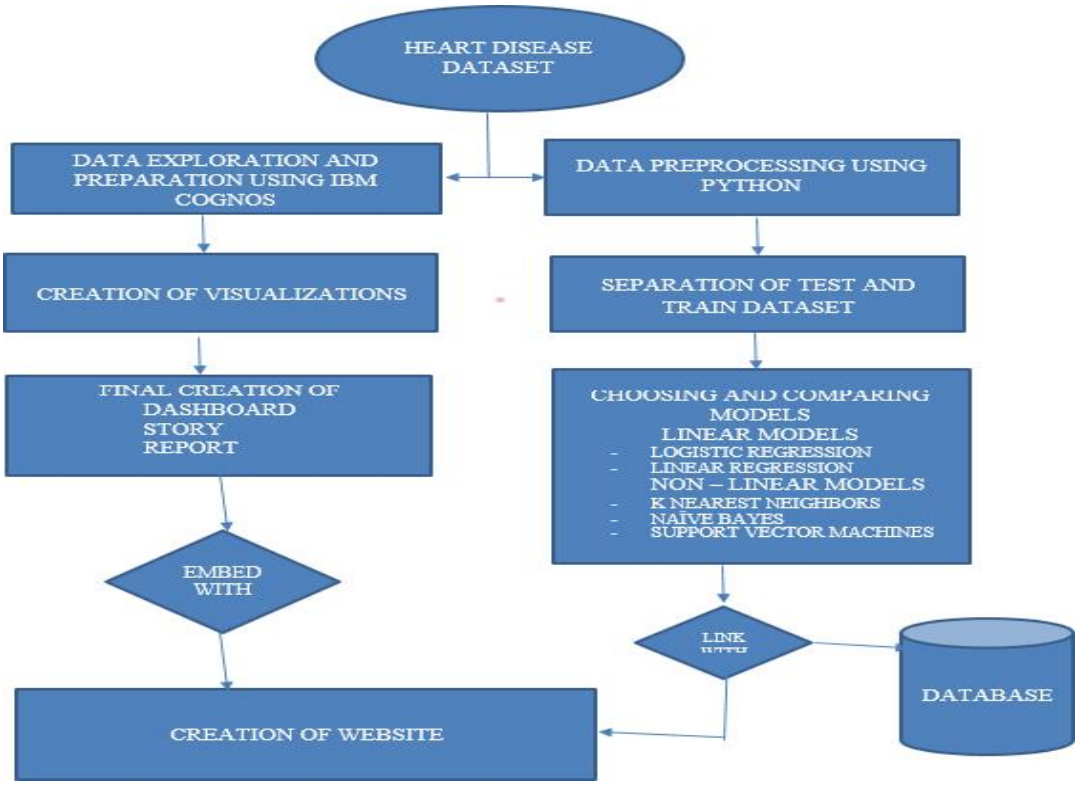


Table-1: Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	The user interacts with the application throughweb UI.	HTML, CSS, python, JS, Bootstrap
2.	Application Logic-1	Logic for login in the application	JS
3.	Application Logic-2	Logic for registration in the application	JS
4.	Application Logic-3	Logic for a process in the application	Python, JS

5.	Cloud Database	Database Service on Cloud	IBM DB2
6.	File Storage	To store files such as prediction report	Local Filesystem
7.	Data Analytics Model	Predictive modeling solutions are a form of data-mining technology that works by analyzing historical and current data and generating a model to help predict future outcomes.	Predictive modeling
8.	Infrastructure (Server / Cloud)	Application Deployment on Local System	Local web server

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Bootstrap	Bootstrap
2.	Security Implementations	Basic HTTP authentication, Session basedauthentication, User Registration, Login Tracking	JS
3.	Scalable Architecture	Here we can add any number of data and user to our database which will process them and store them/ For higher data we can connect to cloud storage and use it. Also deployment of website in cloud will help us scale it .ML model logic will work on any number of population set.	Python, IBM cloud
4.	Availability	Higher compatibility with latest technologies and allows customization	Python , IBM cloud
5.	Performance	Integrated support for unit testing. The security and authentication is added as required . Support for secure cookies (client side sessions) 100% WSGI 1.0 compliant.	JS, Bootstrap

5.3. User Stories

In Agile a user story is a short, informal, plain language description of what a user wants to do within a software product to gain something they find valuable.

User stories typically follow the role-feature-benefit pattern (or template):

- As a [type of user],
- I want [an action]
- so that [a benefit/value]

With user stories you give a development team the context and the why of what they're creating. Doing so helps them understand how they're providing value for the business and to keep the user/customer top of mind.

User stories provide the essence needed to prioritize them.

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
Web User		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
Tab User		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	Medi	Sprint-1

				Gmail Login.	um	
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	USN-6	As a user, I can see all the details easily displayed in dashboard.		High	Sprint-3

6.PROJECT PLANNING & SCHEDULING

6.1. Sprint Planning & Estimation

We have planned to divide the tasks in several sprints in the following way

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Members**
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	Kushagra Kapoor, Uma Maheswari
Sprint-1	Confirmation	USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	Charu Sneha, Uma Maheswari
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email &password	1	High	Priyavarthan, Kushagra Kapoor
Sprint-1	User Interface	USN-4	As a user, I should not need any pre requisites to handle the UI	1	Medium	Charu Sneha, Priyavarthan

Sprint -1	Data Preprocessing	USN-5	As a user, I will cleanse the dataset before building the model by using EDA	2	High	Charu Sneha, Priyavarthan
Sprint-2	Data Visualization	USN-6	As a user I can perform the various data visualizations and view the accuracy of the trained ensembling model.	1	Medium	Priyavarthan, Uma Maheswari
Sprint-2	Prediction	USN-7	As a user, I can check whether there is any heart disease or not	2	High	Uma Maheswari, Kushagra Kapoor
Sprint-3	Dashboard	USN-9	As a user, I Can view all the visualizations Related to heart disease prediction such as whether I have a heart disease or not, BP vs cholesterol etc.,	2	High	Priyavarthan, Kushagra Kapoor
Sprint-3	Present data	USN-8	As a user, will present the data in the IBM cognos analytics platform	2	High	Priyavarthan, Charusneha
Sprint-4	Embedding dashboard, story and report to website	USN-10	As, the user can view the dashboard, story and report after logging to the website.	1	Medium	Kushagra Kapoor, Charu Sneha
Sprint-4	Database Connection to the website	USN-11	As the user, I can store and view all the user details entered in the website	2	High	Uma Maheswari, Charu Sneha
Sprint-4	Deployment in cloud	USN-12	As a user, I will deploy the model into the IBM cloud	2	High	Kushagra Kapoor, Uma Maheswari

