IBM – NALAIYA THIRAN PROJECT

HEART DISEASE VISUALISATION WITH AN INTERACTIVE DASHBOARD

INDUSTRY MENTOR : MAHIDHAR SAUMYA

FACULTY MENTOR : N.GOPALAKRISHNAN

TEAM ID : PNT2022TMID53476

TEAM LEAD : KEERTHANA V

TEAM MEMBER : GAYATHRI S

TEAM MEMBER : JAYASHRI P

TEAM MEMBER : KALAIARASI M S

TABLE OF THE CONTENT

CHAPTER-1

INTRODUCTION

1.1 PROJECT OVERVIEW

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset. We can use that knowledge in our project of Heart Disease Prediction as it will help a lot of people. Cardiovascular diseases are very common these days, they describe a range of conditions that could affect your heart. The World health organization estimates that 17.9 million global deaths from (Cardiovascular diseases) CVDs . It is the primary reason for deaths in adults. Our project can help predict the people who are likely to be diagnosed with a heart disease by help of their medical history. It recognizes who all are having any symptoms of heart disease such as chest pain or high blood pressure and can help in diagnosing disease with less medical tests and effective treatments, so that they can be cured accordingly. This project focuses on mainly three data mining techniques namely: (1) Logistic regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 87.5% which is better than the previous system where only one data mining technique was used. So, using more data mining techniques increased the HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression. The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc. A dataset is selected from the Kaggle. By using this dataset, we predict whether the patient can have a heart disease or not. To predict

this, we use 13 medical attributes of a patient and classify him if the patient is likely to have a heart disease. These medical attributes are trained under three algorithm: Random Forest Classifier. I varied parameters across each model to improve their scores

1.2 PURPOSE

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some advanced data mining techniques are used. In this study, an effective heart disease prediction system (EHDPS) is developed using neural network for predicting the risk level of heart disease. The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction. The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, eg, relationships between medical factors related to heart disease and patterns, to be established. We have employed the multilayer perceptron neural network with backpropagation as the training algorithm. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases. Medical science has made excellent use of technologicalbreakthroughs to raise the standard of healthcare These technological developments have opened the path for precise illness diagnosis and prognosis Machine learning might be a great option for you obtain a high level of accuracywhen it comes to forecasting heart illnesses with the help of algorithms.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEM

Before we did the experiments, we did research on how people explored heart disease prediction so that we can broaden our horizons and learn from them. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. Thus, this paper presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verify that Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

network to detect heart diseases

2.2 REFERENCES

- https://sciencehub.novonordisk.com/congresses.html?cid=pse-2104419060&s_kwcid=AL!10025!3!629038015560!e!!g!!articles%20on%20cardiovascular%20disease&gclid=EAIaIQobChMIw-XSpqy6-wIVVpVLBR0tpQlMEAAYASAAEgK9cfD_BwE HeartAttack/Warning-Signs-of-aHeartAttack_UCM_002039_Article.jsp#.
- achaheart.org/your-heart/educational-qas/living-with-chd/exercise-and-adult-congenital-heart-disease/?gclid=EAIaIQobChMIw-XSpqy6-wIVVpVLBR0tpQlMEAAYAiAAEgIMjPD_BwE
- www.who.int/cardiovascular_diseases/en/.
- https://www.texasheart.org/education/library-learning-resource-center/faq/

2.3 PROBLEM STATEMENT DEFINITION

Visulaization and prediction of Heart Disease with an Interactive Dashboard. The given data consists of records of patients with their features like

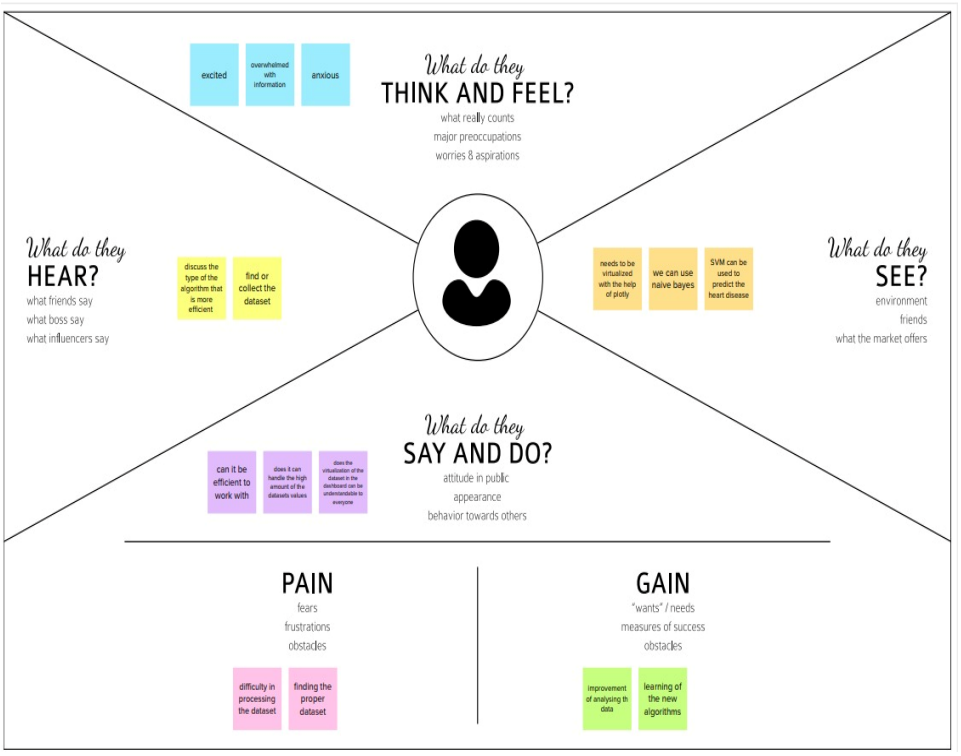
age, sex, ChestPainType, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina and Oldpeak. Each patient has unique patient Id. Your task is to predict the target variable HeartDisease. Classify each patient as either 1: for possibility of heart disease or 0: for normal condition.

