VISUALIZATION AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASH BOARD

NALAIYA THIRAN PROJECT BASED LEARNING

ON
PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

Anukeerthana B	1905006
Bhagyalakshmi T	1905012
Deepika M	1905014
Lavanya B	1905043

BACHELOR OF TECHNOLOGY

IN
INFORMATION TECHNOLOGY

SRI RAMAKRISHNA ENGINEERING COLLEGE
COIMBATORE– 641 022

NOVEMBER 2022

INTERNAL MENTOR

DR. Bhaggiyaraj

Associate Professor

Department of Information Technology

Sri Ramakrishna Engineering College,

Coimbatore– 641 002

INDUSTRY MENTOR

MR. MAHIDHAR, MS. SAUMYA

IBM

ABSTRACT

Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analyzing data that excludes inferences and statistical modelling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Data analytics is considered as a cost-effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analyzing data. In this paper, the risk factors that causes heart disease is considered and predicted using K-means algorithm and the analysis is carried out using a publicly available data for heart disease. The dataset holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain. To predict the heart disease, K-means clustering algorithm is used along with data analytics and visualization tool. The paper discusses the pre-processing methods, classifier performances and evaluation metrics. In the result section, the visualized data shows that the prediction is accurate.

CHAPTER 1

INTRODUCTION

1.1 Project Overview :

Heart disease is one of the most fatal problems in the whole world, which cannot be seen with a naked eye and comes instantly when its limitations are reached. Therefore, it needs accurate diagnosis at accurate time. Health care industry produced huge amount of data every day related to patients and diseases. However this data is not used efficiently by the researchers and practitioners. Today healthcare industry is rich in data however poor in knowledge. There are various data mining and machine learning techniques and tools available to extract effective knowledge from databases and to use this knowledge for more accurate diagnosis and decision making. Increasing research on heart disease predicting systems, it become significant to summarize the completely incomplete research on it. The main objective of this research project is to summarize the recent research with comparative results that has been done on heart disease prediction and also make analytical conclusions.

1.2 Purpose :

The main objective of this project is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data set to assist in the prediction of the heart diseases. It 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.

CHAPTER 2

LITERATURE SURVEY

2.1 Existing System :

A quiet Significant amount of work related to the diagnosis of Cardiovascular Heart disease using Machine Learning algorithms has motivated this project. This project contains a brief literature survey. An efficient Cardiovascular disease prediction has been made by using various algorithms some of them include Logistic Regression, KNN, Random Forest Classifier Etc. It can be seen in Results that each algorithm has its strength to register the defined objectives [1]. The model incorporating had the ability to calculate the decision boundary using the previous and new model of machine learning and deep learning. It facilitated the important and the most basic factors/knowledge such as family history connected with any heart disease. But the accuracy that was obtained in such IHDPS model was far more less than the new upcoming model .heart disease using artificial neural network and other algorithms of machine and deep learning. The risk factors of coronary Heart disease or atherosclerosis is identified by McPherson et al.,[2] using the inbuilt implementation algorithm using uses some techniques of Neural Network and were just accurately able to predict whether the test patient is suffering from the given disease or not. Diagnosis and prediction of Heart Disease and Blood Pressure along with other attributes using the aid of neural networks was introduced by R. Subramanian et al.,[4]. A deep Neural Network was Built incorporating the given attributes related to the disease which were able to produce a output which was carried out by the output perceptron and almost included 120 hidden layers which is the basic and most relevant technique of ensuring a accurate result of having heart disease if we use the model for Test Dataset. The supervised network has been advised for diagnosis of heart diseases [3]. When the testing of the model was done by a

doctor using an unfamiliar data, the model used and trained from the previous learned data and predicted the result thereby calculating the accuracy of the given model.

2.2 References :

[1] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267-72.

[2] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40- 6). IEEE.

[3] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299-303). IEEE.

[4] Kiyasu J Y (1982). U.S. Patent No. 4,338,396. Washington, DC: U.S. Patent and Trademark Office.

## 2.3 Problem Statement Definition :

There are instruments available which can predict heart disease but either they are 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. Since we have a good amount of data in today's world, we can use technologies to analyze the data for hidden patterns.

CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas :

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviours and attitudes. It is a useful tool to helps teams better understand their users.Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges.
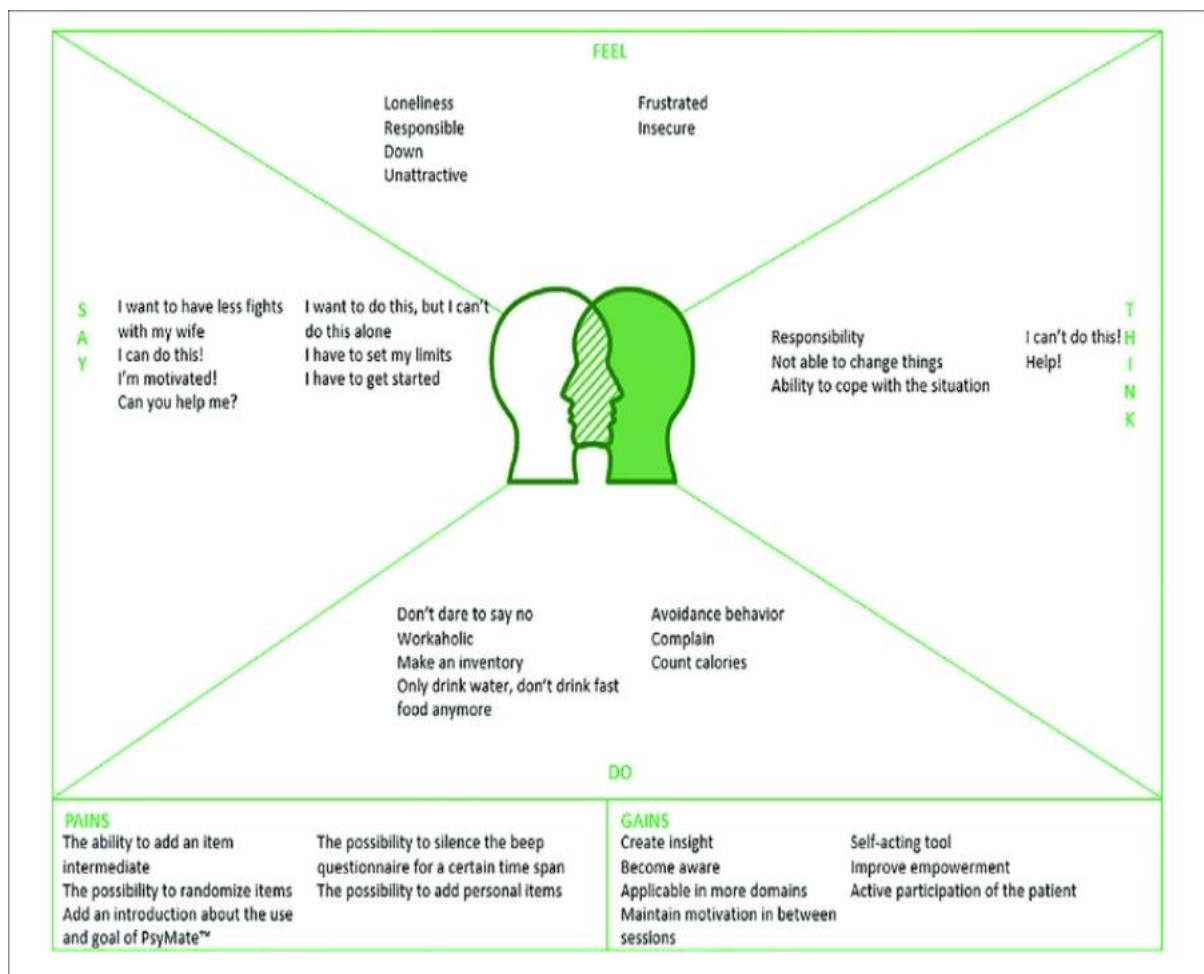
3.2 Ideation & Brainstorming :

Bhagyalakshmi T

Food habits and lifestyle Suggestions to prevent	Predicting affected people
Statistical test for feture selection	data preprocessing and precise outputs

Deepika M

Reduce the death rate due to heart disesae	Performed data analysis to find various result
Frequent updates about the status of health	diagnosis improved by data analysis

Lavanya B

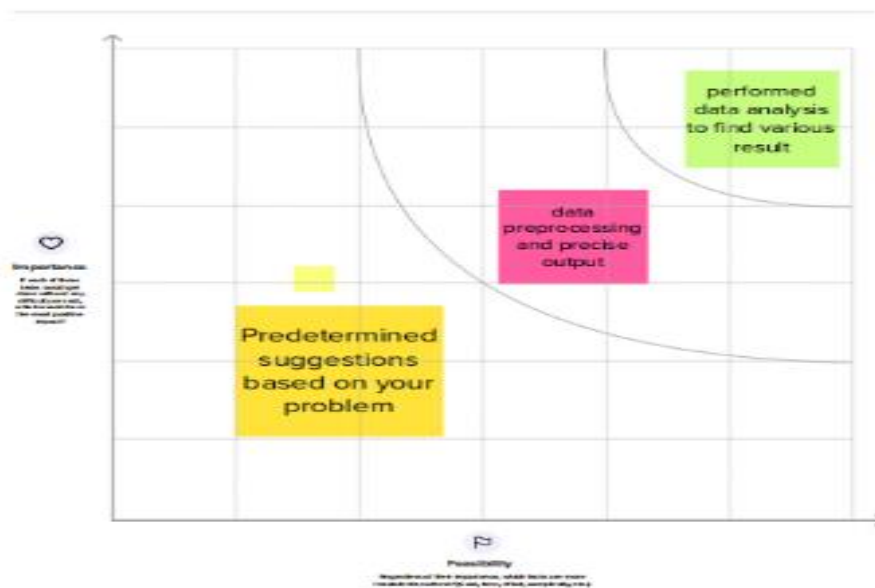|	
Graphical representation predicts heart's health	Predicting the vulnerable people
sensus about the heart diseasem affected people	Hyparameter tunning of various model

Anukeerthana M

Exploratory data analysis	Transperancy about the disease
Early solutions and results for your diseases	Predetermined suggestions based on your problem

3.3 Proposed Solution :

In this system we are implementing effective heart attack prediction system using IBM Cognos Analytics Dashboard. We can give the input as in CSV file to the system. After accessing data set the operation is performed and effective heart attack level is produced. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts. The heart attack prediction system designed to help the identify different risk levels of heart attack like normal, low or high and also giving the prescription details with related to the predicted result.