6.2. Sprint Delivery Schedule

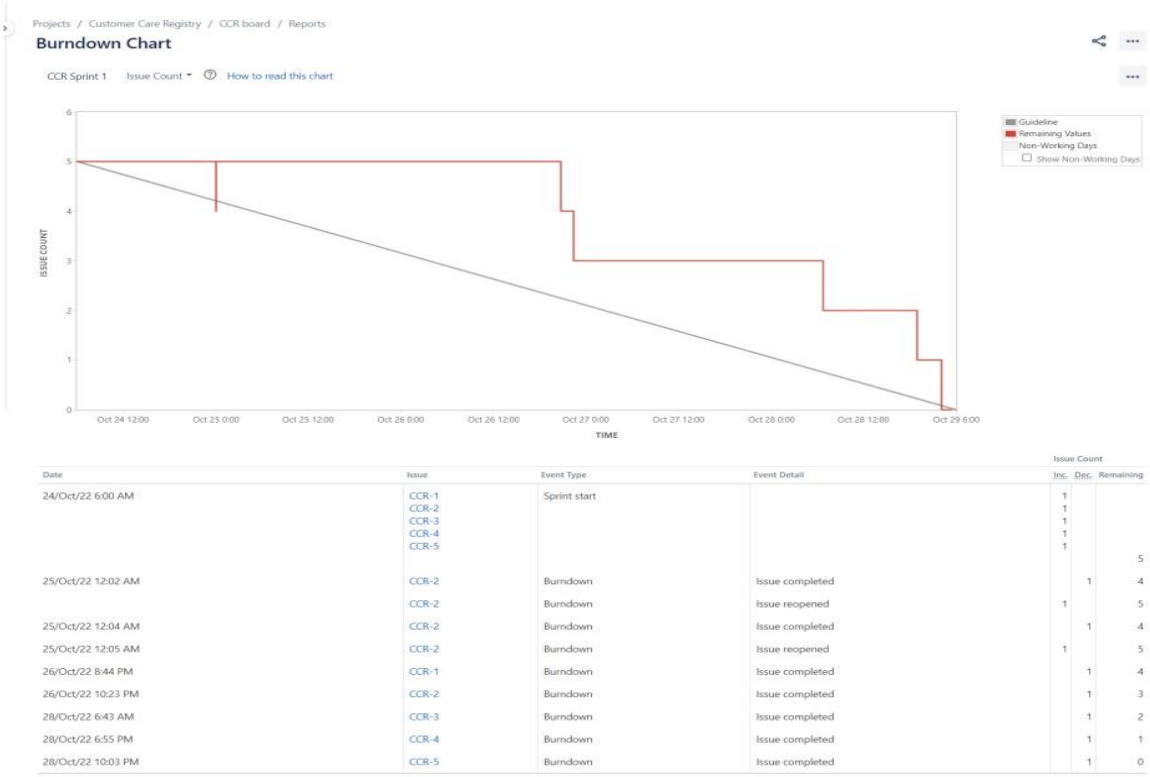
Sprint	Total Story Points	Duration	print Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	27 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	30	01 Nov 2022
Sprint-3	20	6 Days	06 Nov2022	10 Nov 2022	49	07 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	50	18 Nov 2022

Project Tracker, Velocity & Burndown Chart:

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

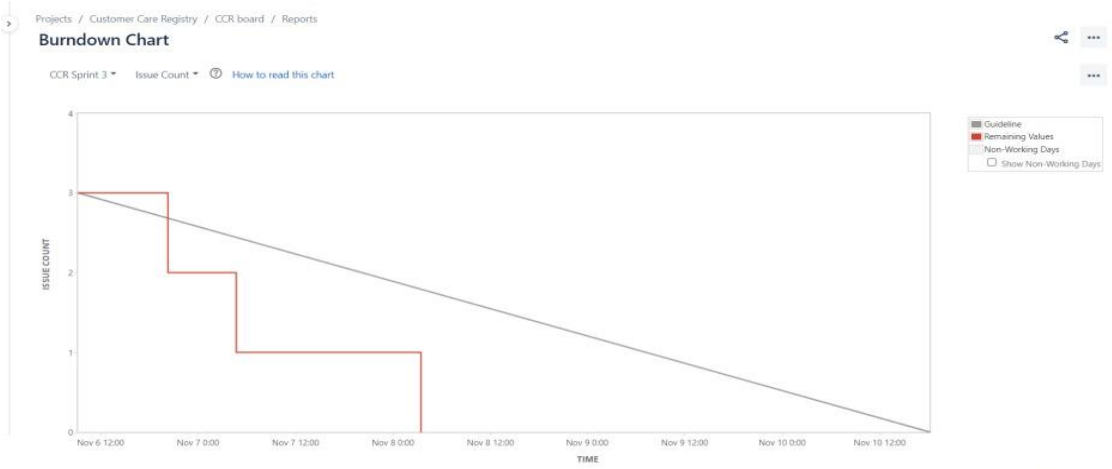
$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

6.3. Reports from Jira



Date	Issue	Event Type	Event Detail	Issue Count Inc.	Dec.	Remaining
24/Oct/22 6:00 AM	CCR-1 CCR-2 CCR-3 CCR-4 CCR-5	Sprint start		1 1 1 1 1		5
25/Oct/22 12:02 AM	CCR-2	Burndown	Issue completed		1	4
	CCR-2	Burndown	Issue reopened	1		5
25/Oct/22 12:04 AM	CCR-2	Burndown	Issue completed		1	4
25/Oct/22 12:05 AM	CCR-2	Burndown	Issue reopened	1		5
26/Oct/22 8:44 PM	CCR-1	Burndown	Issue completed		1	4
26/Oct/22 10:23 PM	CCR-2	Burndown	Issue completed		1	3
28/Oct/22 6:43 AM	CCR-3	Burndown	Issue completed		1	2
28/Oct/22 6:55 PM	CCR-4	Burndown	Issue completed		1	1
28/Oct/22 10:03 PM	CCR-5	Burndown	Issue completed		1	0

