VISULAIZING AND PREDICTING HEART DIEASES WITH AN INTERACTIVE DASHBOARD

NALAIYA THIRAN PROJECT BASED LEARNING On PROFESSIONAL READINESS FOR INNOVATION, EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

Submitted By

SARATHKUMAR B	19110079
SHAJAN SRINIVAAS B	19110082
SHANMUGAM S	19110083
SRIDHAR S	19110089

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

HINDUSTHAN COLLEGE OF ENGINEERING AND TECHNOLOGY

Approved by AICTE, New Delhi, Accredited with 'A' Grade by NAAC

(An Autonomous Institution, Affiliated to Anna University, Chennai)

COIMBATORE – 641 032

November 2022

TABLE OF CONTENTS

1.INTRODUCTION

Heart disease defines a range of conditions that affect human heart. The name "heart disease" is often used commonly with the name "cardiovascular disease". Heart disease is a term that allow to a large number of medical circumstances related to heart. These medical circumstances characterize the irregular health condition that directly affects the heart and all its parts. Heart disease generally allows to some conditions that involve narrowed or blocked blood vessels which can lead to a heart attack, stroke or chest pain. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease .There are various types of cardiovascular disease. The most similar types are heart failure (HF) and Coronary Artery Disease (CAD). The main root cause of heart failure (HF) is occur due to the blockade or narrowing down of coronary arteries. Coronary arteries also supply blood to the heart. Data mining is a non trivial extraction of implicit, previously unknown potential useful information called as knowledge from the medical data using complex algorithms. Big data (BD) can be referred as huge record of information set. Big Data and Data Mining are two various things. The task carried out by these two methods are similar focusing on collecting the huge amount of data, handling them and preparing report on the data by taking out the information which is knowledgeable. Data Mining is basically an activity of observing the patterns in the data which is relevant and with particular information by using Big Data. The useful patterns with hidden patterns, unknown correlations are analytically handled for making knowledgeable decision.

1.1 Project Overview

Healthcare analytics refers to the use of vast amounts of collected data to provide organizations with actionable insights. These insights are developed through analytical disciplines to drive fact-based decision

making. In turn, these decisions improve planning, management, measurement and learning. As healthcare organizations around the world are challenged to reduce costs, improve coordination with care teams, provide more with less, and focus on improving patient care, analytics will be especially important. Primary care physician and nursing shortages are requiring overworked professionals to be even more productive. Plus, new businesses entering the market and new approaches to healthcare delivery will increase competition in the industry. Building analytics competencies can help healthcare organizations harness big data to create actionable insights that can be used by healthcare providers, hospital and health system leaders, and those in government health and human services to improve outcomes and deliver value for the people they serve. As tumultuous as the current healthcare environment is, it's expected to become even more complex over the next several years. Challenges such as evolving market dynamics, increasing governmental regulation and more demanding consumers will require smarter, more informed decisions from organizations so they can remain.

1.2 Purpose

The healthcare industry generates a tremendous amount of data but struggles to convert that data into insights that improve patient outcomes and operational efficiencies. Data analytics in healthcare is intended to help providers overcome obstacles to the widespread application of data-derived intelligence:

● Making healthcare data easier to share among colleagues and external partners, and easier to visualize for public consumption.

● Providing accurate data-driven forecasts in real time to allow healthcare providers to respond more quickly to changing healthcare markets and environments.

● Enhancing data collaboration and innovation among healthcare organizations to convert analytics-ready data into business-ready information by automating low impact data management tasks The tools used in analytics fall into three general categories.

● Software that acquires the data from sources that include patient surveys, case files, and machine-to-machine data transfers.

● Programs that clean, validate, and analyze the data in response to a specific research question Along with collecting, analyzing, and interpreting data, analytics software must secure the data and the analysis results while ensuring that the healthcare professionals who'll benefit from the insights have ready access to the information in a form that they can easily use in their work. Data Analytics is the process of examining raw datasets to find trends, draw conclusions and identify the potential for improvement. Healthcare analytics uses current and historical data to gain insights, macro and micro, and support decisionmaking at both the patient and business level.

2. LITERATURE SURVEY

[1] Bo Jin, Chao Che et al. (2018) proposed a "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling" model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the

sequential nature of clinical records. Advantages: Get data of patients known to have heart disease. This dataset contains information related to heart diseases like blood sugar, cholesterol and other medical information about the individual. Disadvantages: HD prediction model can be trained only 303 data of HD patient due to difficult of collecting Nepalese heart patient data but in future, we will collect large data and train model with their high accuracy. [2] Regarding the above issues, we are proposing a web-based HDPS that is one of the best solutions to efficiently and accurately predict the HD patients. The proposed system eliminates the various testing of HD and supports the decision making of doctors. This system can accept a singleton query and display the clear output of the presence of HD level. This system is useful for any hospital and clinic to evaluate the patient getting a HD. It is reduced the number of tests and provide an efficient output of patient HD. It supports to make the decision of doctors that consulates with their patients easily. Advantages: With the consideration of WHO statistical facts, the most powerful causes of death globally are a HD. It seemed to the negligence of patients as well as doctors to increase a HD patient. Some of the difficulties to execute the doctor's decision and lack of application to clearly diagnosis of HD become the cause of human death. Disadvantages: The main problem will be built a classifier model that will predict if it is getting HD or not it means how to train the classifier model so that the system will get exact if the patient is possessing HD or not that is the main challenge of this project. [3] The proposed system will support the healthcare systems as well as health-related application to expand their services with efficiently and accurately providing results. It mitigates the time to checkup of doctors.With the consideration of WHO statistical facts, the most powerful causes of death globally are a HD. It seemed to the negligence of patients as well as doctors to increase a HD patient. Some of the difficulties to execute the doctor's decision and lack of application to clearly diagnosis of HD become the cause of human death. Advantages: In the survey of this project, the interview of HD doctors will difficult to conduct due to their busy time schedule.

Thereby it will make hard to identify the right attribute of the HD. Disadvantages:Due to lack of spare time of doctor, we were unable to collect more data about HD. lack of time constraint, we were unable to explore other features to include the system and explore different heart disease type prediction. [4] We were able to research or investigate our topic by conducting literature review as domain and technical research. At domain research, we achieved 13 important attributes of HD that every researcher had done their research to predict the HD. we also achieved an appropriate two algorithm to provide high accuracy in heart disease dataset. Advantages: This system resolves the real-environment problem. It is successfully predicted the presence of HD in patient. It is also store and manage the prediction report of heart patient by doctor account. Admin user can handle create doctor account, mange doctor account and view the report of the patient. Disadvantages: For the implementation of the HDPS, we faced the problem of accuracy of algorithms due to the result has three-level, that's why data is distributed and give low accuracy. So, we chose Kaggle dataset of HD which have two results. Again, we faced the problem of low accuracy of DT and NB. [5] We made the line of research boundary to complete this project by planning the system requirement and planning. System requirement and planning consists of aim, objective, deliverables, and target audience of the system. We conducted the literature review by studying various journal papers and articles to understand the way of solving the problem as well as collecting important attributes of HD that lead role of the HD in patient. we will also apply this model into a mobile app to easily test ourselves HD. we will integrate smart wear to the hospital and police emergency system to save the life of the patient at the emergency condition. Advantages: After identifying the problem statement, we strived to find a way of solving the problem through ML. After that, we conducted requirement analysis and planning of the system. Disadvantages: The overall testing was a success but there was unit testing had a problem. These were hard to solve the displaying error message of the system. In the future, we can solve that proble.

2.2 References

Asthana, S., 2018. Essential libraries for Machine Learning in Python. [Online] Available at: https://medium.freecodecamp.org/essentiallibraries-for-machine-learningin-python-82a9ada57aeb [Accessed 08 02 2019].

1 Bahrami, B. a. S. M., 2015. Prediction and Diagnosis of Heart Disease by Data Mining Techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(2), pp. 164-168. Chala Beyene, P. K., 2018. "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques. International Journal of Pure and Applied Mathematics, 118(8), pp. 165-174. Dangare Chaitrali S., P. A. S. S., 2012. Improved Study of Heart Disease Prediction System using Data Mining Classification Technique. International Journal of Computer Applications (0975 – 888), 47(10), pp. 44-48.