3.4 Problem Solution Fit :

1. CUSTOMER SEGMENT(S) CS	**6. CUSTOMER LIMITATIONS** EG. BUDGET, DEVICES CL	**5. AVAILABLE SOLUTIONS** PLUSES & MINUSES AS
People who have heart diease Hospitals Clinics Any medical support field who prepare data of diesease	This solution provides only the visualization of the dashboard by the symptoms given my the customer	The Customers can prefer over amanual data visualization and prediction,which is very tedious job and requires the Knowledge over the AI/ML
2. PROBLEMS / PAINS + ITS FREQUENCY PR	**9. PROBLEM ROOT / CAUSE** RC	**7. BEHAVIOR** + ITS INTENSITY BE
Chest Pain Pain between the Shoulder blades clammy skin,cold sweat or sweating anxiety feeling of impending doom	Reason of increase in heart diesease will not be rootly identified There is a chance of Identifying every heart heart disease as same Difficulty of predicting heart disease will not be rootly identified Will Not Have Proper Idea of relation between in heart diseases as same	We must have a knowledge of difference between datasets used for comparision Customers need to collect more number of datasets in order to obtain more accurate result Generation of Legitimate and reliable dataset
3. TRIGGERS TO ACT TR Insufficient ways of handling huge amount of datasets and inferring the root cause of the heart disease cannot be found out	**10. YOUR SOLUTION** SL By using cognos Analysis using AI/ML and predict heart diseases and related disease by the ultimate power Cognos Analytics Tool we can create a proper dashboard for the customers to work with and visualize and analyze the heart disease on their work with limited knowledge	**8. CHANNELS of BEHAVIOR** CH ONLINE Visualizing the datasets and Exploring the data
4. EMOTIONS BEFORE / AFTER EM Before:It creates huge knowledge for proper or accurate reason for heart disease After:It creates a large chance of understanding Heart root cause of it.		OFFLINE Collection of the datasets Filtering the datasets

Left margin labels: Define CS, fit into CL | Focus on PR, tap into BE, understand RC | Identify strong TR & EM

Right margin labels: Explore AS, differentiate | Focus on PR, tap into BE, understand RC | Extract online & offline CH of BE

REQUIREMENT ANALYSIS

4.1 Functional Requirement :

- In this system have two actors are there admin and user
- Admin has to Explore and Visualzing the user data
- User can upload a dataset, they can explore and visualize the data using IBM cognitive Dashboard.
- In user part they have to visualize their data themselves.
- Then user has to input patient data, and then based on user input data will check and it will give the output whether Heart Disease or not

4.2 Non Functional Requirement :

- Reliability

  The structure must be reliable and strong in giving the functionalities. The movements must be made unmistakable by the structure when a customer has revealed a couple of enhancements. The progressions made by the Programmer must be Project pioneer and in addition the Test designer.

- Maintainability

  The system watching and upkeep should be fundamental and focus in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employments are running without lapses.

- Performance

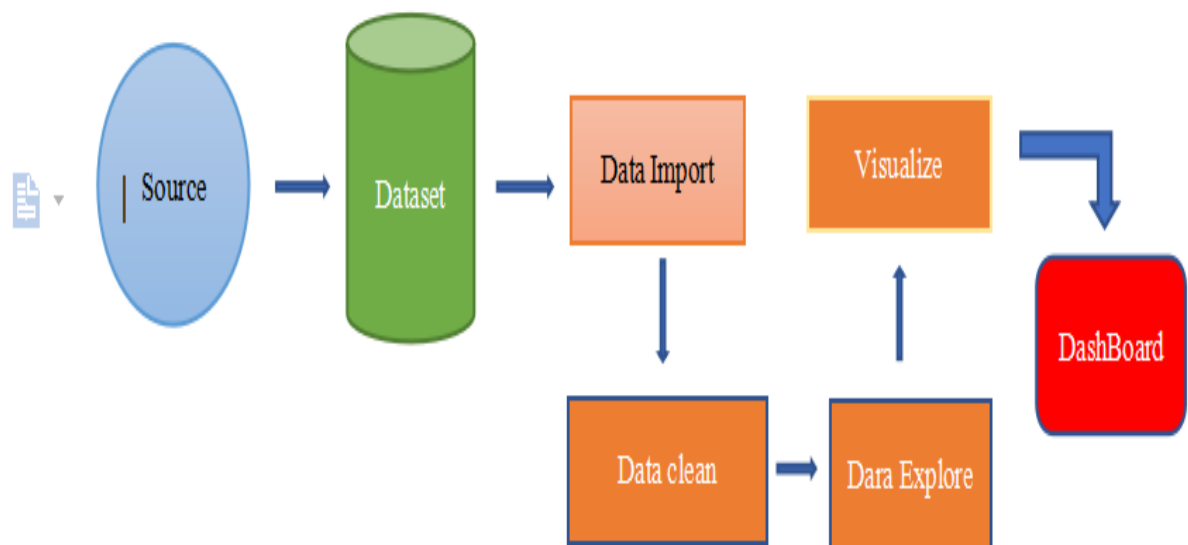 The framework will be utilized by numerous representatives all the while.

