A NAALAIYA THIRAN PROJECT

PARISUTHAM INSTITUTE OF TECHNOLOGY AND SCIENCE

THANJAVUR

A PROJECT REPORT

VISUALIZING AND PREDICTING HEART DISEASE USING DATA ANALYTICS

Submitted by

TEAM ID PNT2022TMID46792

N.DEVENDRAN	**-**	**821319104007**
R.RAGUVARAN	**-**	**821319104027**
N.PARTHA SARATHI	**-**	**821319104022**
S.KESAVAN	**-**	**821319104017**
T.PERIYASAMY	**-**	**821319104023**

LIST OF FIGURES

LIST OF ABBREVATIONS

ML	Machine Learning
AI	Artificial Intelligen
NN	Neural Networks
SVM	Support Vector
XG	ExtremeGradient
DT	Decision Tree
FR	Features Reduction
KNN	K-Nearest Neighbour

CHAPTER 1

INTRODUCTION

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of+ heart disease plays a vital role in making decisions on lifestyle changes in highrisk patients and in turn reduces the complications.

Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

1.1 PROJECT OVERVIEW

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three classification algorithms namely Naïve Bayes, Decision Tree, and Random Forest are used at different levels of evaluations.

Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the three algorithms are evaluated at numerous levels and types of evaluation strategies.This will provide researchers and medical practitioners to establish a better.

1.2 PURPOSE

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients everyday in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise.

CHAPTER 2

LITERATURE SURVEY

Paper Title : Prediction and Analaysis of Heart Disease Using Data Mining Technique
Author : Anusha N.B, Chaitra.K ,
Publication : IRJET,may-2020,Volume:07 , Issue:05
Methodology: K-Nearest Neighbor(KNN), Navie bayes , Support Vector Machine (SVM)

In this study we have used an R studio rattle to perform Heart Disease classification of the Cleveland UCI repository. It provides an easy-to-use visual representation of the dataset, working environment and building the predictive analytics. SVM (Support Vector Machine) is a supervised machine learning algorithm that is mainly used to classify data into different classes. Unlike most algorithms, SVM makes use of a hyperplane, which acts like a decision boundary between the various classes. Weighted KNN is a modified version of KNN. A Naive Bayes' classifier may be a term addressing a simple probabilistic classification supported applying Bayes' theorem.

Paper title : Heart Attack Prediction and Visualization Using Machine learning.
Authors : Megha Banerjee, Reetodeep hazra, Suvranil saha, Megha bushan. Publication : IJIREEICE , July 2021,Volume -09, Issue 07
Methodology : Logistic regression, Gaussian naïvebayes , Random forest algorithm In this project, 4 machine learning algorithms are used namely Decision Tree, Random Forest, Gaussian Naive Bayes and Logistic Regression.

A Decision Tree (DT) represents a tree like structure where each number considered being a branch with an outcome. DT is a fundamental component of Random Forest, which are among the most powerful ML algorithms available today. . While building subsets of data for trees, the word "random" comes into the picture.

A subset of data is made by randomly selecting x number of features (columns) and y number of examples (rows) from the original dataset of n features and m examples. Random forests are more stable and reliable than just a decision tree.

Paper title : Big Data Analytics in Heart Disease Prediction
Authors : Ahmed Imail, Samir Abdlerazek ,I.M.El-Henawy.
Publication : JATIT & LLS , June 2020
Methodology : SVM , Features reduction (FR).

The proposed methodology is using cloud structure to process medical data and to try to support doctors to decide in the diagnosis of heart diseases using SVM. This methodology is an efficient system because the proposed system applies a selection method based on the main features of the given dataset to classify the heart disease from the user profile in the cloud. The proposed framework used clusters from MapReduce. The streaming data is analyzed then can be stored on the cosmos DB.The Features Reduction (FR) is a methodology of selection that selects features with a highclass correlation (output).When a dataset is provided as data mining input with many classification features, the first target is cleaning the dataset and improving the

2.1 EXISTING PROBLEM

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

2.2 REFERNCES

1. Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-8

2. Dangare C S & Apte S S (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

3. Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334-43.

4. Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies, 6(1), 637-9.

5. Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259-64). IEEE. ICCRDA
2020 IOP Conf. Series: Materials Science and Engineering 1022 (2021) 012072 IOP
Publishing doi:10.1088/1757-899X/1022/1/012072 9

2.3 PROBLEM DEFINITION

Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients everyday in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise.

Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

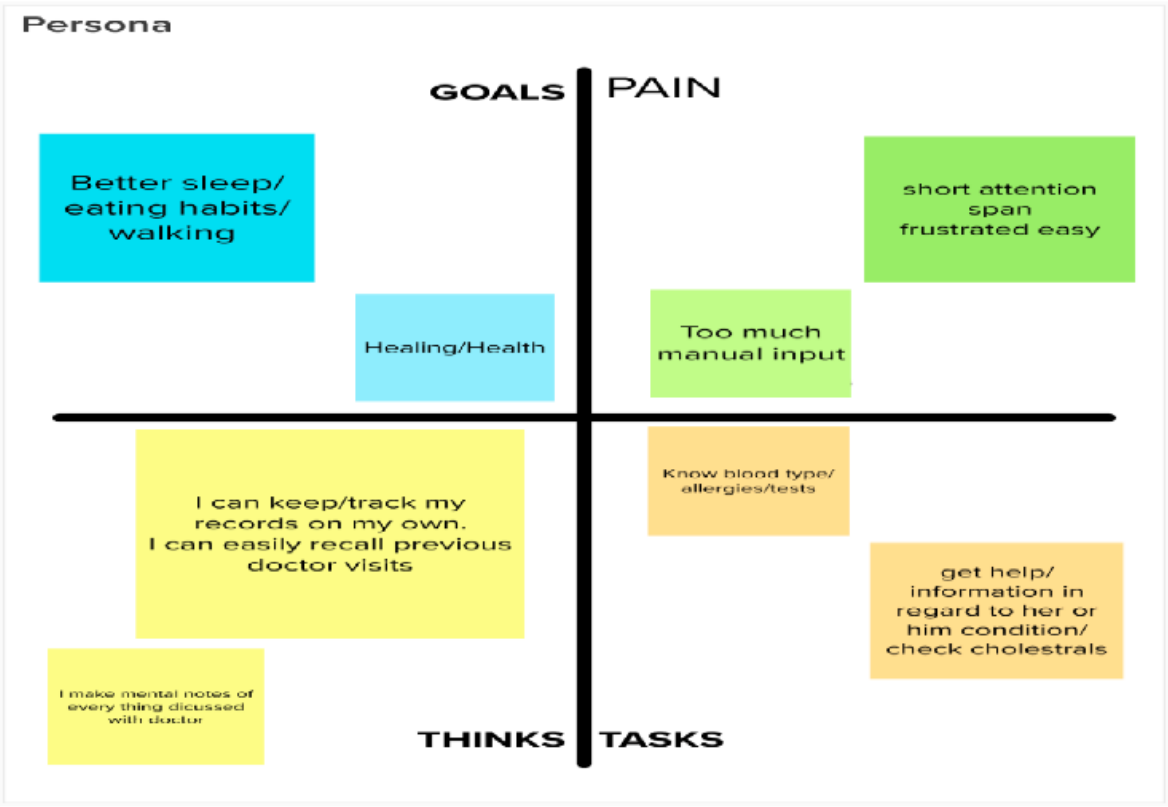
CHAPTER 3

IDEATION & PROPOSED SOLUTION

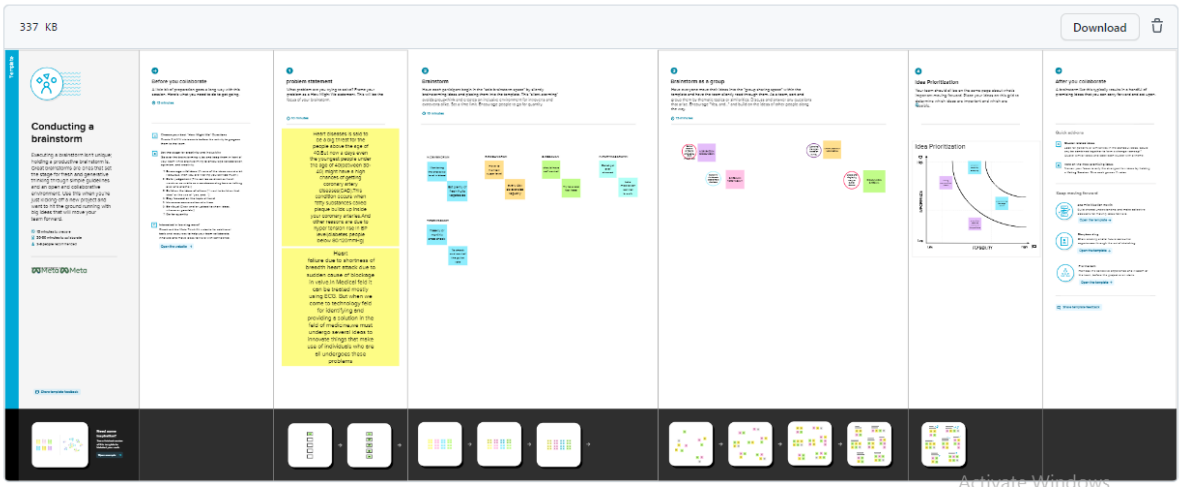
The working of the system starts with the collection of data and selecting the important attributes. Then the required data is preprocessed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1. Collection of Dataset
2. Selection of attributes
3. Data Pre-Processing
4. Balancing of Data
5. Disease Prediction

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAIN STROMING



3.3 PROPOSED SOLUTION

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	¢ To develop an interactive dashboard to predict the heart disease accurately with few tests and attributes the presence of heart disease.
2.	Idea / Solution description	¢ Analyzing data and identifying the heart disease using Cognos analysis.
3.	Novelty / Uniqueness	¢ Hoping to achieve maximum accuracy to provide prior treatment to the patients and reduce the fatality rate.
4.	Social Impact / Customer Satisfaction	Reduces the biases and mistakes caused by the decisions of doctors based on their intuitions and experiences
5.	Business Model (Revenue Model)	1. Data security. 2. Easy to use. 3.Constant updates according to necessity
6.	Scalability of the Solution	1. Can be used in any platform (Windows, mac, etc..) 2. performance of the system.t affect the 3. Scalable dataset.

