

Visualizing and Predicting Heart Diseases with an Interactive Dash Board

Team Id: PNT2022TMID12648

Project Report

1. INTRODUCTION

1.1 Project Overview

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. This project aims to create an interactive Dashboard using IBM Cognos Tool and dataset to predict which patients are most likely to suffer from a heart disease in the near future using the features given.

1.2 Purpose

Heart disease (HD) is a major cause of mortality in modern society. Medical diagnosis is an extremely important but complicated task that should be performed accurately and efficiently. Cardiovascular disease is difficult to detect due to several risk factors, including high blood pressure, cholesterol, and an abnormal pulse rate. Based on the analytics we can analyze which patients are most likely to suffer from heart disease in the near future and based on the patient details we will make decisions to cure them.

2. LITERATURE SURVEY

2.1 Existing problem

Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analysing 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 analysing data..

2.2 Problem Statement Definition

Who does the problem affect? People with unhealthy lifestyles, stress, depression, age above 40 and when their ancestors got heart disease (since heart disease is hereditary). When does the issue occur? The issue occurs for people with unhealthy lifestyles and age above 40. Where is the issue occurring? The issue is originating from an unhealthy lifestyle. It mostly occurs in the blood valves of the heart. What would happen if we didn't solve the problem? If we don't solve the problem, many people will die at a young age. The death rate due to heart disease will

increase rapidly. Why is it important to fix the problem? We should predict the problem before giving treatment to the patients. As the problem is predicted early, we can solve it easily and early.

2.3 Reference

Heart Disease Prediction using Exploratory Data Analysis” R. Indrakumari, T.Poongodi, Soumya Ranjan Jena

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, Kmeans 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.

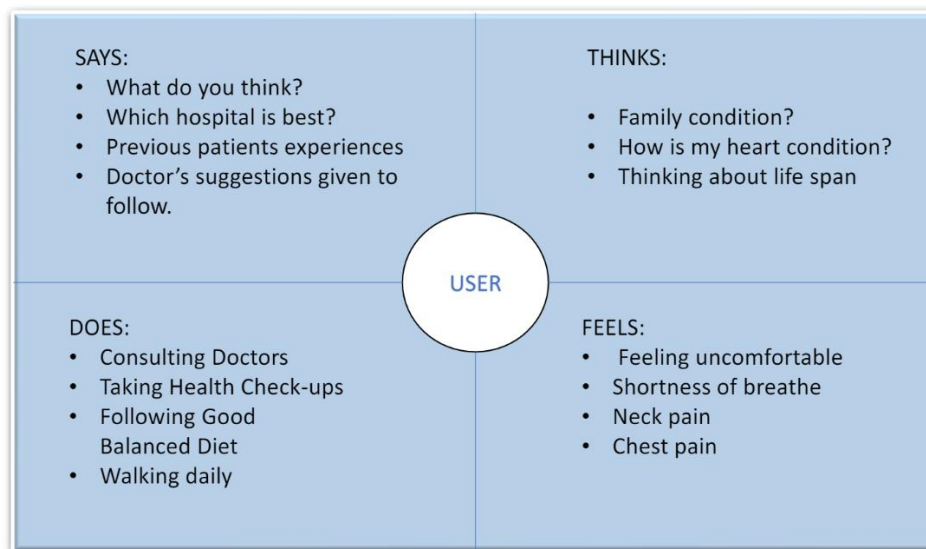
Prediction of heart disease at early stage using data mining and big data analytics: A survey N. K. Salma Banu, Suma Swamy

Several studies have been carried out for developing prediction model using individual technique

and also by combining two or more techniques. This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers. Into practice.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

Step-1: Team Gathering, Collaboration and Select the Problem Statement

Template



Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

1. 10 minutes to prepare
2. 15 minutes to collaborate
3. 10 minutes to communicate

4

Before you collaborate

A little bit of preparation goes a long way with any session. Here's what you need to do to get going.

→ 10 minutes

12

Team gathering

Invite who should participate in the session and send all invitees. Make sure you have the space to get started.

13

Set the goal

Focus about the problem you're trying to solve in the brainstorming session.

14

Invite lead to use the facilitation tool

Use the Facilitation tool to ensure a happy and productive session.