Date	Issue	Event Type	Event Detail	Issue Count Inc.	Dec.	Remaining
31/Oct/22 12:01 AM	CCR-10 CCR-11 CCR-12	Sprint start		1 1 1		3
31/Oct/22 11:28 PM	CCR-10	Burndown	Issue completed		1	2
01/Nov/22 11:34 PM	CCR-12	Burndown	Issue completed		1	1
	CCR-12	Burndown	Issue reopened	1		2
01/Nov/22 11:45 PM	CCR-11	Burndown	Issue completed		1	1
04/Nov/22 12:17 AM	CCR-12	Burndown	Issue completed		1	0
	CCR-10 CCR-11 CCR-12	Sprint ended by Bala Abinesh Surya				0 0 0 0



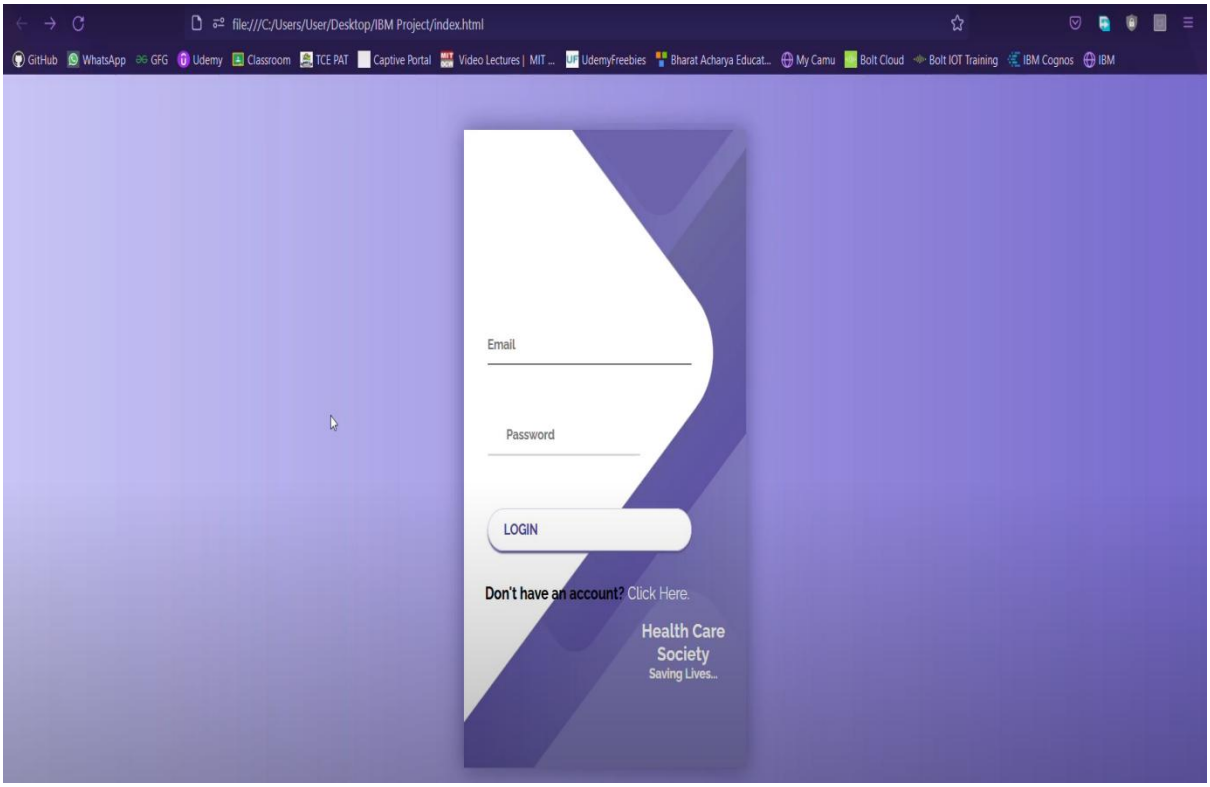
Date	Issue	Event Type	Event Detail	Issue Count Inc.	Dec.	Remaining
06/Nov/22 9:11 AM	CCR-13 CCR-14 CCR-15	Sprint start		1 1 1		 3
06/Nov/22 8:17 PM	CCR-14	Burndown	Issue completed		1	2
07/Nov/22 7:30 PM	CCR-15	Burndown	Issue completed		1	1
08/Nov/22 6:16 PM	CCR-13	Burndown	Issue completed		1	0
08/Nov/22 6:18 PM	CCR-13 CCR-14 CCR-15	Sprint ended by Bala Abinesh Surya				0 0 0 0

