

VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASH BOARD

NALAIYA THIRAN PROJECT REPORT-2022

Team Details	
Team ID	PNT2022TMID34621
Team Leader	Bibishek G S Steephen
Team Member 1	Bibin S
Team Member 2	Ajay A J
Team Member 3	Gino M
Team Member 4	Ajay E
Faculty Mentor	Ajin M

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.

The application is fed with various details and the heart disease associated with those details. The application allows user to share their heart related issues. It then processes user specific details to check for various illness that could be associated with it. Here we use some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's details. Based on result, the person can contact doctor accordingly for further treatment. The system allows user to view doctor's details too. The system can be used for free heart disease consulting online.

1.2 Purpose

The main goal of this project is to provide a tool for doctors to detect heart disease as early stage. This in turn will help to provide effective treatment to patients and avoid severe consequences. 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 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.

2.2 References

“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, 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.

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.

2.3 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.

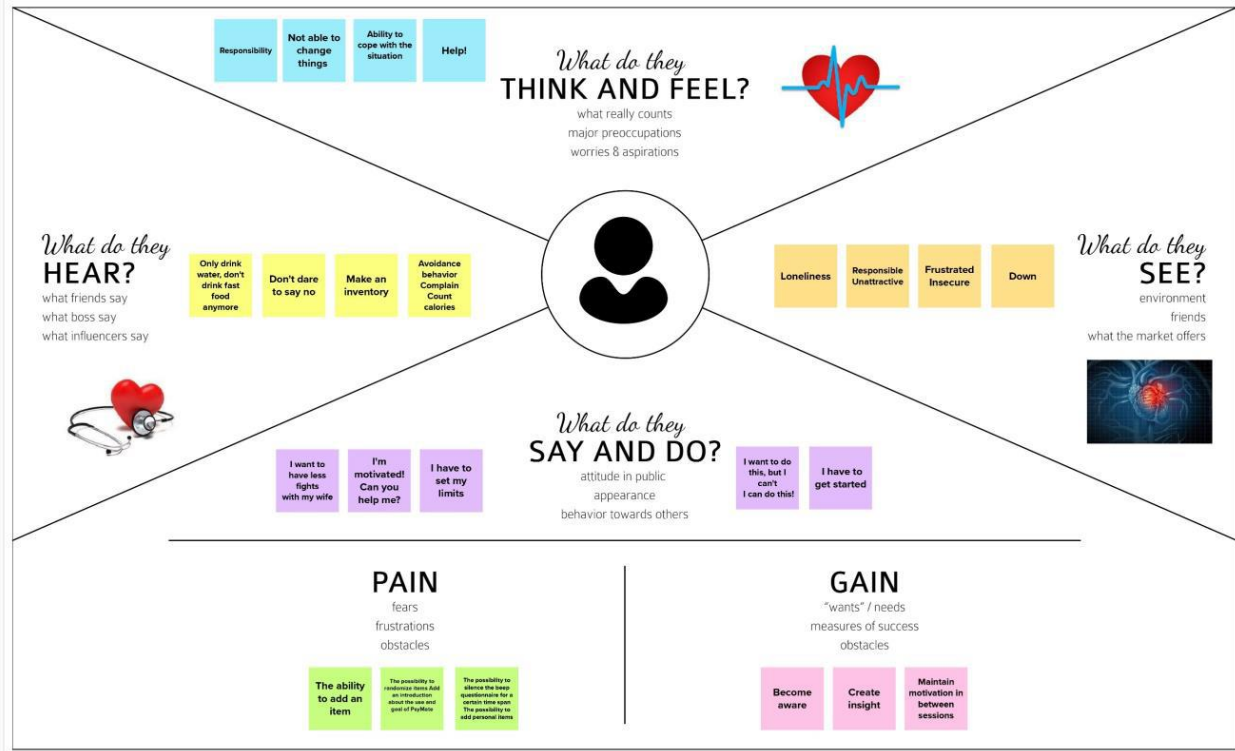
Why is it important to fix the problem?

We should predict the problem before giving treatment to the patients. As the problem is predicted early, we can solve it easily and early.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

Empathy map



3.2 Ideation & Brainstorming

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



Step-2: Brainstorm, Idea Listing and Grouping

3 Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

DEFINITION

- Implementing the code
- Visualize the dataset
- Loading the dataset
- Understand the dataset

QUESTION

- By making use of IBM cloud
- Achieving a high accuracy
- Keeping up with current trends
- Collecting the project datasets

ANSWER

- Using open or closed source software
- Utilizing the customer's data
- By using new coding techniques
- Connecting the complex flows

MORE INFORMATION

- Understanding the flow
- Creating a data visualization interface
- Utilizing machine learning
- Using business analytics

3 Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you can break it up into smaller sub-groups.

20 minutes

Data Visibility

- Charts
- Bar
- Maps
- Table
- Graphs

Expectations from the dashboard

- Report status
- Report trends

Customer

- Meet a need
- Report associated with need
- Easy to use

Accessibility

- Easy access for anyone
- Easy understandable

Requirements

- Real-time data
- Push
- Real-time
- Real-time
- Real-time
- Real-time

Pros

- Real-time
- Real-time
- Real-time
- Real-time
- Real-time
- Real-time

Tools

- IBM Cognos
- IBM Cloud

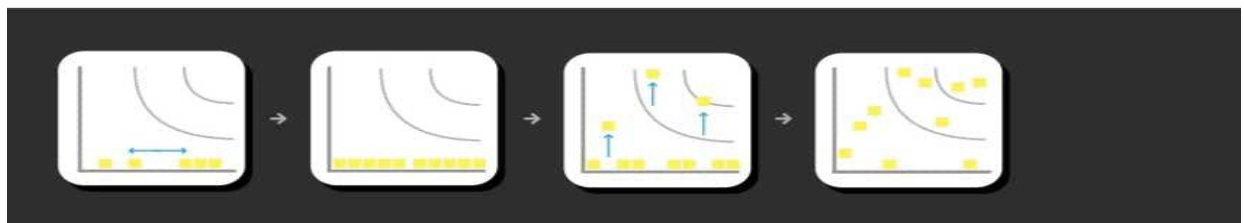
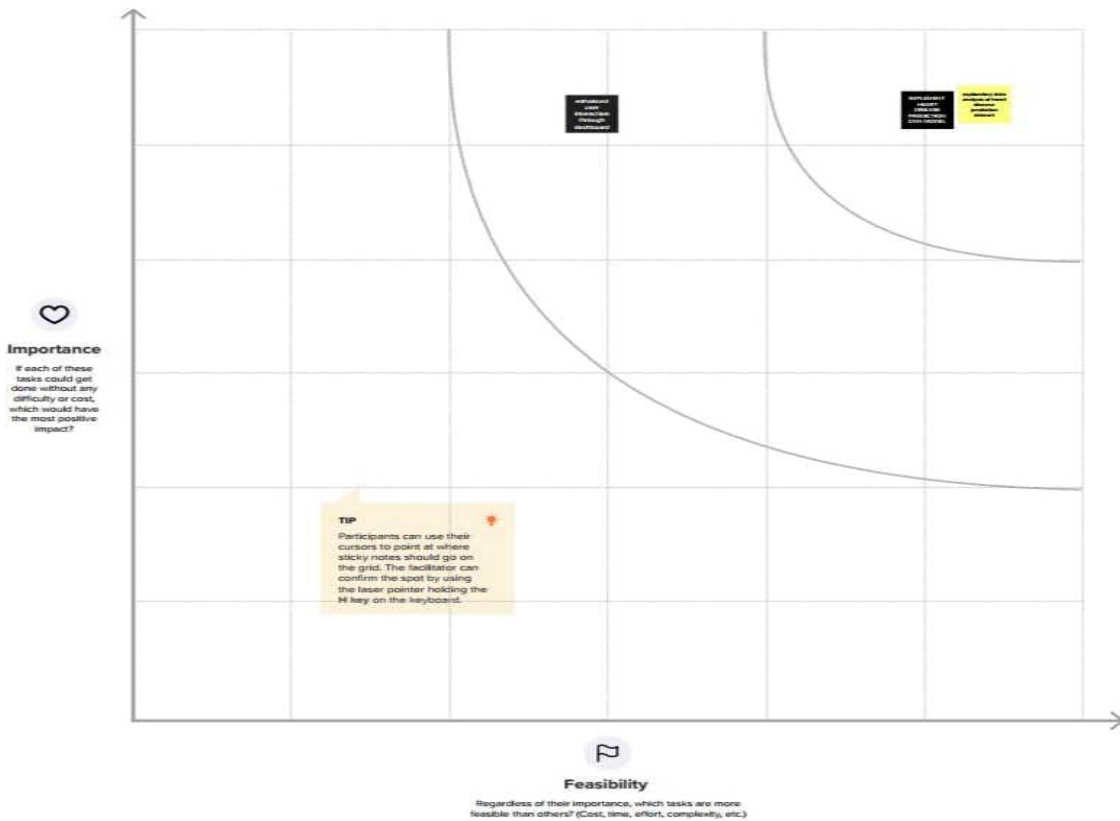
Step-3: Idea Prioritization