→ 10 minutes

5

Define your problem statement

What problem are you trying to solve? Frame your problem as a Pain-Might-Be statement. This will be the focus of your brainstorm.

→ 10 minutes

PROBLEM

The leading cause of death in the developed world is heart disease. Therefore, there needs to be something done to help prevent the risk of having a heart attack or stroke.

15

Key rules of brainstorming

To get an emotional production session

16

Keep it light

17

Encourage wild ideas

18

Build on others

19

Go for quantity

20

One idea at a time

21

No criticism

22

Remember the goal

23

Brainstorm session recap

After a 10-minute session, you should have a list of ideas. Review them and select the best ones.

→ 10 minutes

Step-2: Brainstorm, Idea Listing and Grouping

AJITH

Identify the parameters involved in predicting a heart disease

collection of datasets available from various sources

choosing all the feasible models and model evaluation

weight updation and parameter tuning

GOWTHAM

experimental exploration of various features and relations in the dataset

accuracy estimation

re-train models based on changes in data distribution should be known to serve the most updated model in production

Any changes in the downstream inputs of the ML system should be immediately notified to quickly check for any ML performance deterioration

NIRANJAN V

Feature generation code for both training and inference should be the same.

After a model is trained but before it actually serves the real requests, a system needs to inspect it and verify that its quality is sufficient

get more insights about what could go wrong and then continue improving our model with continuous integration

compare performance of the model with field expert

VENKAT RAM

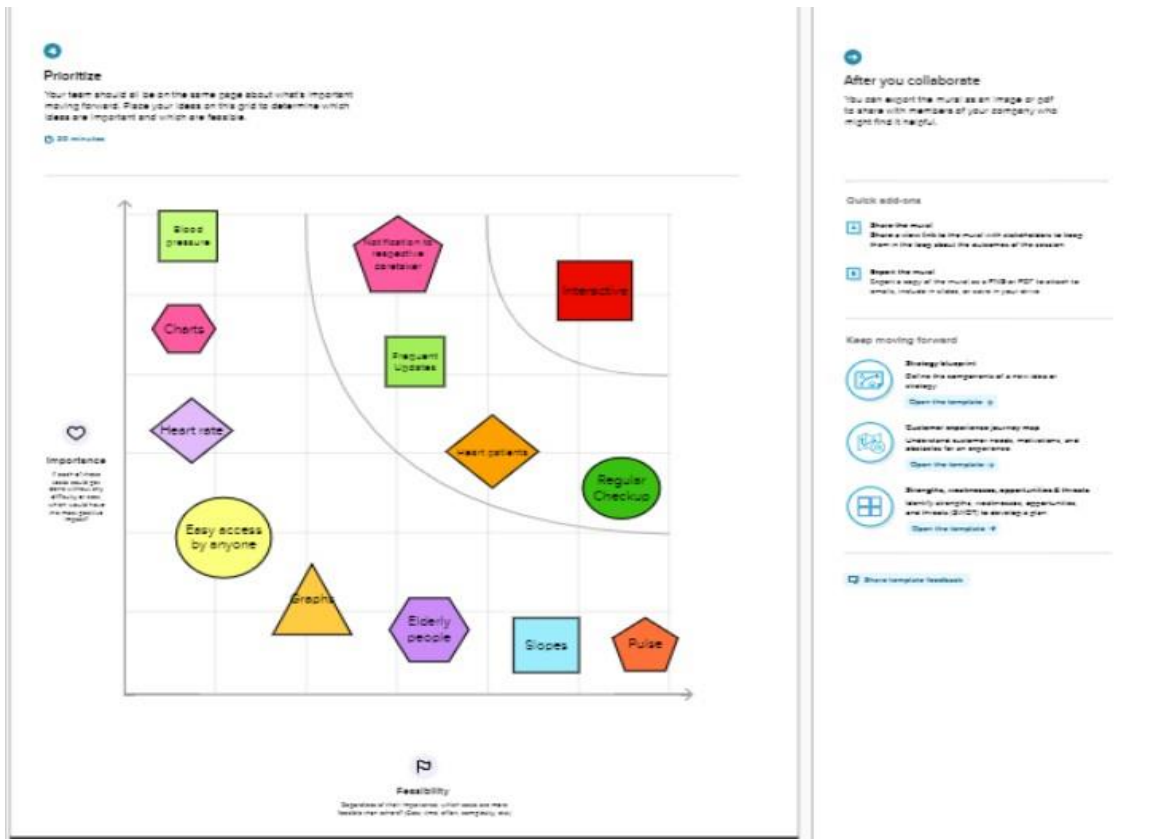
Diagnose high bias and/or high variance and act in consequence