3.4 PROPOSED SOLUTION FIT

CHAPTER 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	Cognos Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

4.2 NON-FUNCTIONAL REQUIREMENTS:

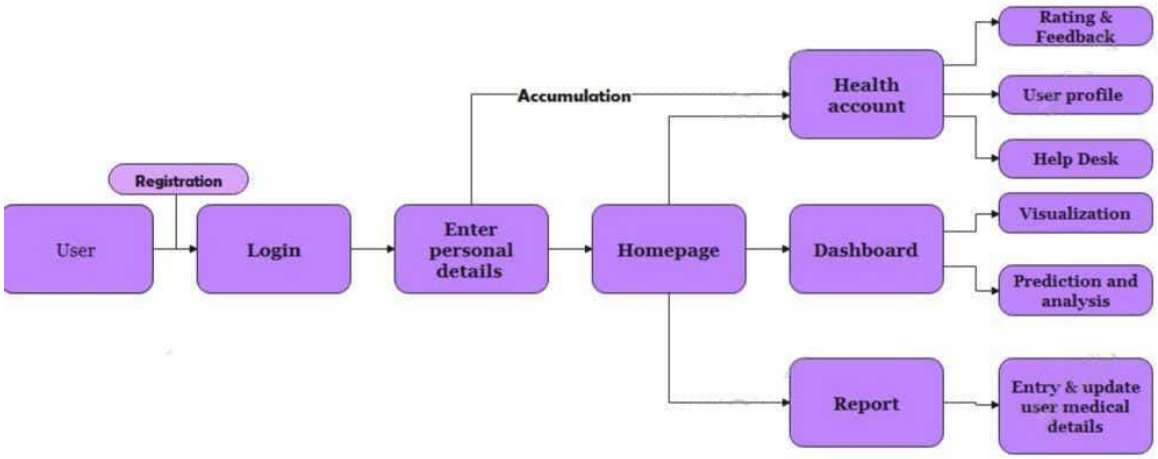
Following are the non-functional requirements of the proposed solution.

NFR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The application will have a simple and user- friendly graphical interface.
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
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.
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

CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS



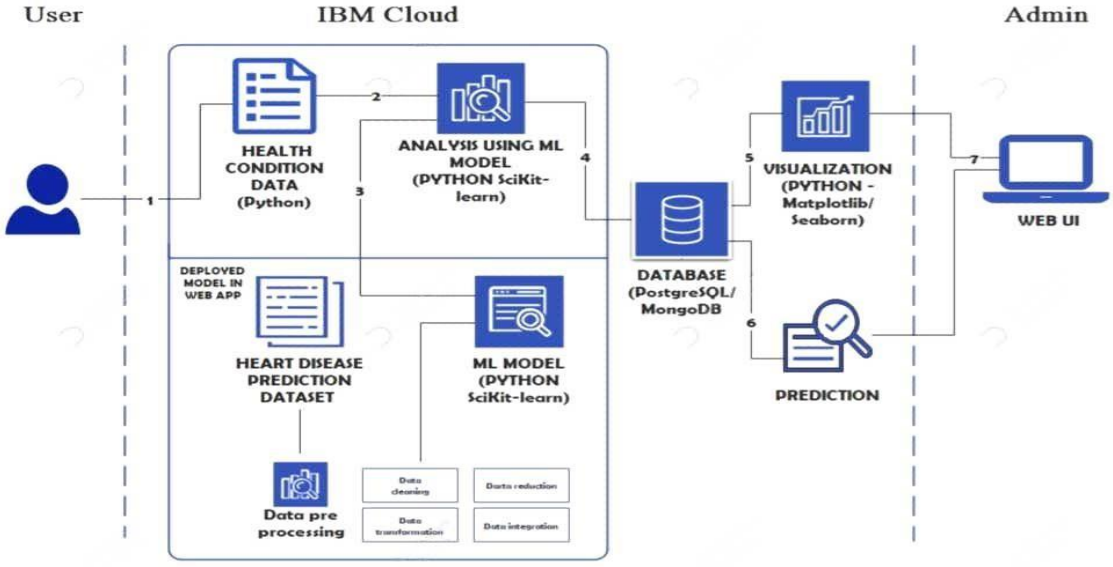
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Data Flow Diagram for Heart Disease Prediction Dashboard:
Flow:

1. User creates an account in the application.
2. User enters the medical records in the dashboard.
3. User can view the visualizations of trends in the form of graphs and charts
4. for his/her medical records with the trained dataset.

5.2 SOLUTIONS AND TECHNICAL ARCHITECTURE


user
Collection of dataset
01
collecting the raw data like chest pain,BP,cholestrol,ST depression,EKG results using user input (from smart watch or app)
Attribute selection
02
Identifying the attributes pertaining the risk of heart disease
Data pre-processing
03
It incudes data cleaning,data transformation,data balancing,integration, reduction,data mining,Normalizing,handling missing values,outlier detection & removal
Classification techniques
04
Various machine learning algorithms like SVM, Naive Bayes,k nearest neighbour, Decision Tree, Random Forest, Logistic Regression, Adaboost, Xg-boost,ANN are used for classification
using data analytics & heart disease prediction
05
identification of heart disease with highest accuracy using suitable algorithm and decision making by using analytics
visualization
05
visualization using interactive dashboard


User
IBM Cloud
Admin
HEALTH CONDITION DATA (Python)
ANALYSIS USING ML MODEL (PYTHON SciKit-learn)
DEPLOYED MODEL IN WEB APP
HEART DISEASE PREDICTION DATASET
ML MODEL (PYTHON SciKit-learn)
Data pre processing
DATABASE (PostgreSQL/ MongoDB
VISUALIZATION (PYTHON - Matplotlib/ Seaborn)
PREDICTION
WEB UI
1
2
3
4
5
6
7

S.No	Component	Description	Technology
1.	Importing data set	Data Import lets you upload data from external sources and combine it with data you collect via Analytics	Python, numpy, pandas.
2.	Data Cleaning	Data cleaning is a process by which inaccurate, poorly formatted, or otherwise messy data is organized and corrected	Python, numpy, pandas
3.	Data Preprocessing	Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure	Python, numpy, scipy, pandas
4.	Training data	Training data is the subset of original data that is used to train the machine learning model,	Numpy, scipy, pandas
5.	Testing data	Test data is data which has been specifically identified for use in tests, typically of a computer program.	Numpy, scipy, pandas
6.	Machine learning model	A machine learning model is a file that has been trained to recognize certain types of patterns. You train a model over a set of data, providing it an algorithm that it can use to reason over and learn from those data	Numpy, scipy, pandas, sklearn
7.	Improve model performance	Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right.	sklearn
8.	Checking accuracy	A data accuracy check, sometimes called a data sanity check, is a set of quality validations that takeplace before using data.	Sklearn

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / Dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email & password	I can access my account / Dashboard whenlogged in	High	Sprint-1
Customer (Web user)	Dashboard	USN-4	User can view his/her complete medical analysis and accuracy of disease prediction	I can view my medical analysis in the dashboard	High	Sprint-2
		USN-5	User can view the accuracy of occurrence of heart disease	I can view the accuracy of heart disease in the dashboard	High	Sprint-2
Customer Care Executive	Helpdesk	USN-6	As a customer care executive, he/she can view the customer queries.	I can post my queries in the dashboard	Medium	Sprint-3
		USN-7	As a customer care executive, he/she can answer the customer queries.	I can get support from helpdesk	High	Sprint-3
Administrator	User Profile	USN-8	As an admin, he/she can update the health details of users.	I can view my updated health details.	High	Sprint-4
		USN-9	As an admin, he/she can add or delete users.	I can access my account / Dashboard when loggedin	High	Sprint-4