2 Dwivedi, A. k., 2016. Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold crossvalidation Springer Edrawsoft, 2019. SSADM Diagram Software - Structured Systems Analysis and Design Methodology. [Online] Available at: https://www.edrawsoft.com/SSADM.php [Accessed 11 06

ACS (American College of Surgeons) Fact Sheet. The National Trauma Registry of the American College of Surgeons (Draft). Chicago: American College of Surgeons, 1992.

3. Adler, N.E., Boyce, T., Chesney, M.A., et al. Socioeconomic Inequalities in Health. No Easy Solution. Journal of the American Medical Association 269:3140-3145, 1993. [PubMed].

4. AHCPR (Agency for Health Care Policy and Research). AHCPR Purpose and Programs. Rockville, Md.: Department of Health and Human Services, Public Health Service, AHCPR, 1990. a.

5. AHCPR. Program Note. Medical Treatment Effectiveness Research. Rockville, Md.: Department of Health and Human Services, Public Health Service, AHCPR, 1990. b.

6. Allen, A. Uneasy Access: Privacy for Women in a Free Society. Totowa, N.J.: Rowman and Allanheld, 1987.

7. AMA (American Medical Association). Principles of Medical Ethics, Current Opinions of the Council on Ethical and Judicial Affairs of the AMA Including the Principles of Medical Ethics and Rules of the Council on Ethical and Judicial Affairs. 1992.

8. AMPRA (American Medical Peer Review Association). Managed Competition and the Role of Quality Oversight. Washington, D.C.: AMPRA, 1993

9. Anderson, H.J. Managing Data in a New Era. Medical Claims Management 1(4):29- 36, 1993. [PubMed]

10. Andrussier S.E., The Freedom of Information Act in 1990: More Freedom for the Government; Less Information for the Public. Duke Law Journal 729:753-801, 1991.

2.2 Problem Statement Definition

The healthcare Application-e-Hospital within the state has a unique opportunity to leverage data analytics to conduct scientific research, clinical trials, develop personalized and genetic medicine, and use medical data to chart out public policies. Data in healthcare comes from various sources such as biometric, patient records, prescription, and machines. Big data analytics can be used on the data emanating from all

these sources to generate actionable insights, predict outcomes, and plan treatment protocols for effective public health Heart disease defines a range of conditions that affect human heart. The name "heart disease" is often used commonly with the name "cardiovascular disease". Heart disease is a term that allow to a large number of medical circumstances related to heart. These medical circumstances characterize the irregular health condition that directly affects the heart and all its parts. Heart disease generally allows to some conditions that involve narrowed or blocked blood vessels which can lead to a heart attack, stroke or chest pain. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease .There are various types of cardiovascular disease. The most similar types are heart failure (HF) and Coronary Artery Disease (CAD). The main root cause of heart failure (HF) is occurdue to the blockade or narrowing down of coronary arteries. Coronary arteries also supply blood to the heart. Data mining is a non trivial extraction of implicit, previously unknown potential useful information called as knowledge from the medical data using complex algorithms. Big data (BD) can be referred as huge record of information set. Big Data and Data Mining are two various things. The task carried out by these two methods are similar focusing on collecting the huge amount of data, handling them and preparing report on the data by taking out the information which is knowledgeable. Data Mining is basically an activity of observing the patterns in the data which is relevant and with particular information by using Big Data. The useful patterns with hidden patterns, unknown correlations are analytically handled.

3.IDEATION & PROPOSED SOLUTION

- Analysis carried out states that the multiple regression testing provides the maximum accuracy of 98.51%
- Our system will include the multiple linear regression techniques to maximize the favourable output
- It would also be facilitated as a web or android application which third party users can make use of model which is trained
- The model in the proposed system will be trained with data extracted from Kaggle so that to provide maximum

3.1 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to helps teams better understand their users. Empathy mapping is a simple workshop activity that can be done with stakeholders, marketing and sales, product development, or creative teams to build empathy for end users. For teams involved in the design and engineering of products, services, or experiences, an empathy mapping session is a great exercise for groups to "get inside the heads" of users.

Template

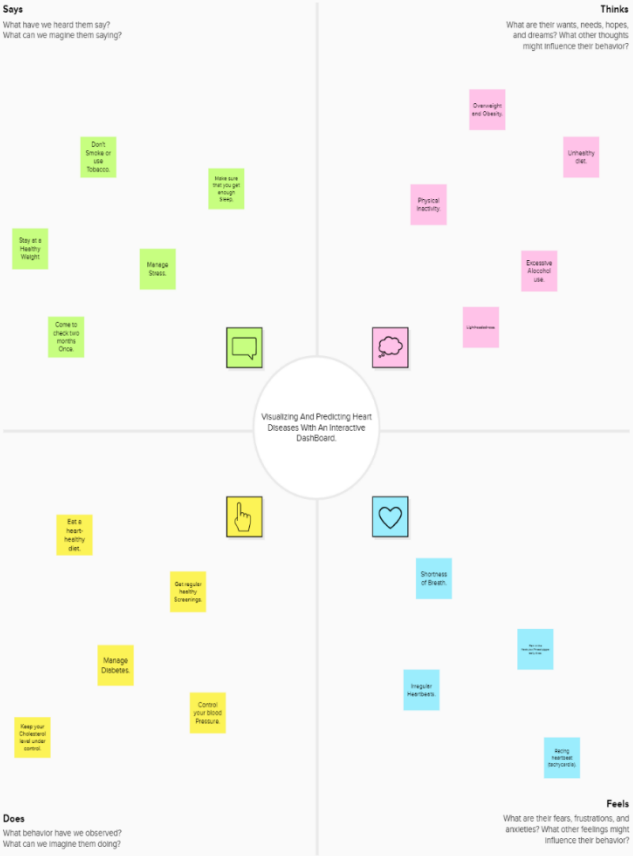
Empathy map

Use this framework to develop a deep, shared understanding and empathy for other people. An empathy map helps describe the aspects of a user's experience, needs and pain points, to quickly understand your users' experience and mindset.

Share template feedback

Build empathy

The information you add here should be representative of the observations and research you've done about your users.