Manually analyze miss classified records and look for patterns

Extract significant variables

Use of cross-validation technique to increase the accuracy of the model

Step-3: Idea Prioritization



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The leading cause of death in the developed world is heart disease. As a result, work must be done to reduce the risks of having a heart attack or stroke. It is infeasible for a common man to frequently undergo tests for ECG and so on. Hence, it requires a replacement that is both convenient and dependable.
2.	Idea / Solution description	The proposed solution proposes an interactive dashboard for visualizing 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.
3.	Novelty / Uniqueness	Makes recommendations to the user based on that person's medical analysis.
4.	Social Impact / Customer Satisfaction	It helps with disease prediction at an early stage and frequently alerts the user to their current health status. Both the user and the doctor can benefit from the system's improved decision-making regarding cardiac disease
5.	Business Model (Revenue Model)	Can be deployed by Hospitals or NFOs, so that it makes the analysis in a fast manner.
6.	Scalability of the Solution	The solution can work effectively on long and small datasets. It can also be changed to predict various other diseases depending on the dataset

3.4 Problem Solution fit:

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS <ul style="list-style-type: none"> Hospitals Clinics WHO Any medical related agencies those prepare medicines or any kind of solutions inferring over the data of diseases. 	6. CUSTOMER CONSTRAINTS CC <p>The unawareness over the AI/ML technologies, collaborative dashboards, network connection, lack of data.</p>	5. AVAILABLE SOLUTIONS AS <p>The customers can prefer over a manual data visualization and prediction, which is very tedious job and requires the knowledge over the technologies of AI/ML.</p> <p>Hard mathematical formulae were created and the results were being calculated manually.</p>	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS JALP <p>Quality of Data: The quality of data should be accurate and reliable. Obviously, the outcome will solely depend on the data we put into the prediction. If the data is skewed, then the prediction which is dependent on it, will be skewed as well.</p>	9. PROBLEM ROOT CAUSE PRC <ul style="list-style-type: none"> Difficulty of predicting a heart disease. Will not have a proper idea of relation between similar heart diseases. There is a chance of identifying every heart diseases as same. Reason of increase in heart disease will not be rootly identified. 	7. BEHAVIOUR BE <ul style="list-style-type: none"> Generation of legitimate and reliable datasets. Customers need to collect more number of datasets in order to obtain more accurate result. Must obtain knowledge of difference between datasets that is used for comparison. 	
Identify strong TR & EM	3. TRIGGERS TR <ul style="list-style-type: none"> Insufficient ways of handling huge amounts of datasets and inferring the root cause of the heart disease cannot be found out. Similarity of heart disease has not been identifiable. 	10. YOUR SOLUTION YS <p>With the notable technology of AI/ML we are able to visualize and predict heart diseases and related diseases, by the ultimate power Cognos Analytics Tool we will be able to properly create a dashboard for the customers to work with and visualize and analyze the heart disease on their work with limited knowledge.</p>	8. CHANNELS of BEHAVIOR CH <p>8.1 ONLINE Visualizing the datasets. Exploration of data.</p> <p>8.2 OFFLINE Cleansing of datasets. Collection and noting the datasets.</p>	Fit 3 TR, Success Assured
	4. EMOTIONS: BEFORE / AFTER EM <p>Before -> It creates a huge ambiguity in knowing the proper or accurate reasons for a heart disease.</p> <p>After -> There is a large chance understanding of the heart disease and root cause of it, which makes a better solution and finding a preventive way over it.</p>			

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	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	Generation Report	User can view his/her health report and can make decisions accordingly

