PROJECT TITLE:

Visualizing and Predicting Heart Diseases with an Interactive Dash Board

TEAM ID:

PNT2022TMID12687

TEAM MEMBERS:

1. Lovely M - 19L128 **(Team Lead)**
2. Vishali R - 19L153
3. Suganth M - 19L147
4. Indra Priyadharshini P - 19L121

1.INTRODUCTION

1.1Project Overview

Predicting and diagnosing heart disease is the biggest challenge in the medical industry and it is based on factors like physical examination, symptoms and signs of the patient. Factors which influence heart diseases are cholesterol level of the body, smoking habit, and obesity, family history of diseases, blood pressure and working environment. Machine learning algorithms play a vital and accurate role in predicting heart disease. The advancement of technologies allows machine language to pair with big data tools to handle unstructured and exponentially growing 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.

1.2Purpose

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.Using the dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

2.LITERATURE SURVEY

2.1Existing problem

A major challenge faced by health care organizations, such as hospitals and medical centers, is the provision of quality services at affordable costs.1 The quality service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of both numerical and categorical data. Before further processing, cleaning and filtering are applied on these records in order to filter the irrelevant data from the database.2 The proposed system can determine an exact hidden knowledge, ie, patterns and relationships associated with heart disease from a historical heart disease database. It can also answer the complex queries for diagnosing heart disease; therefore, it can be helpful to health care practitioners to make intelligent clinical decisions. Results showed that the proposed system has its unique potency in realizing the objectives of the defined mining goals.

2.2References

1. Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. Int J Comput Sci Net Secur. 2008;8:343–350.

2. Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach. Int J Adv Sci Eng Technol. 2014;2:88–92

3. Gudadhe M, Wankhade K, Dongre S. Decision support system for heart disease based on support vector machine and Artificial Neural Network. Computer and Communication Technology (ICCCT), 2010 International Conference on; 2010. pp. 741–745