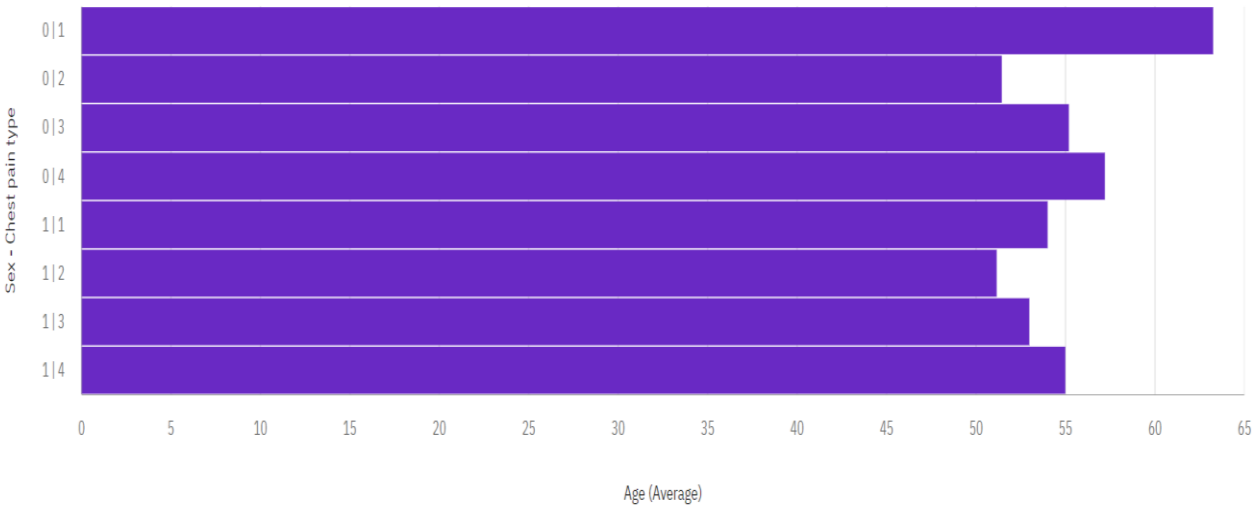# CHAPTER 5

PROJECT DESIGN

5.1 Data Flow Diagram :