CHAPTER 6

PROJECT PLANNING AND SCHEDULING

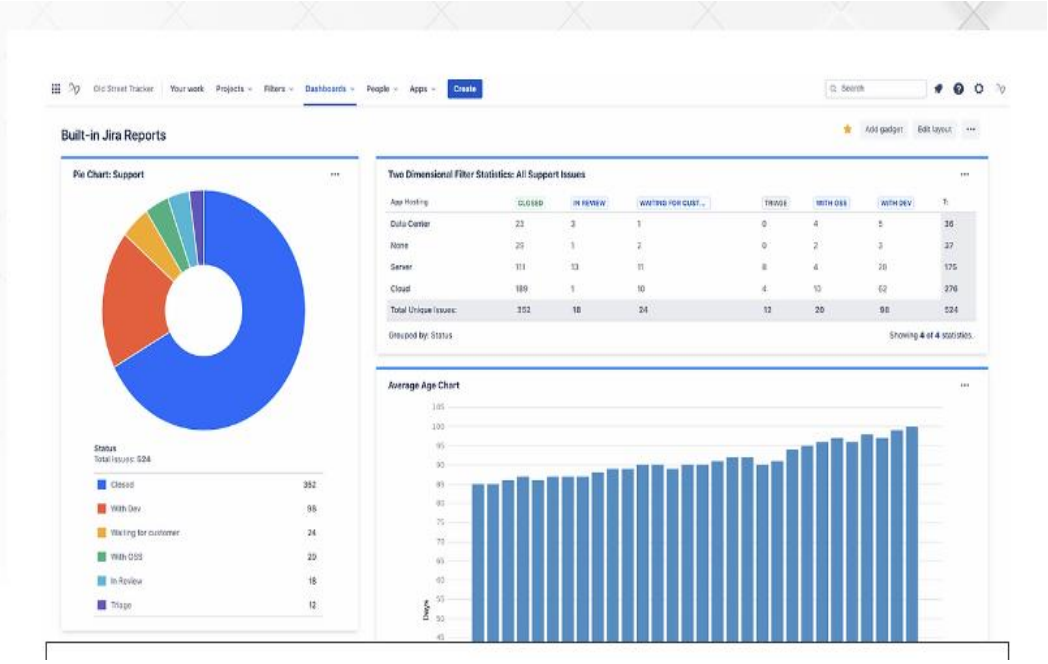
6.1 SPRINT PLANNING AND ESTIMATION

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	4
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	-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

6.2 SPRINT DELIVERY AND SCHEDULE

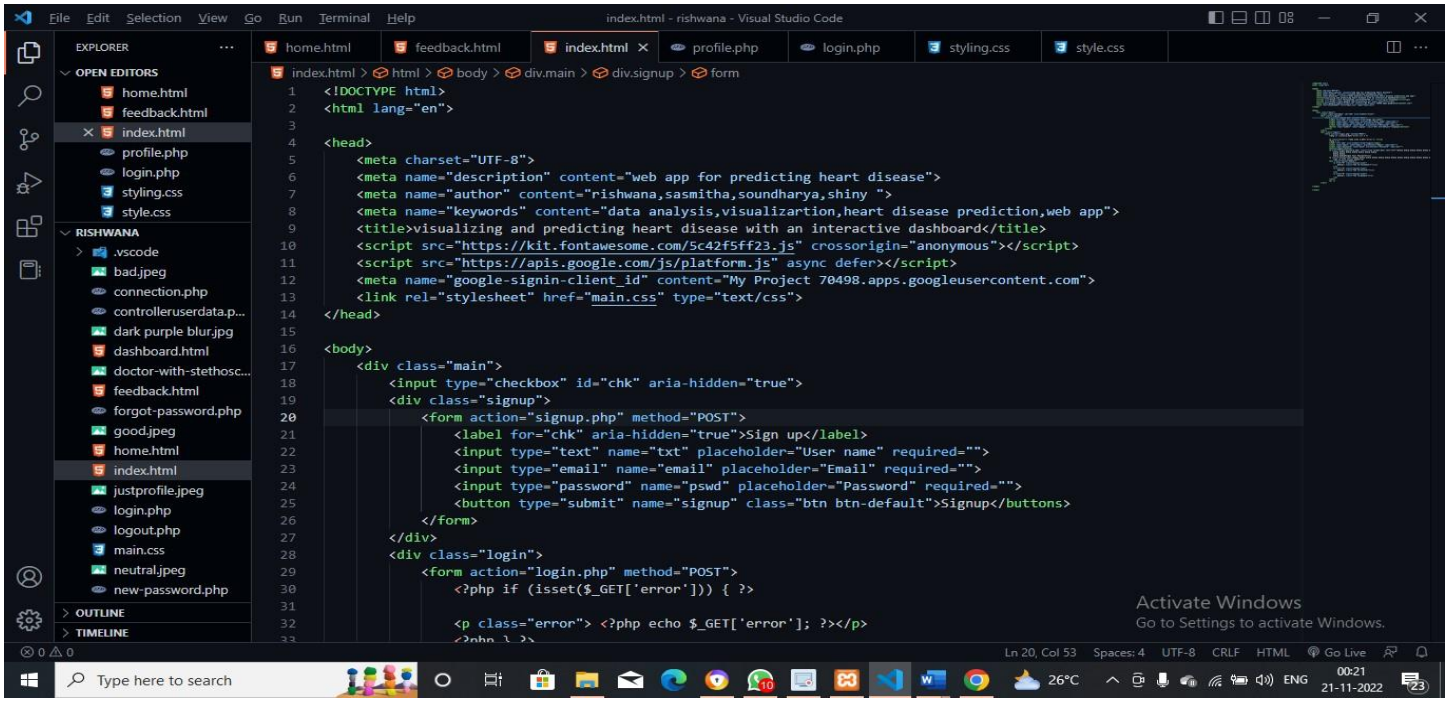
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	30 Oct 2022	04 Nov 2022	17	04 Nov 2022
Sprint-3	20	6 Days	05 Nov 2022	11 Nov 2022	18	11 Nov 2022
Sprint-4	20	6 Days	12 Nov 2022	17 Nov 2022	19	17 Nov 2022