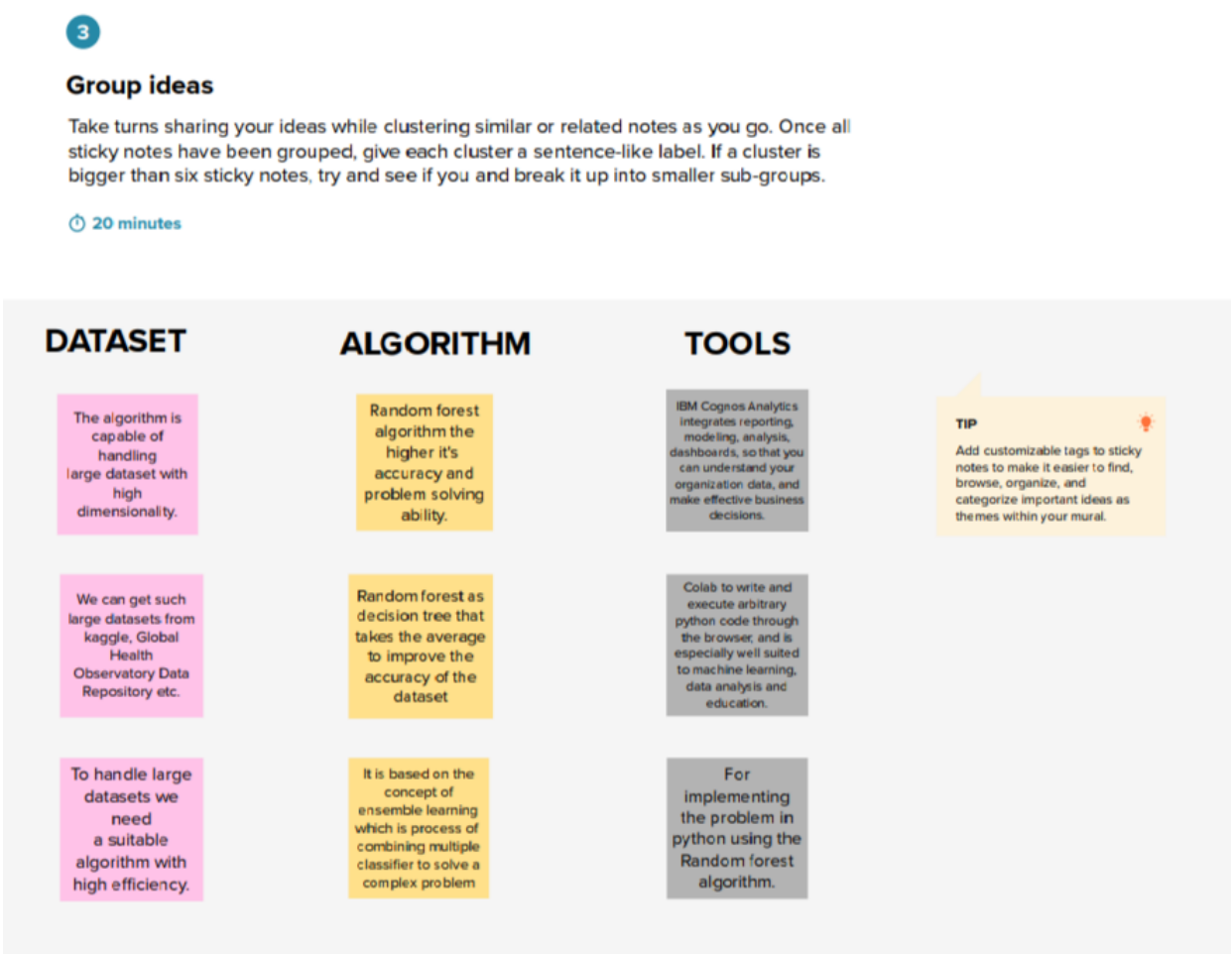
2.3Problem Statement Definition

Data analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analysing data.