CHAPTER-3

IDEATION AND PROPOSE SOLUTION

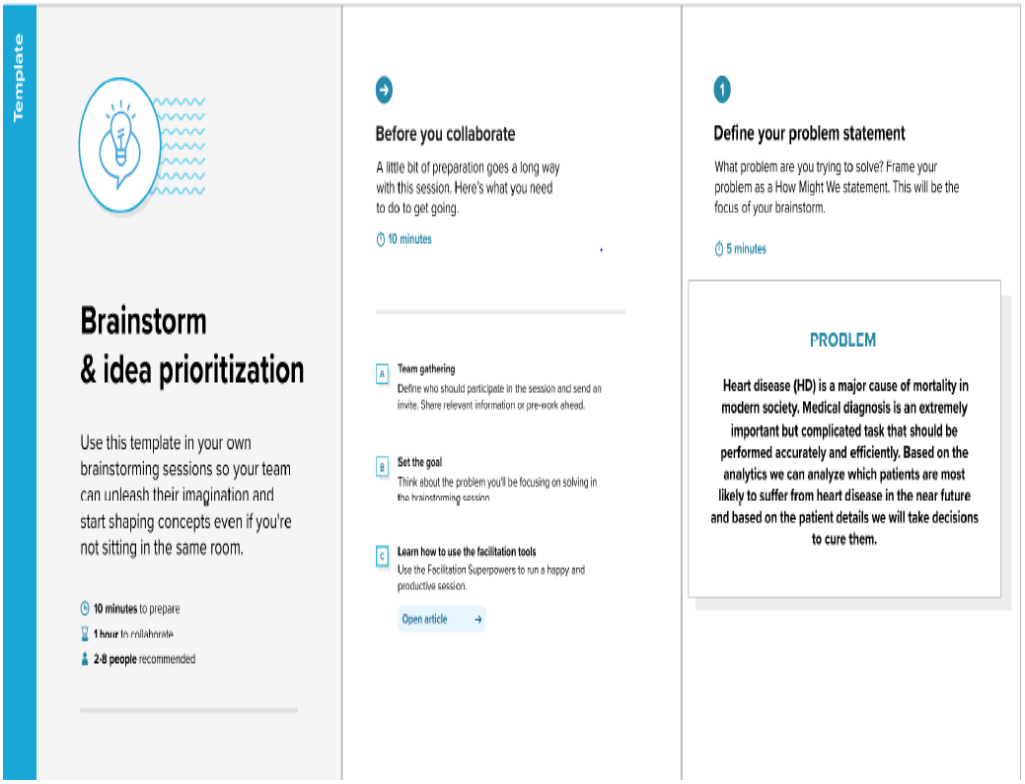
3.1 EMPATHY MAP CANVAS

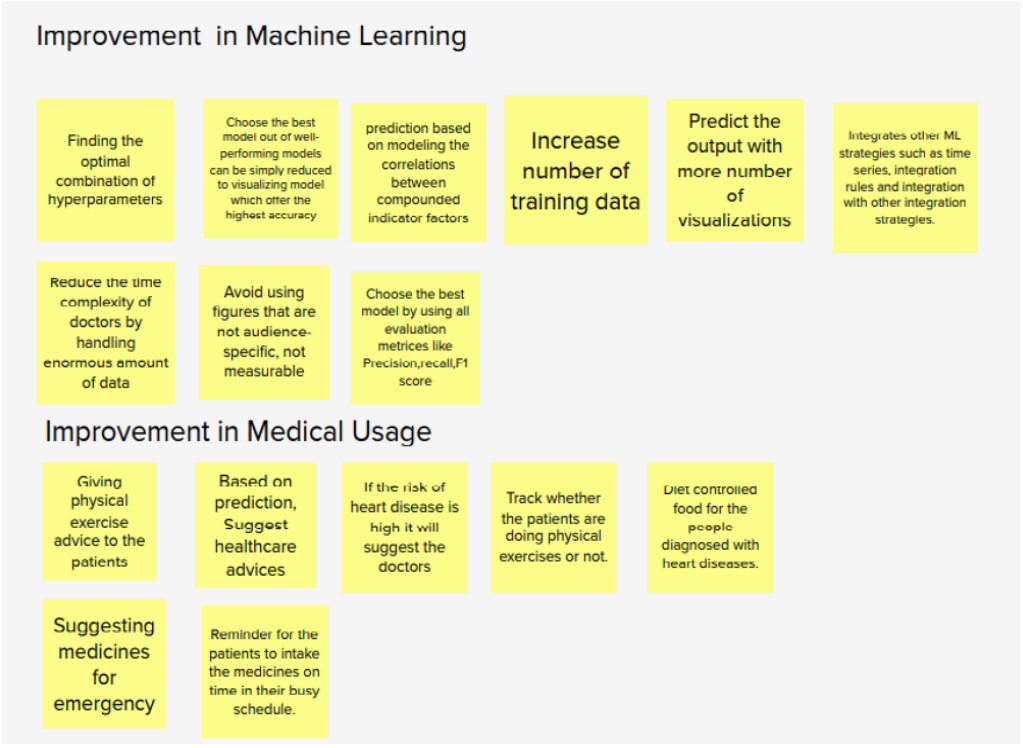
An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy map was originally created by Dave Gray and has gained much popularity within the agile community.

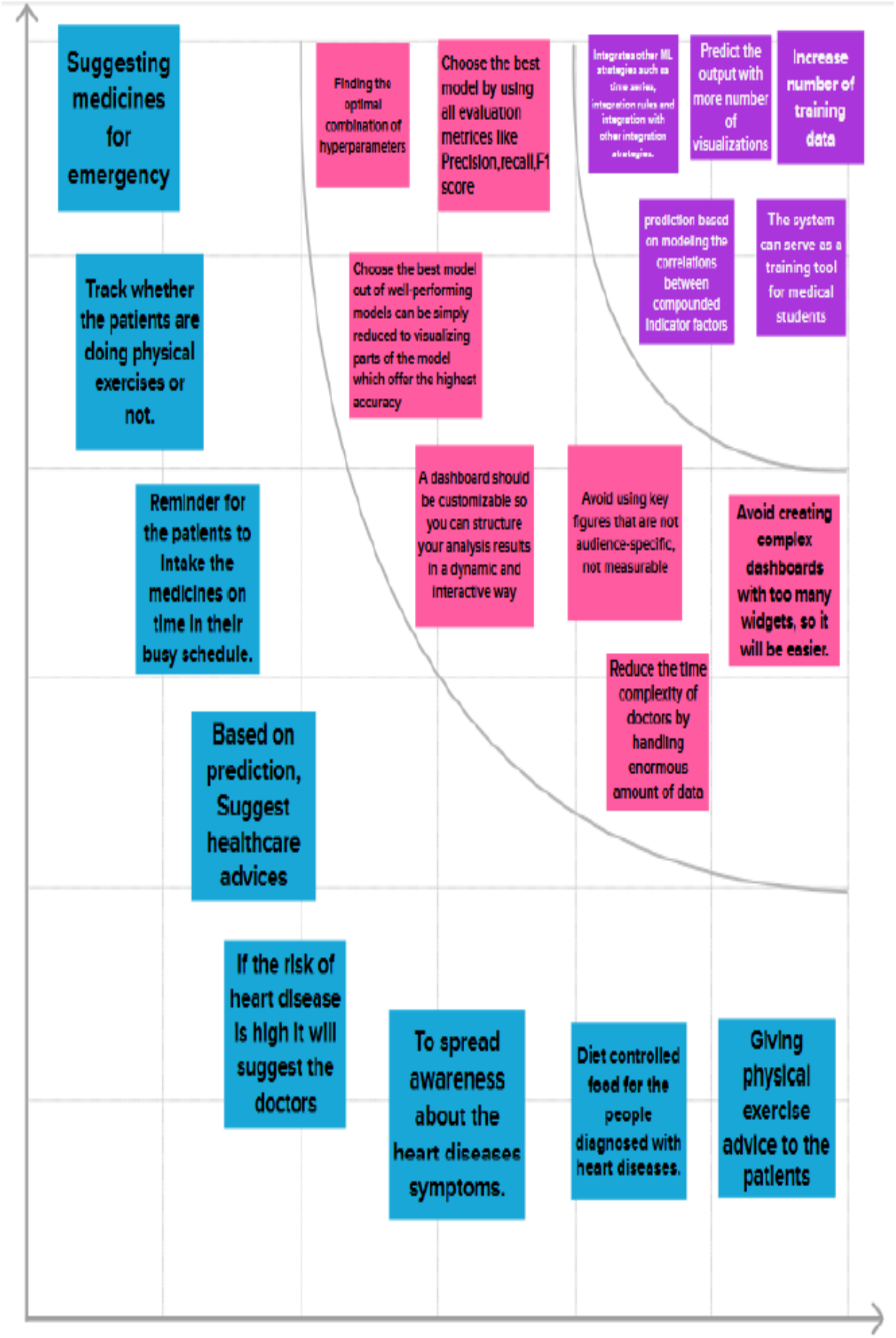


3.2 IDEATION AND BRAIN STROMING

A group problem-solving technique that involves the spontaneous contribution of ideas from all members of the group . The mulling over of ideas by one or more individuals in an attempt to devise or find a solution to a problem.