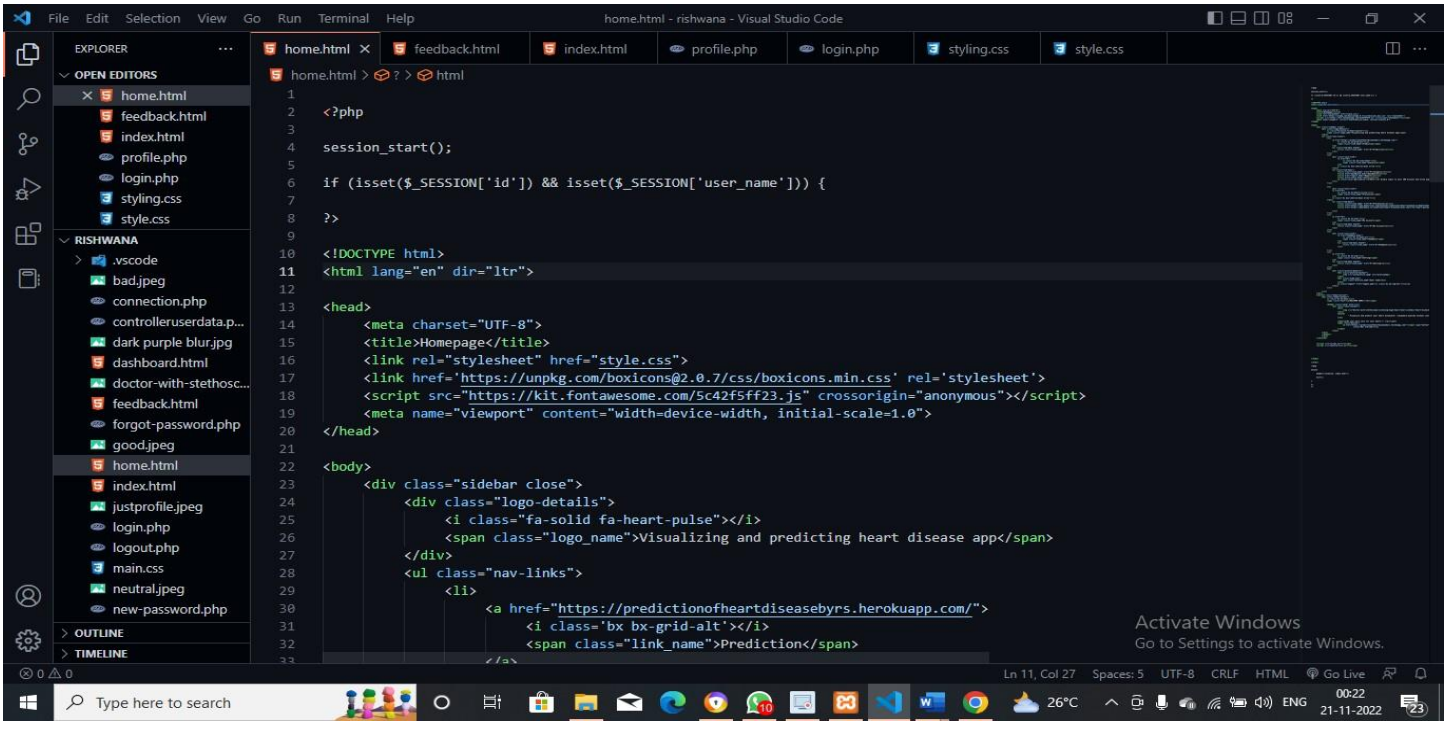
6.3 REPORTS FROM JIRA



CHAPTER 7

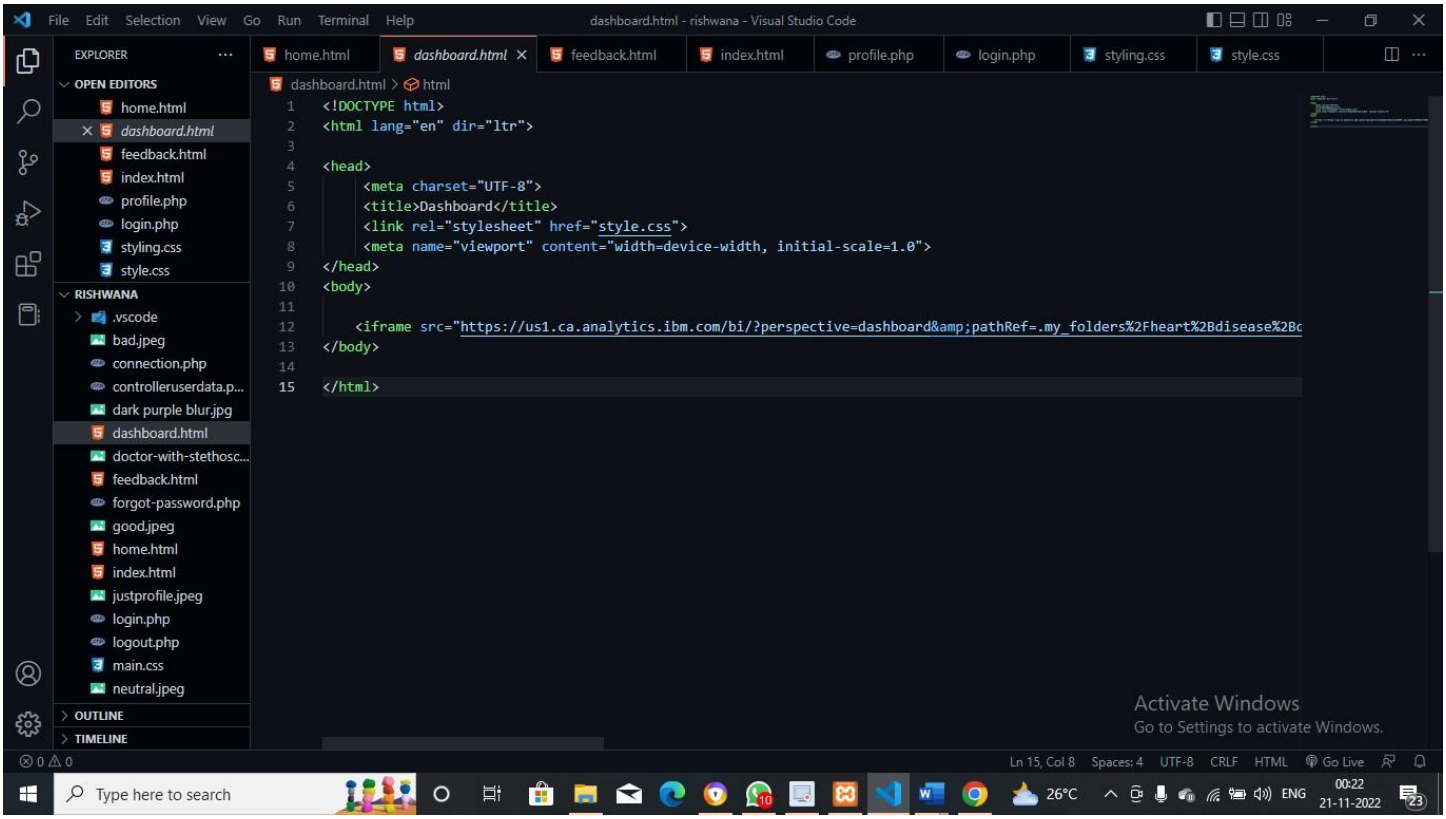
CODING AND SOLUTIONING





7.1.1

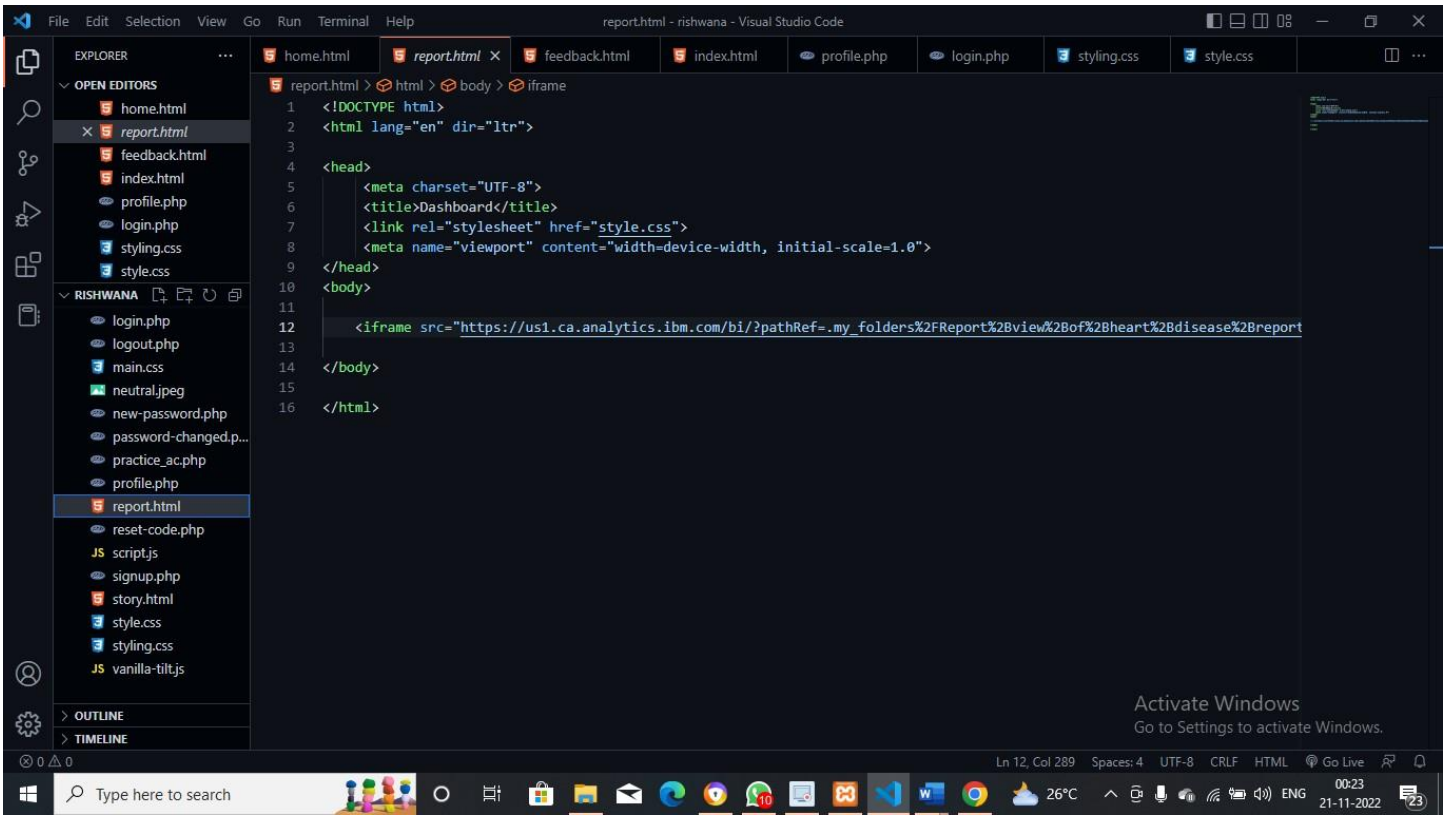
Dashboard



Using Cognos Analytics, dashboard is created which shows the relation between attributes and how they are responsible for chances of heart disease. The dashboard is incorporated in website using iframe. It is mandatory to have an IBM account to view the dashboard. As soon as the page is loaded, it asks to sign in to the IBM account. Once signed in, user can view the dashboard. Dashboard has multiple tabs, each containing a chart of relation between attributes. The above code shows how dashboard is included in the website.