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



3.3 Proposed Solution

- The prime novelty of the solution is the fusion of highly efficient algorithms, that eliminates the disadvantage of every algorithm when employed individually and also provides a higher level of accuracy in the prediction. Another innovation is employed in the dashboard by providing diet and fitness related suggestions to the user based on his/her medical reports and history. In addition to it, the patient is given a list of hospitals closer to the patient's locality and severity of the disease.
- It helps with disease prediction at an early stage and alerts the user about his/her current health status. Heart disease can be cured by a mix of medication, lifestyle modifications, and occasionally, surgery. The system helps the user as well as the doctor to make better decisions. Complex questions related to heart diseases can be answered by extracting hidden knowledge, i.e., patterns and relationships from the heart disease database.
- This interactive dashboard for heart disease prediction can be installed in hospitals and healthcare facilities. Predicted outcomes can be utilized to avoid expensive surgeries.
- It can be used in educational institutions, industries and all types of workplaces to monitor the employees' health conditions and thereby helping them lead a healthier life.
- The proposed solution works efficiently in both smaller and larger datasets.
- This predictive model can be used to detect diseases in other internal organs too.

3.4 Problem Solution Fit

<p>1. CUSTOMER SEGMENT(S) CS</p> <ul style="list-style-type: none"> • Senior citizens • Hospitals • Pharmaceutical agencies • Smokers • Alcoholics • Diabetes patients • Hypercholesterolemia patients • Hypertension patients • Thrombosis patients • Obese persons • Peripheral artery disease patients • Angina patients 	<p>6. CUSTOMER CONSTRAINTS CC</p> <ul style="list-style-type: none"> • Instant network connectivity • Presence of good-condition communication devices like smartphones and laptops • Financial constraints to consult specialists • Lack of awareness about heart disease • Complex and expensive scanning methodologies • Psychological problems • Lack of hope in treatment 	<p>5. AVAILABLE SOLUTIONS AS</p> <ul style="list-style-type: none"> • Manual data visualization and prediction are very tedious • Consult doctors (heart specialists), but it requires financial stability • Quit smoking • Restrain from alcohol • Practice a healthy lifestyle with daily exercises and a nutritious diet plan • Take cholesterol tests periodically
<p>2. JOBS-TO-BE-DONE / PROBLEMS J&P</p> <ul style="list-style-type: none"> • The data used for prediction should be accurate and reliable. • If data is skewed, then the prediction is also skewed • Predictions should be done based on various metrics such as blood pressure, cholesterol levels, heartbeat rates, etc. that require complex integration • Risk of lives depends on further medical support • Timely alerts help in the prevention of the sudden onset of cardiac arrests <p>Focus on J&P, tap into BE, understand</p>	<p>9. PROBLEM ROOT CAUSE RC</p> <ul style="list-style-type: none"> • Difficulty in predicting heart disease at earlier stages • Lack of awareness about physical fitness • Genetic problems • Lifestyle and eating habits • A buildup of fatty plaques in the arteries is the most common cause of coronary artery disease. • Obesity • Alcohol and Smoking habits • Stress, anxiety, depression and psychological problems 	<p>7. BEHAVIOUR BE</p> <ul style="list-style-type: none"> • Look up on the internet to find answers • Visit healthcare specialists • Take advice from friends and family • Physical activity helps to lower the risk of heart disease. • Adopting a healthy diet can help in improving blood pressure and cholesterol and also reduces the risk of diabetes. • Reduction of intake of alcohol and cigarettes • Get quality sleep • Prioritizing mental peace • Develop unwanted mental trauma and anxiety about the aftermath of disease onset • Falling into wrong assumptions and choosing the instant solutions that have worse side effects <p>Focus on J&P, tap into BE, understand</p>

4. REQUIREMENT ANALYSIS

4.1 Functional Requirement

User Registration:

Enables user to make registration for the application through Gmail.

User Confirmation:

Once after registration, the user will get confirmation via., Email.

Visualizing Data:

User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics.

Generation Report:

User can view his/her health report and can make decisions accordingly.

4.2 Non-Functional Requirements**Usability:**

Users can access the application with a minimal spec Devices and no additional dependencies are required to access the application.

Security:

The cloud vendor provides security to the deployed application.

Reliability:

The User's credentials are protected by the secured database provided by the cloud vendor. So, the customers can trust the application.

Performance:

Since the application is deployed in the Docker container, Docker provides a smooth performance to the clients/customers.

Availability:

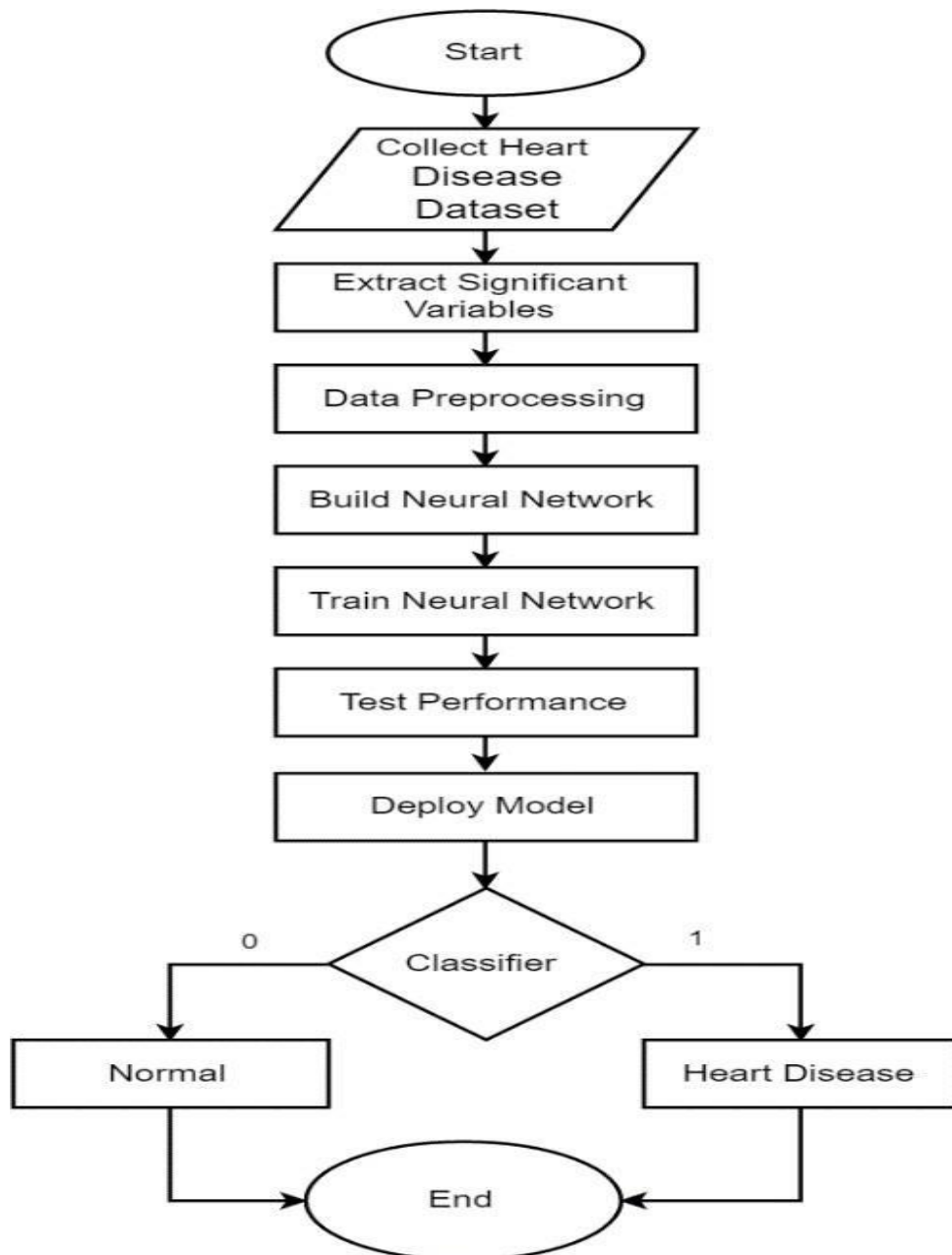
Since the application is deployed in cloud. The cloud vendor provides the availability of the application.