3.IDEATION & PROPOSED SOLUTION

3.1Empathy Map Canvas

3.2Ideation & Brainstorming

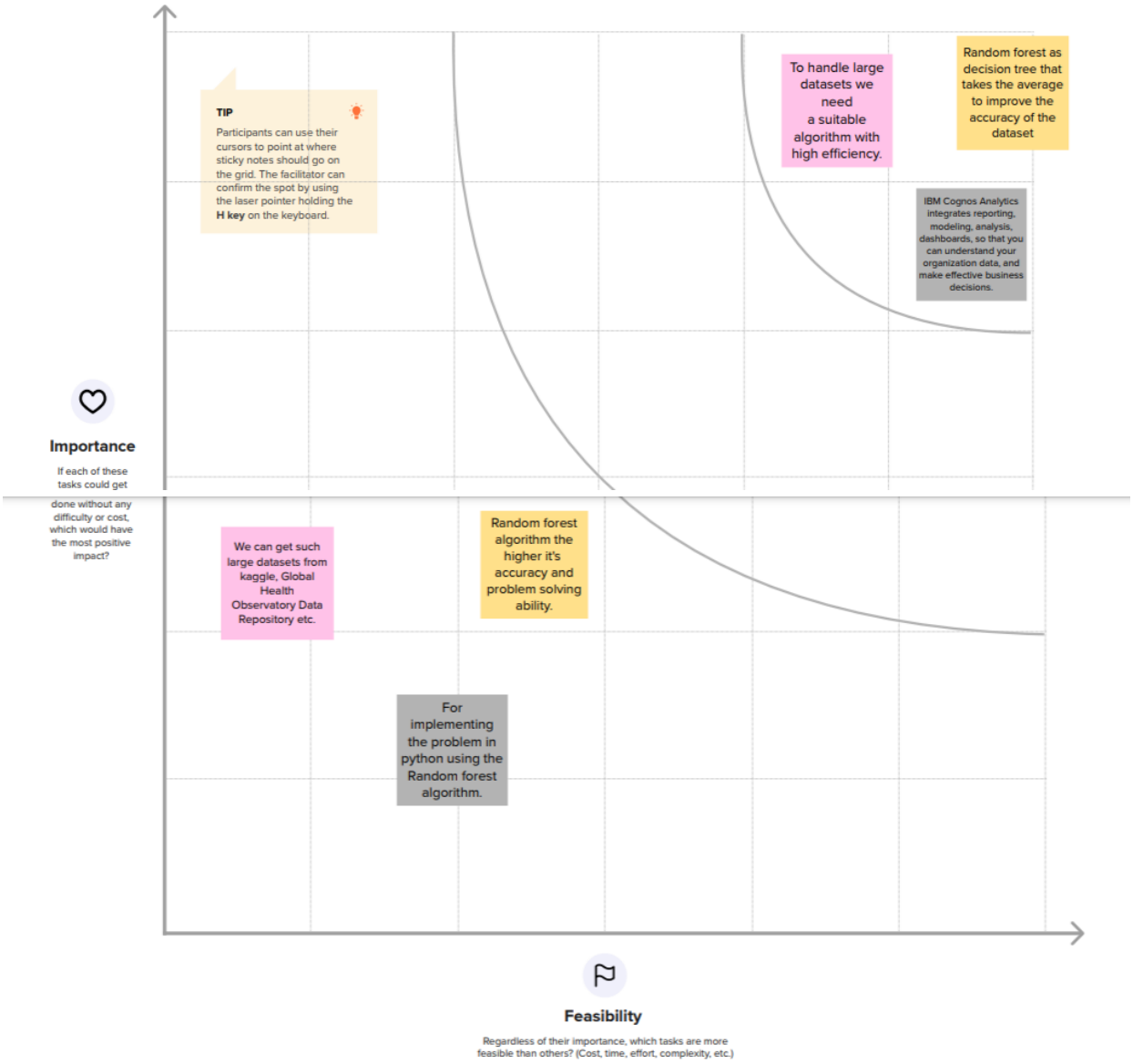


4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

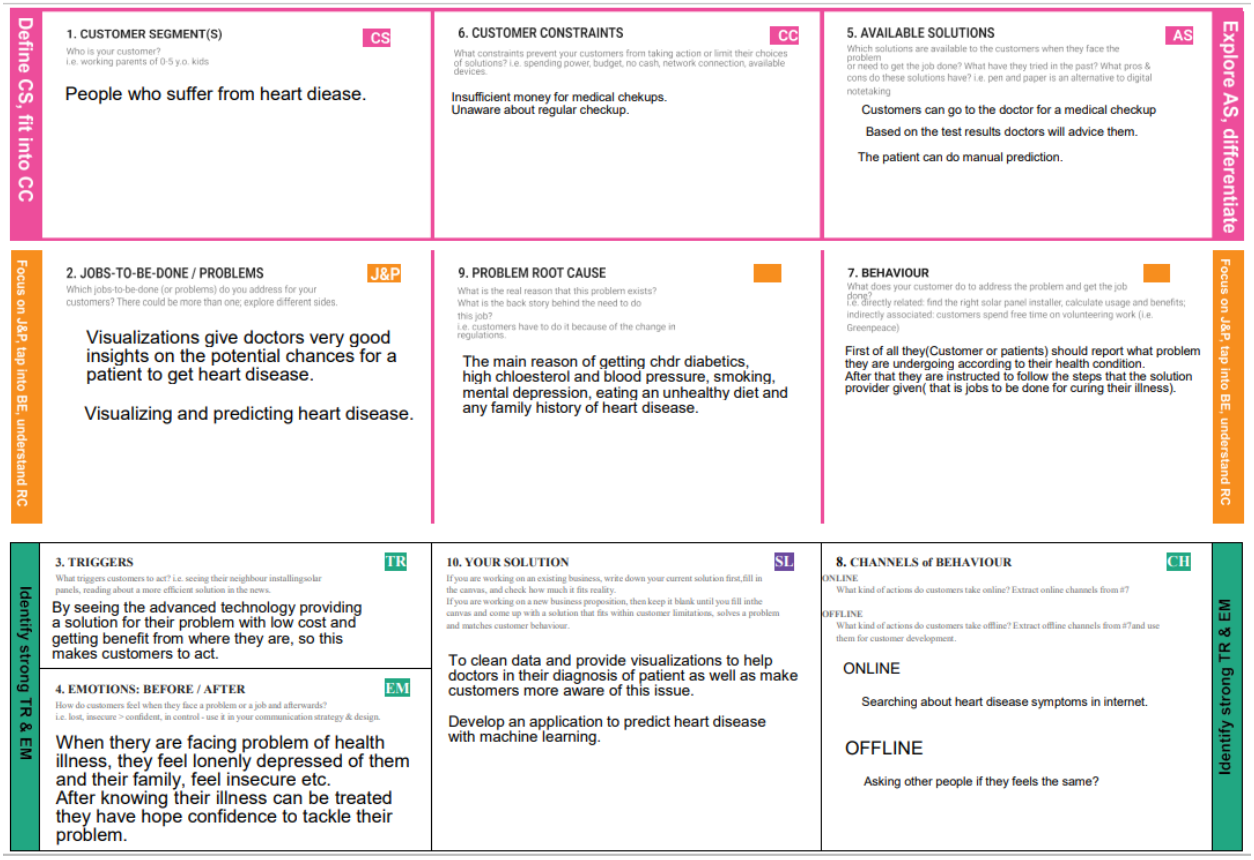
20 minutes



3.3Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to besolved)	1. To develop an interactive dashboard to predict the heartdisease accurately withfew tests and attributes the presence of heart disease.
2.	Idea / Solution description	1. Analyzing data and identifying the heartdisease using Cognosanalytics tools.
3.	Novelty / Uniqueness	1. Hoping to achievemaximum accuracy to provide prior treatment to the patientsandreduce the fatality rate.
4.	Social Impact/ Customer Satisfaction	1. Saving lives, User friendlyinteractive dashboard. 2. Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions andexperiences.
5.	Business Model (Revenue Model)	1. Data security. 2. Easy to use. 3. Constant updates according tonecessity.
6.	Scalability of the Solution	1. Can be used in any platform (Windows, mac, etc.,) 2. Adding new feature doesn't affect the performance of thesystem. 3. Scalable dataset.