Gain

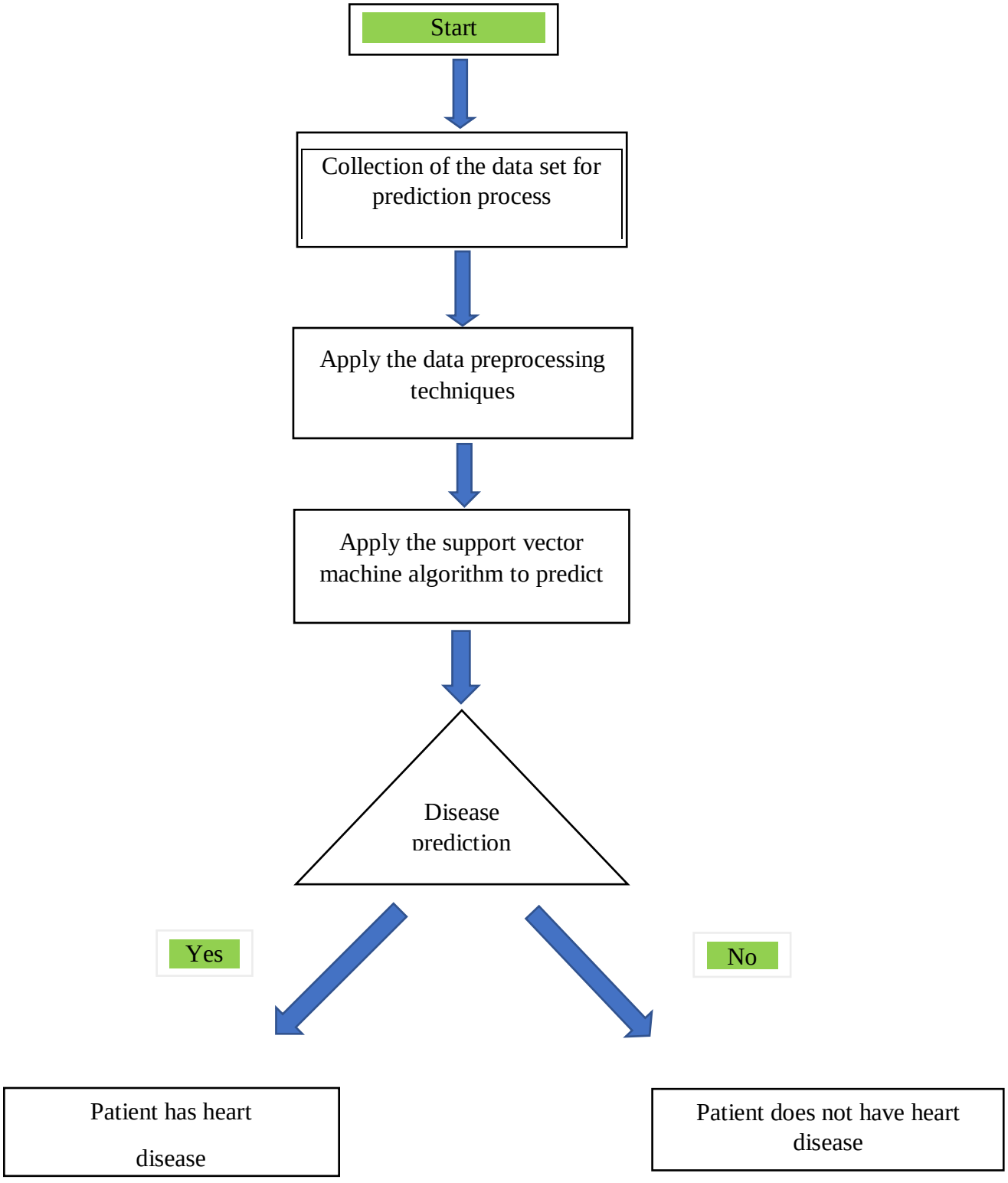
Limit Unhealthy fats.
Limit or reduce Salt.
Eat more vegetables and fruits.
Choose low-fat protein sources.
Allow yourself an Occasional treat.

Pain

Rapid or irregular Heartbeat.
Pain on both Arms and shoulders.
Chest pain if heart failure is caused by a HeartAttack.

3.2 Ideation & Brainstorming

The Ideation First Burst Map is a visual map of the initial ideas and solutions that are identified following research and empathy building during the Inspiration phase of design thinking.

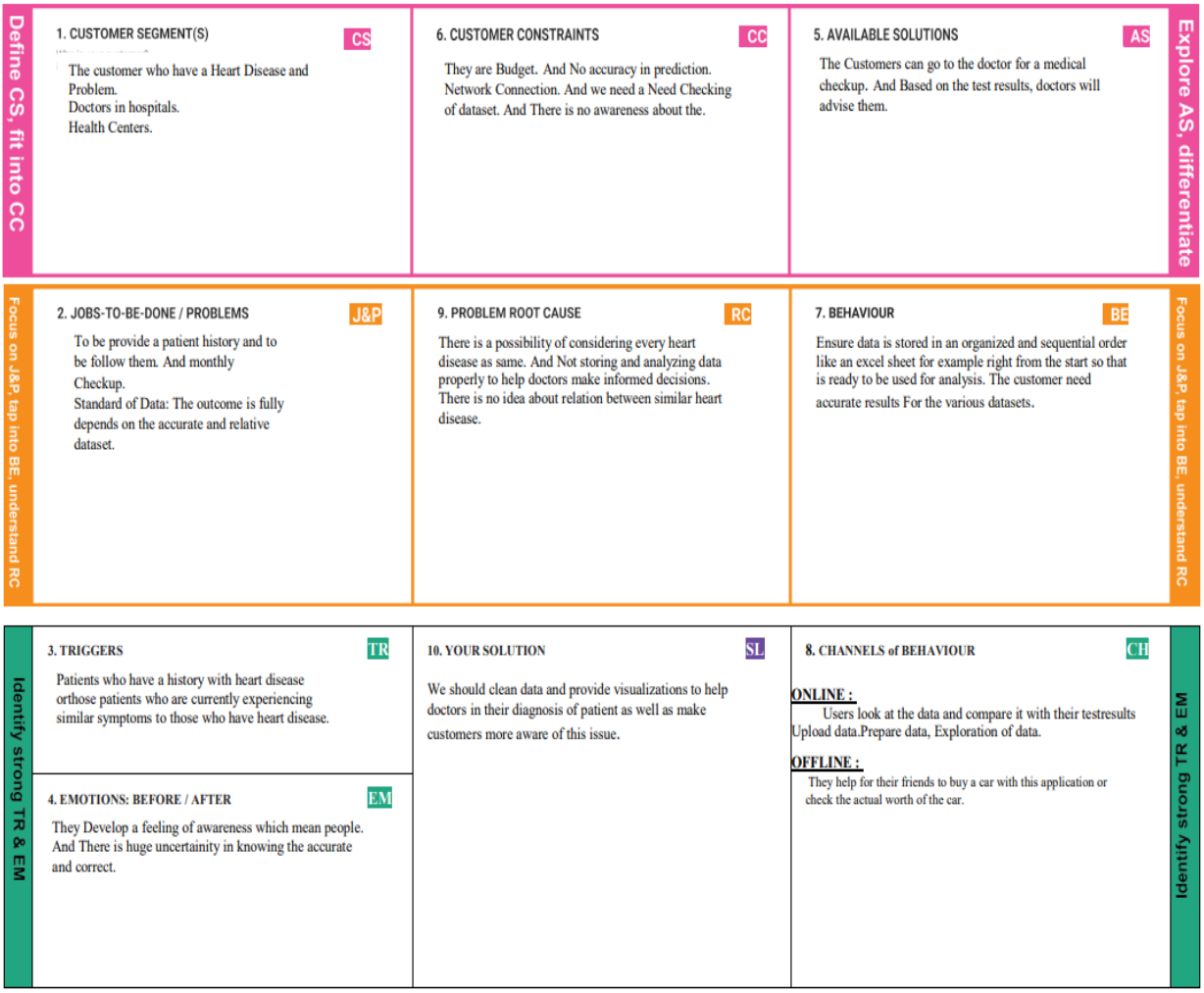


3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem statement	In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product.
2.	Idea/solution description	Algorithms like random forest classification will be used to provide an optimal accuracy for the problem which will be provided to the end user as a form of live web application facilitated with IBM watson.

3.	Novelty/uniqness	Customer provided data will also be used to train the model to provide more accurate results.
4.	Social Impact / Customer Satisfaction	End user feels that it is trustworthy as it is not dependent other factors which may lead to the misusing of the information and getting scammed as a result.
5.	Business Model (Revenue Model)	Calculating the price of the secondhand cars while the process of providing leasing documents bounds to generate certain income.
6.	Scalability of the Solution	The model which is framed is bound to be scalable as it is equipped with datasets which is recently framed.