Scalability:

If the number of customers are increased, the application can be scaled through the cloud vendor.

5. PROJECT DESIGN

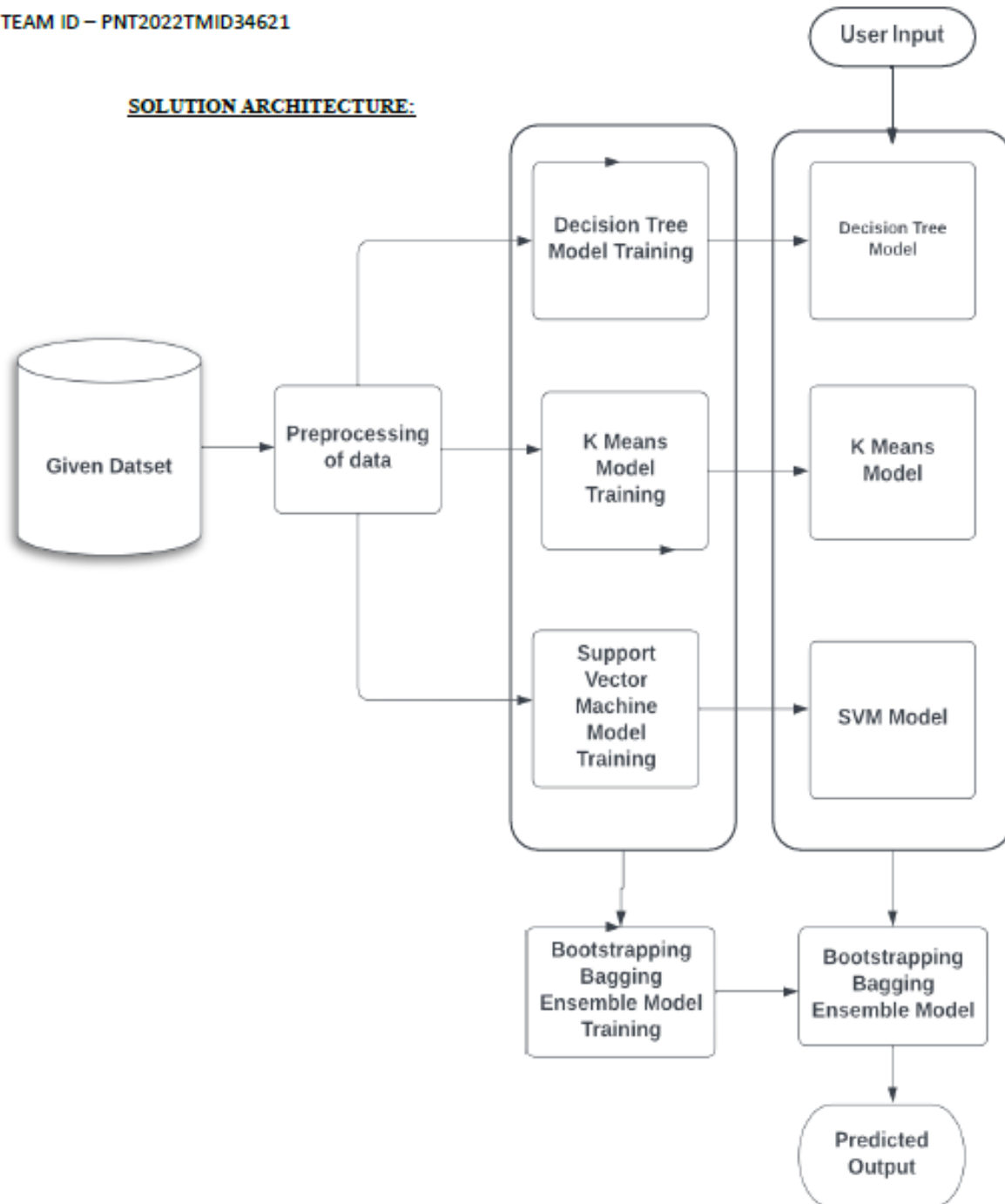
5.1 Data Flow Diagrams



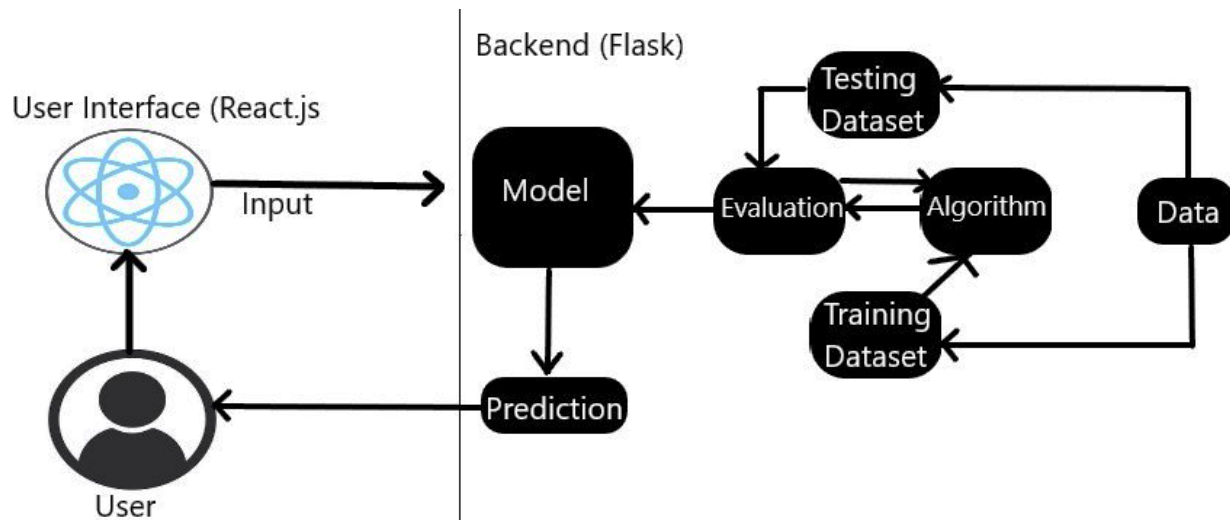
5.2 Solution & Technical Architecture

TEAM ID – PNT2022TMID34621

SOLUTION ARCHITECTURE:



Technical Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

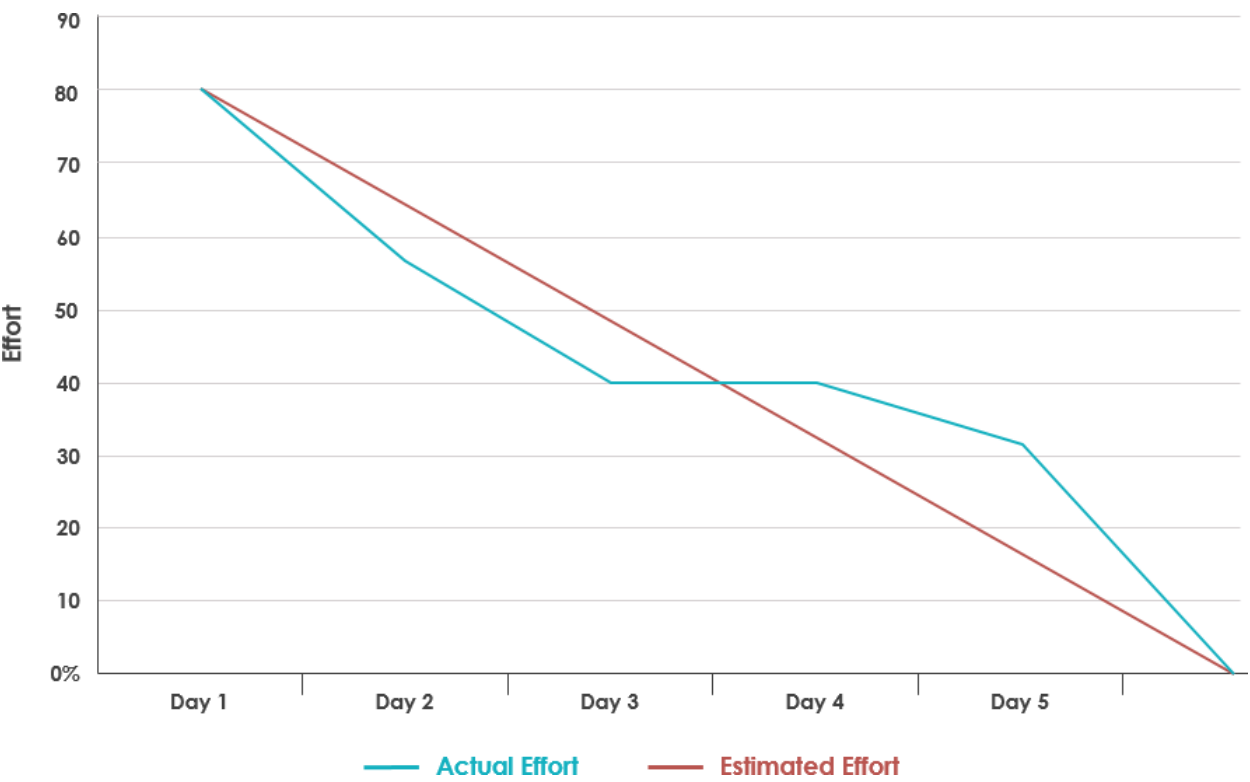
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Story points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the Dashboard by entering Email, password, and confirm password.	I can access my account / dashboard.	10	High	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen
		USN-2	As a user, I can register for the application through Gmail.	I can register & access the dashboard with gmail login.	10	High	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen
Sprint-2	Login	USN-3	After Registration Login page will appear, the user can login using the login credentials.	I can register & access the dashboard with gmail login.	20	High	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen
Sprint-3	Cognos Dashboard	USN-4	The user is allowed to view or	I can view the profile.	10	Medium	Bibin, Gino, Ajay.A.J,

			update his/her profile.				Ajay.E, Bibishek.G.S.Steephen
Sprint-4	Classified result	USN-6	Home - Analyze your Heart.	I can predict the heart condition.	5	High	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen
		USN-7	The user will have to fill in the 13 required fields for the system to predict a heart disease.	As a user, I can enter the datas in the specified fields.	10	High	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen
		USN-8	The report is generated based on the condition.	The user is able to view	5	Medium	Bibin, Gino, Ajay.A.J, Ajay.E, Bibishek.G.S.Steephen

6.2 Sprint Delivery Schedule

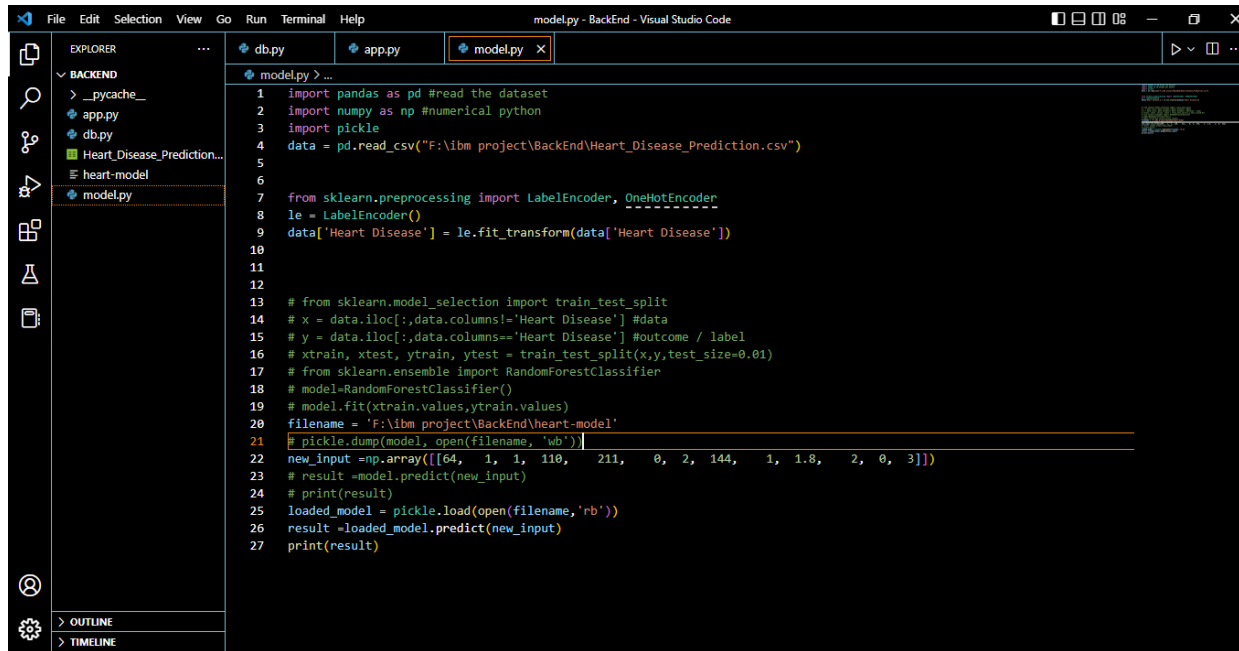
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	28 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	02 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	09 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	16 Nov 2022