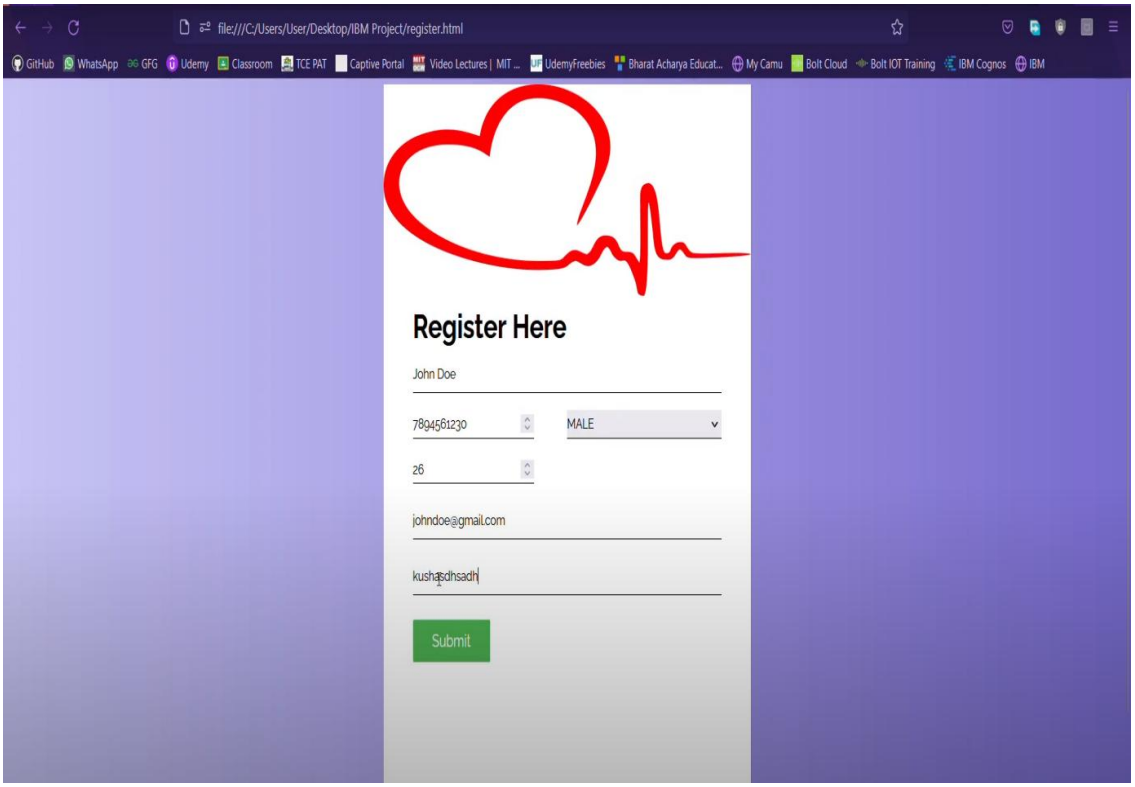
7. CODING & SOLUTIONING

7.1. Feature 1

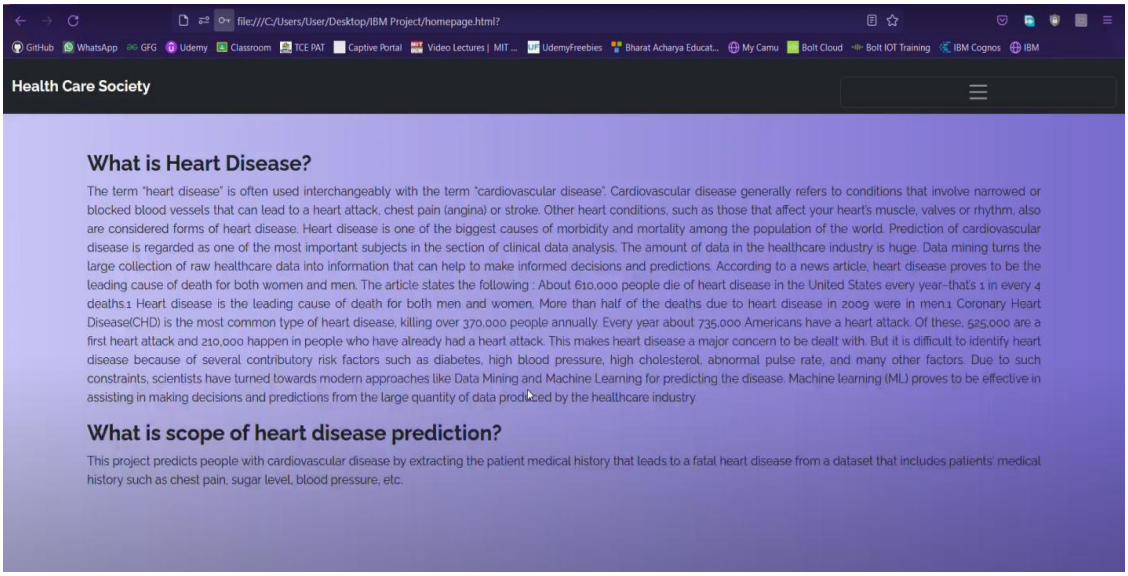
Website:

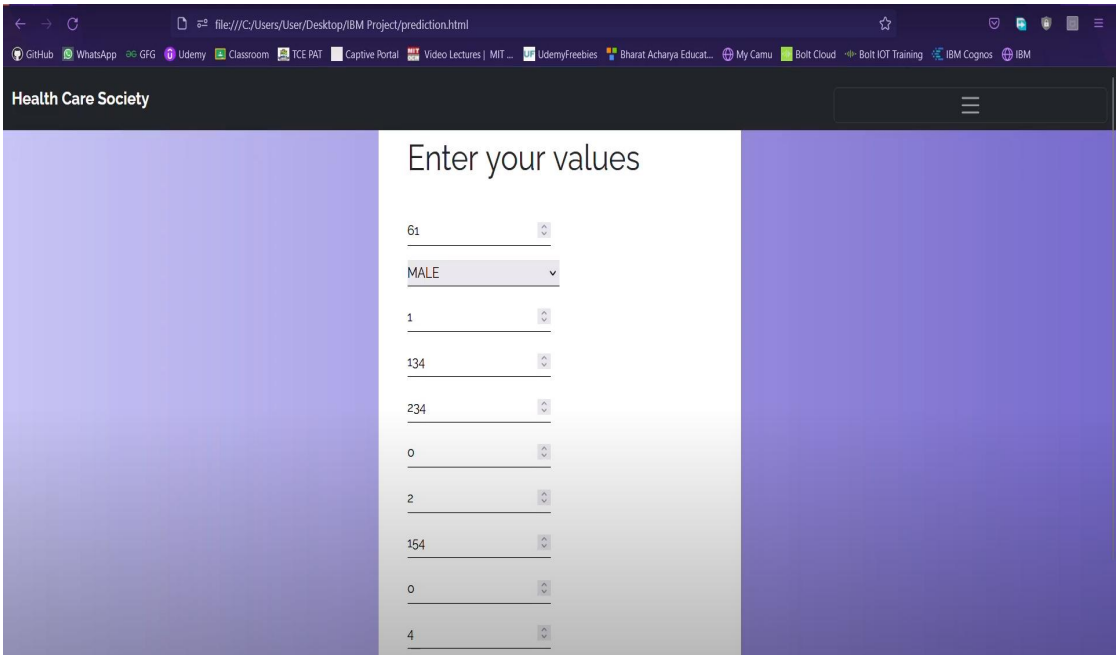
A user can register in a website by providing their personal details such as name, age and email to create the account and can login to a website through registered Email and password.