3.4 Problem Solution Fit



4.REQUIREMENTS ANALYSIS

4.1 Functional Requirements

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
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Dashboard	Mandetory field for analyzing the price
FR-4	Result	The Price will be shown based on the given details

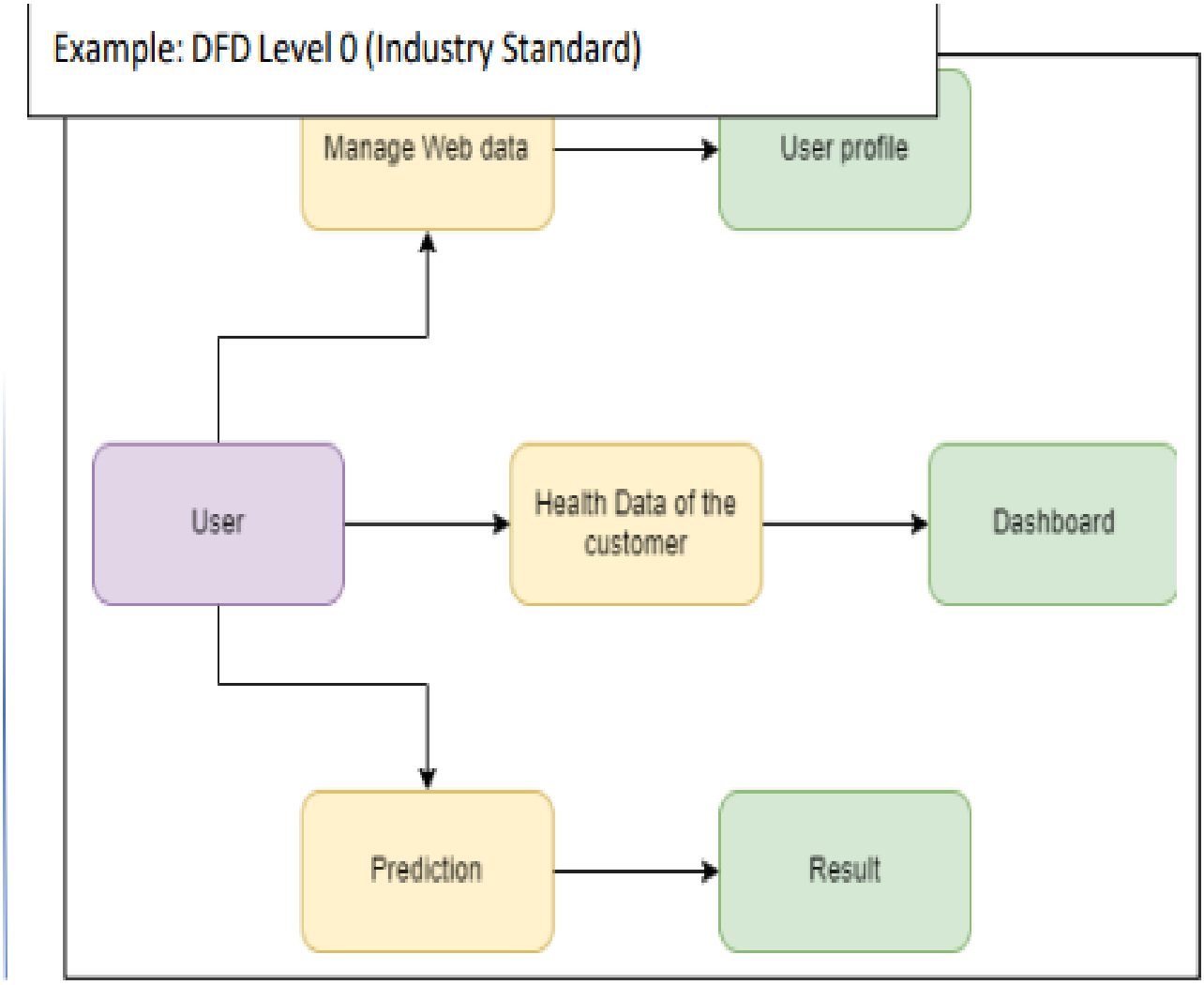
4.2 Non-Functional Requirements

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

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The usability of the Application is far more better than any other application
NFR-2	Security	The information about the car or user must be secured
NFR-3	Reliability	The result will be shown without any failures
NFR-4	Performance	Processing of data won't take much time,so that the performance speed will be fast.
NFR-5	Availability	Any user can use it without any cost.
NFR-6	Scalability	The performance will be fast , so that the user get the fast result because of this possible of scalability is low.

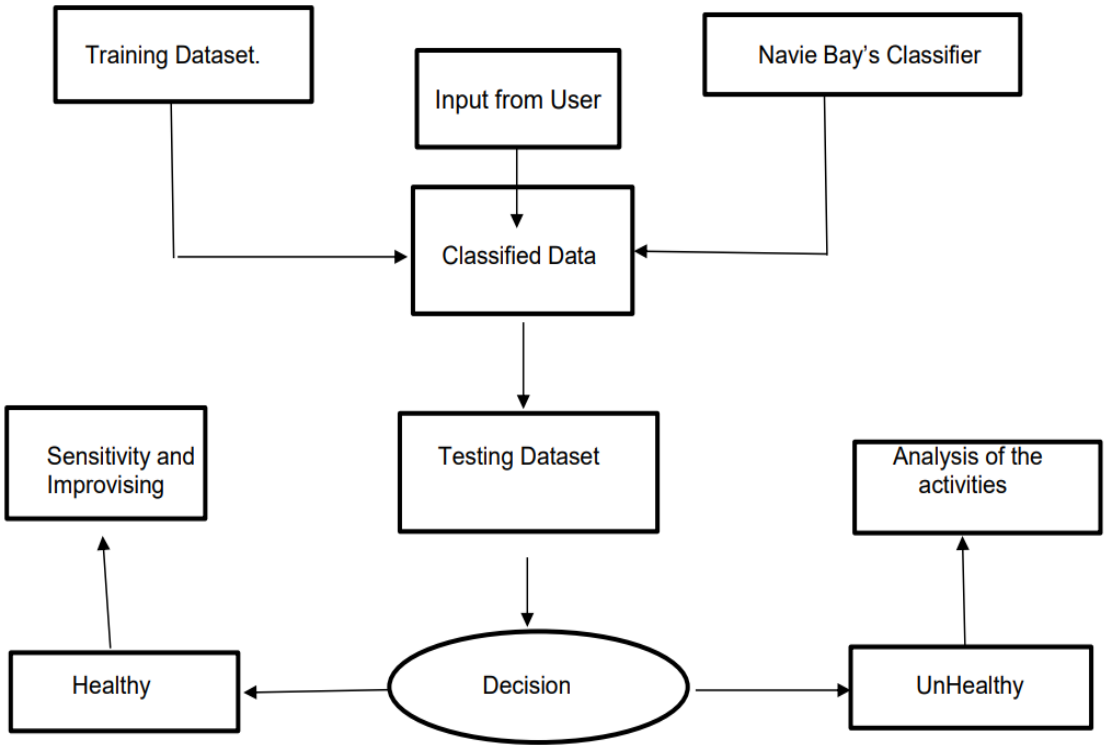
5.PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2



Guidelines:

1. Include all the processes (As an application logic / Technology Block)
2. Provide infrastructural demarcation (Local / Cloud)
3. Indicate external interfaces (third party API's etc.)
4. Indicate Data Storage components / services
5. Indicate interface to machine learning models (if applicable

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	Collect Dataset	ML is a data hunger technology, it depends heavily on data, without data, it is impossible for a machine to learn. It is the most crucial aspect that makes algorithm training possible.	We store the dataset as a Excel.
2.	Pre-Process the data	1. Handling the null values. 2. Handling the categorical values if any. 3. Normalize the data if required. 4. Identify the dependent and independent variables. 5. Split the dataset into train and test sets.	Python

3.	Model Building	There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms can be chosen according to the objective. As the dataset which we are using is a REgression dataset so you can use the following algorithms Multi Linear Regression Random Forest Regression / Classification Decision Tree Regression / Classification K-Nearest Neighbors Support Vector Machine	Python
4	Application Building	After the model is built, we will be integrating it into a web application so that normal users can also use it to know the resale price of the care. In the application, the user provides the parameter values affecting the resale value.	HTML,CSS,Python flask
5	File Storage	File storage is required to store the dataset(Excel)	Local Filesystem

6	Machine Learning Model	Different regression models can be used to know the performance and choose whichever works better.	•Multi Linear Regression •RandomForest RegressionClassification •DecisionTree Regression Classification •K-Nearest Neighbors •Support Vector Machine

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	**Flask:** It is a framework of microweb that is written in Python language and is classified as a microframework because it does not need any particular libraries and tools. Database abstraction layer, form validation and other such components with third-party libraries providing functionalities are all absent in flask.	Python Flask

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	IBM Cognos / Python .
2.	Data Set	The data set prepared for hospitals health care	Python .
3.	IBM Cognos	Data analytics platform	IBM Watson service
4.	Data Import	Data set is imported in IBM cognos	IBM Watson Assistant
5.	Data Cleaning	Data is cleaned by using some mathematical techniques such as mean,mode etc.to clean the null and missing data.	IBM Assistant
6.	Data Exploration	Cleaned data can be explored.	IBM Cognos
7.	Story Card	Data is explored and story card was prepared for visual representation	IBM Cognos
8.	IBM Cloud	Storage of data	IBM DB2
9.	Data Explored and Deployed	Purpose of External API to explored and deployed	Data deployed to user by UI
10.	Admin	Purpose of Data set model	Recognition of data set model etc.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source	Open source model is used for the data set	Python
2.	Security Implementations	Security for our data set	SHA 256, SHA 1
3.	Scalable Architecture	Health care service utilizes the relational patient data and big data analytics to tailor the medication recommendations	Python
4.	Availability	The availability of technology used in data analytics	Python-Anaconda distribution and jupyter notebook is available and open source application
5.	Performance	The performance of the application and its efficiency	Python and other languages is that Python is usually interpreted. Interpreted languages Tend to perform worse than compiled languages, each command takes up a greater number of machine instructions.