3.4Problem Solution fit



4.REQUIREMENT ANALYSIS

4.1Functional requirement

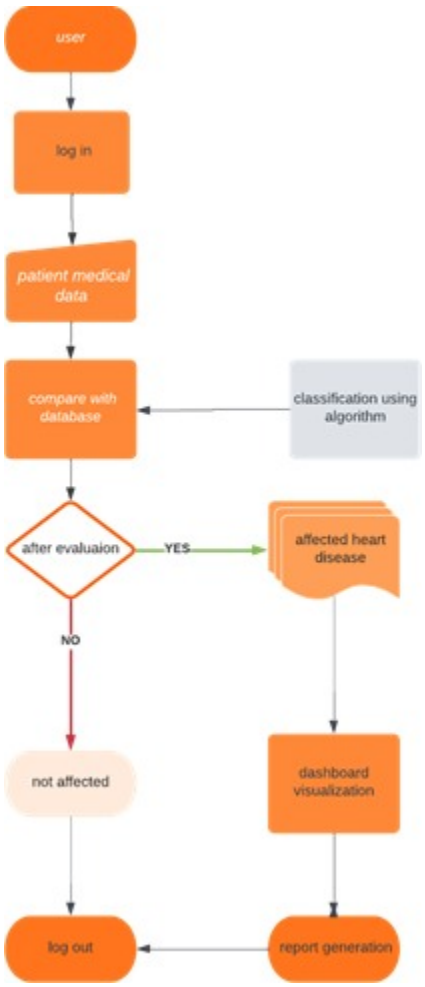
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

4.2Non-Functional requirements

NFR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user- friendlygraphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performedwith just a few clicks
NFR-2	**Security**	For security of the application the technique knownas database replication should be used so that all the important data should be kept safe. Incase of crash, the system should be able to backup and recover the data
NFR-3	**Reliability**	The application has to be consistent at every scenario and has to work without failure in anyenvironment
NFR-4	**Performance**	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
NFR-5	**Availability**	The application has to be available 24 x 7 for users without any interruption
NFR-6	**Scalability**	The application can withstand the increase in the no. of users and has to be able to develop higherversions