Burndown Chart



7. CODING & SOLUTIONING

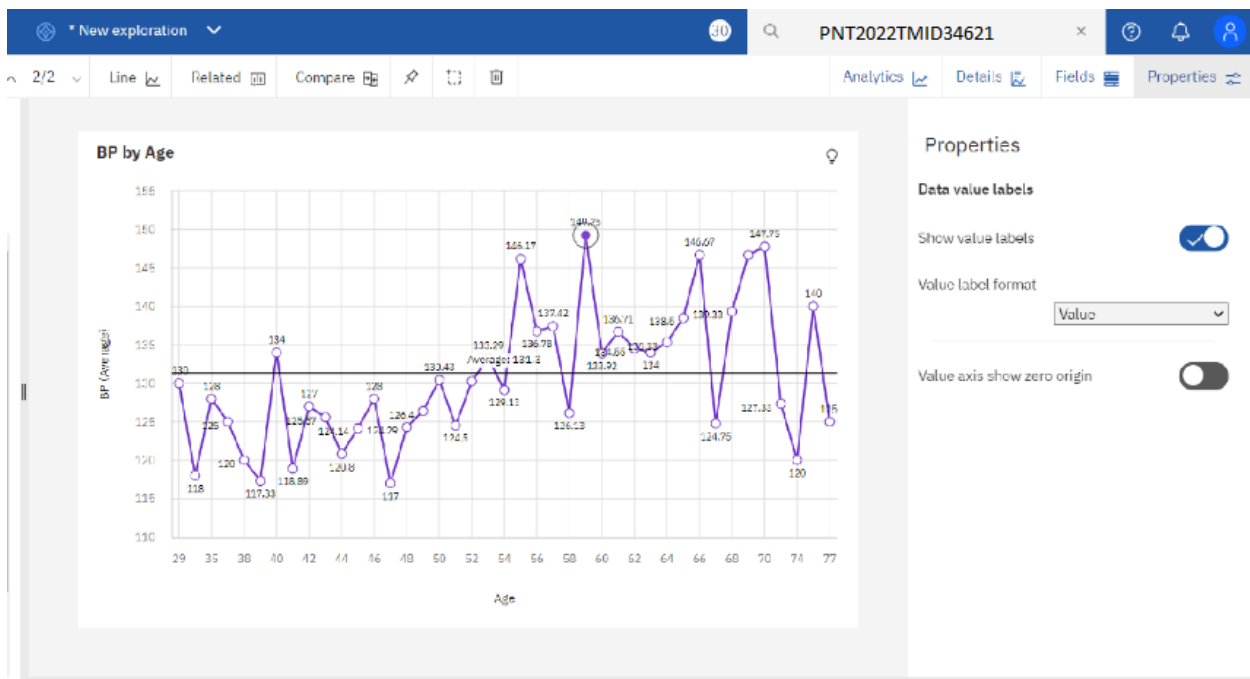
7.1 Machine Learning



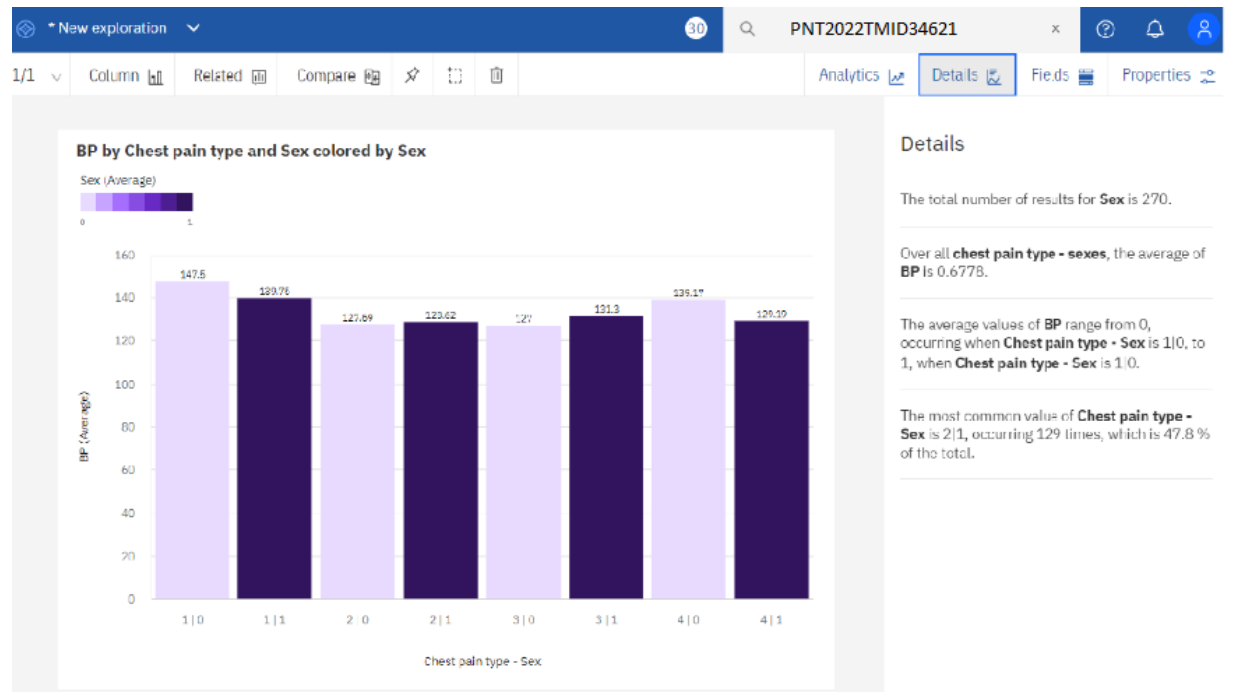
```
1 import pandas as pd #read the dataset
2 import numpy as np #numerical python
3 import pickle
4 data = pd.read_csv("F:\ibm project\BackEnd\Heart_Disease_Prediction.csv")
5
6
7 from sklearn.preprocessing import LabelEncoder, OneHotEncoder
8 le = LabelEncoder()
9 data['Heart Disease'] = le.fit_transform(data['Heart Disease'])
10
11
12
13 # from sklearn.model_selection import train_test_split
14 # x = data.iloc[:,data.columns!='Heart Disease'] #data
15 # y = data.iloc[:,data.columns=='Heart Disease'] #outcome / label
16 # xtrain, xtest, ytrain, ytest = train_test_split(x,y,test_size=0.01)
17 # from sklearn.ensemble import RandomForestClassifier
18 # model=RandomForestClassifier()
19 # model.fit(xtrain.values,ytrain.values)
20 filename = 'F:\ibm project\BackEnd\heart-model'
21 # pickle.dump(model, open(filename, 'wb'))
22 new_input = np.array([[64, 1, 1, 110, 211, 0, 2, 144, 1, 1.8, 2, 0, 3]])
23 # result =model.predict(new_input)
24 # print(result)
25 loaded_model = pickle.load(open(filename, 'rb'))
26 result =loaded_model.predict(new_input)
27 print(result)
```