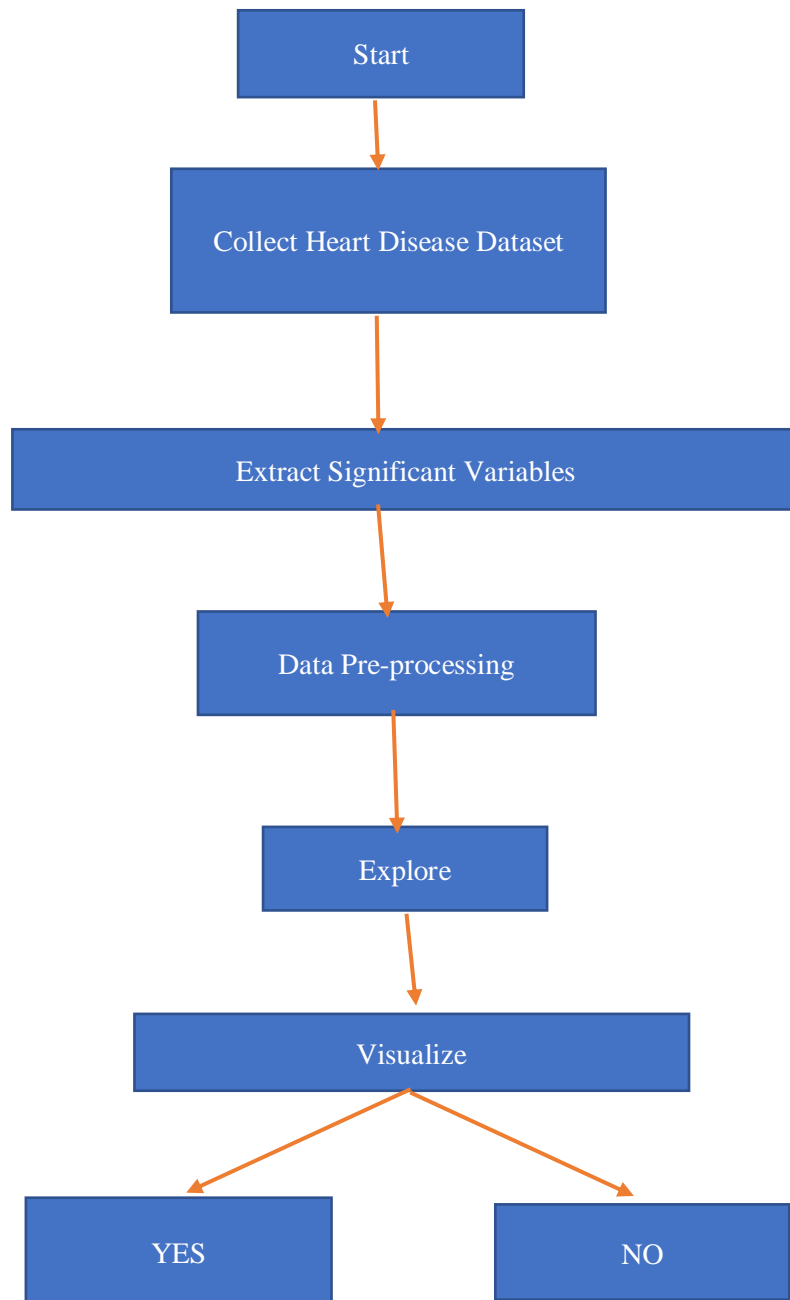
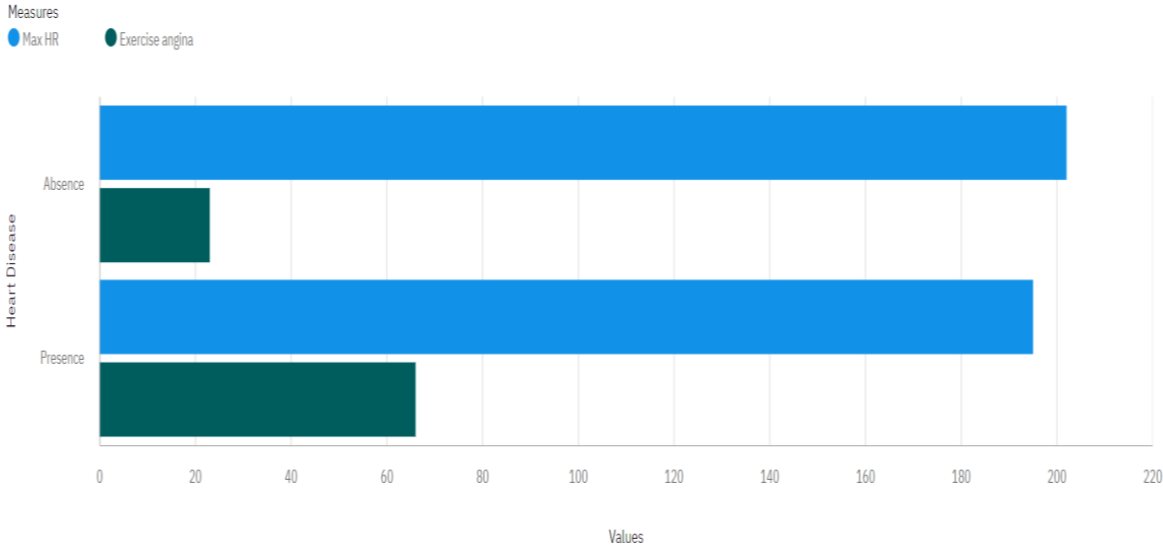## 5.2 Solution & Technical Architecture :