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.
10 minutes to prepare
1 hour to collaborate
2-8 people recommended
Before you collaborate
A little bit of preparation goes a long way with this session. Here's what you need to do to get going.
10 minutes
A Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.
B Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.
C Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.
Open article
1
Define your problem statement
What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.
5 minutes
PROBLEM
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. 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 take decisions to cure them.


Improvement in Machine Learning
Finding the optimal combination of hyperparameters
Choose the best model out of well-performing models can be simply reduced to visualizing model which offer the highest accuracy
prediction based on modeling the correlations between compounded indicator factors
Increase number of training data
Predict the output with more number of visualizations
Integrates other ML strategies such as time series, integration rules and integration with other integration strategies.
Reduce the time complexity of doctors by handling enormous amount of data
Avoid using figures that are not audience-specific, not measurable
Choose the best model by using all evaluation metrices like Precision,recall,F1 score
Improvement in Medical Usage
Giving physical exercise advice to the patients
Based on prediction, Suggest healthcare advices
If the risk of heart disease is high it will suggest the doctors
Track whether the patients are doing physical exercises or not.
Diet controlled food for the people diagnosed with heart diseases.
Suggesting medicines for emergency
Reminder for the patients to intake the medicines on time in their busy schedule.


Suggesting medicines for emergency
Finding the optimal combination of hyperparameters
Choose the best model by using all evaluation metrices like Precision,recall,F1 score
Integrates other ML strategies such as time series, integration rules and integration with other integration strategies.
Predict the output with more number of visualizations
Increase number of training data
prediction based on modeling the correlations between compounded indicator factors
The system can serve as a training tool for medical students
Track whether the patients are doing physical exercises or not.
Choose the best model out of well-performing models can be simply reduced to visualizing parts of the model which offer the highest accuracy
A dashboard should be customizable so you can structure your analysis results in a dynamic and interactive way
Avoid using key figures that are not audience-specific, not measurable
Avoid creating complex dashboards with too many widgets, so it will be easier.
Reminder for the patients to intake the medicines on time in their busy schedule.
Reduce the time complexity of doctors by handling enormous amount of data
Based on prediction, Suggest healthcare advices
If the risk of heart disease is high it will suggest the doctors
To spread awareness about the heart diseases symptoms.
Diet controlled food for the people diagnosed with heart diseases.
Giving physical exercise advice to the patients

3.3 PROPOSED SOLUTION

Visualizing And Predicting Heart Diseases With An Interactive Dashboard

- 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 database contains of 14 fields. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (nopresence) to 4.
- Use this dataset to predict which patients are most likely to suffer from a heart diseasein the near future using the features given.
- Training the model to predict the heart disease possibility in the future

THE DATA DICTIONARY IS AS FOLLOWS

S.NO	FIELD NAME
1	Age
2	Sex
3	Chest pain type
4	BP
5	Cholesterol
6	FBS over 120
7	EKG results
8	Max HR
9	Exercise angina

10	ST depression
11	Slope of ST
12	Number of vessels fluro
13	Thallium
14	Heart Disease

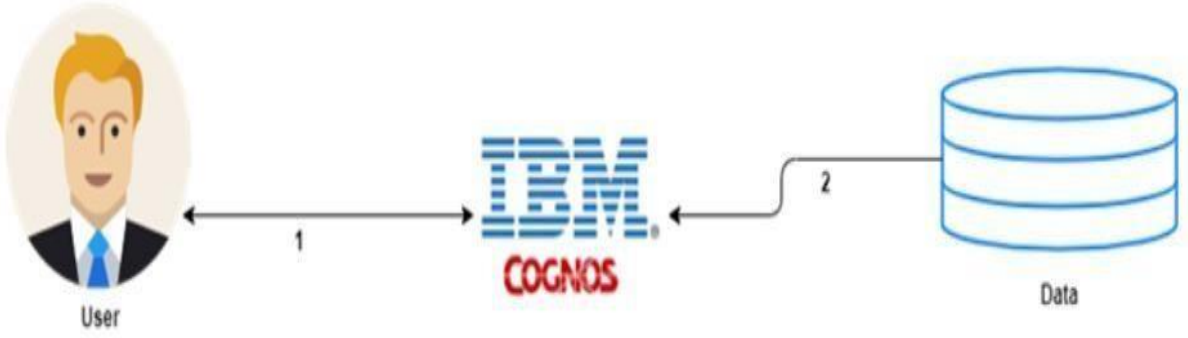
NOVELTY

- Heart diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. In this project, we have developed andresearched about models for heart disease prediction through the various heartattributes of the patient and we are going to create the interactive dashboard through which we can analyse the heart diseases based on age, sex, blood pressure of a person , etc. Dataset available publicly in Kaggle Website, further evaluating the resultsusing confusion matrix and cross-validation. And training the model using the random forest model.

FEASIBILITY OF THE IDEA

- Know fundamental concepts and can work on IBM Cognos Analytics, Gain a broad understanding of plotting different visualizations to provide a suitable solution. We consider a dataset which is having 14 fields by using that we are going to do explorations and building visualizations so that we can analyze the heart diseases of the patient.

BUSINESS MODELS :



SCALABILITY OF THE SOLUTION:

- We are going to do explorations and visualizations ,Exploration of bp versus chest pain type and gender ,Exploration of max heart rate during the chest pain, Exploration of BP by age Exploration of cholestrol by age and gender these are the explorations we are going to use.
- Average age for different chest pain types, Average exercise angina during chest pain ,BP variation with respect to age, Effort of existing heart disease on average of exercise angina , Average age for different types of chest pain in existing heart diseases , Maximum heart rate in existing heart disease by exercise angina these all are the visualizations.