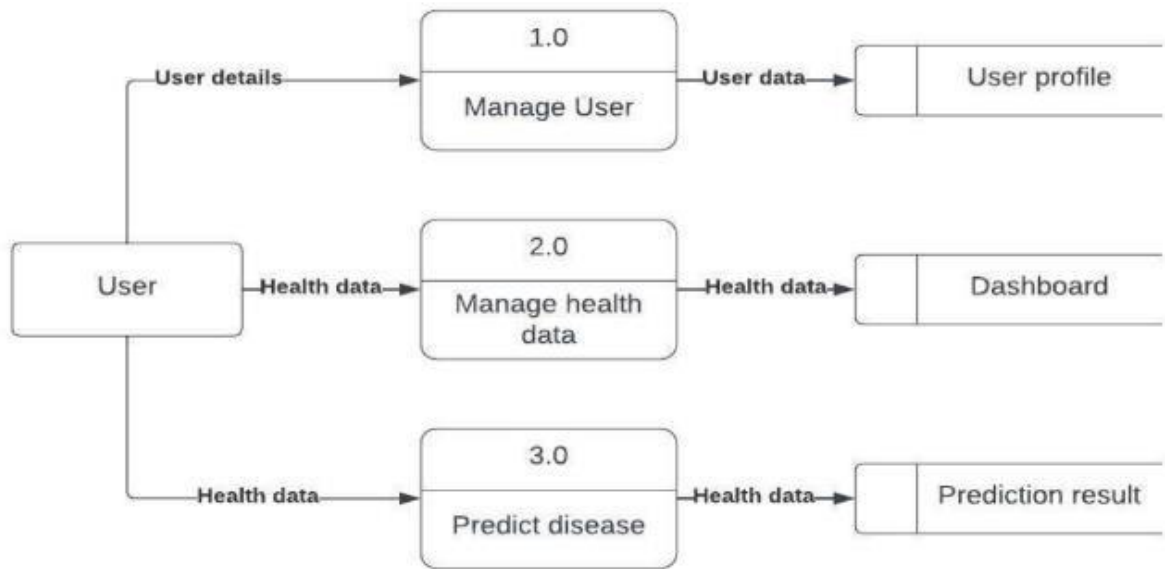
4.2 Non-Functional requirements

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

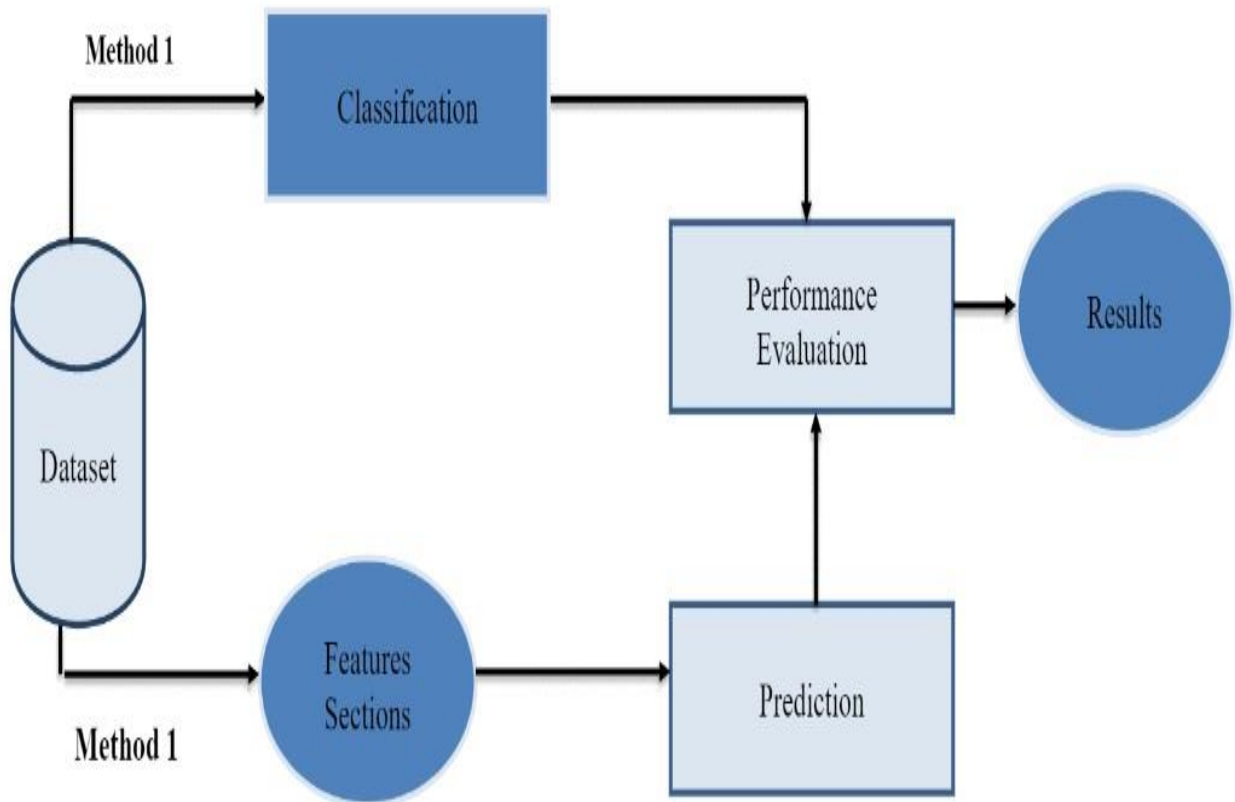
FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and userfriendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	Security	For security of the application the technique known as database replication should be used so that all the important data should be kept safe. In case 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 any environment
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 Higher versions

PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

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	3	High	1
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	3	High	3
Sprint-1		USN-3	As a user, I can register for the application through Gmail	3	Medium	1
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email & password	6	High	5
Sprint-2	Dashboard	USN-5	Attractive dashboard For the Application	3	Medium	3
Sprint-2		USN-6	Profile - view & update your profile	5	Low	2
Sprint-2		USN-7	Home - Analyze your Heart problem	2	High	4
Sprint-2		USN-8	The user will have to fill in the below 13 fields for the system to predict a disease -Age in year -Gender -Chest pain Type -Fasting Blood Sugar -Resting Electrographic Results -Exercise Induced Angina -Trust Blood Pressure	7	High	2

Sprint-3	Support	USN-9	Get feedback from users	10	Medium	3
Sprint-3		USN-10	Responds to user queries via telephone,email etc.	3	Medium	2
Sprint-3		USN-11	The team must respond immediately to the queries based on the priority	5	High	5
Sprint-4	System Requirements	USN-12	Hardware Requirement 3. Laptop or PC • i5 processor system or higher • 4 GB RAM or higher • 128 GB ROM or higher 4. Mobile • (12.0 and above)	5	Low	2
Sprint-4		USN-13	Software Requirement 2. Laptop or PC	8	Medium	4

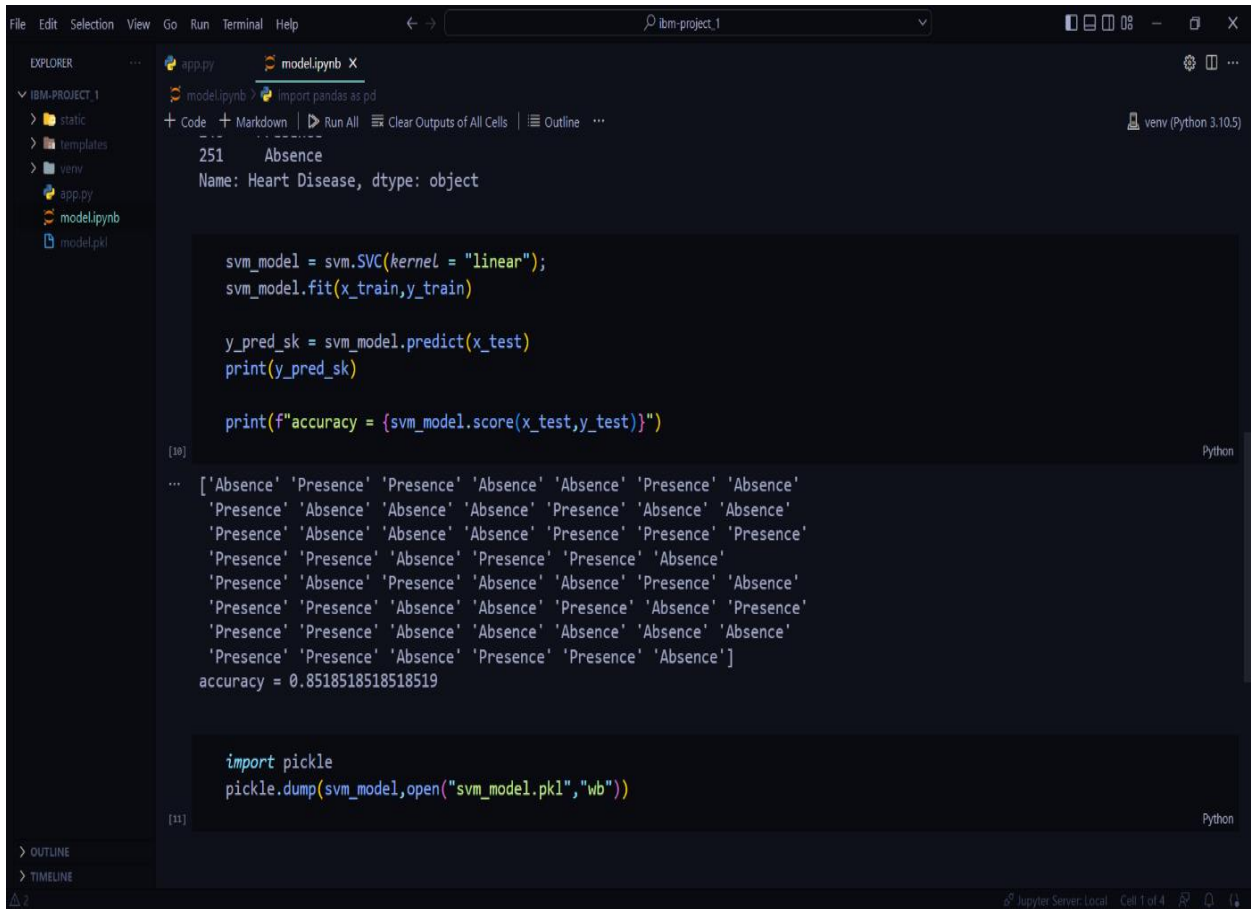
6.2 Sprint Delivery Schedule

Sprint	Total Points	Story	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	30 Oct 2022	04 Nov 2022	17	04 Nov 2022
Sprint-3	20		6 Days	05 Nov 2022	11 Nov 2022	18	11 Nov 2022
Sprint-4	20		6 Days	12 Nov 2022	17 Nov 2022	19	17 Nov 2022

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

7.1 Feature : Prediction of heart diseases with the help of interactive dashboards we have used SVM machine learning model for the prediction

Code :