CHAPTER 9

RESULT

Age by Sex and Chest pain type
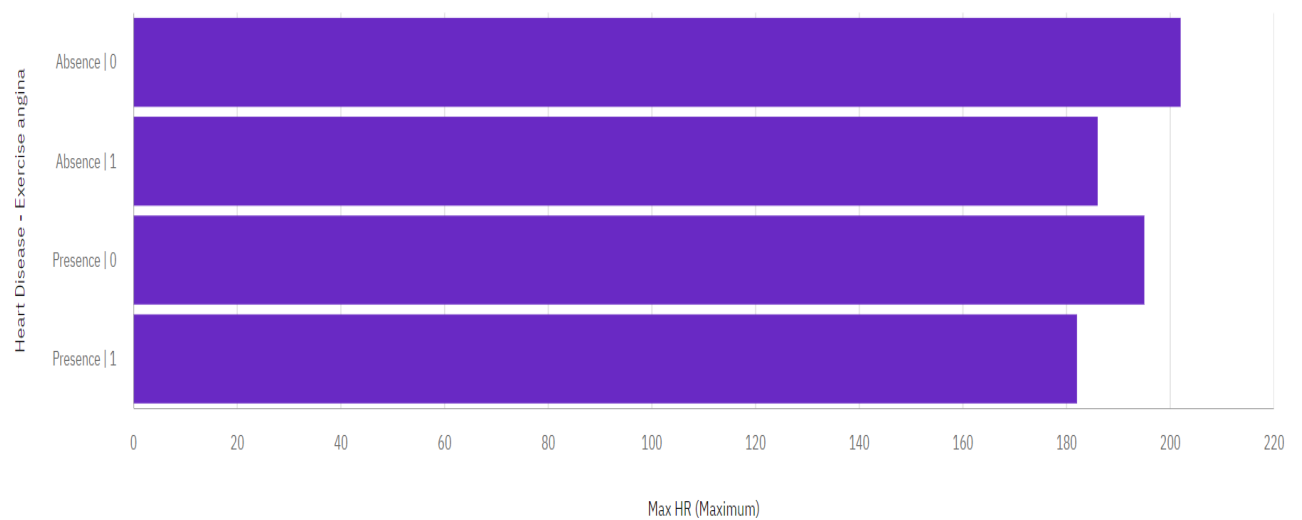

Max HR and Exercise angina by Heart Disease

Measures
- Max HR
- Exercise angina

Cholesterol by Age colored by Sex

Sex
● 0 ● 1
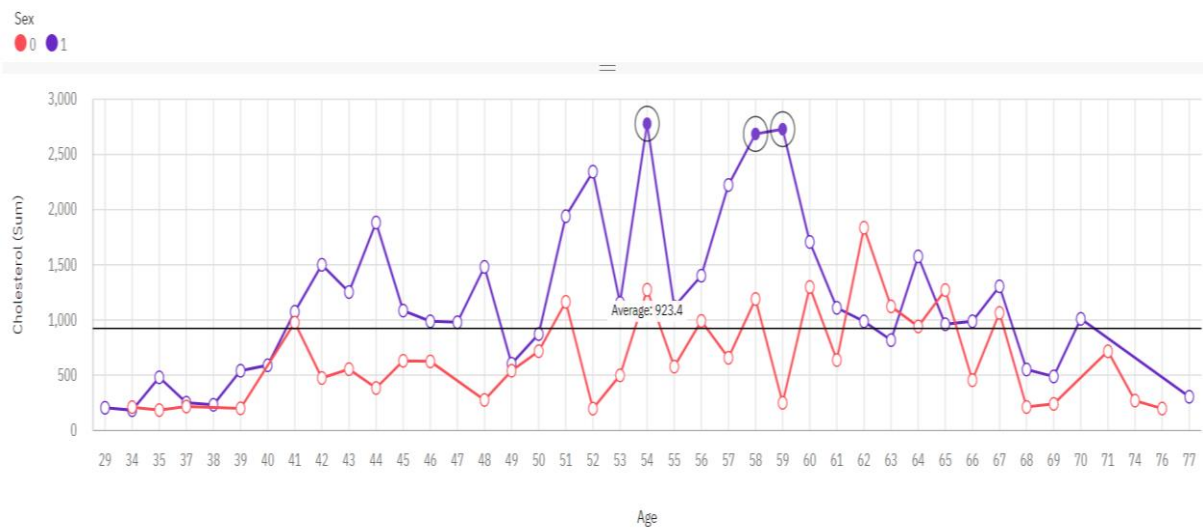