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	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
		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
		3	As a user, I can register for the application through Facebook.	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		4	As a user, I can register for the application through Gmail.		Medium	Sprint-1
Customer (Web user)	Login	5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard	6	User can able to view only his medical records.	I can view it in Dashboard	High	Sprint-2
		7	User can able to view the possibilities of occurrence of heart disease.	I can view it in the analysis reports.	High	Sprint-2
Customer Care Executive	Helpdesk	8	Able to view the queries	I can able to post queries on dashboard	Medium	Sprint-3
		9	Able to answer queries	I can able to view the answers for those queries	High	Sprint-3
Administrator	User Profile	10	Able to update the users medical records	I can view my updated health details.	High	Sprint-4
		11	Able to add or delete users	I can access my accounts when logged in.	High	Sprint-4
		12	Able to manage the user details	I can view the organized data of myself.	High	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

TITLE	DESCRIPTION	SUBMITTED DATE
Literature Survey and Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc.	17 September 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	17 September 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	21 September 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	27 October 2022

Problem Solution Fit	Prepare problem - solution fit document.	27 October 2022
Solution Architecture	Prepare solution architecture document.	27 October 2022

Functional Requirements	Prepare the functional requirement document.	28 October 2022
Technology Architecture	Prepare the technology architecture diagram.	28 October 2022
Data Flow Diagram & User Stories	Draw the data flow diagrams and User Stories submit for review.	28 October 2022
Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application.	31 October 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	03 November 2022
Project Development - Delivery of Sprint-1, 2, 3 & 4	Develop & submit the developed code by testing it.	IN PROGRESS

6.2 Sprint Delivery Schedule

Product Backlog, Sprint Schedule, and Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	**registration**	USN-1	Collect Dataset	1	Low	**Srither**
Sprint-1	**confirmation**	USN-2	Import Required Libraries	2	Low	**Shajan sri nivaas**
Sprint-2		USN-3	Read and Clean data sets	2	Low	**Srither**
Sprint-1		USN-4	Split data into independent and dependent variables	3	Medium	**Sarath kumar**
Sprint-1	**Login**	USN-5	Apply using regression model	3	Medium	**Srither**
Sprint-1	**User interface**	USN-6	Build python flask application and html page	5	High	**Shanmugam**
Sprint-3	**Dashboard**	USN-7	Execute and test	5	High	**Shajan sri nivaas**
Sprint-4	**Present data**	USN-8	Train machine learning model	5	High	**Sarath kumar**

Sprint-4		USN-9	Integrate Flask	5	High	**shanmugam**

Project Tracker, Velocity & Burndown Chart:

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

Velocity:

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

Sprint 1: 1 user stories x 20 story points=20

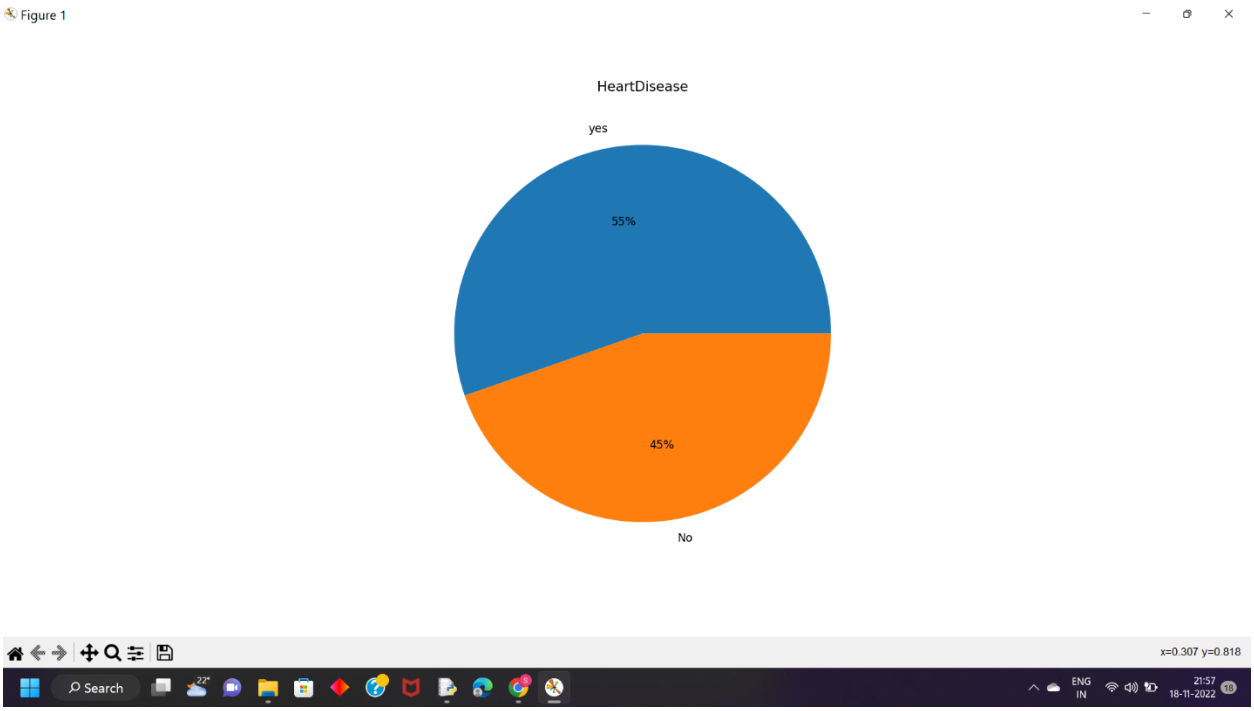
Sprint 2: 1 user stories x 20 story points=20