3.4 PROBLEM 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	Registration for application through Gmail
FR-2	User Confirmation	User gets Confirmation via Email
FR-3	Visualizing data	User can view the Visualization of the available data
FR-4	Predicted Result	User can view the Predicted result based on the medical details

4.2 NON FUNCTIONAL REQUIREMENT

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

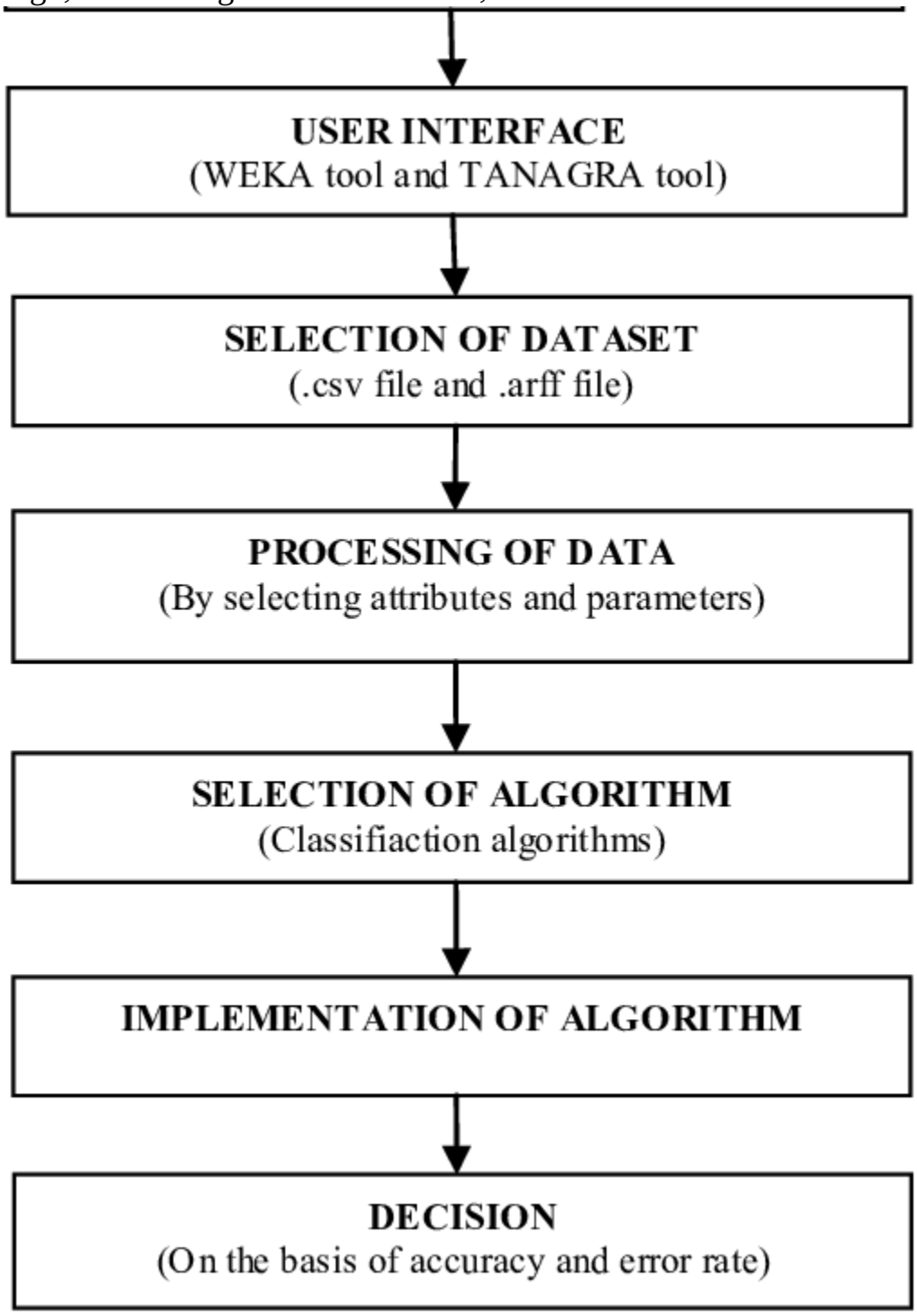
FR NO.	NON-FUNCTIONAL REQUIREMENT	DESCRIPTION
NFR-1	**Usability**	The program will have a straightforward and user-friendly graphical user interface. The application's functionality will be simple for users to comprehend and utilize.
NFR-2	**Security**	Data replication
NFR-3	**Reliability**	The application must be reliable in any environment and consistent in every scenario.
NFR-4	**Performance**	Response time and data submission speed affect how well an application performs.
NFR-5	**Availability**	The application has availability of 24x7 for users
NFR-6	**Scalability**	The program can handle an increase in the number of users.