7.1.2

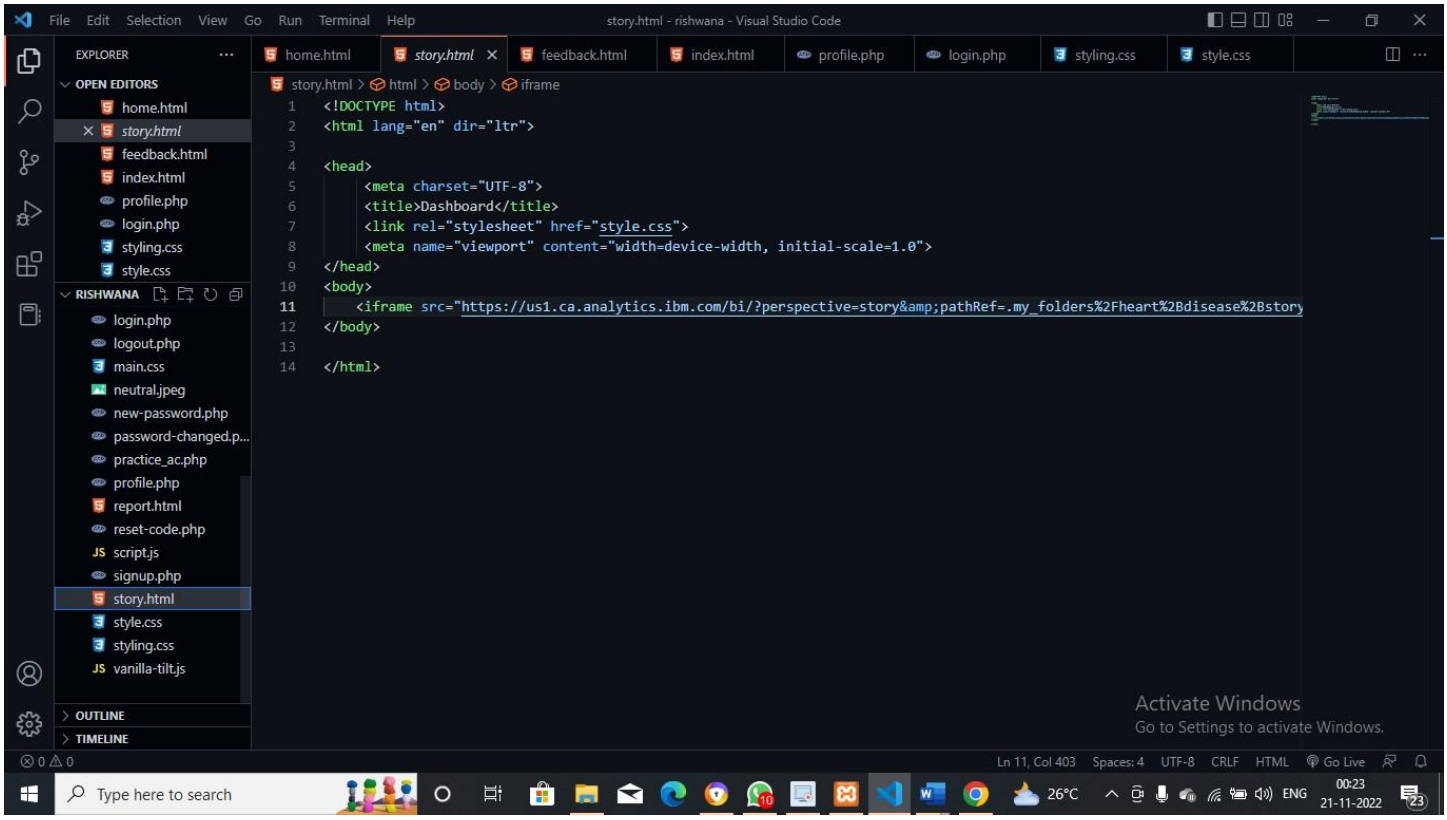
Report



Using Cognos Analytics, Report is created which shows the relation between attributes and how they are responsible for chances of heart disease. The Report is incorporated in website using iframe. It is mandatory to have an IBM account to view the Report. As soon as the page is loaded, it asks to sign in to the IBM account. Once signed in, user can view the Report. The above code shows how Report is included in the website.

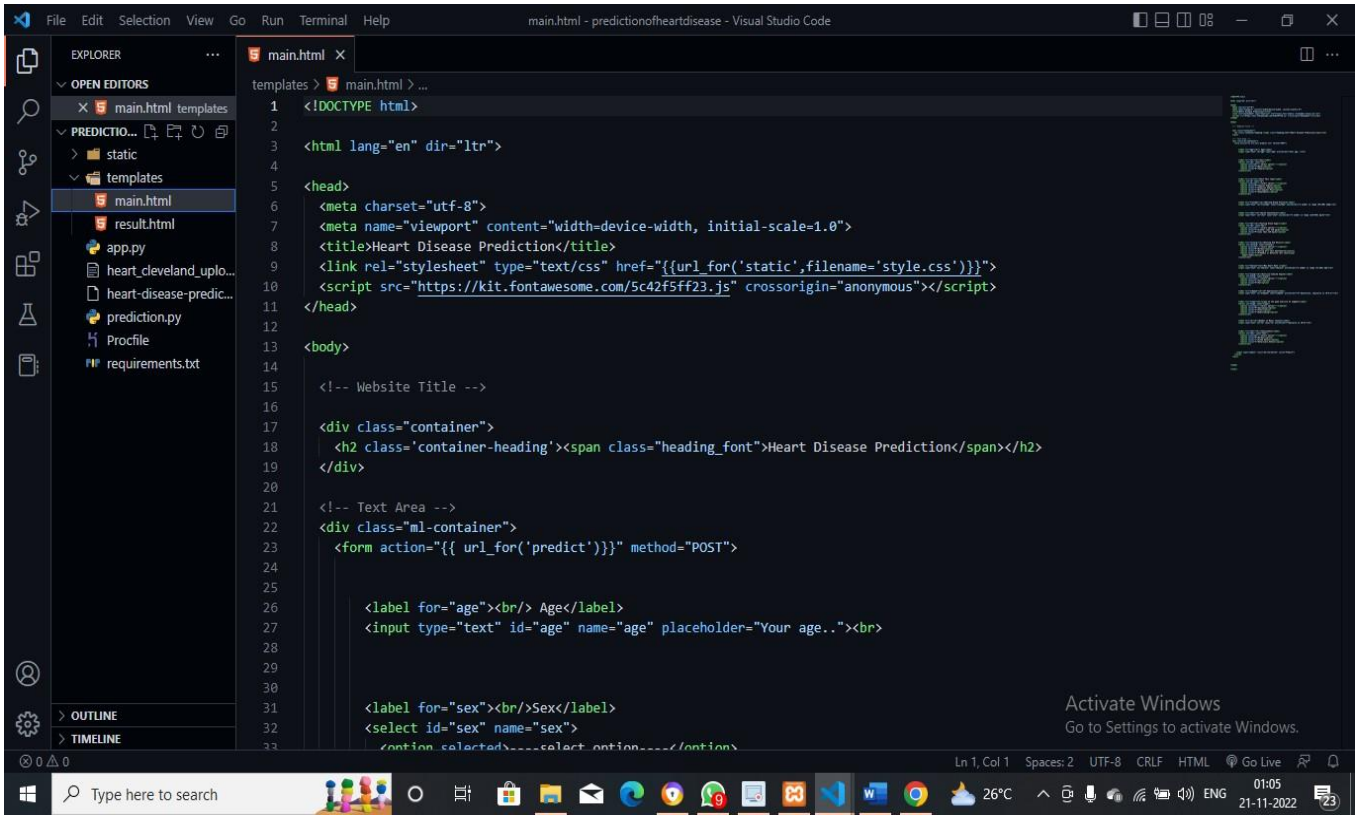
7.1.3

Story



Using Cognos Analytics, Story t is created which shows the relation between attributes and how they are responsible for chances of heart disease. The story is incorporated in website using iframe. It is mandatory to have an IBM account to view the story. As soon as the page is loaded, it asks to sign in to the IBM account. Once signed in, user can view the story. story has multiple scenes, each containing a chart of relation between attributes. The above code shows how story is included in the website.