Sprint 3: 1 user stories x 20 story points=20

Sprint 4: 1 user stories x 20 story points=20

Total = 80

Average Sprint Velocity is 80/4=20



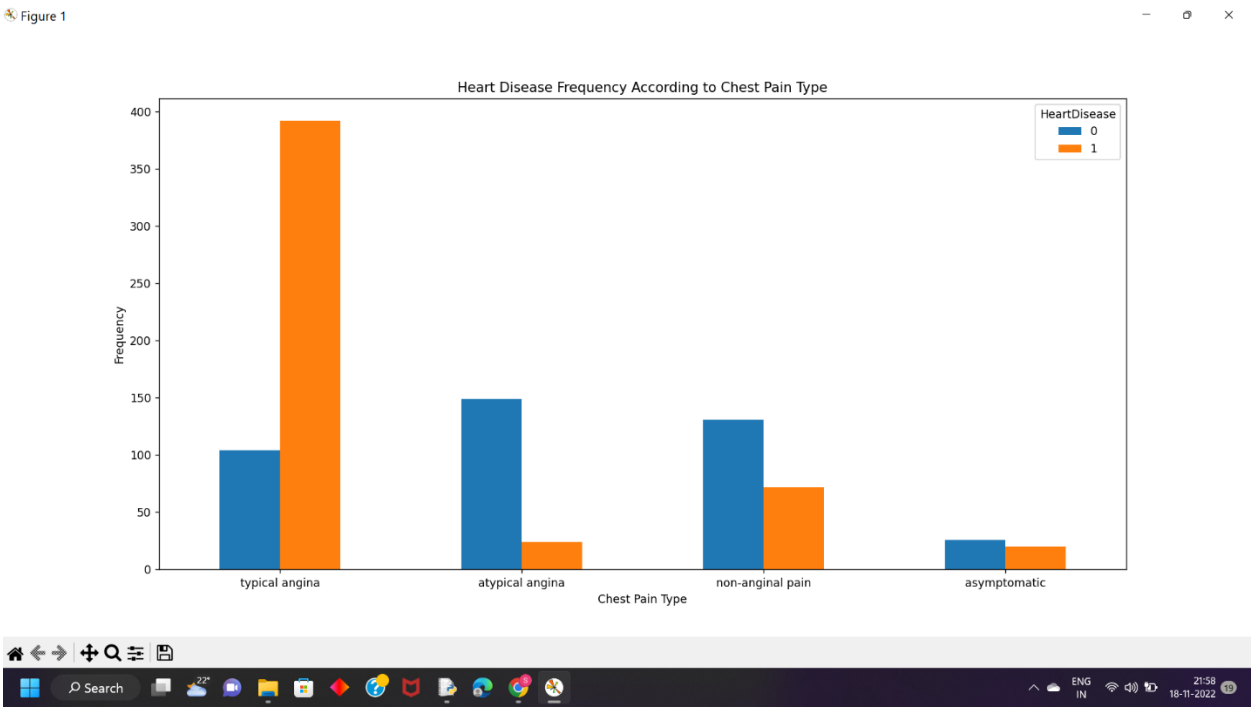


Figure 1
Heart Disease Frequency According to Chest Pain Type
HeartDisease
0
1
Frequency
400
350
300
250
200
150
100
50
0
typical angina
atypical angina
non-anginal pain
asymptomatic
Chest Pain Type
Search
ENG
IN
21:58
18-11-2022

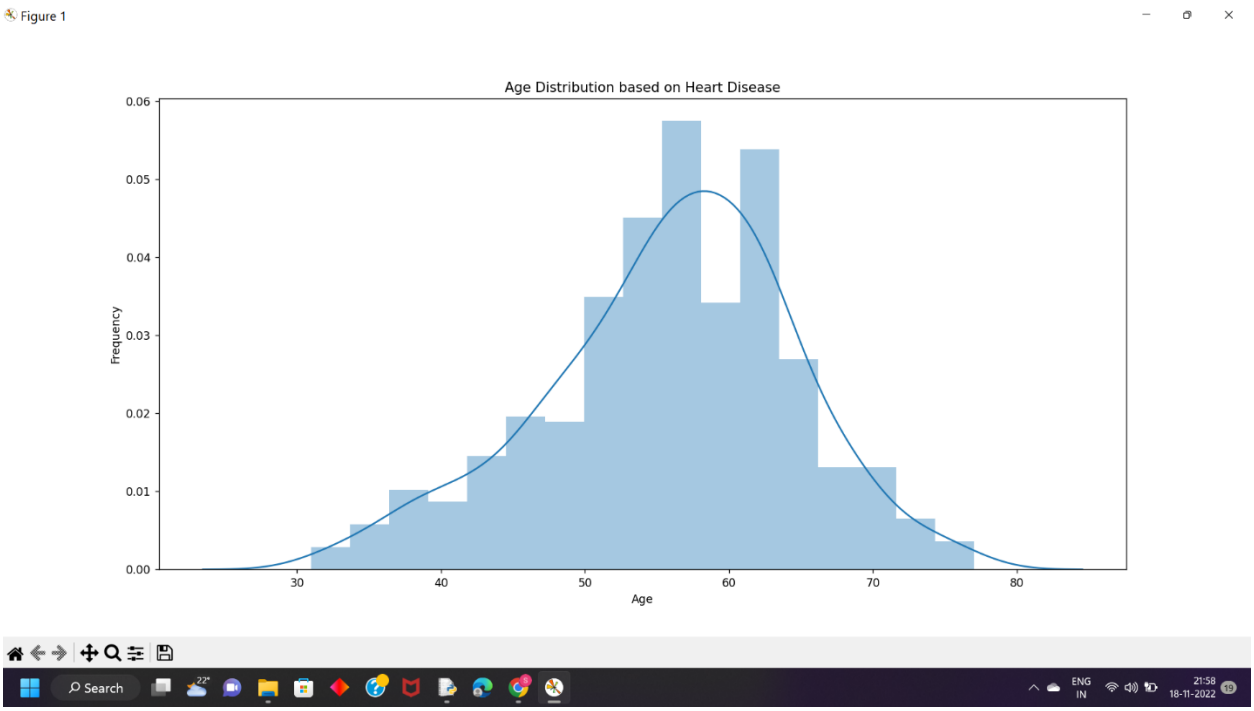
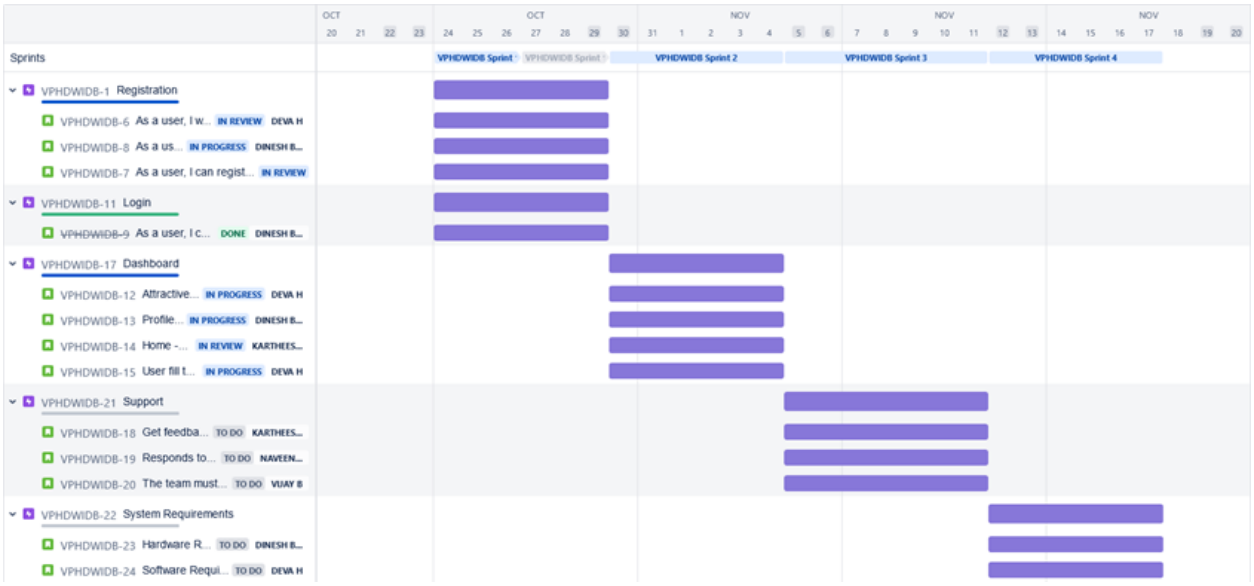


Figure 1
Age Distribution based on Heart Disease
Frequency
0.06
0.05
0.04
0.03
0.02
0.01
0.00
30
40
50
60
70
80
Age
Search
ENG
IN
21:58
18-11-2022

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down.