CHAPTER-5

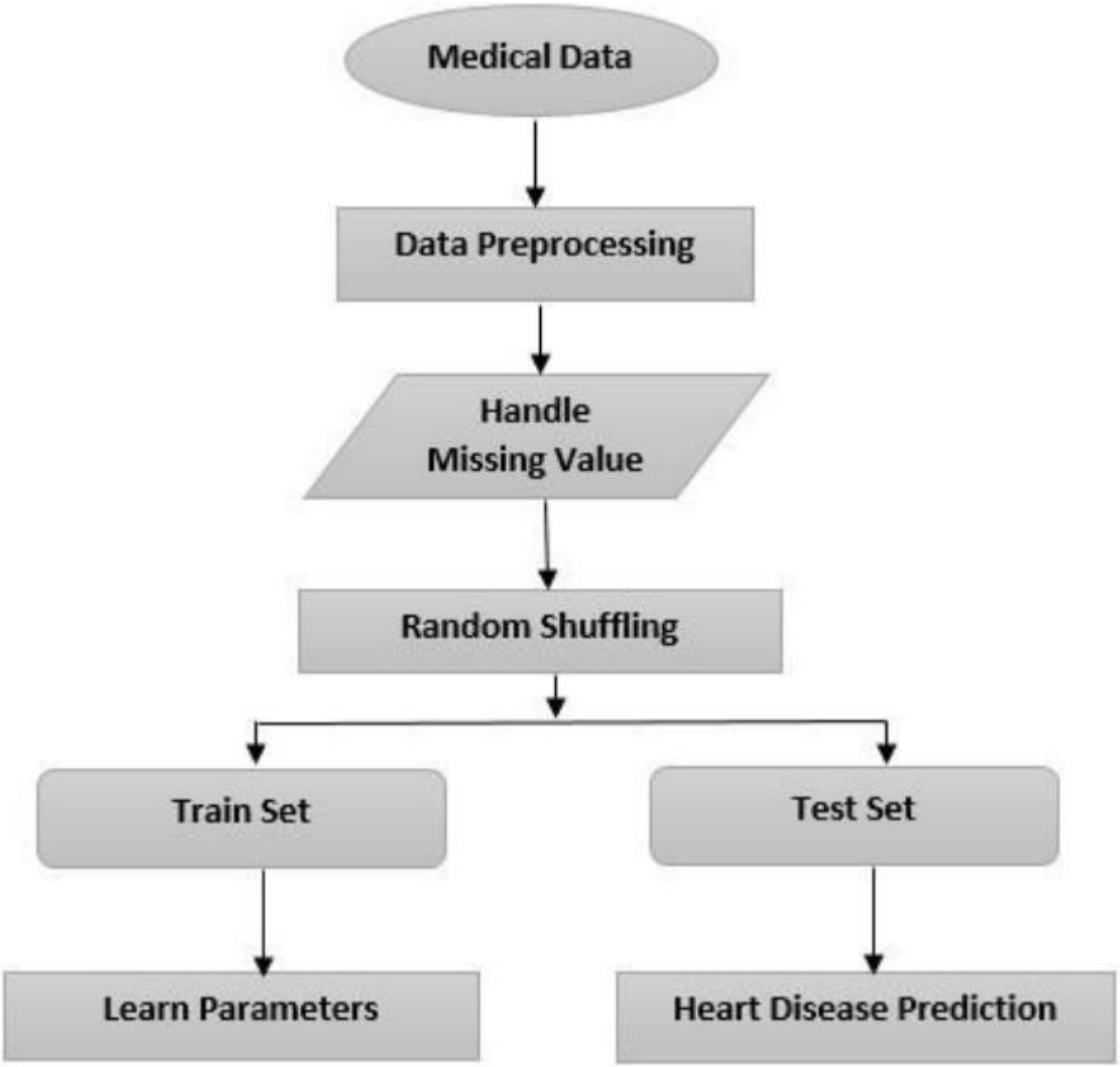
PROJECT DESIGN

5.1 DATA FLOW DIAGRAM

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



5.2 SOLUTION AND TECHNICAL ARCHITECTURE



Heart disease is the main cause of death in the developed world.Therefore, efforts must be made to reduce the likelihood of suffering a heart attack or stroke. Using the provided attributes, this dataset can identify which patients are most likely to have a heart condition in the near future. One of the leading causes of morbidity and mortality among the global population is heart disease. One of the most crucial

topics in the clinical data analysis subsection is the prediction of cardiovascular disease. In the healthcare sector, there is an enormous amount of data. The vast amount of unprocessed healthcare data is transformed via data mining into knowledge that may be used to make forecasts and educated judgments. The dataset consists of 270 individual's data. There are 14 columns in the dataset, which are described below.

IMPORT DATASET

read_csv() is used to read the CSV data with the pandas package, and then with the sklearn package we can work with some models for the prediction process .

5.3 USER STORIES

USER TYPE	FUNCTIONAL REQUIREMENT (EPIC)	USER STORY NUMBER	USER STORY / TASK
Customer (Web user)	Registration	USN-1	As a user, I can register for the application by entering my email, And password, and confirming mypassword.
		USN-2	As a user, I will receive a Confirmation email once I haveregistered for the application
	Login	USN-3	As a user, I can log into the application with the user-id and password
Customer (Web user)	Dashboard	USN-4	As a user, I can give medical data in the application
Customer care	Helpdesk	USN-5	As a staff, I can view the Customer queries
staff			
		USN-6	As a staff, I can answer the customer queries
Administrator	Profile	USN-7	As an admin, I can add or delete Users
		USN-8	As an admin, I can manage user details

ACCEPTANCE CRITERIA	PRIORITY	RELEASE
I can access my account / dashboard	High	Sprint-1
I can receive a confirmation email &click confirm	High	Sprint-1
I can access my account / dashboard after logging into the application	High	Sprint-1
I can view the visualization of trained medical data with user data	High	Sprint-2
I can post queries	Low	Sprint-3
I can get support	Low	Sprint-3
I can access the account / dashboard after logging into an application	High	Sprint-4
	High	Sprint-4

CHAPTER-6

PROJECT PLANNING & SCHEDULING

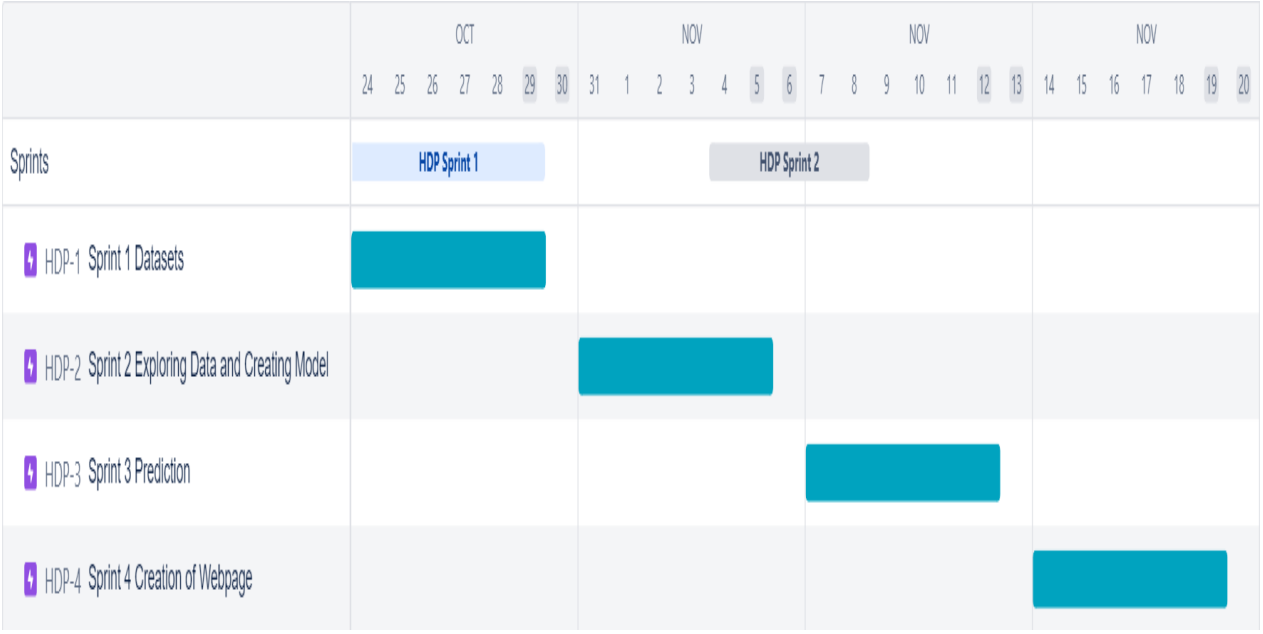
6.1 SPRINT PLANNING AND ESTIMATION

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.	2	High	M.S.Kalaiarasi V.Keerthana
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	2	High	M.S.Kalaiarasi V.Keerthana
Sprint-1		USN-3	As a user, I can register for the application through Email, Google account and mobile number	2	Medium	S.Gayathri P.Jayasri
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email &password	2	High	P.Jayasri V.Keerthana
Sprint-2	Dashboard	USN-5	As a user, I can update my profile and medical records for analysis	5	High	S.Gayathri M.S.Kalaiarasi
Sprint-2		USN-6	As a user, I can view the accuracy of occurrence of heart disease through the report generation	3	High	V.Keerthana S.Gayathri
Sprint-3	Helpdesk	USN-7	As a user, I can post my queries	3	Medium	P.Jayasri M.S.Kalaiarasi
Sprint-3		USN-8	As a customer care executive, he/she can view and answer the customer queries.	5	High	V.Keerthana S.Gayathri
Sprint-4	User profile	USN-9	As an admin, he/she can update the health details of the users	1	High	V.Keerthana
Sprint-4		USN-10	As an admin, he/she can add or delete users	2	High	P.Jayasri M.S.Kalaiarasi
Sprint-4		USN-11	As an admin, he/she can manage the user details	5	High	S.Gayathri P.Jayasri