5.PROJECT DESIGN

5.1Data Flow Diagrams

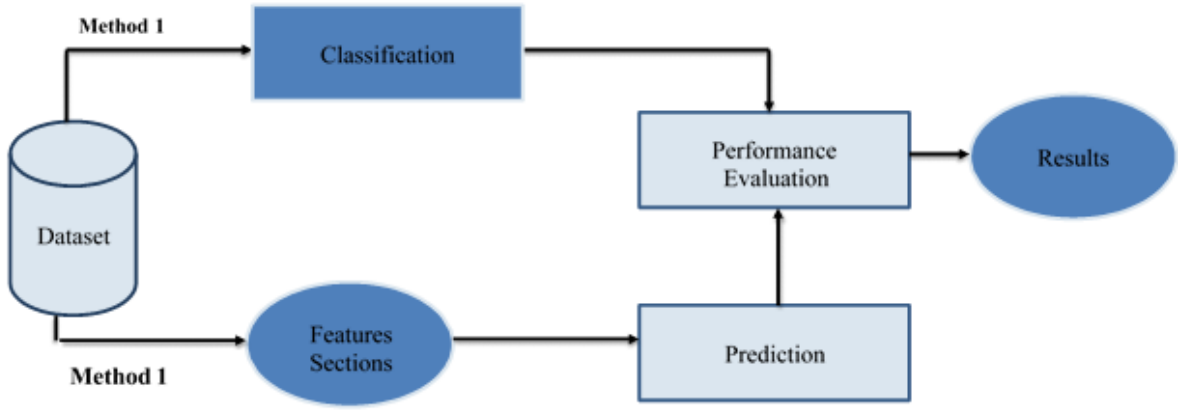


5.2 Solution & Technical Architecture

SOLUTION:

The proposed solution proposes an interactive dashboard for visualising and forecasting heart disorders, in which the user may observe his/her medical report analysis as well as the projected end result. IBM Cognos will be used to create the dashboard. Machine learning Algorithms will be used to forecast cardiac disease.

ARCHITECTURE:



5.3User 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 Gmail	I can receive conformation email	Medium	Sprint-1
Customer (Web user)	Login	USN-4	As a user, I can log into the application by entering email &	I can access my account using my details	High	Sprint-1

			password			
	Dashboard	USN-5	User can view his/her complete medical analysis and accuracy of disease prediction	I can view my medical analysis and accuracy	High	Sprint-2
	Dashboard	USN-6	User can view the accuracy of occurrence of heart disease through report generation	I can view the accuracy of heart disease in the dashboard	high	Sprint-2
Customer Care Executive	Helpdesk	USN-7	As a customer care executive, he/she can view the customer queries.	I can post my queries in the dashboard	Medium	Sprint-3