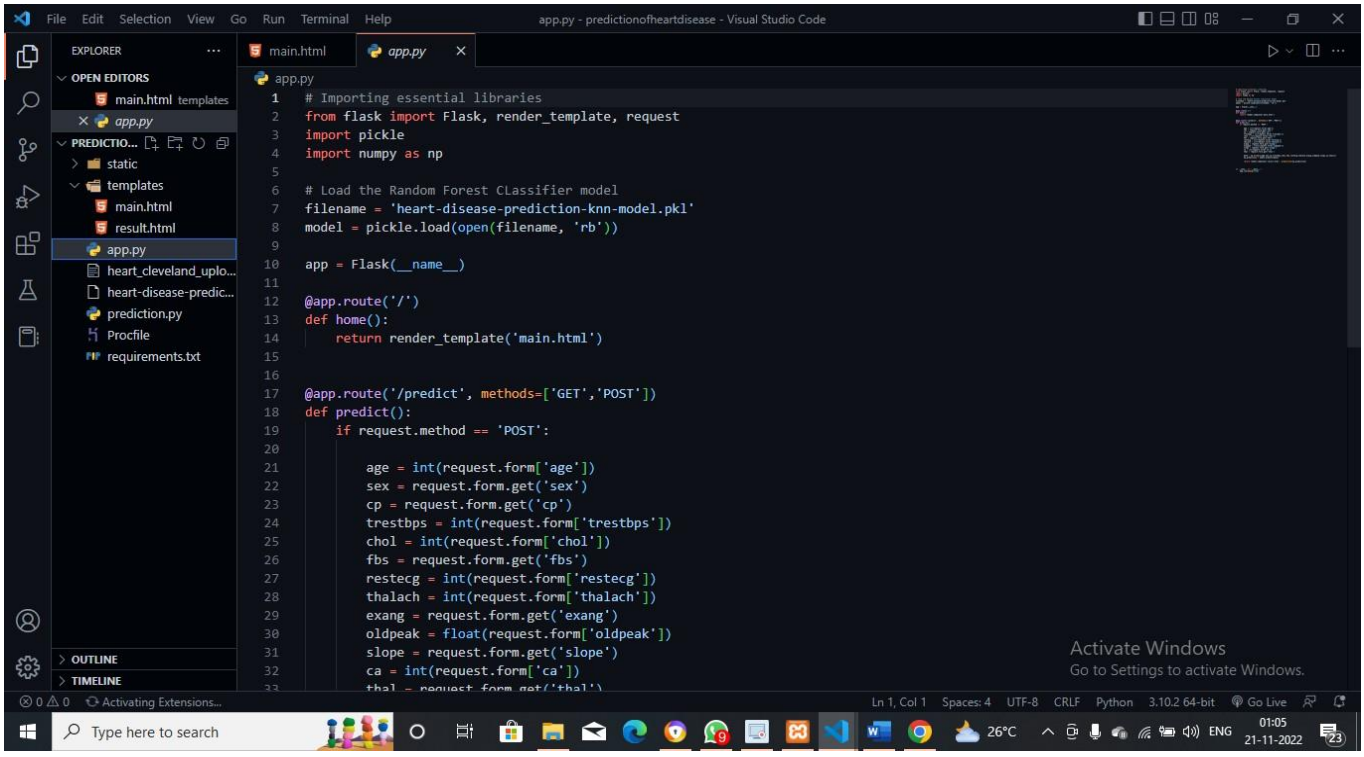
7.1.4


main.html - predictionofheartdisease - Visual Studio Code
EXPLORER
OPEN EDITORS
main.html templates
PREDICTIO...
static
templates
main.html
result.html
app.py
heart_cleveland_uplo...
heart-disease-predic...
prediction.py
Procfile
requirements.txt
OUTLINE
TIMELINE
templates > main.html > ...
<!DOCTYPE html>
<html lang="en" dir="ltr">
<head>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<title>Heart Disease Prediction</title>
<link rel="stylesheet" type="text/css" href="{{url_for('static',filename='style.css')}}">
<script src="https://kit.fontawesome.com/5c42f5ff23.js" crossorigin="anonymous"></script>
</head>
<body>
<!-- Website Title -->

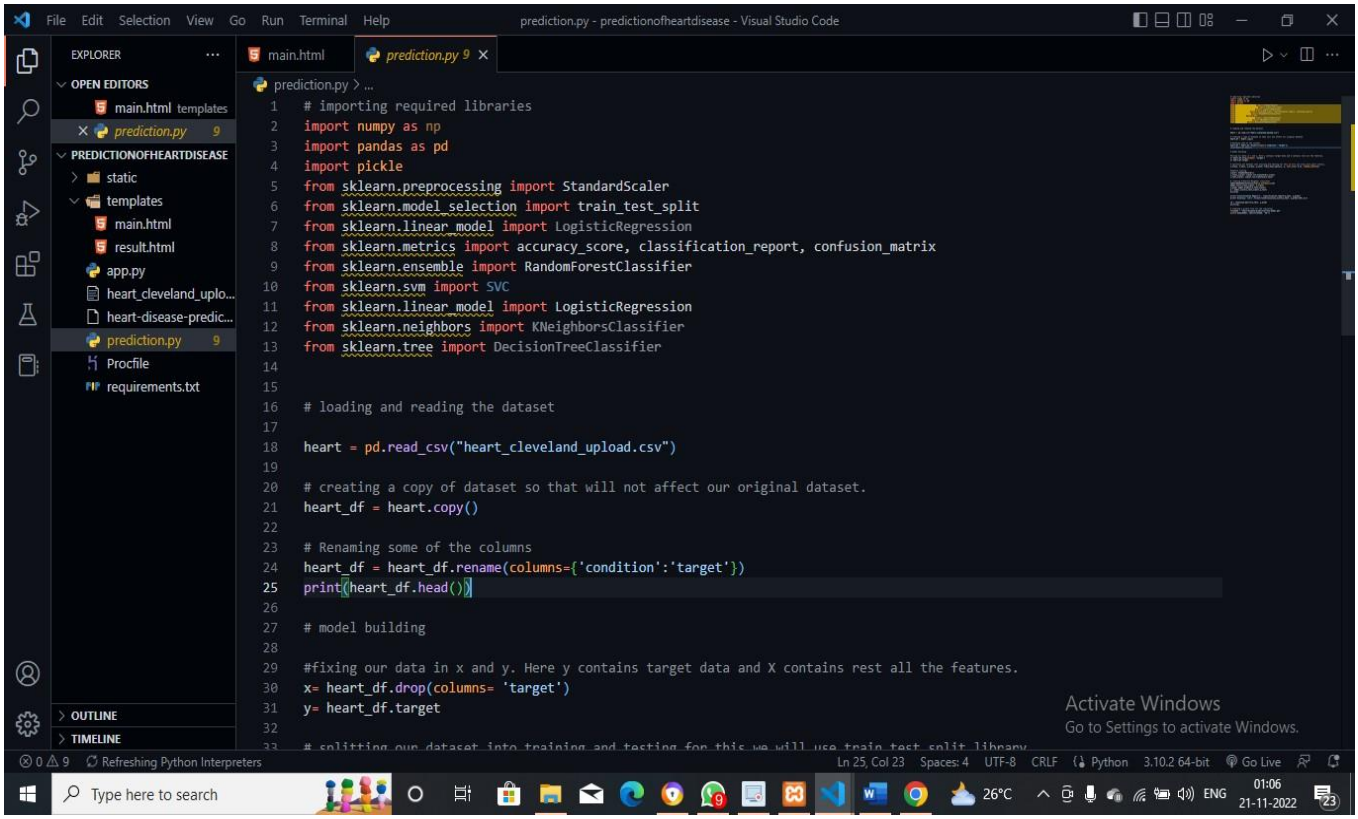
<h2 class='container-heading'>Heart Disease Prediction</h2>

<!-- Text Area -->

<form action="{{ url_for('predict')}}" method="POST">
<label for="age">
 Age</label>
<input type="text" id="age" name="age" placeholder="Your age..">
<label for="sex">
Sex</label>
<select id="sex" name="sex">
Activate Windows
Go to Settings to activate Windows.
Type here to search
26°C
ENG
01:05
21-11-2022