6.2 SPRINT DELIVERY AND SCHEDULE

SPRINT	TOTAL STORY POINTS	DURATION	SPRINT START DATE	SPRINT ENDDATE (PLANNED)	STORY POINTS COMPLETED (AS ON PLANNED END DATE)	SPRINT RELEASE DATE (ACTUAL)
Sprint-1	8	6 Days	24 Oct 2022	29 Oct 2022	8	06 Nov 2022
Sprint-2	8	6 Days	31 Oct 2022	05 Nov 2022	8	09 Nov 2022
Sprint-3	8	6 Days	07 Nov 2022	12 Nov 2022	8	12 Nov 2022
Sprint-4	8	6 Days	14 Nov 2022	19 Nov 2022	8	19 Nov 2022

6.3 REPORTS FROM JIRA



CHAPTER-7

CODING AND SOLUTIONING

7.1 FEATURES 1

FRONT END

- CSS
- BOOTSTRAP

FRONT END CODE

main.py

```
<!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 Predictor</title>

  <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='style.css') }}">

  <script src="https://kit.fontawesome.com/5f3f547070.js"
crossorigin="anonymous"></script>

  <link href="https://fonts.googleapis.com/css2?family=Pacifico&display=swap"
rel="stylesheet">

</head>

<body>

  <!-- Website Title -->

  <div class="container">

    <h2 class='container-heading'><span class="heading_font">Heart Disease
Predictor</span></h2>
```

```
  <div class='description'>
    <p>A Machine Learning Web Application that predicits chances of having heart Disease or not, Built with Flask and Deployed using Heroku.</p><br>
      <p>(Note:This model is 82.67% accurate)</p>
  </div>
 </div>
 <!-- Text Area -->
 <div class="ml-container">
  <form action="{{ url_for('predict')}}" method="POST">
<label for="age">Age</label>
      <input type="text" id="age" name="age" placeholder="Your age.."><br>
      <label for="sex">Sex</label>
      <select id="sex" name="sex">
        <option selected>----select option----</option>
        <option value="1">Male</option>
        <option value="0">Female</option>
      </select><br>
      <label for="cp">Chest Pain Type</label>
      <select id="cp" name="cp">
        <option selected>----select option----</option>
        <option value="1">Typical Angina</option>
        <option value="2">Atypical Angina</option>
        <option value="3">Non-anginal Pain</option>
        <option value="4">Asymtomatic</option>
      </select><br>
```

```
<label for="trestbps">Blood Pressure</label>

<input type="text" id="trestbps" name="trestbps" placeholder="A number in range [94-200] mmHg"><br>

<label for="chol">Cholesterol</label>

<input type="text" id="chol" name="chol" placeholder="A number in range [126-564] mg/dl"><br>

<label for="fbs">Fasting Blood Sugar</label>

<select id="fbs" name="fbs">

  <option selected>----select option----</option>

  <option value="1">Greater than 120 mg/dl</option>

  <option value="0">Less than 120 mg/dl</option>

</select><br>

<label for="restecg">EKG Results</label>

<select id="restecg" name="restecg">

  <option selected>----select option----</option>

  <option value="0">Normal</option>

  <option value="1">Having ST-T wave abnormality</option>

  <option value="2">Probable or definite left ventricular

    hypertrophy</option>

</select><br>

<label for="thalach">Max Heart Rate </label>

<input type="text" id="thalach" name="thalach" placeholder="A number in range [71-202] bpm"><br>

<label for="exang">Exercise-induced Angina</label>

<select id="exang" name="exang">

  <option selected>----select option----</option>
```

```
        <option value="1">Yes</option>
        <option value="0">No</option>
      </select><br>
      <label for="oldpeak">ST depression</label>
      <input type="text" id="oldpeak" name="oldpeak" placeholder="ST depression, typically in [0-6.2]"><br>
      <label for="slope">slope of the peak exercise ST segment</label>
      <select id="slope" name="slope">
        <option selected>----select option----</option>
        <option value="1">Upsloping</option>
        <option value="2">Flat</option>
        <option value="3">Downsloping</option>
      </select><br>
      <label for="ca">Number of Major vessels</label>
      <input type="text" id="ca" name="ca" placeholder="Typically in [0-3]"><br>
       <label for="thal">Thalassemia</label>
      <select id="thal" name="thal">
        <option selected>----select option----</option>
        <option value="3">Normal</option>
        <option value="6">Fixed Defect</option>
        <option value="7">Reversible Defect</option>
      </select><br>
    <input type="submit" class="my-cta-button" value="Predict">
  </form>
</div>
```

```
  <!-- Footer -->
  <div class='footer'>
    <div class="contact">
      <a target="_blank" href="https://github.com/asthasharma98/Heart-Disease-Prediction-Deployment"><i
          class="fab fa-github fa-lg contact-icon"></i></a>
      <a target="_blank" href="https://www.linkedin.com/in/astha-sharma-47266b11b/"><i
          class="fab fa-linkedin fa-lg contact-icon"></i></a>
    </div>
    <p class='footer-description'>Made by Astha Sharma.</p>
  </div>
</body>
</html>
```