7.2 Dashboard

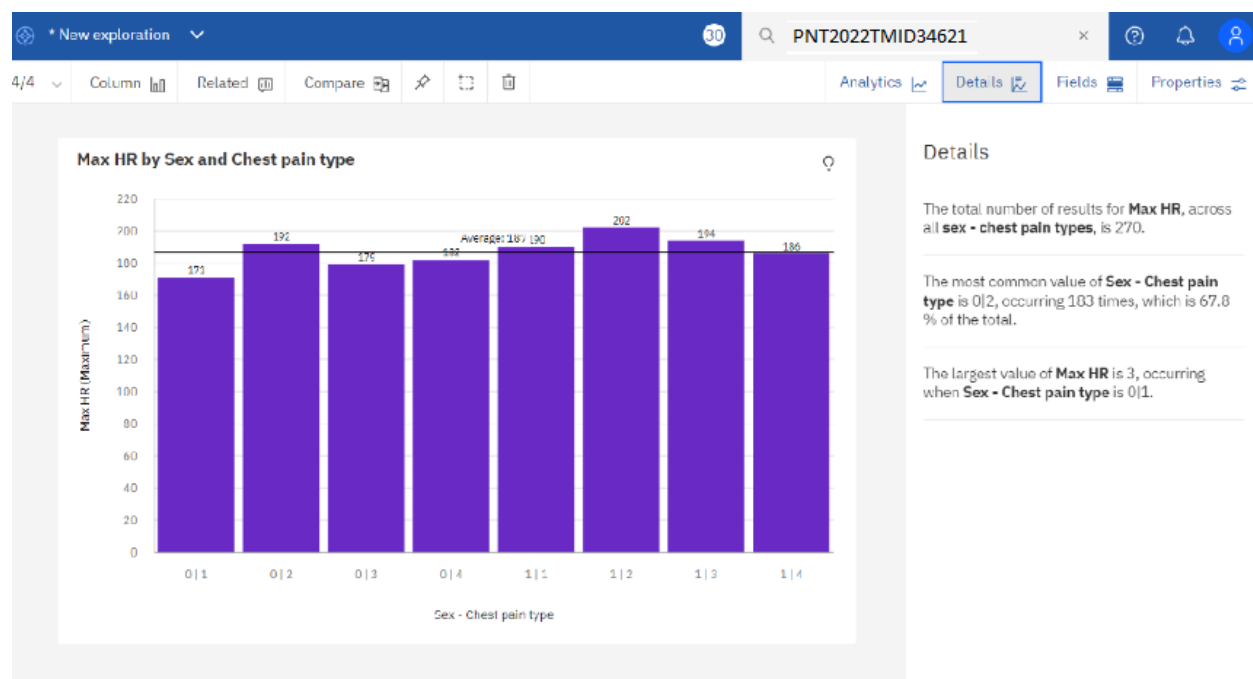
Average BP during chest pain



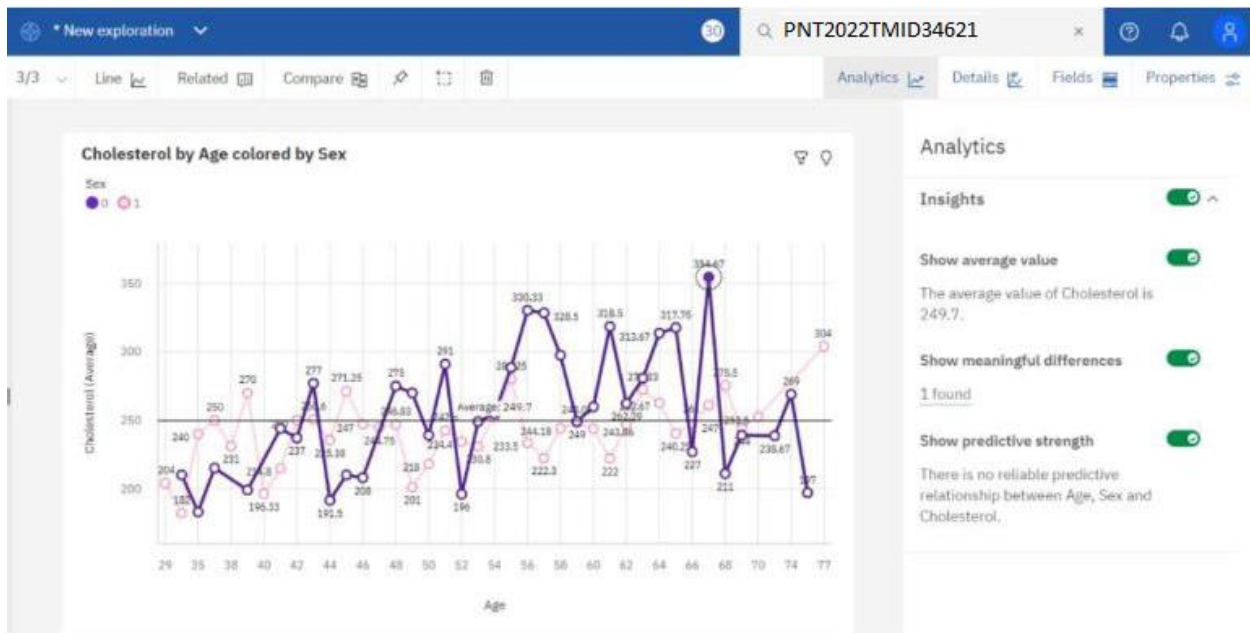
Exploration of BP vs Chest Pain Type and Gender:



Exploration of Max Heart Rate During the Chest Pain:



Exploration Of Cholesterol by age and Gender:

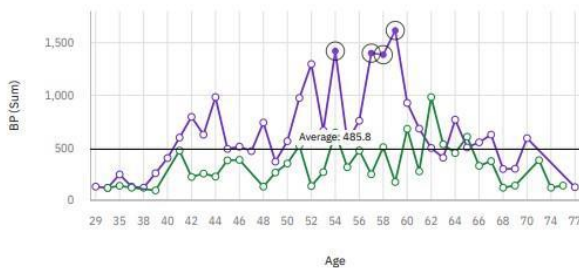


Dashboard Showing Different Types of Visuals:

Tab 8

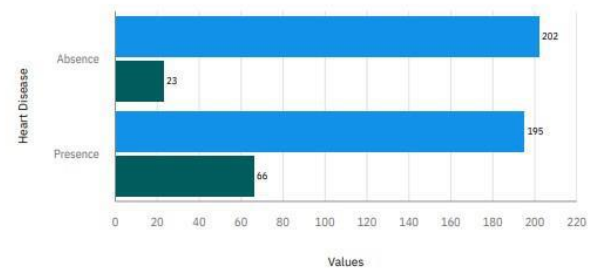
BP by Age colored by Sex

Sex: 0 1



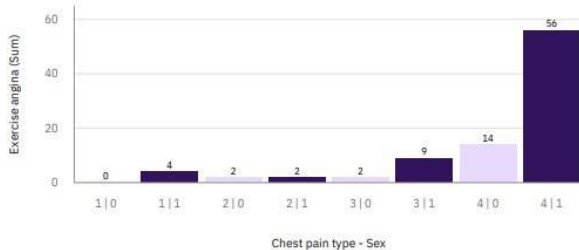
Max HR and Exercise angina by Heart Disease

Measures: Max HR, Exercise angina



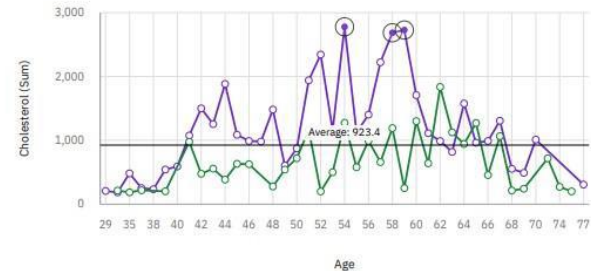
Exercise angina by Chest pain type and Sex colored by Sex

Sex (Average): 0 1

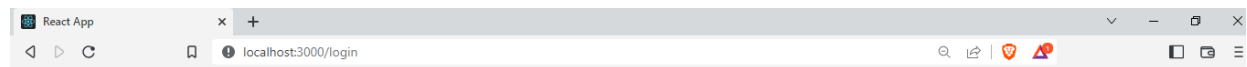
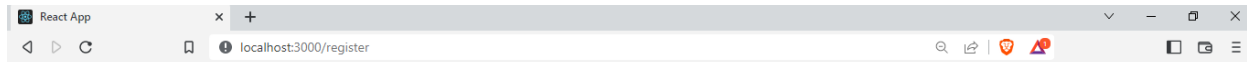


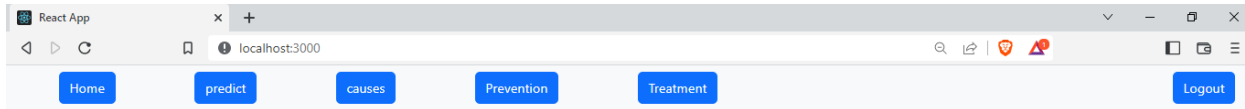
Cholesterol by Age colored by Sex

Sex: 0 1



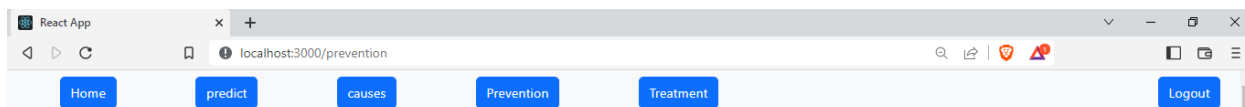
8. TESTING





Heart Disease Prediction Tool

Heart Disease Prediction Tool provides you the option of diagnosing whether you have heart disease or not without visiting a hospital.



Prevention tips

The same lifestyle changes used to manage heart disease may also help prevent it. Try these heart-healthy tips:

Eat a healthy, balanced diet

A low-fat, high-fibre diet is recommended, which should include plenty of fresh fruit and vegetables (5 portions a day) and whole grains. You should limit the amount of salt you eat to no more than 6g (0.2oz) a day as too much salt will increase your blood pressure. 6g of salt is about 1 teaspoonful.