		USN-8	As a customer care executive, he/she can answer the customer queries	I can get support from helpdesk	High	Sprint-3
User Type	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Acceptance criteria**	**Priority**	**Release**
Administrator	User profile	USN-9	As an admin, he/she can update the health details of users.	I can view my updated health details	High	Sprint-4
		USN-10	As an admin, he/she can add or delete users.	I can access my account / Dashboard when logged in	High	Sprint-4
		USN-11	As an admin, he/she can manage the user details.	I can view the organized data of myself.	High	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1Sprint Planning & Estimation

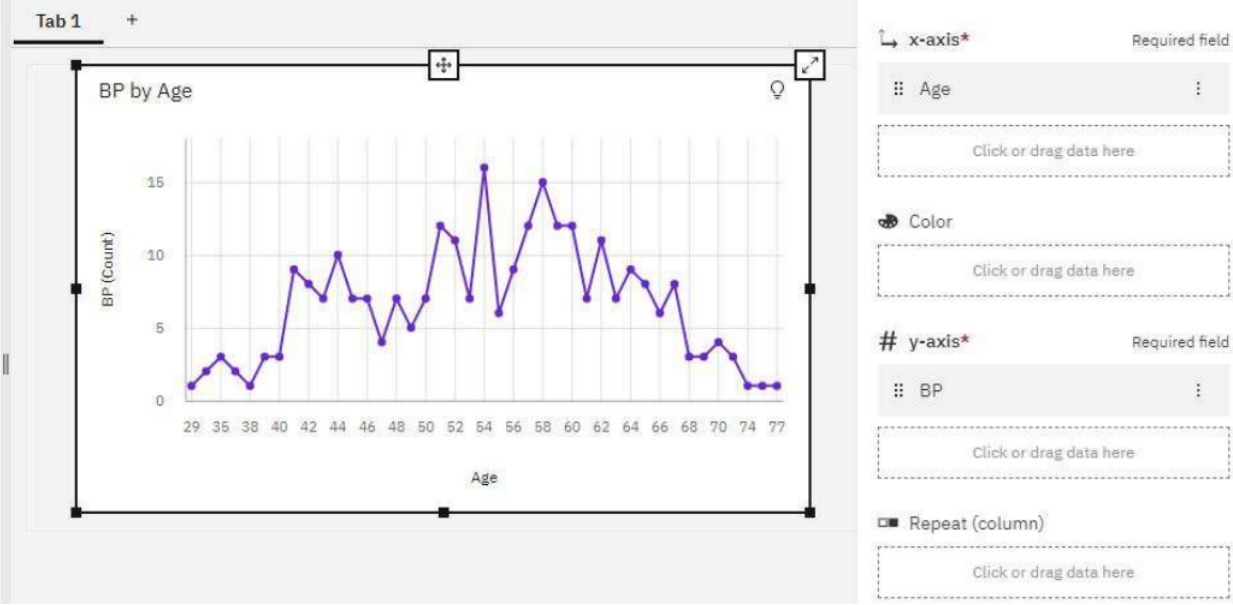
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority
Sprint-1	Registration	USN-1	As a health care provider I can create account in IBM cloud and the data arecollected.	20	High
Sprint-2	Analyze	USN-2	As a health care provider all the data that are collected is cleaned and uploaded in the database or IBM cloud.	20	Medium
Sprint-3	Dashboard	USN-3	As a health care provider I can use my account in my dashboard for uploading dataset.	10	Medium
Sprint-3	Visualization	USN-4	As a health care provider I can prepare data for Visualization.	10	High
Sprint-4	Visualization	USN-5	As a health care provider I canpresent data in my dashboard.	10	High