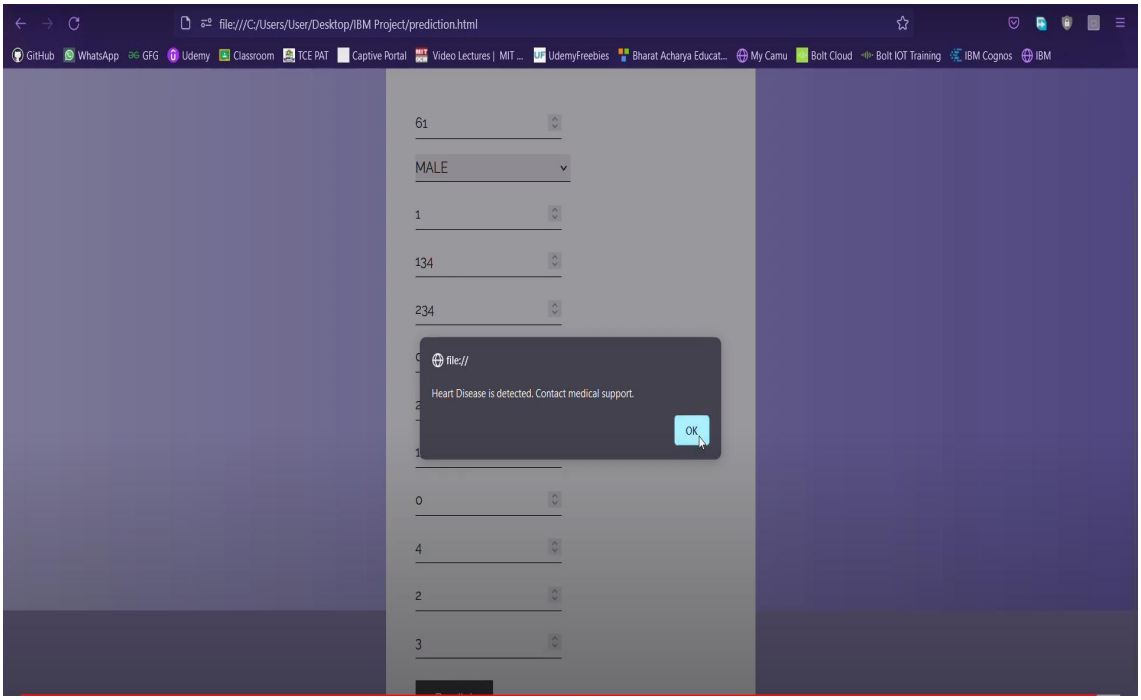

file:///C:/Users/User/Desktop/IBM Project/index.html
GitHub WhatsApp GFG Udemy Classroom TCE PAT Captive Portal Video Lectures | MIT ... UdemyFreebies Bharat Acharya Educat... My Camu Bolt Cloud Bolt IOT Training IBM Cognos IBM
Email
Password
LOGIN
Don't have an account? Click Here.
Health Care Society
Saving Lives...


file:///C:/Users/User/Desktop/IBM Project/register.html
GitHub WhatsApp GFG Udemy Classroom TCE PAT Captive Portal Video Lectures | MIT ... UdemyFreebies Bharat Acharya Educat... My Camu Bolt Cloud Bolt IOT Training IBM Cognos IBM
Register Here
John Doe
7894561230
MALE
26
johndoe@gmail.com
kushagdhsadh
Submit

This is the home page of our website in which a short description is provided about the heart disease prediction and its scope