6.3 Reports from JIRA



7. CODING & SOLUTIONING

7.1 Feature Code 1

```
import pandas as pd
import numpy as np
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

df=pd.read_csv('heart.csv')    #to read the file
##print(df.head())

# Create a plot to display the percentage of the positive and negative heart disease
labels = ['yes', 'No']
values = df['HeartDisease'].value_counts().values

plt.pie(values, labels=labels, autopct='%1.0f%%')
plt.title('HeartDisease')
```

```
plt.show()


# Display chest pain types based on the Heart Disease
pd.crosstab(df.ChestPainType,df.HeartDisease).plot(kind = "bar", figsize = (8, 6))
plt.title('Heart Disease Frequency According to Chest Pain Type')
plt.xlabel('Chest Pain Type')
plt.xticks(np.arange(4), ('typical angina', 'atypical angina', 'non-anginal pain', 'asymptomatic'), rotation = 0)
plt.ylabel('Frequency')
plt.show()



# Get min, max and average of the age
print('Min age: ', min(df['Age']))
print('Max age: ', max(df['Age']))
```

7.2 Feature Code 2

```
# Display age distribution based on heart disease
sns.distplot(df[df['HeartDisease'] == 1]['Age'], label='Have heart disease')
sns.distplot(df[df['HeartDisease'] == 2]['Age'], label = 'Do not have heart disease')
plt.xlabel('Age')
plt.ylabel('Frequency')
```

```
plt.title('Age Distribution based on Heart Disease')
plt.show()



# Get min, max and average of the age of the people do not have heart diseas
print('Min age of people who do not have heart disease: ', min(df[df['HeartDisease'] == 1]['Age']))
print('Max age of people who do not have heart disease: ', max(df[df['HeartDisease'] == 1]['Age']))



le=LabelEncoder()
df['Age'] = le.fit_transform(df['Age'])
df['Sex'] = le.fit_transform(df['Sex'])
df['ChestPainType'] = le.fit_transform(df['ChestPainType'])
df['RestingBP'] = le.fit_transform(df['RestingBP'])
df['Cholesterol'] = le.fit_transform(df['Cholesterol'])
df['FastingBS'] = le.fit_transform(df['FastingBS'])
df['RestingECG'] = le.fit_transform(df['RestingECG'])
df['MaxHR'] = le.fit_transform(df['MaxHR'])
df['ExerciseAngina'] = le.fit_transform(df['ExerciseAngina'])
df['Oldpeak'] = le.fit_transform(df['Oldpeak'])
df['ST_Slope'] = le.fit_transform(df['ST_Slope'])
```

```
NB = GaussianNB()

x=df.drop(columns=['HeartDisease'])
y=df['HeartDisease']      #to create the variable
print(x)
print(y)



x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_s
tate=4)   #split the val
print(x_test)
print(y_test)



NB.fit(x_train, y_train)  #train the data
y_pred=NB.predict(x_test)
print(y_pred)
print(y_test)
print('ACCURACY is', accuracy_score(y_test,y_pred))

##import pickle
##pickle.dump(NB,open('model.pkl','wb'))
```

```
##testPrediction = NB.predict([[50,1,0,145,0,1,1,139,1,0.7,1]])
##if testPrediction==1:
##    print(testPrediction,"The Patient Have Heart Disease,please consult the Doctor")
##else:
##    print(testPrediction,"The Patient Normal")
##39,1,2,120,339,0,1,170,0,0,2
##36,1,1,120,166,0,1,138,0,0,2
##51,0,0,120,0,1,1,127,1,1.5,2
```

8 TESTING

8.1 Test Cases

Missing values

The trained ML model requires 4 feature inputs for predicting the output. Failing which, the model throws invalid Input error. All the fields in the html form have been marked required using CSS and thus user must input all fields.

Output: User must input all the fields, failing which, form shows warning message "this field needs to be filled".Thus, there can be no errors in model prediction.

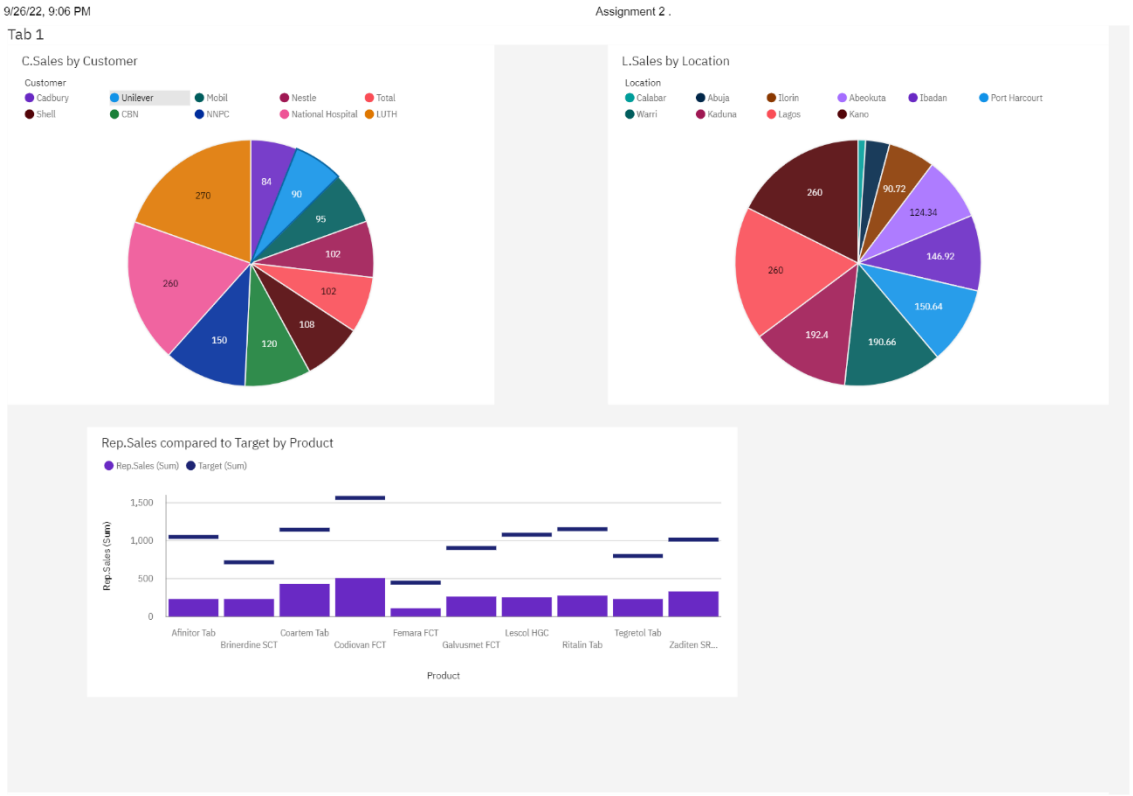
Invalid Input

The trained ML model requires only numerical input for all 4 features. Thus, if user uses symbols such as comma while input, model may throw error. To overcome the same, preprocessing script is deployed in backend which removes all unwanted characters like comma, whitespaces etc. so that model gets required input.

Output:

Due to python preprocessing script, model will get the desired input and thus will give accurate prediction.

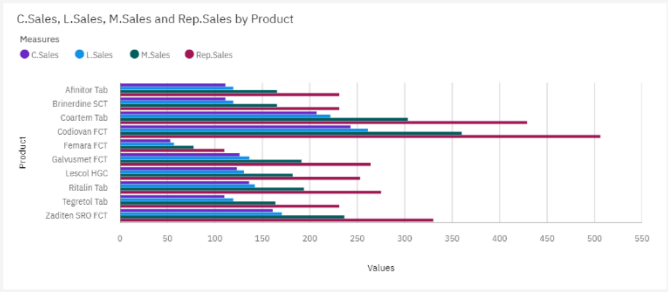
8.2 User Acceptance Testing



Tab 2

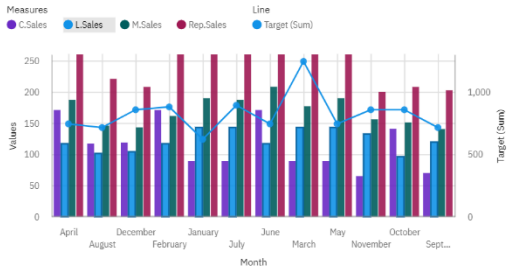

Target compared to Received Inventory by Supplier colored by Supplier
Supplier
Astra Zeneca
Novartis
Pfizer
Sanofi
Target (Sum)
600,000
400,000
200,000
0
Received Inventory (Sum)
Astra Zeneca
Novartis
Pfizer
Sanofi
Supplier


Target compared to Inventory Stock by Warehouse Locations colored by Warehouse Locations
Warehouse Locations
Amuwo
Apapa
Calabar
Ikeja
In Transit
Kaduna
Kano
Port Harcourt
Target (Sum)
80,000
60,000
40,000
20,000
0
Inventory Stock (Sum)
Amuwo
Apapa
Calabar
Ikeja
In Transit
Kaduna
Kano
Port Harcourt
Warehouse Locations