Heart Disease for Chest pain type and Sex

Heart Disease	1	2	3	4	Summary
0	4	16	32	35	87
1	16	26	47	94	183
Summary	20	42	79	129	270

Max HR by Heart Disease and Exercise angina

BP by Age colored by Sex
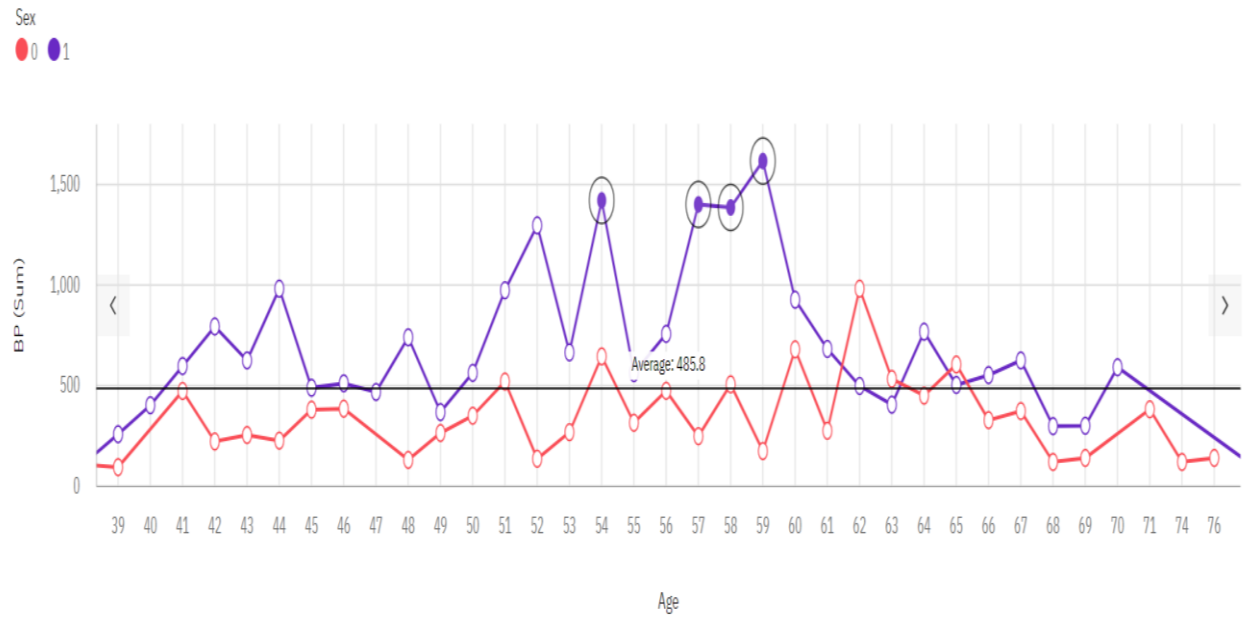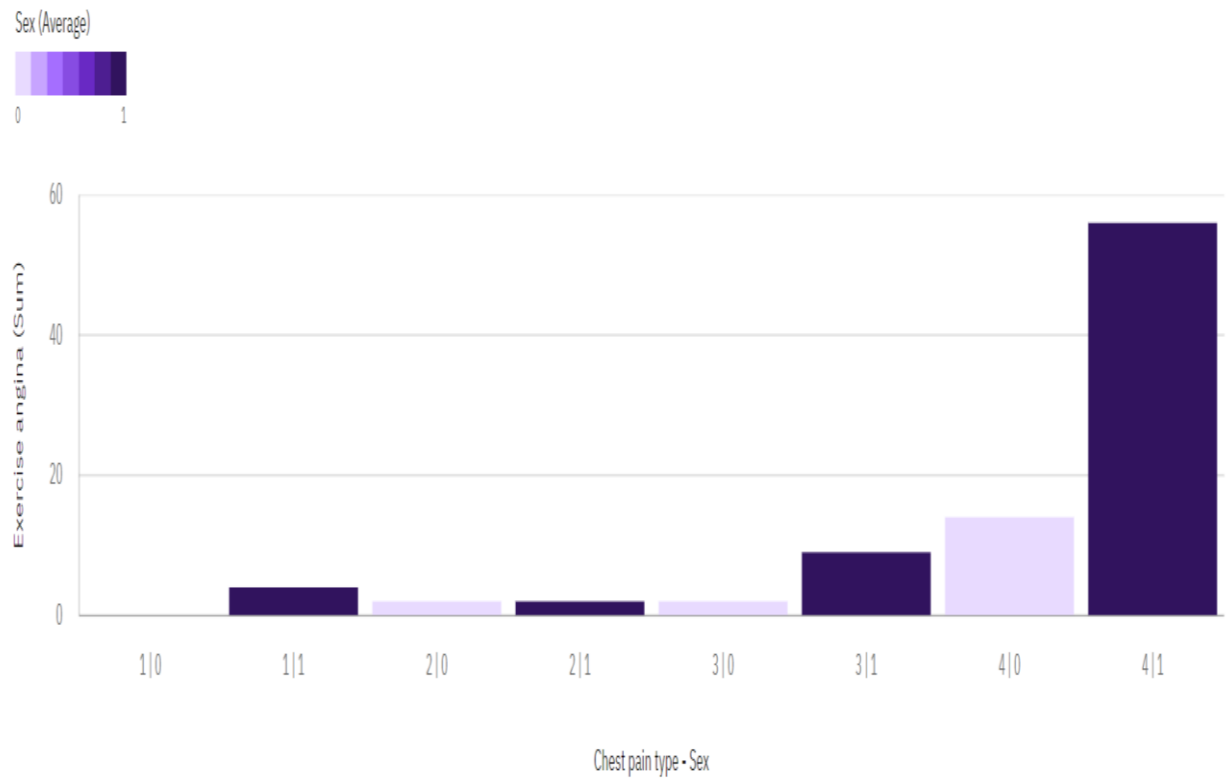
Sex
● 0 ● 1