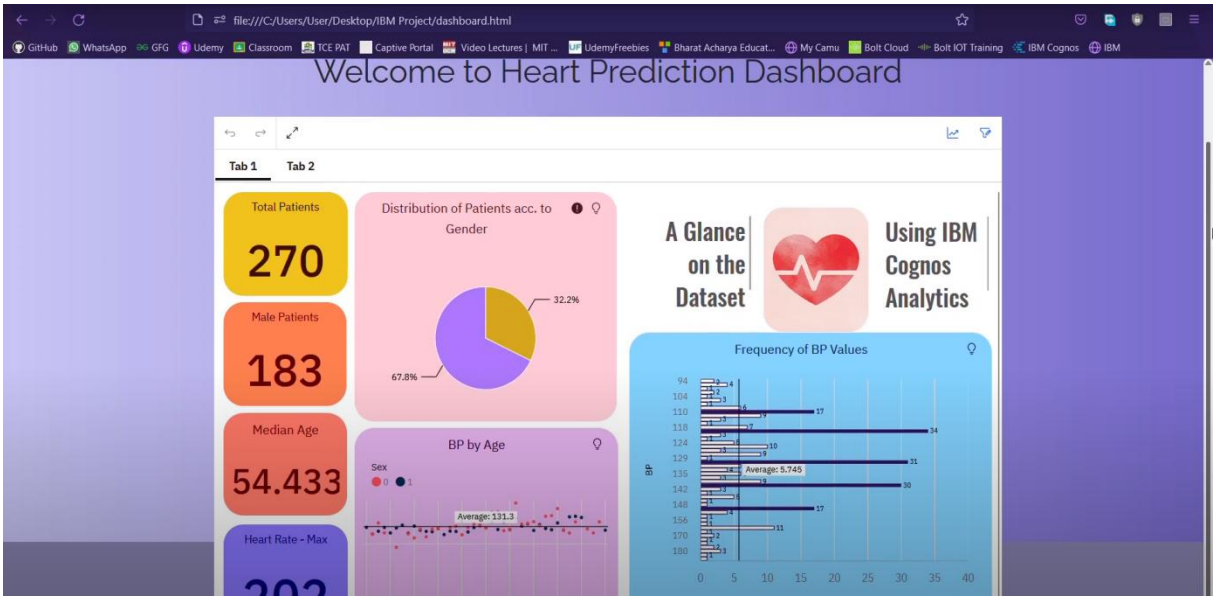


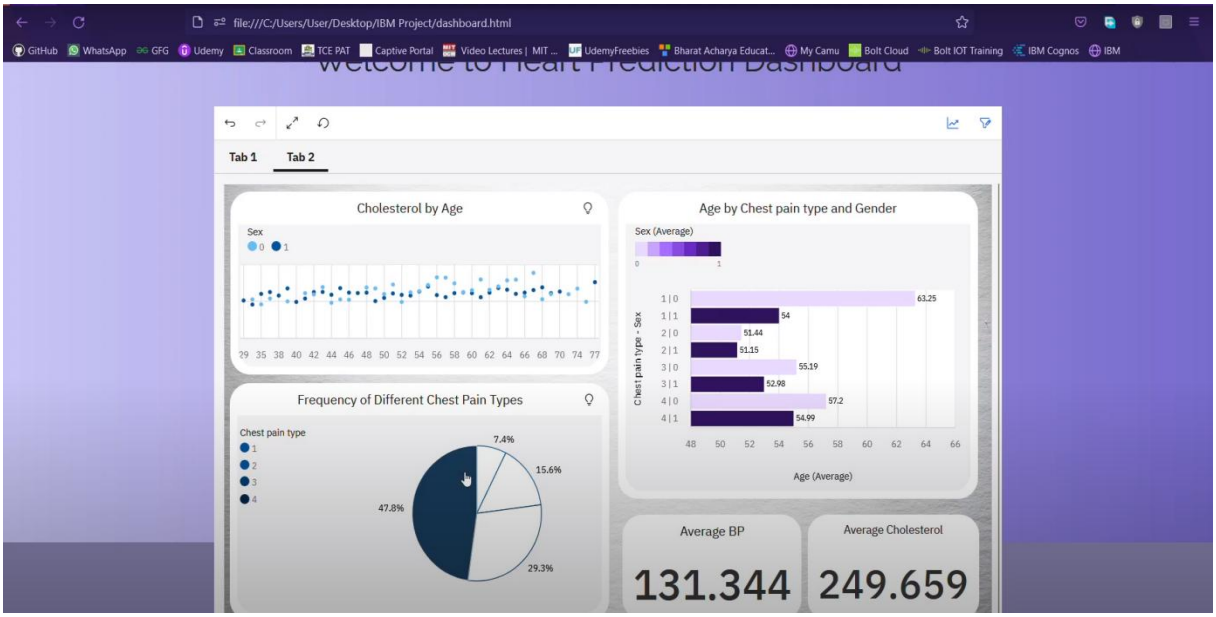
In this page, the user can enter all their medical record details such as BP, Cholestrol etc., and then it displays whether the user has heart disease or not. If any patient has heart disease, it will display as **"Heart Disease is detected. Contact medical Support"**

7.3. Feature 2

Dashboard:

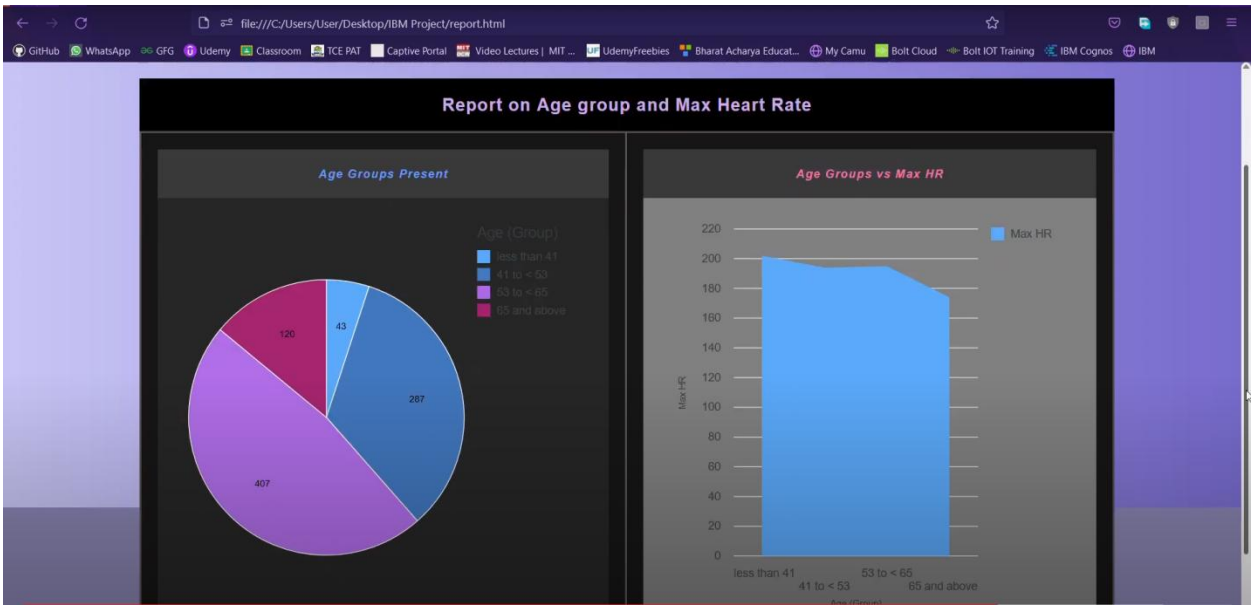
After logging into website, in tab1 it displays all the visualizations about the Heart Disease prediction such as total no. of patients having heart disease and number of male and female patients and average age of the patients having heart disease etc.,





7.4. Creation of Report using IBM cognos

It displays the report of various visualizations such as Age vs BP, BP vs Cholesterol, Age Groups vs Max HR etc., individually in which we classified the age into 4 groups.



7.5. Feature 3

ML Model for heart Disease Prediction:

Heart Disease prediction dataset is taken and some data preprocessing techniques are performed and model is trained by using various ML algorithms and highest accuracy model Logistic regression with the accuracy rate of 81.196 is taken to predict the whether the person has a heart disease or not.

The model is trained in Google Colab using required python libraries.