The screenshot shows a Jupyter Notebook environment with a dark theme. The Explorer panel on the left shows a project named 'IBM-PROJECT_1' with files 'app.py', 'model.ipynb', and 'model.pkl'. The main area displays the code for training and testing an SVM model. The code includes importing pandas, defining training and testing data, fitting the SVM model with a linear kernel, predicting on test data, and saving the model using pickle. The output of the prediction shows a list of 'Absence' and 'Presence' labels, and the accuracy is 0.8518518518518519.

```
File Edit Selection View Go Run Terminal Help
IBM-PROJECT_1
static
templates
venv
app.py
model.ipynb
model.pkl

model.ipynb X
model.ipynb > import pandas as pd
+ Code + Markdown | Run All | Clear Outputs of All Cells | Outline ...
251 Absence
Name: Heart Disease, dtype: object

svm_model = svm.SVC(kernel = "linear");
svm_model.fit(x_train,y_train)

y_pred_sk = svm_model.predict(x_test)
print(y_pred_sk)

print(f"accuracy = {svm_model.score(x_test,y_test)}")

[10] Python
... ['Absence' 'Presence' 'Presence' 'Absence' 'Absence' 'Presence' 'Absence'
      'Presence' 'Absence' 'Absence' 'Absence' 'Presence' 'Absence' 'Absence'