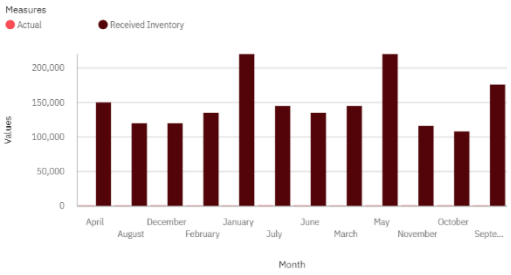

C.Sales, L.Sales, M.Sales and Rep.Sales by Product
Measures
C.Sales
L.Sales
M.Sales
Rep.Sales
Afinitor Tab
Brinerdine SCT
Coartem Tab
Codiovan FCT
Femara FCT
Galvusmet FCT
Lescol HGC
Ritalin Tab
Tegretol Tab
Zaditen SRO FCT
Product
0
50
100
150
200
250
300
350
400
450
500
550
Values

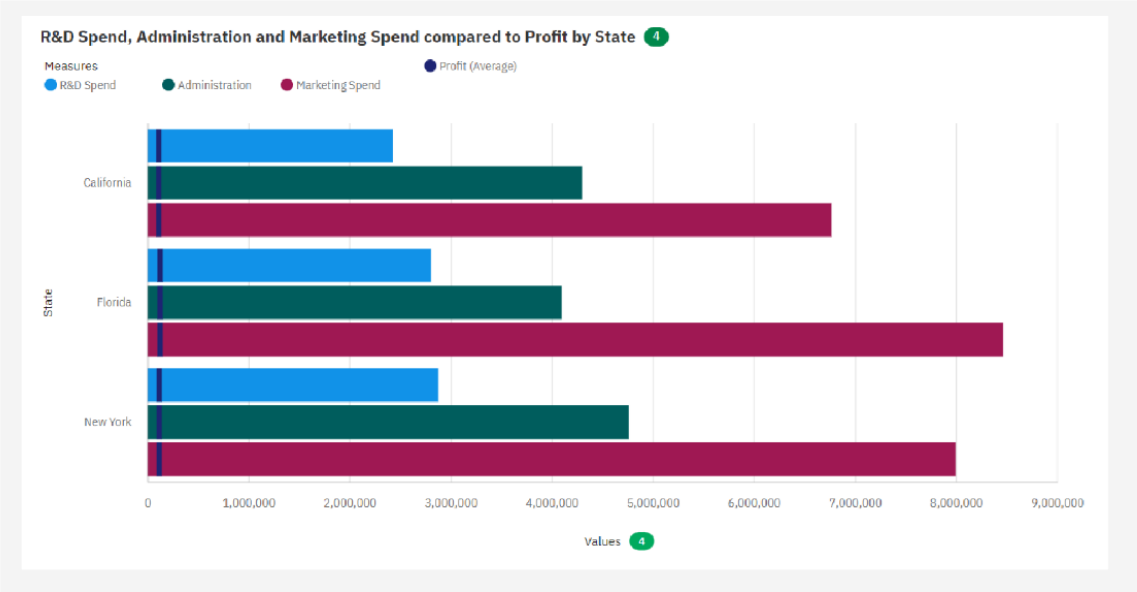
Tab 3

Target and C.Sales, L.Sales, M.Sales and Rep.Sales by Month



Actual and Received Inventory by Month




R&D Spend, Administration and Marketing Spend compared to Profit by State
4
Measures
Profit (Average)
R&D Spend
Administration
Marketing Spend
California
Florida
New York
State
0
1,000,000
2,000,000
3,000,000
4,000,000
5,000,000
6,000,000
7,000,000
8,000,000
9,000,000
Values
4

9.RESULTS

9.1 Performance Metrics

PERFORMANCE METRICS Health care data originate from many sources, including electronic health records (EHRs), medical imaging, payor records, pharmaceuticals, wearables, and medical devices. This data dramatically differs from that of other industries in that it streams in higher volumes and velocities. Thanks to technological advances in analytics, hospitals can store and analyze this data using software tools to make smarter and cost-effective decisions.The analyzed data is used in several applications, resulting in increased cost savings: Operational efficiency: Data collected from admissions and discharges is used to analyze staff efficiency and productivity during varying patient volumes. This analysis can lead to more efficient use of personnel resources while improving patient care. Proactive medical care: Big data garnered from electronic health records (EHRs) such as clinical data, conditions, and diagnoses are used to study more effective treatments for patients. Being proactive with inpatient treatment lowers the duration of hospital stays, decreasing costs for both the health care facility and the patient. Medical equipment maintenance: Critical medical equipment, such as MRI scanners, requires preventative maintenance to ensure proper operation 24/7. Data from sensors in the machines can predict when it’s time to replace critical components and prevent sudden, costly breakdown.

ADVANTAGE & DISADVANTAGE

Advantages: In the survey of this project, the interview of HD doctors will difficult to conduct due to their busy time schedule. Thereby it will make hard to identify the right attribute of the HD. Disadvantages:Due to lack of spare time of doctor, we were unable to collect more data about HD. lack of time constraint, we were unable to explore other features to include the system and explore different heart disease type prediction. We were able to research or investigate our topic by conducting literature review as domain and technical research. At domain research, we achieved 13 important attributes of HD that every researcher had done their research to predict the HD. we also achieved an appropriate two algorithm to provide high accuracy in heart disease dataset. Advantages: This system resolves the real-environment problem. It is successfully predicted the presence of HD in patient. It is also store and manage the prediction report of heart patient by doctor account. Admin user can handle create doctor account, mange doctor account and view the report of the patient. Disadvantages: For the implementation of the HDPS, we faced the problem of accuracy of algorithms due to the result has three-level, that's why data is distributed and give low accuracy. So, we chose Kaggle dataset of HD which have two results. Again, we faced the problem of low accuracy of DT and NB. We made the line of research boundary to complete this project by planning the system requirement and planning. System requirement and planning consists of aim, objective, deliverables, and target audience of the system. We conducted the literature review by studying various journal papers and articles to understand the way of solving the problem as well as collecting important attributes of HD that lead role of the HD in patient. we will also apply this model into a mobile app to easily test ourselves HD. we will integrate smart wear to the hospital and police emergency system to save the life of the patient at the emergency condition. Advantages: After identifying the problem statement, we strived to find a way of solving the problem through ML. After that, we

conducted requirement analysis and planning of the system. Disadvantages: The overall testing was a success but there was unit testing had a problem. These were hard to solve the displaying error message of the system. In the future, we can solve that problem.

- Research Predictive analytics also can be influential for medical research purposes. In drug trials, experiments are conducted on a small group of subjects to ensure that the drug will be successful when implemented on a larger scale. Researchers with access to the patient's data can effectively study the impact of any specific treatment over patient populations. All this research is mostly done by collecting and then analysing large sets of data. The outcomes of these types of research can be improvement in treatment procedures.

REFERENCE

[1] Asthana, S., 2018. Essential libraries for Machine Learning in Python. [Online] Available at: https://medium.freecodecamp.org/essentiallibraries-for-machine-learningin-python-82a9ada57aeb [Accessed 08 02 2019].

[2] Bahrami, B. a. S. M., 2015. Prediction and Diagnosis of Heart Disease by Data Mining Techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(2), pp. 164-168. Chala Beyene, P. K., 2018. "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques. International Journal of Pure and Applied Mathematics, 118(8), pp. 165-174. Dangare Chaitrali S., P. A. S. S., 2012. Improved Study of Heart Disease Prediction System using Data Mining Classification Technique. International Journal of Computer Applications (0975 – 888) , 47(10), pp. 44-48.

[3] Dwivedi, A. k., 2016. Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold crossvalidation. Springer. Edrawsoft, 2019. SSADM Diagram Software - Structured Systems Analysis and Design Methodology. [Online] Available at: https://www.edrawsoft.com/SSADM.php [Accessed 11 06 2019].