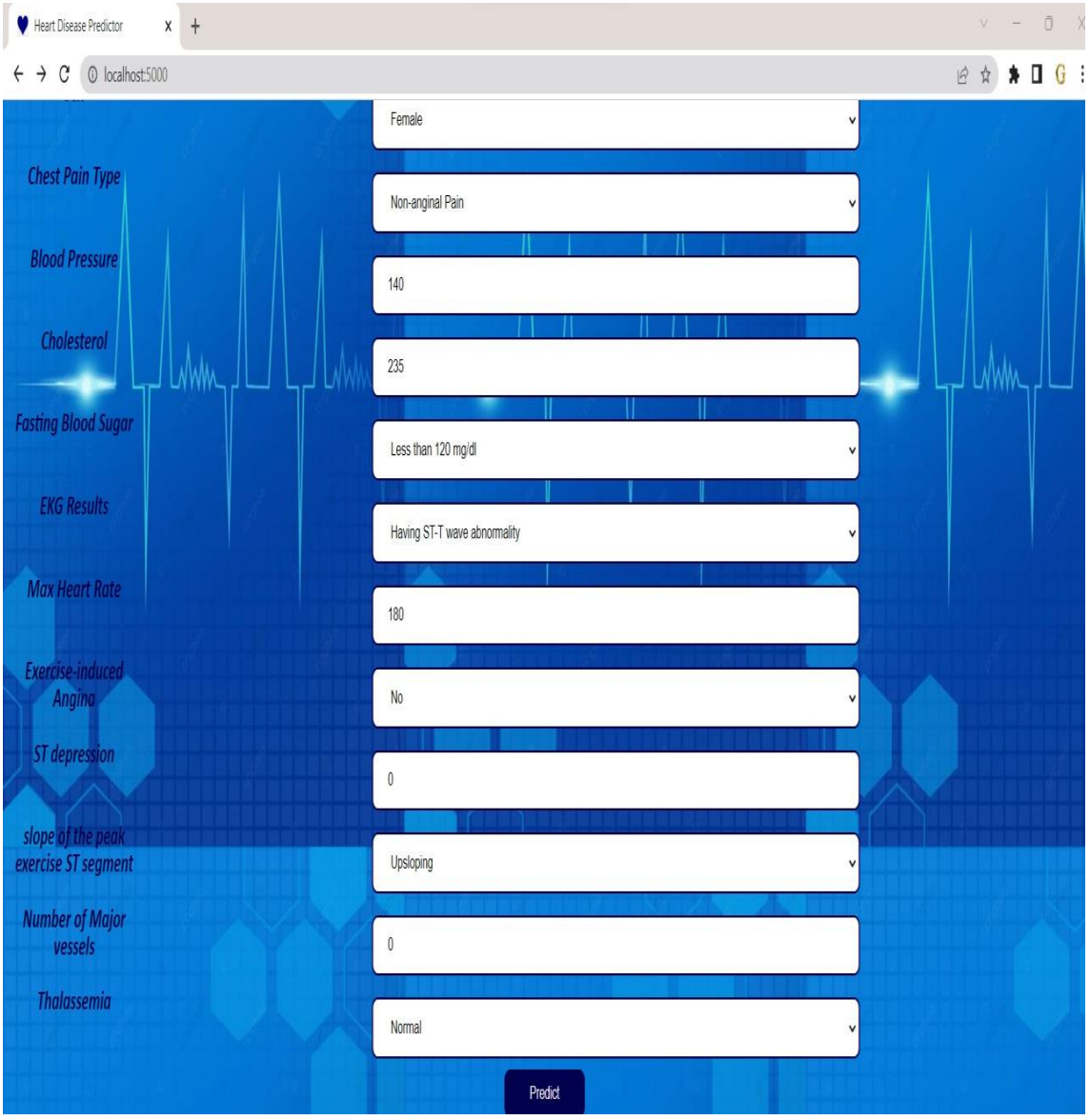
result.py

```
<!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 Predictor</title>
                <link rel="shortcut icon" href="{{ url_for('static', filename='diabetes-favicon.ico') }}">
                <link rel="stylesheet" type="text/css" href="{{ url_for('static', filename='style.css') }}">
                <script src="https://kit.fontawesome.com/5f3f547070.js" crossorigin="anonymous"></script>
```

```
            <link
href="https://fonts.googleapis.com/css2?family=Pacifico&display=swap"
rel="stylesheet">

        </head>

    <body>

        <!-- Website Title -->

            <div class="container">

          <h2 class='container-heading'><span class="heading_font">Heart Disease
Predictor</span></h2>

          <div class='description'>

                    <p>A Machine Learning Web App, Built with Flask, Deployed using
Heroku.</p>

            </div>

        </div>

            <!-- Result -->

            <div class="results">

                {% if prediction==1 %}

                    <h1>Prediction: <span class='danger'>Oops! You have Chances
of Heart Disease.</span></h1>

                {% elif prediction==0 %}

                    <h1>Prediction: <span class='safe'>Great! You DON'T chances
have Heart Disease.</span></h1>

                {% endif %}

            </div>

            <!-- Footer -->

      <div class='footer'>

          <div class="contact">
```

```
            <a target="_blank" href="https://github.com/asthasharma98/Heart-Disease-Prediction-Deployment"><i class="fab fa-github fa-lg contact-icon"></i></a>

            <a target="_blank" href="https://www.linkedin.com/in/astha-sharma-47266b11b/"><i class="fab fa-linkedin fa-lg contact-icon"></i></a>

        </div>

        <p class='footer-description'>Made by Astha Sharma.</p>

    </div

        </body>

</html>
```

SCREENSHOTS



7.2 FEATURE 2

BACK END CODE:

App.py

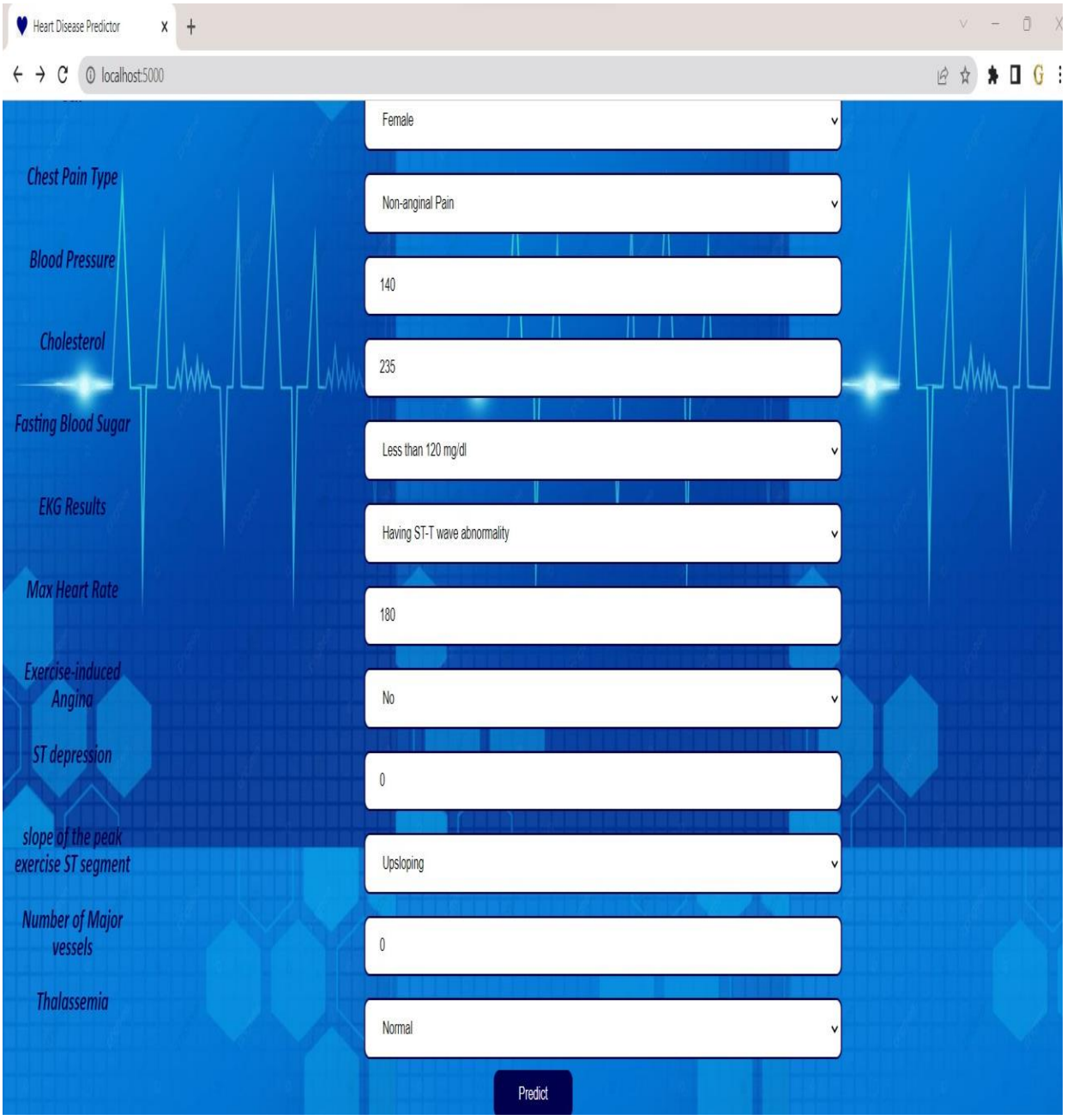
```
# Importing essential libraries
from flask import Flask, render_template, request
import pickle
import numpy as np

# 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'])
thal = request.form.get('thal')
data =
np.array([[age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal]])
my_prediction = model.predict(data)
return render_template('result.html', prediction=my_prediction)
if __name__ == '__main__':
app.run(debug=False)
```