      'Presence' 'Absence' 'Absence' 'Absence' 'Presence' 'Presence' 'Presence'
      'Presence' 'Presence' 'Absence' 'Presence' 'Presence' 'Absence'
      'Presence' 'Absence' 'Presence' 'Absence' 'Absence' 'Presence' 'Absence'
      'Presence' 'Absence' 'Absence' 'Absence' 'Presence' 'Absence' 'Presence'
      'Presence' 'Presence' 'Absence' 'Absence' 'Absence' 'Absence' 'Absence'
      'Presence' 'Presence' 'Absence' 'Presence' 'Presence' 'Absence']
accuracy = 0.8518518518518519

import pickle
pickle.dump(svm_model,open("svm_model.pkl","wb"))

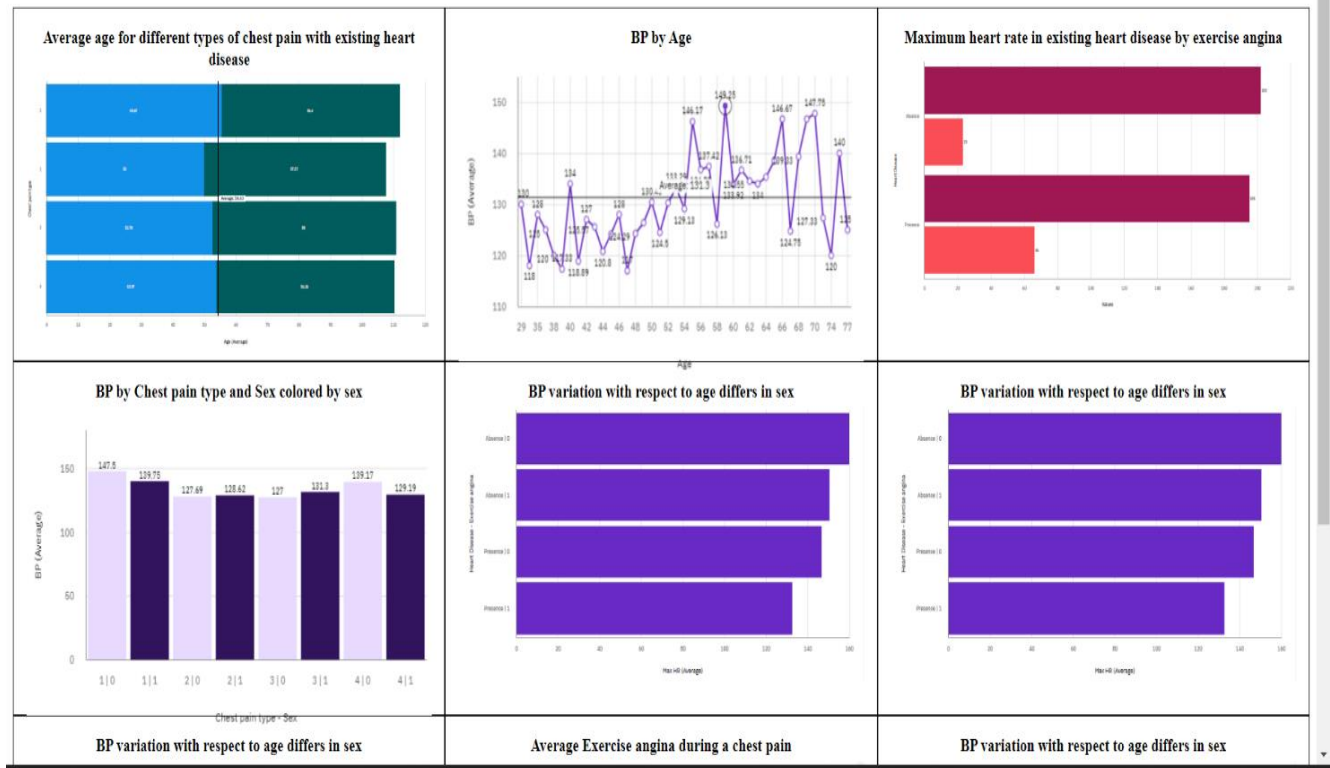
[11] Python

> OUTLINE
> TIMELINE
Jupyter Server: Local Cell 1 of 4
```

7.2 Dashboards:

[Home](#)

Visualization of trained dataset



8. Testing

8.1 User acceptance Testing

Testing a case where user has heart disease

Home [logout](#)

Enter Medical results below

51
0
3
120
295
12
111
157
0
0.6
1
0
3

Submit

Home [logout](#)

The Predicted result for the entered medical details : Presence

Testing a case where user does not have heart disease:

[Home](#) [logout](#)

The Predicted result for the entered medical details : Absence

[Home](#) [logout](#)