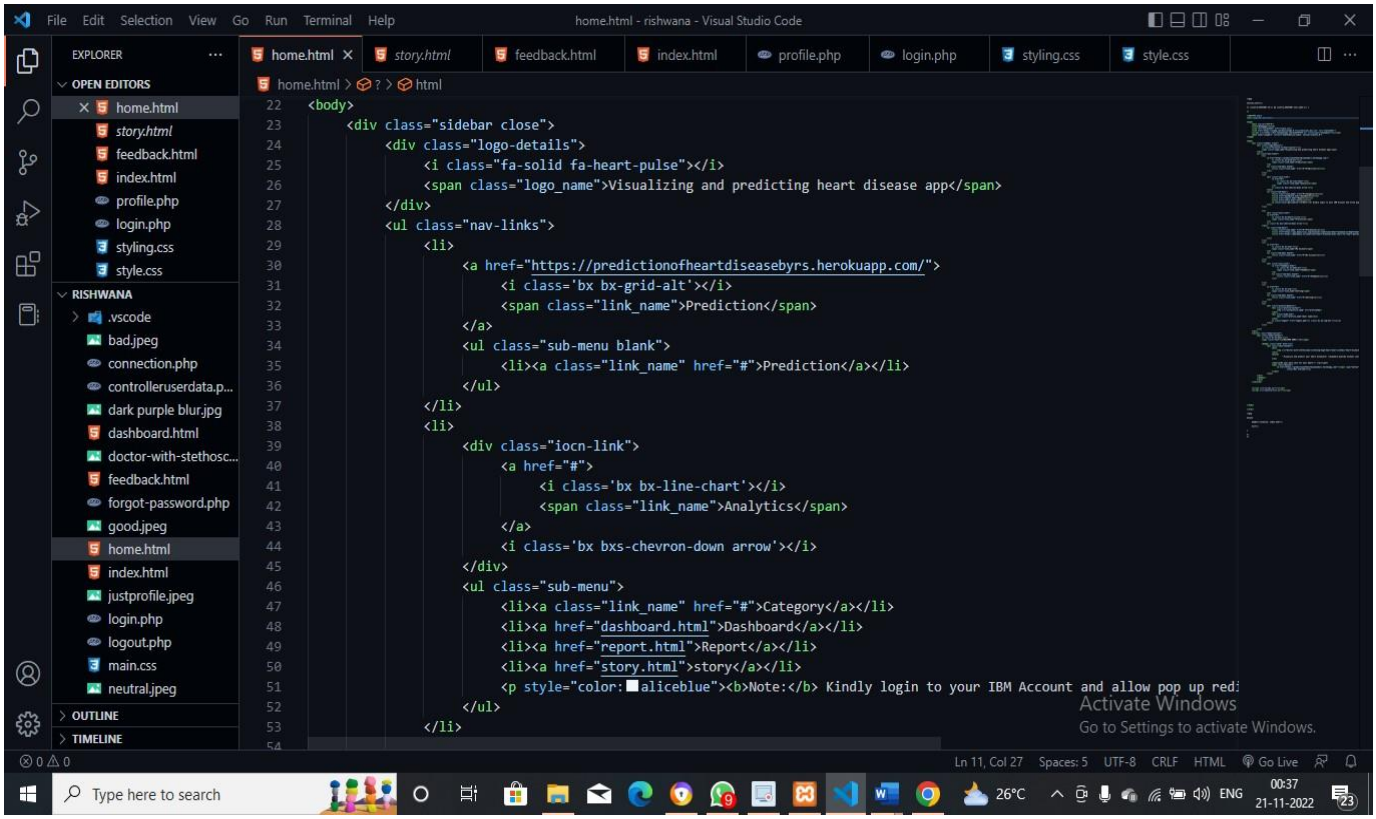

app.py - predictionofheartdisease - Visual Studio Code
main.html
app.py
EXPLORER
OPEN EDITORS
main.html templates
app.py
PREDICTIO...
static
templates
main.html
result.html
app.py
heart_cleveland_uplo...
heart-disease-predic...
prediction.py
Procfile
requirements.txt
OUTLINE
TIMELINE
# Importing essential libraries
from flask import Flask, render_template, request
import pickle
import numpy as np
# Load the Random Forest CLassifier model
filename = 'heart-disease-prediction-knn-model.pkl'
model = pickle.load(open(filename, 'rb'))
app = Flask(__name__)
@app.route('/')
def home():
return render_template('main.html')
@app.route('/predict', methods=['GET','POST'])
def predict():
if request.method == 'POST':
age = int(request.form['age'])
sex = request.form.get('sex')
cp = request.form.get('cp')
trestbps = int(request.form['trestbps'])
chol = int(request.form['chol'])
fbs = request.form.get('fbs')
restecg = int(request.form['restecg'])
thalach = int(request.form['thalach'])
exang = request.form.get('exang')
oldpeak = float(request.form['oldpeak'])
slope = request.form.get('slope')
ca = int(request.form['ca'])
Activate Windows
Go to Settings to activate Windows.
Type here to search
26°C
ENG
01:05
21-11-2022

7.1.5



The above code shows how user input is got as form and how it is processed and given as input to machine learning model. Which in turn gives if heart disease is present or absent. This model is created using flask and deployed on Heroku cloud. That link is embedded in the main web app.

7.1.6



The above code shows how heart disease prediction app which is deployed on Heroku cloud is included in the website.

7.1.7

CHAPTER 8

TESTING

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

8.1 TYPES OF TESTS

8.1.1 Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

8.1.2 Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

8.1.3 Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

7.1.8

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

8.1.4 System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test.

8.1.5 White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

8.1.6 Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as

most other kinds of tests, must be written from a definitive source document, such a[,s specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot “see” into it.

7.1.9

8.2 Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

8.2.1 Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

8.2.2 Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

8.2.3 Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

8.3 Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

8.4 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

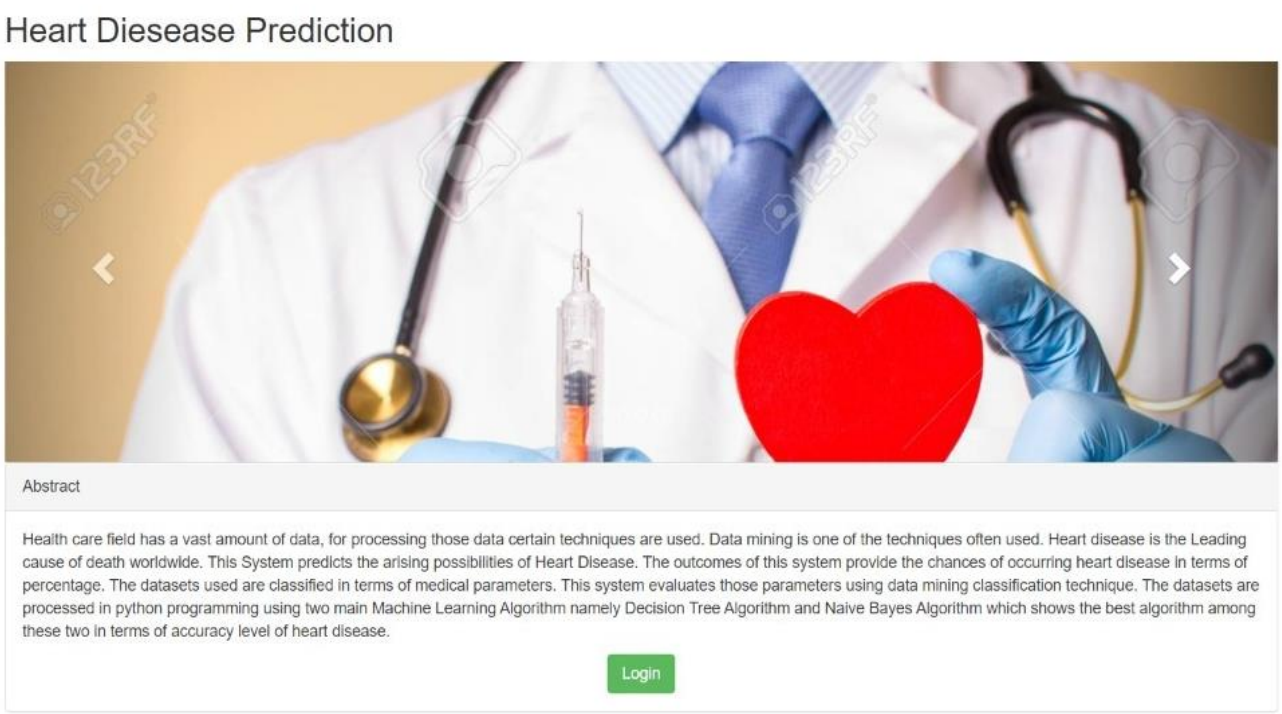
Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7.1.10