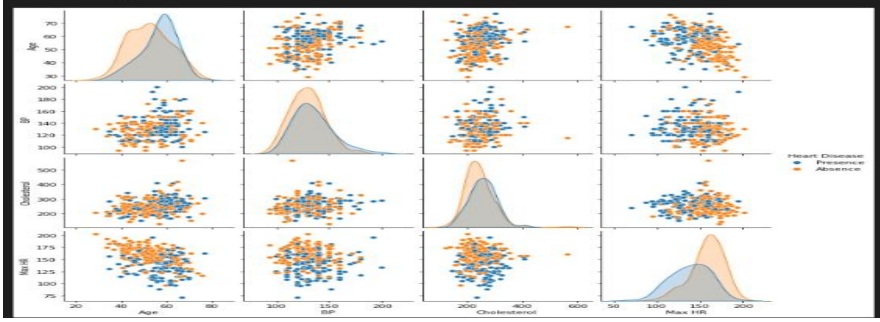
```
n_var_data[n_var_data['Cholesterol'] > 500]
```

	Age	BP	Cholesterol	Max HR	Heart Disease
1	67	115	564	160	Absence

A potential outlier, with cholesterol value greater than 500 is checked and observed to be usual

```
sns.pairplot(n_var_data, hue = 'Heart Disease')
```

```
<seaborn.axisgrid.PairGrid at 0x7f2c6148ec50>
```



Retesting the model with the selected Feature

Here, we will check whether the performance of model with only selected variable which is better than previous model with all variables.

```
data = df[df.columns & selected_feature]

X_train, X_test, y_train, y_test = train_test_split(data, y, test_size=0.33)
model = LogisticRegression()
model.fit(X_train, y_train)
r_sq = model.score(data, y)
print(f"coefficient of determination: {r_sq}")
```

```
coefficient of determination: 0.8185185185185185

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: FutureWarning: Index.__and__ operating as a set operation is deprecated, in the future this will be a logical operation matching Series.__and__.  Use index.intersection(other) instead
  """Entry point for launching an IPython kernel.
```

- The accuracy score of model with the selected variables is not significantly different than all variables.
- The number of variables is reduced from 13 to 4 with the feature selection

The number of variables is reduced from 13 to 4 with the feature selection

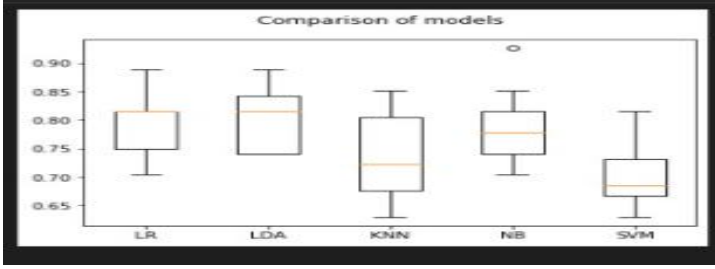
```
# prepare models
models = []
models.append(('LR', LogisticRegression()))
models.append(('LDA', LinearDiscriminantAnalysis()))
models.append(('KNN', KNeighborsClassifier()))
models.append(('NB', GaussianNB()))
models.append(('SVM', SVC()))
```

```
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'

for name, model in models:
    kfold = KFold(n_splits=10, random_state=7, shuffle = True)
    cv_results = cross_val_score(model, data, y, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

```
LR: 0.796296 (0.053029)
LDA: 0.800000 (0.052899)
KNN: 0.740741 (0.075903)
NB: 0.788889 (0.062085)
SVM: 0.696296 (0.054433)
```

```
# boxplot comparison of all the models
fig = plt.figure()
fig.suptitle('Comparison of models')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```

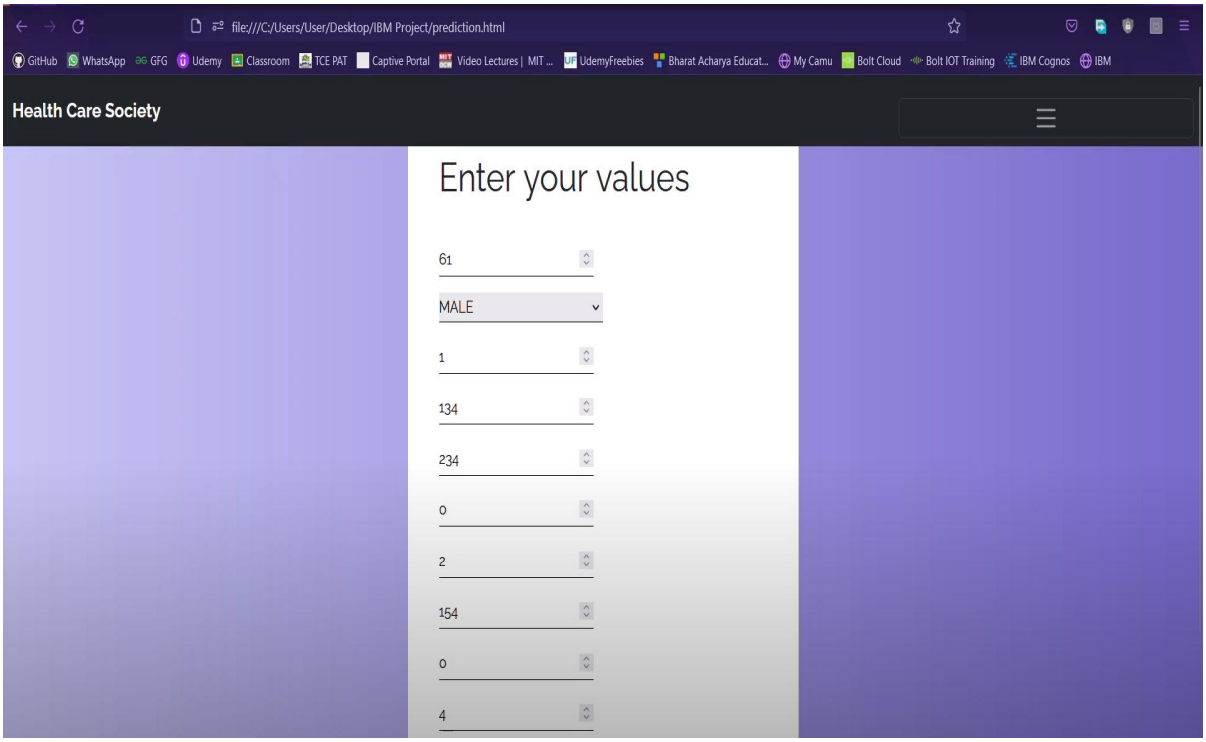


Here the box plot is built to visualize the comparison between the accuracy of prediction for various ML algorithms such as Logistic regression, Linear Discriminant Analysis, Naïve Bayes, K-Nearest Neighbors, Support Vector Machine etc.,