Be more physically active

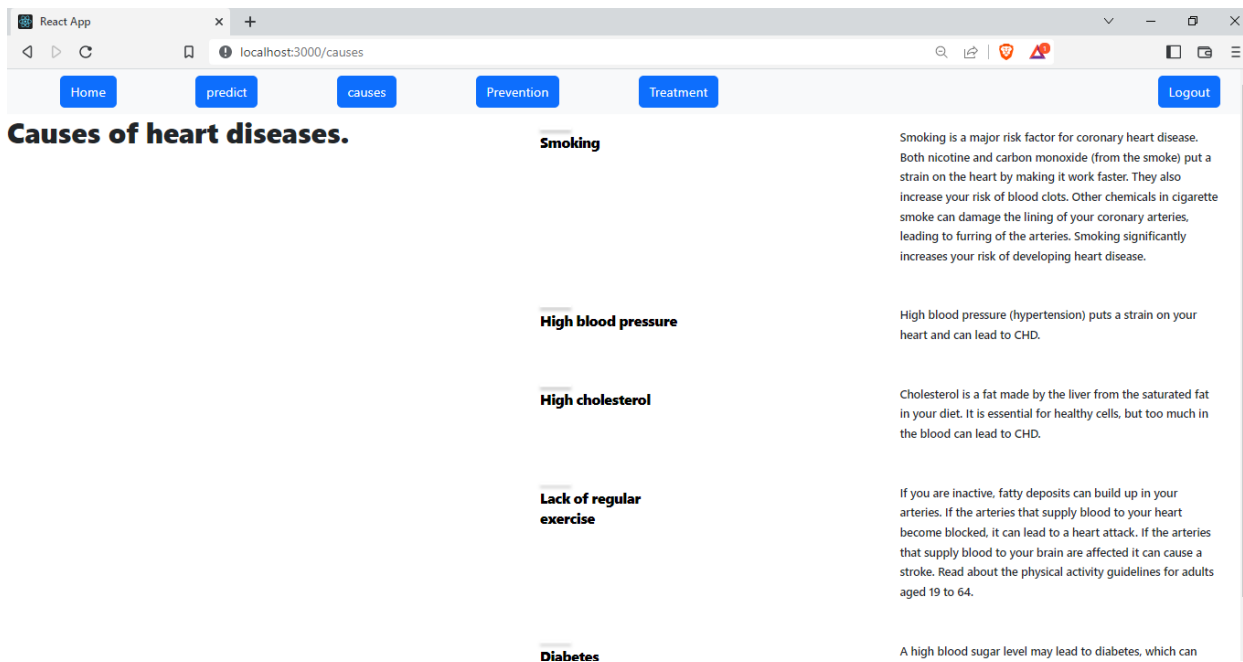
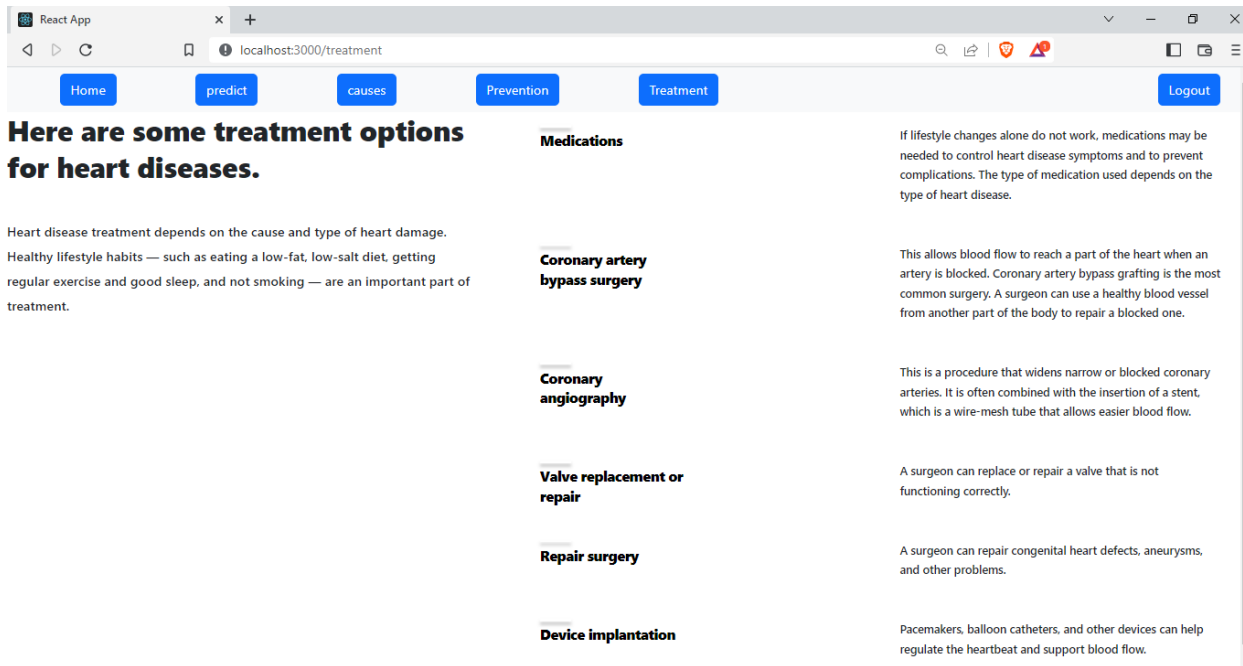
Combining a healthy diet with regular exercise is the best way of maintaining a healthy weight. Having a healthy weight reduces your chances of developing high blood pressure. Regular exercise will make your heart and blood circulatory system more efficient, lower your cholesterol level, and also keep your blood pressure at a healthy level. Exercising regularly reduces your risk of having a heart attack. The heart is a muscle and, like any other muscle, benefits from exercise.

Keep to a healthy weight

A GP or practice nurse can tell you what your ideal weight is in relation to your height and build. Alternatively, find out what your body mass index (BMI) is by using our BMI calculator.

Keep your blood pressure under

You can keep your blood pressure under control by eating a healthy diet low in saturated fat, exercising regularly and, if needed, taking medicine to lower your blood pressure. Your