Enter Medical results below

12

1

0

11

11

12

11

157

1

12

1

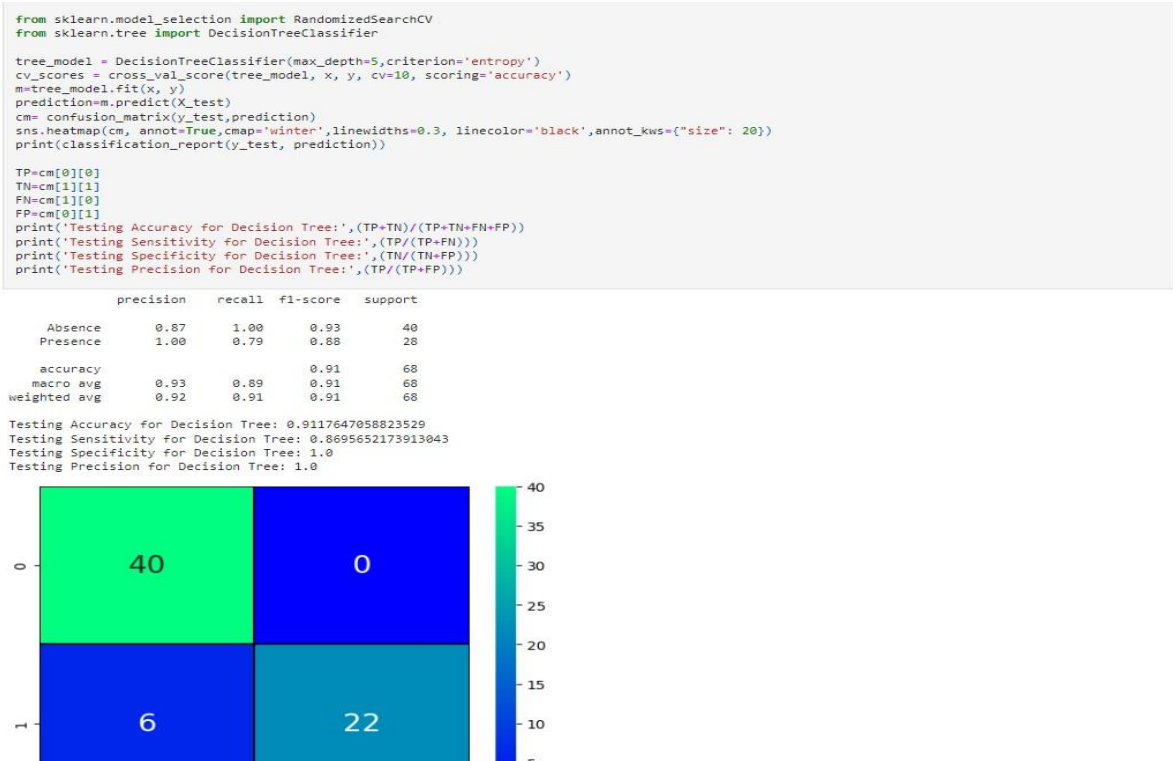
12

13

Submit

9. Result

9.1 Performance Metrics :The confusion matrix below shows the performance metrics of the machine learning model.



10. Advantages Disadvantages

Advantages:

- This is one of the fastest ways to determine if a person is likely to suffer from a heart disease or not.
- Useful for medical practitioners to easily classify their patients.
- User Friendly
- Easy to understand
- Secure
- Dashboard provides insightful informations

Disadvantages:

- Needs work
- Users need to know all the fields
- Does Not take null value as input

11. Conclusion

Complications of heart disease include heart attack and stroke. You can reduce the risk of complications with early diagnosis and treatment. So the suggestion that we get from the website might help save patients. It is always to get treated in the early stages of heart disease.

12. Future Scope

Like the saying goes “Prevention is better than cure”. We have to look into methods to prevent heart diseases altogether other than just predicting it in early stages. To use this website we need to take a lot of tests beforehand. So it would be better if we require less attributes and still give an effective result

13. Appendix

Demo Video link: [Visualizing and Predicting Heart Diseases with an Interactive Dashboard PNT2022TMID12648](#)

Git-hub link: <https://github.com/IBM-EPBL/IBM-Project-13864-1659534203/tree/main/Final%20Deliverables>