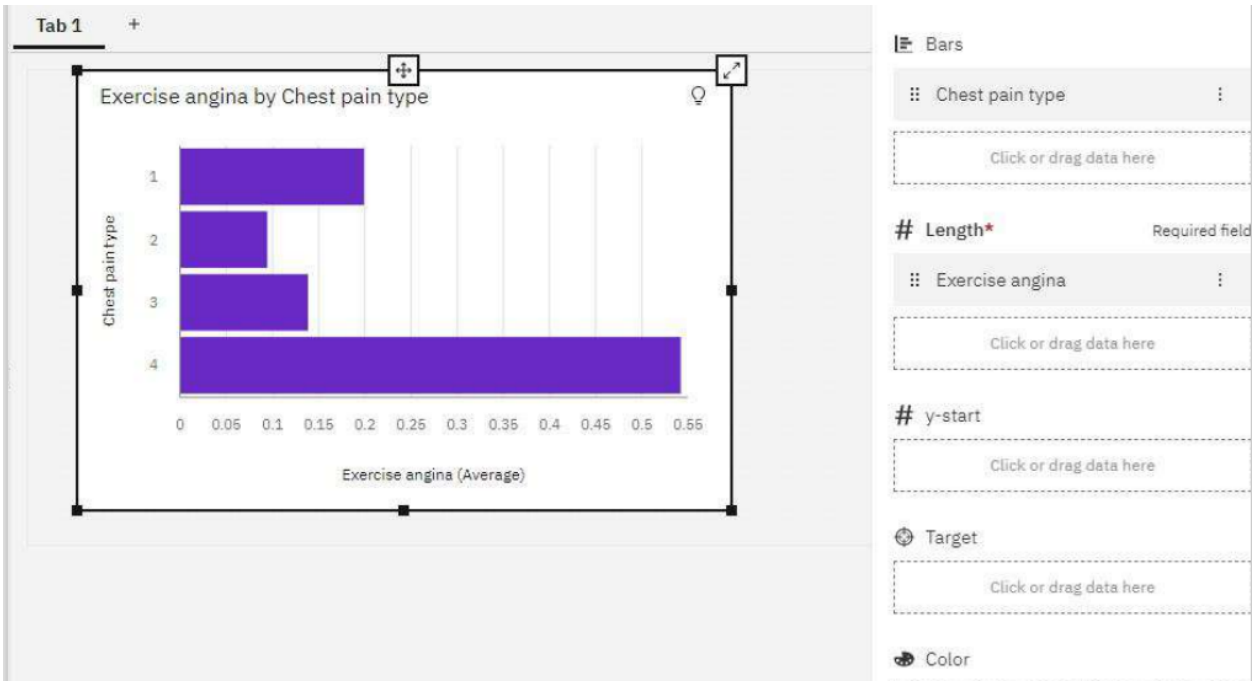
6.2Sprint Delivery Schedule

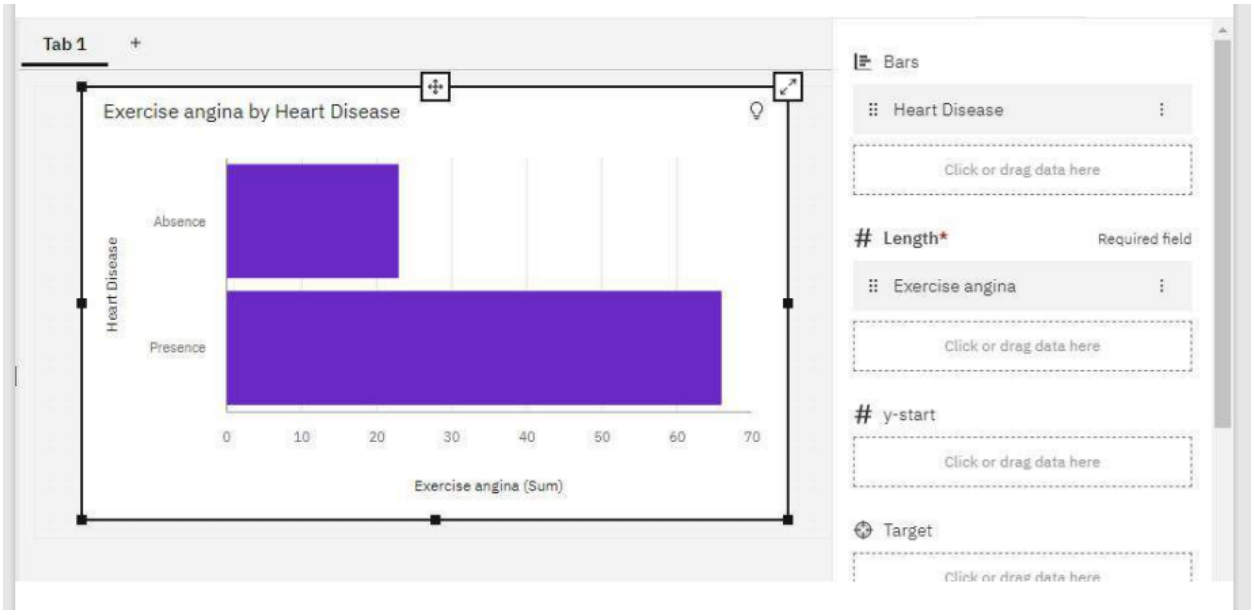
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date(Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

7.CODING & SOLUTIONING (Explain the features added in the project along with code)