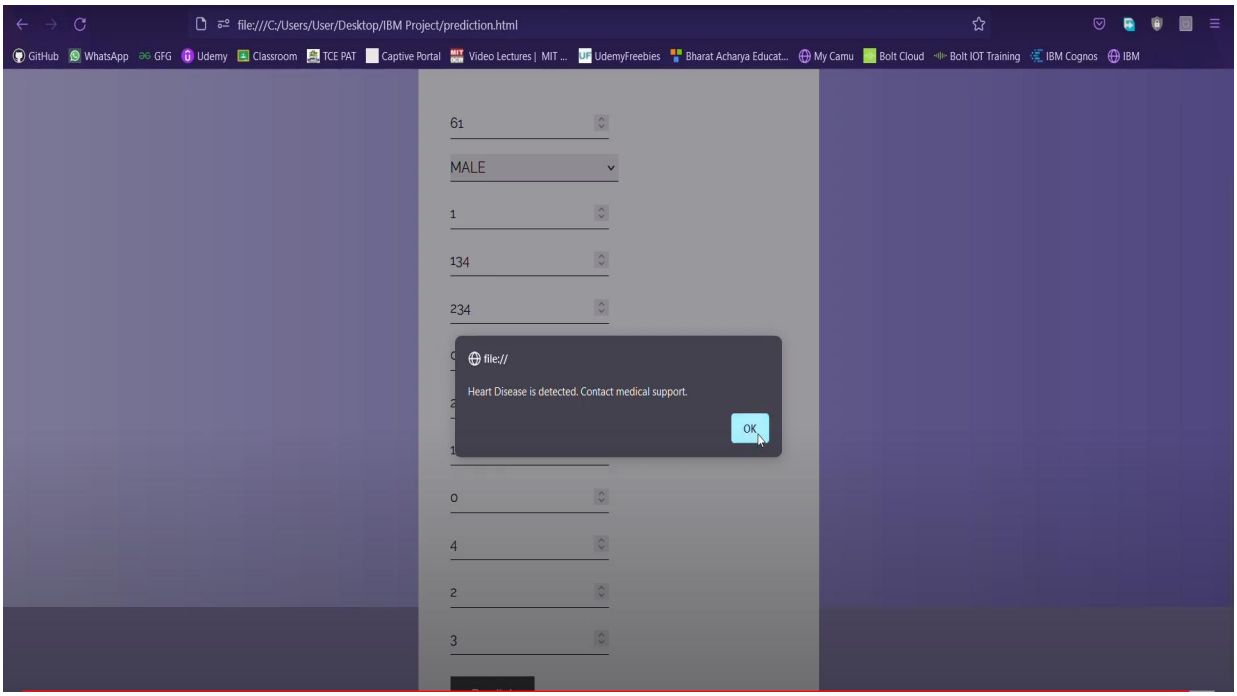
8. TESTING

8.1. Test Cases

Testing the model by collecting various user inputs of their medical details and checking the accuracy of prediction in a website.



8.2. User Acceptance Testing



When the user enters all the required medical details such as BP, Cholesterol, Thallium, EKG, FBS etc., and then it displays whether the user has heart disease or not. If any patient has heart disease, it will display the popup window as **"Heart Disease is detected. Contact medical Support"**

9. RESULTS

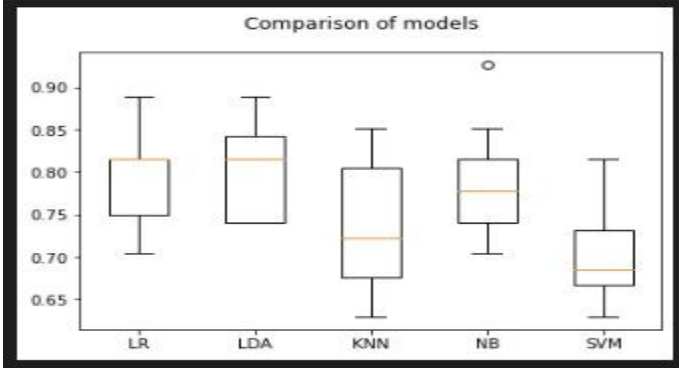
9.1. Performance Metrics

```
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'

for name, model in models:
    kfold = KFold(n_splits=10, random_state=7, shuffle = True)
    cv_results = cross_val_score(model, data, y, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

```
LR: 0.796296 (0.053029)
LDA: 0.800000 (0.052899)
KNN: 0.740741 (0.075903)
NB: 0.788889 (0.062085)
SVM: 0.696296 (0.054433)
```

```
# boxplot comparison of all the models
fig = plt.figure()
fig.suptitle('Comparison of models')
ax = fig.add_subplot(111)
plt.boxplot(results)
ax.set_xticklabels(names)
plt.show()
```



10. ADVANTAGES & DISADVANTAGES

ADVANTAGES:

1. Reduce the time complexity of doctors.
2. Cost effective for patients
3. The advantage of this model are high performance and accuracy rate
4. . It is very flexible and high rates of success are achieved
5. The application when implemented using Ensembling model (combination of Logistic, LDA, SVM, Naïve Bayes, Random Forest) has more accuracy rate when compare to other algorithm. In this system, we achieve around 80%

DISADVANTAGES:

1. Does not provide suggestions to user.
2. Does not accept any null values as a input.

11. CONCLUSION

Heart diseases when aggravated spiral way beyond control. Heart diseases are complicated and take away lots of lives every year. When the early symptoms of heart diseases are ignored, the patient might end up with drastic consequences in a short span of time. Sedentary lifestyle and excessive stress in today's world have worsened the situation. If the disease is detected early then it can be kept under control. It is always better to get treated in the early stages of heart disease. However, our website which is user friendly helps the users in prediction of heart disease and the suggestions that we get from the website might help save them.

12. FUTURE SCOPE

Like the saying goes “Prevention is better than cure”**.** In future an intelligent system may be developed that can lead to selection of proper treatment methods for a patient diagnosed with heart disease. A lot of work has been done already in making models that can predict whether a patient is likely to develop heart disease or not. There are several treatment methods for a patient once diagnosed with a particular form of heart disease.

13. APPENDIX

Demo link:

https://drive.google.com/file/d/1Kdmm2TkaYI6l07Um9lqEw0_nAtZ9zWf5/view?usp=sharing

Cloud deployment link: https://health-care-app-ibm.netlify.app/