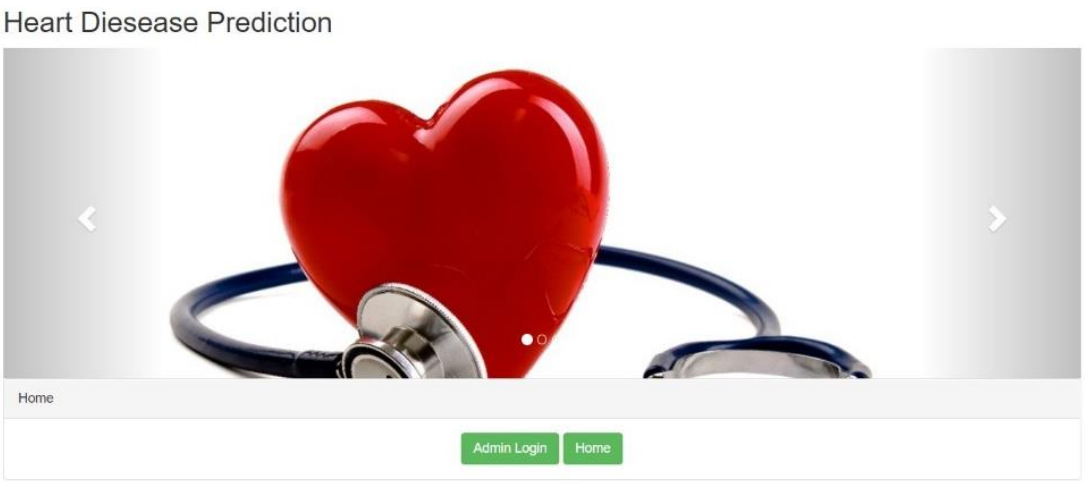
CHAPTER 9

RESULT

Homepage



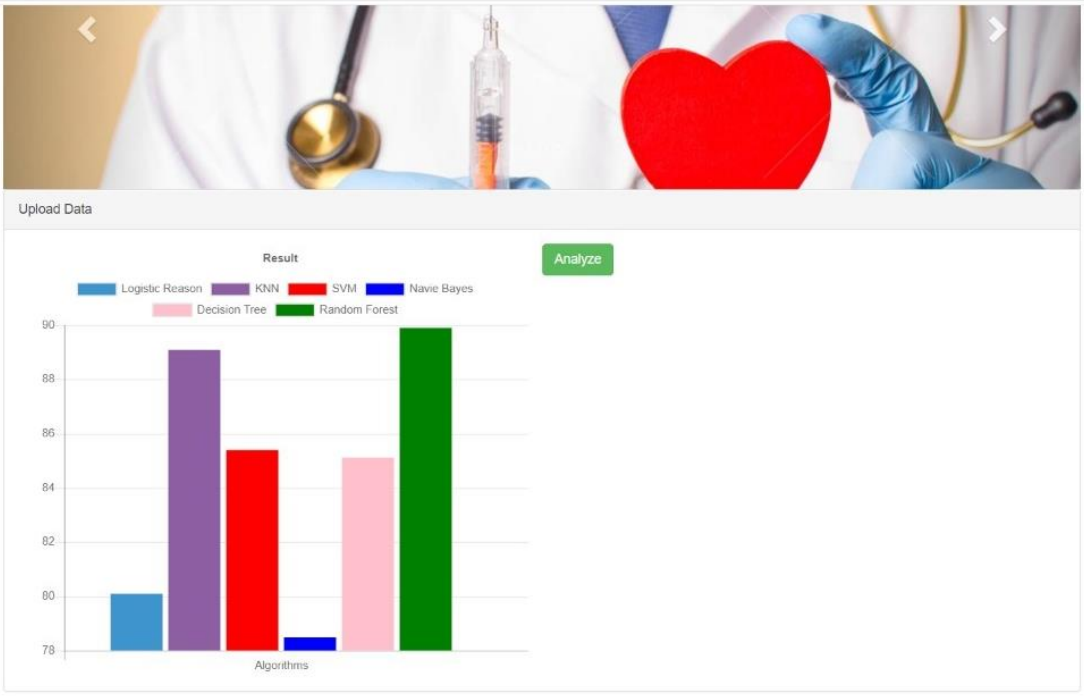
Dashboard



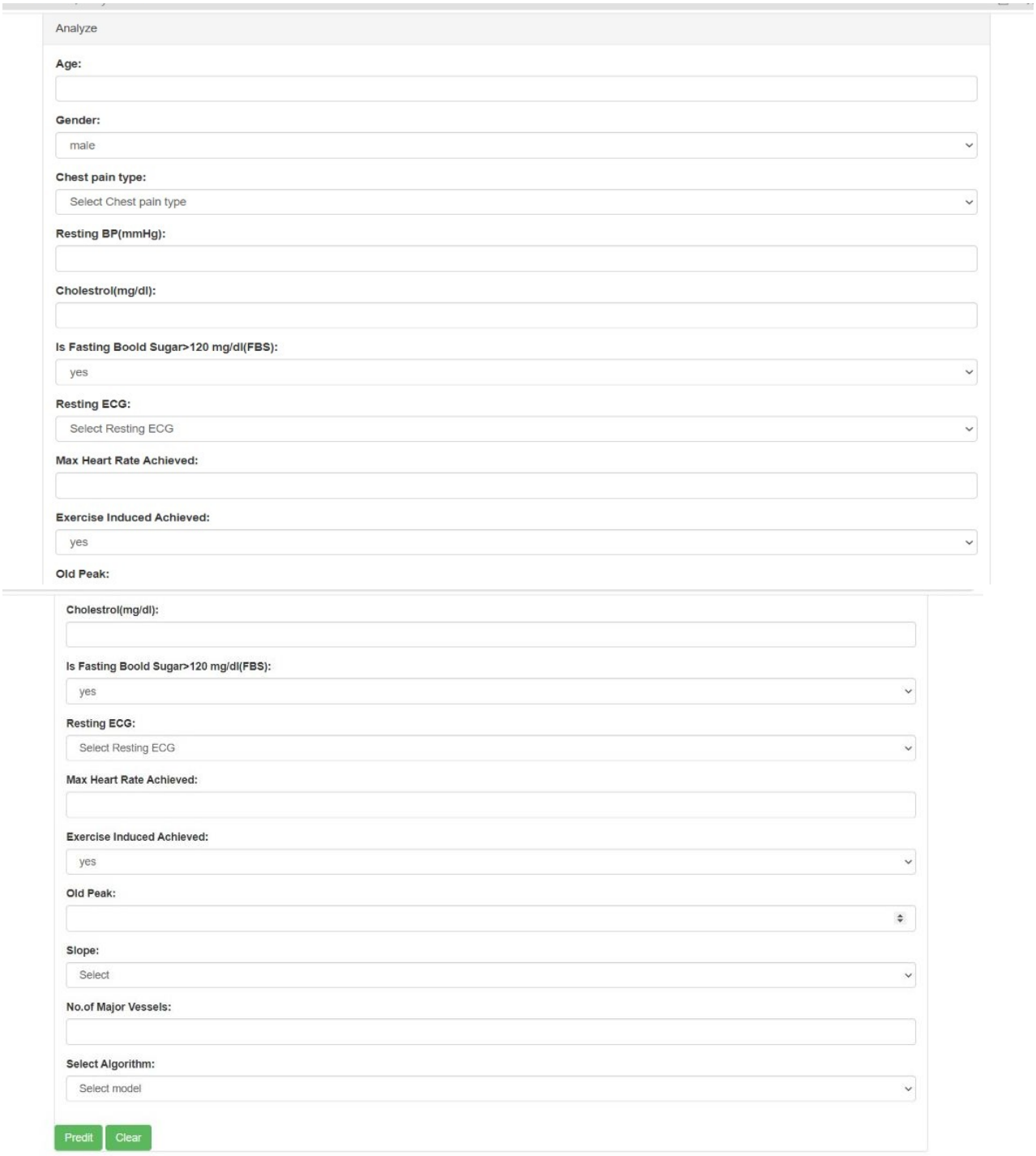
7.1.11


Heart Diesease Prediction
Admin Login
User Name:
Admin
Password:
Remember me
Submit

Report

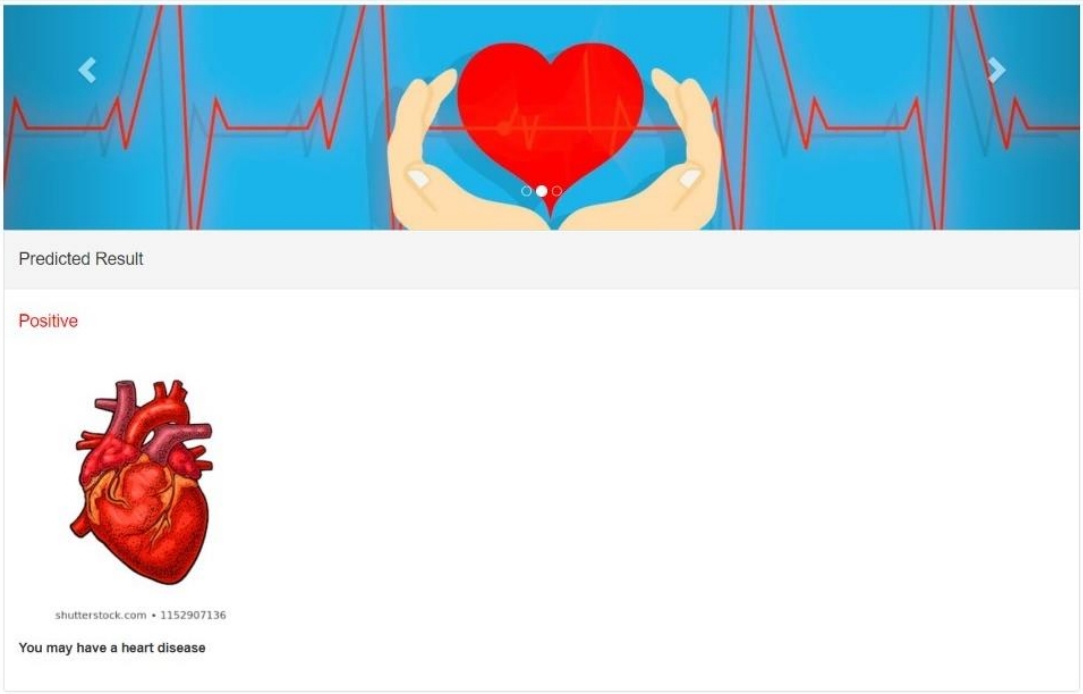

Upload Data
Result
Analyze
Logistic Reason
KNN
SVM
Navie Bayes
Decision Tree
Random Forest
90
88
86
84
82
80
78
Algorithms

7.1.12
Prediction Form



7.1.13

Prediction Result



9.1 PERFORMANCE METRICS

Several standard performance metrics such asaccuracy. precision and error in classification have been considered for the computation of performance efficiency of this model.

7.1.14

CHAPTER 10
ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Increase in exercise tolerance.
- Reduction in body weight.
- Reduction in blood pressure.
- Reduction in bad (LDL and total) cholesterol.
- Increase in good (HDL) cholesterol.
- Increase in insulin sensitivity.

DISADVANTAGES:

- Heart Attack. A heart attack, or myocardial infarction, usually tops the list of cardiovascular diseases in the United States statistically and anecdotally.
- Stroke.
- Heart Failure.
- Arrhythmia.
- Heart Valve Complications.

7.1.15

CHAPTER 11

CONCLUSION

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the seven different machine learning algorithms used to measure the performance are SVM, Decision Tree, Random Forest, Naïve Bayes, Logistic

Regression, Adaptive Boosting, and Extreme Gradient Boosting applied on the dataset.

The expected attributes leading to heart disease in patients are available in the dataset which contains 76 features and 14 important features that are useful to evaluate the system are selected among them. If all the features taken into the consideration then the efficiency of the system the author gets is less. To increase efficiency, attribute selection is done. In this n features have to be selected for evaluating the model which gives more accuracy. The correlation of some features in the dataset is almost equal and so they are removed.

All the seven machine learning methods accuracies are compared based on which one prediction model is generated. Hence, the aim is to use various evaluation metrics like confusion matrix, accuracy, precision, recall, and f1-score which predicts the disease efficiently. Comparing all seven the extreme gradient boosting classifier gives the highest accuracy of 81%.

7.1.16

CHAPTER 12

FUTURE SCOPE

Using more accurate dataset with more necessary parameters, the accuracy of prediction can be increased. In collaboration with hospitals, doctors can be suggested with contact information. The dashboard can be expanded to have more charts and relations. Can also be connected to smart watch that helps to notify hospitals nearby if the user gets a sudden heart attack.

7.1.17

CHAPTER 13

APPENDIX

GITHUB LINK(source code):https://github.com/IBM-EPBL/IBM-Project-36091-1660292665

PROJECT DEMO LINK : https://drive.google.com/file/d/15InrGdApI-cBV8czD2TxOVkUeS5oE4iO/view?usp=drivesdk