React App

localhost:3000/predict

Home predict causes Prevention Treatment Logout

Age

Sex ☐ Female ☐ Male

Chest pain type

BP

Cholesterol

FBS over 120

EGX results

Max HR

Exercise angina

ST Depression

Slope of ST

Number of vessels fluro

Thallium

CHECK

8.1 User Acceptance Testing

Testing a case where user has heart disease

React App

localhost:3000/predict

Home predict causes Prevention Treatment Logout

Age

Sex ☐ Female ☒ Male

Chest pain type

BP

Cholesterol

FBS over 120

EGX results

Max HR

Exercise angina

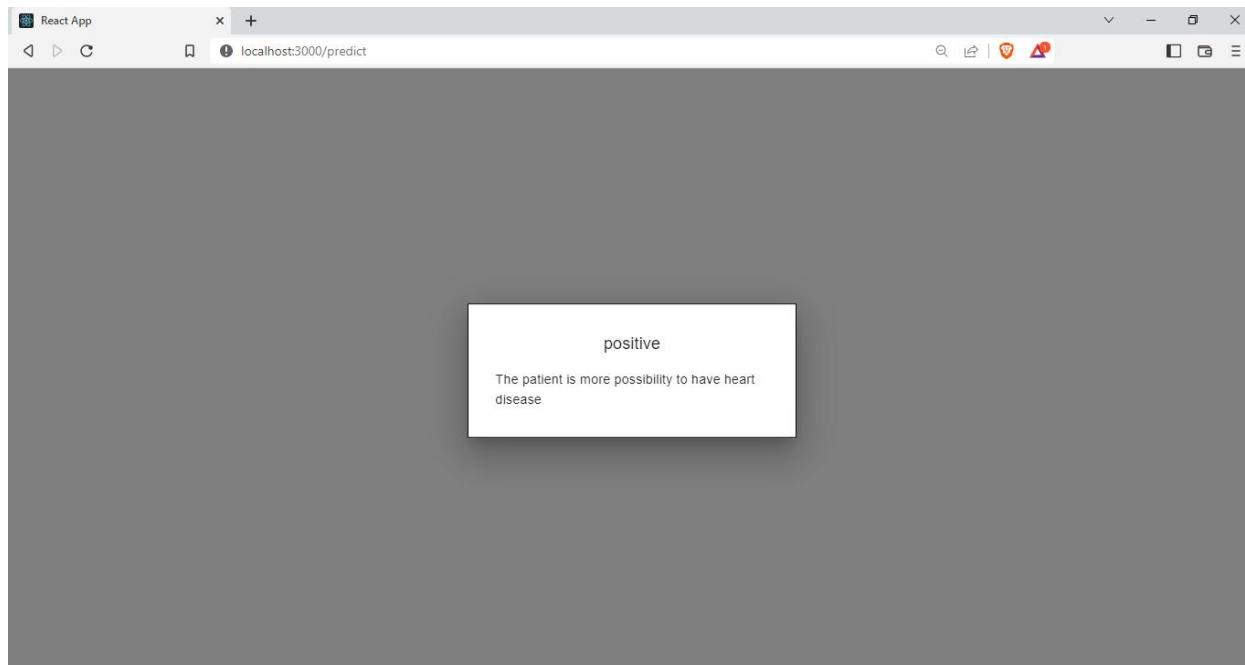
ST Depression

Slope of ST

Number of vessels fluro

Thallium

CHECK

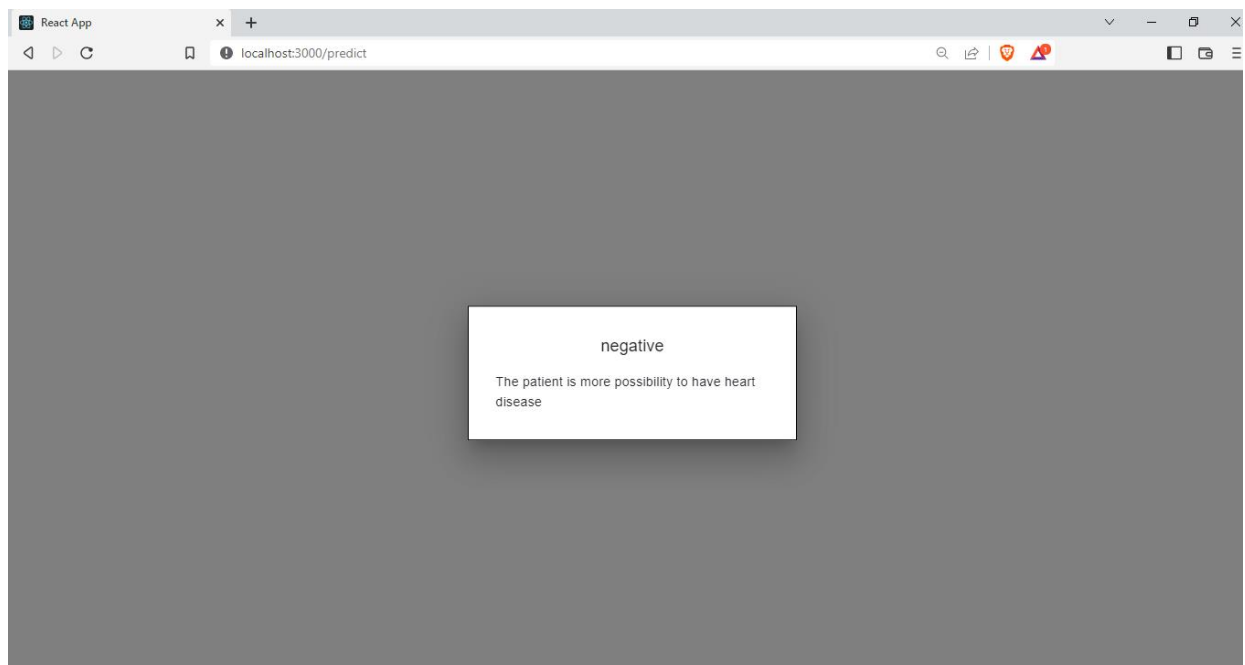


Testing a case where user does not have heart disease

A screenshot of the web application's input form. The browser address bar shows localhost:3000/predict. The page has a light gray background with a navigation bar at the top containing buttons for "Home", "predict", "causes", "Prevention", "Treatment", and "Logout". The form consists of several input fields arranged in a grid:

- Age**: A text input field containing the number "67".
- Sex**: Two radio button options, "Female" (selected) and "Male".
- Chest pain type**: A text input field containing the number "3".
- BP**: A text input field containing the number "115".
- Cholesterol**: A text input field containing the number "564".
- FBS over 120**: A text input field containing the number "0".
- EGX results**: A text input field containing the number "2".
- Max HR**: A text input field containing the number "160".
- Exercise angina**: A text input field containing the number "0".
- ST Depression**: A text input field containing the number "1.6".
- Slope of ST**: A text input field containing the number "2".
- Number of vessels fluro**: A text input field containing the number "0".
- Thallium**: A text input field containing the number "7".

At the bottom center of the form is a blue button labeled "CHECK".



9. RESULTS

9.1 Performance Metrics

The confusion matrix below shows the performance metrics of the machine learning model.

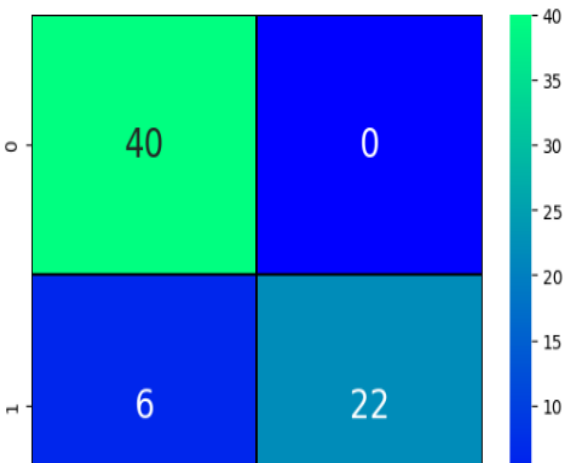
```
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier

tree_model = DecisionTreeClassifier(max_depth=5,criterion='entropy')
cv_scores = cross_val_score(tree_model, x, y, cv=10, scoring='accuracy')
m=tree_model.fit(x, y)
prediction=m.predict(X_test)
cm= confusion_matrix(y_test,prediction)
sns.heatmap(cm, annot=True,cmap='winter',linewidths=0.3, linecolor='black',annot_kws={"size": 20})
print(classification_report(y_test, prediction))
```

```
TP=cm[0][0]
TN=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]
print('Testing Accuracy for Decision Tree:',(TP+TN)/(TP+TN+FN+FP))
print('Testing Sensitivity for Decision Tree:',(TP/(TP+FN)))
print('Testing Specificity for Decision Tree:',(TN/(TN+FP)))
print('Testing Precision for Decision Tree:',(TP/(TP+FP)))
```

	precision	recall	f1-score	support
Absence	0.87	1.00	0.93	40
Presence	1.00	0.79	0.88	28
accuracy			0.91	68
macro avg	0.93	0.89	0.91	68
weighted avg	0.92	0.91	0.91	68

```
Testing Accuracy for Decision Tree: 0.9117647058823529
Testing Sensitivity for Decision Tree: 0.8695652173913043
Testing Specificity for Decision Tree: 1.0
Testing Precision for Decision Tree: 1.0
```



10. ADVANTAGES & DISADVANTAGES

10.1 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 information.

10.2 Disadvantages

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

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

Source Code:

<https://github.com/IBM-EPBL/IBM-Project-49273-1660817397/tree/main/Final%20Deliverables>

GitHub & Project Demo Link:

<https://drive.google.com/file/d/1HWYqEY6cHOqc46PWtDTOKezXJy01Nf4b/view?usp=sharing>