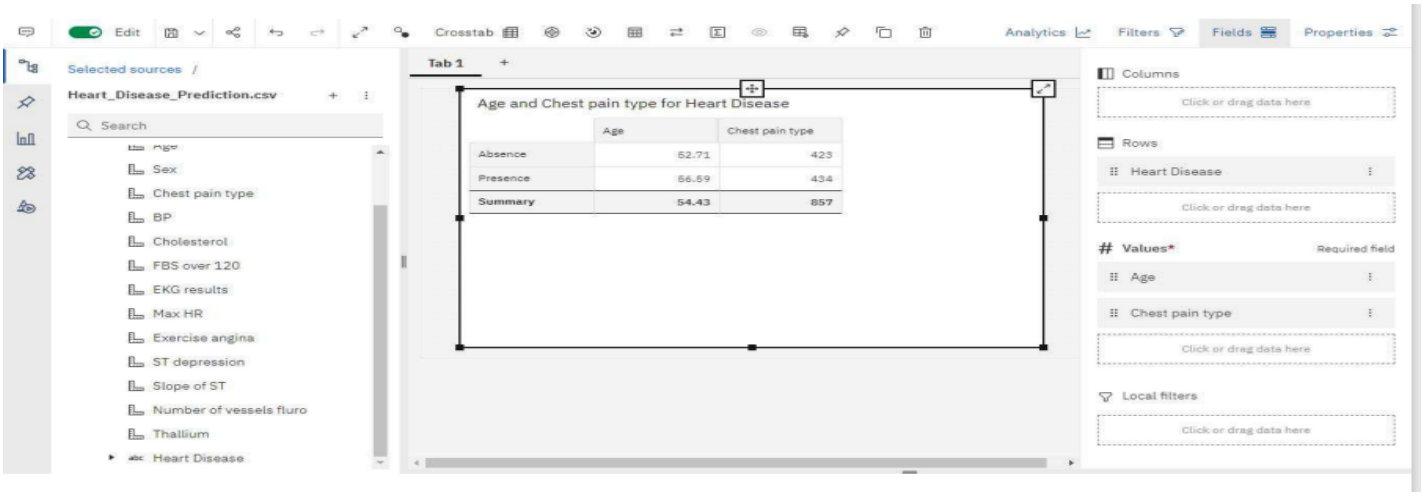

Age by Chest pain type
60
50
40
30
20
10
0
Age (Average)
3
Age (Average): 53.87
1
2
3
4


Tab 1
+
BP by Age
15
10
5
0
BP (Count)
29 35 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 74 77
Age
x-axis*
Required field
Age
Click or drag data here
Color
Click or drag data here
y-axis*
Required field
BP
Click or drag data here
Repeat (column)
Click or drag data here


Tab 1
+
Exercise angina by Chest pain type
Chest pain type
1
2
3
4
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
Exercise angina (Average)
Bars
Chest pain type
Click or drag data here
Length*
Required field
Exercise angina
Click or drag data here
y-start
Click or drag data here
Target
Click or drag data here
Color


Tab 1
+
Exercise angina by Heart Disease
Heart Disease
Absence
Presence
0
10
20
30
40
50
60
70
Exercise angina (Sum)
Bars
Heart Disease
Click or drag data here
Length*
Required field
Exercise angina
Click or drag data here
y-start
Click or drag data here
Target

7.1Feature 1



8.TESTING

Many different tests are used to diagnose heart disease. Besides blood tests and a chest X-ray, tests data set to diagnose heart disease can include:

- Electrocardiogram (ECG or EKG).
- Holter monitoring
- Echocardiogram
- Exercise tests or stress tests.
- Cardiac catheterization.
- Heart (cardiac) magnetic resonance imaging (MRI) scan.

9.RESULTS

The result of the data analysis to identify the necessary hidden patterns for predicting heart diseases are presented in the early stage throught this model . Based on the data that are collect from the patients throught varies test considered to predict the heart disease are age, chest pain type, blood pressure, blood glucose level, The pre-processed heart disease data set is then composed by K-means algorithm.The aim of this project is to predict heart disease using data mining techniques and machine learning algorithms.The random forest algorithm was used in the study to predict the occurrence of heart disease in patients. A set of important feature scores and rules were identified in diagnosing heart disease are consulted to confirm the validity of these rules. The experiments performed throught the open dataset, widely used for heart disease research yielded the highest confidence score of 98% in predicting heart disease.

10.ADVANTAGES & DISADVANTAGES

ADVANTAGES:

As we are using the pre-existing data of the patient or person, it predicts the chances of heart diseases and classies the patient based on the risk level. Due to this prediction, the patients can consult to the respective doctor as a precautionary measure and can be mental free from the thoughts of their problem

DISADVANTAGES:

Those with heart failure can develop swelling, dizziness, and other symptoms that can affect their ability to complete daily tasks. A person with diagnosed heart disease must also live with the stress of knowing they have a long-term illness that could result in a cardiac event, such as heart attack or stroke. Patients with history of CVD may experience various physical and emotional symptoms such as fatigue, edema, and sleeping difficulties that limit their physical and social activities which will in turn result in poor quality of life.

11.CONCLUSION

With the increasing number of deaths due to heart diseases, it has become mandatory to develop a system to predict heart diseases effectively and accurately. The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart disease using UCI machine learning repository dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 90.16% for prediction of heart disease. In future the work can be enhanced by developing a web application based on the Random Forest algorithm as well as using a larger dataset as compared to the one used in this analysis which will help to provide better results and help health professionals in predicting the heart disease effectively and efficiently.

12.FUTURE SCOPE

In this work, we are analyzing the data and predicting the risk of heart disease by merely a history or a data set collected from the person. But, if we want to predict the heart disease from this, there may be risks of being prone to one before the prediction. So, we would like to develop or find some algorithm to analyze data that has been streamingly collected from the wearable devices like smart watches, fitness bands, healthcare meters and analyze the data in real time to visualize the condition of the heart to the person.

13.APPENDIX

Source Code

GitHub & Project Demo Link