Exercise angina by Chest pain type and Gender

Sex (Average)

0   1

Exercise angina by Chest pain type and Gender

Sex (Average)

0 — 1

Exercise angina (Sum)

60

40

20

0

1 | 0   1 | 1   2 | 0   2 | 1   3 | 0   3 | 1   4 | 0   4 | 1

Chest pain type - Sex

Cholesterol by Age colored by Sex

Sex

● 0   ● 1

Cholesterol (Sum)

3,000

2,000

Average: 923.4

1,000

0

29 35 38 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 74 77

Age

BP by Age colored by Sex

Sex

● 0   ● 1

BP (Sum)

1,500

1,000

Average: 485.8

500

0

48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71

Age

Max HR and Exercise angina by Heart Disease

Measures

● Max HR   ● Exercise angina

Heart Disease

Absence

Presence

0   20   40   60   80   100   120   140   160   180   200   220

Values
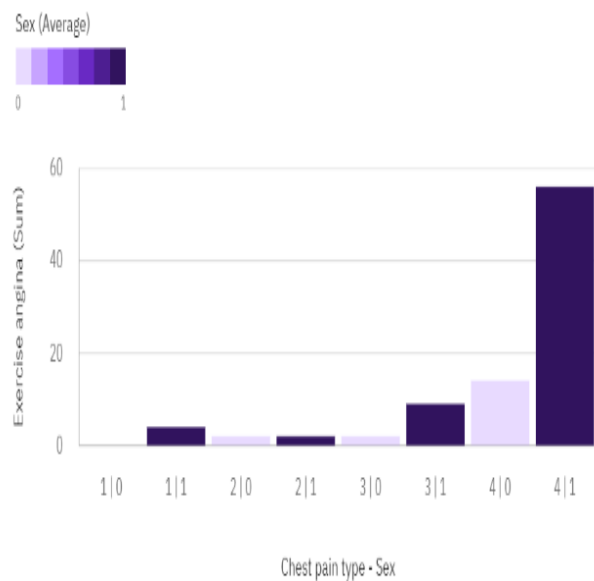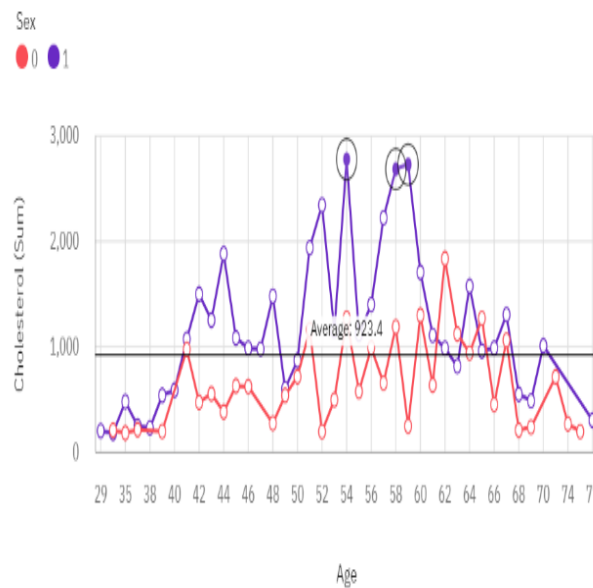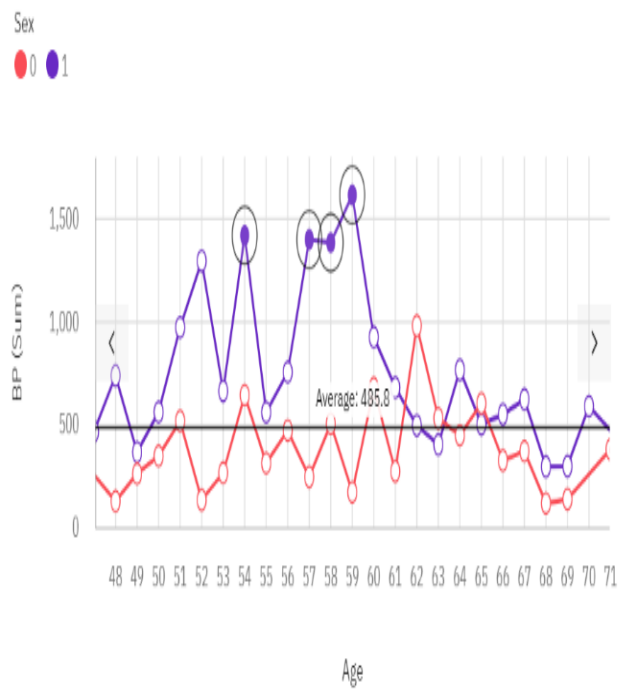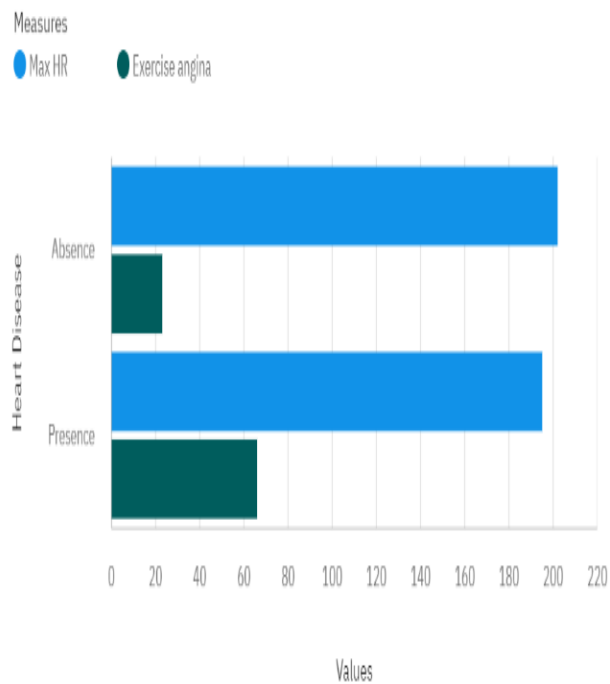
CHAPTER 10

ADVANTAGES & DISADVANTAGES

ADVANTAGES :

1. Increased accuracy for effective heart disease diagnosis.

2. Handles roughest(enormous) amount of data and featu selection.

3. Reduce the time complexity of doctors.

4. Cost effective for patients.

DISADVANTAGES :

1. Prediction of cardiovascular disease results is not accurate.

2. Data mining techniques does not help to provide effective decision making.

3. Cannot handle enormous datasets for patient records.

CHAPTER 11

CONCLUSION

Heart stroke and vascular disease are the major cause of disability and premature death. Chest pain is the key recognize the heart disease.Data analytics in healthcare reported in various literature are highlighted particularly on clinical data, Pharmaceutical data, Patient behaviour, sentiment data, viral and Global Infectious Disease Surveillance.

The most important factor, however, in the development and application of big data, mobile phone or tablet use and landline use in patient privacy and to safeguard the patients ability to direct and discover the use of his or her health care information.

CHAPTER 12

FUTURE SCOPE

Global healthcare industry is under significant pressure to reduce costs and more efficiently manage resources while improving patient care.

Rising costs, chronic illness, an again population and a shortage of professional are forcing massive changes in the healthcare industry.

To gain insight into how they can improve service while reducing costs, healthcare payers and providers are turning to data Analytics.

CHAPTER 13

APPENDIX

Source code :

- **Visualizing and Predicting Heart Diseases**

- **Interactive Dash Board**

Github link : https://github.com/Bhagyalakshmi-19/Naalayathiran-B8

Project demo link : https://youtu.be/2SDQTZoRNK8