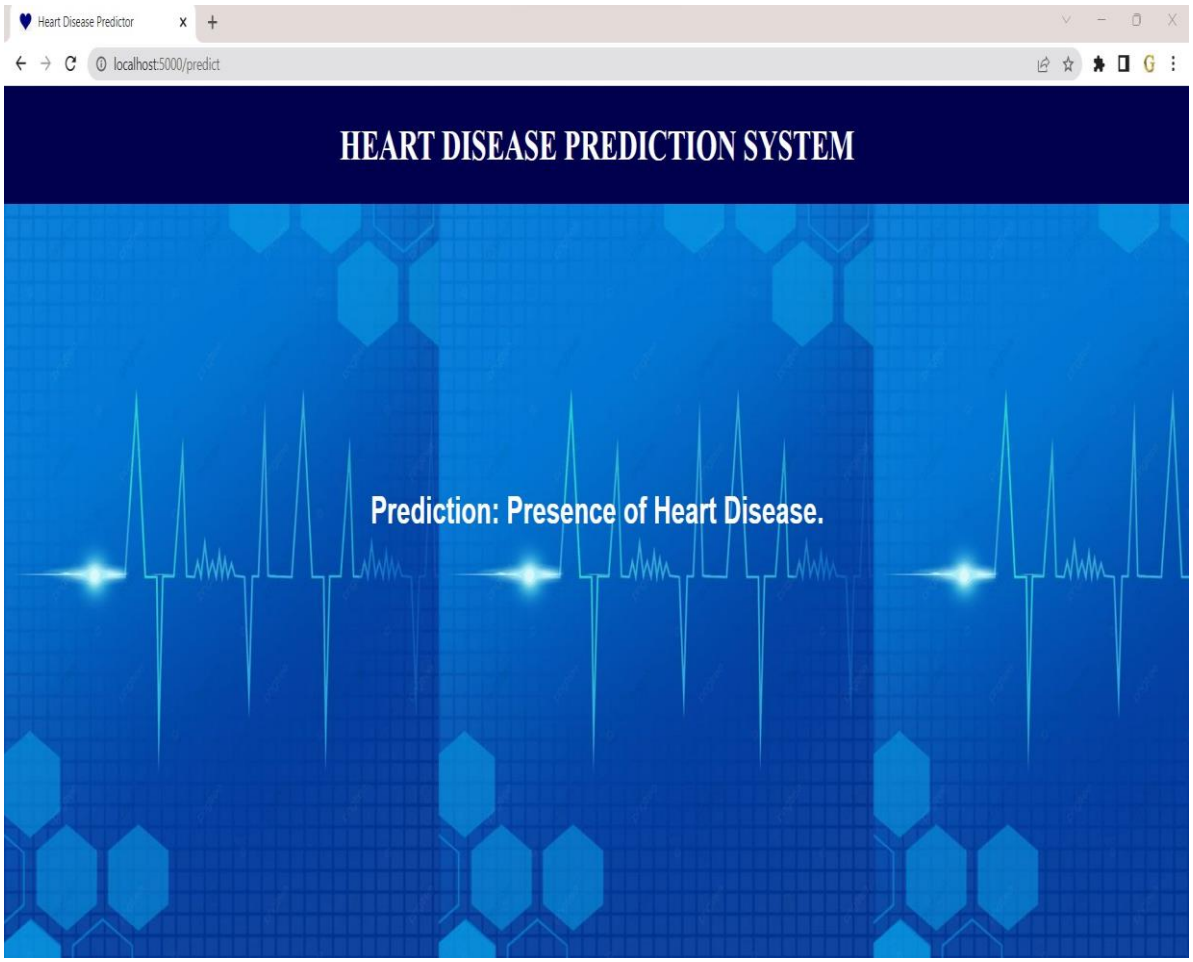
prediction.py

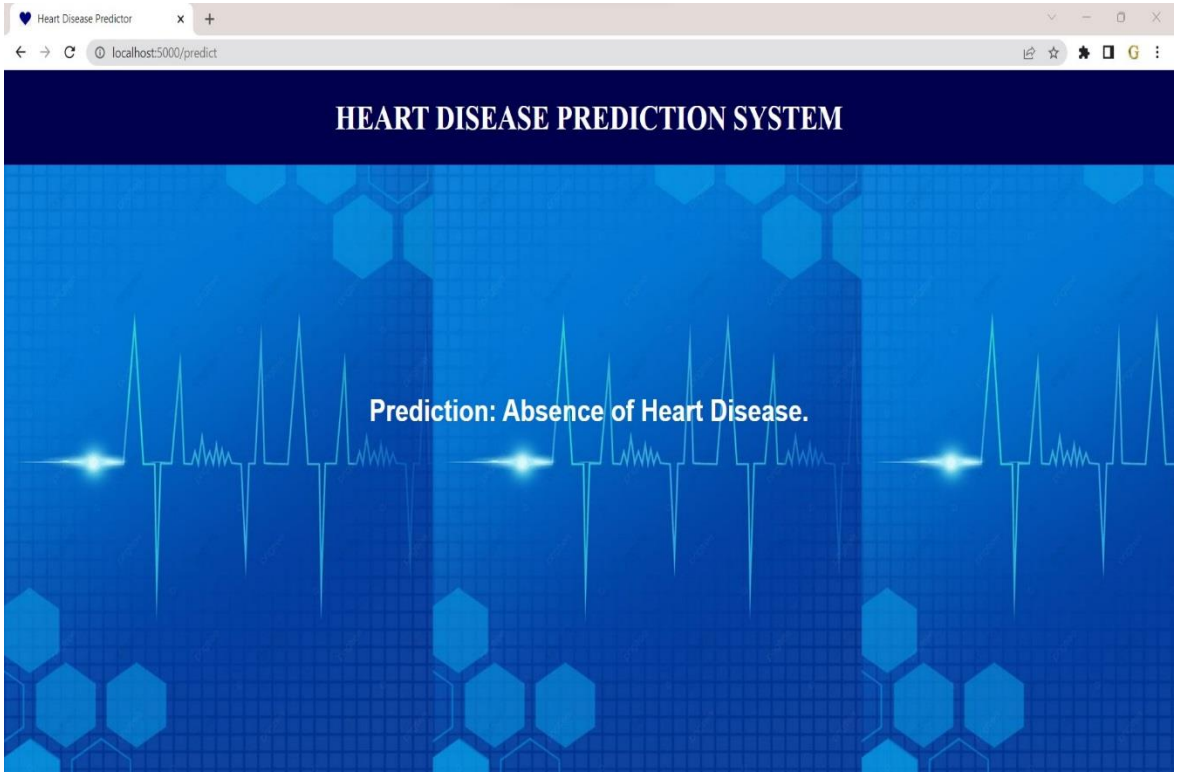
```
# importing required libraries
import numpy as np
import pandas as pd
import pickle
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier

# loading and reading the dataset

heart = pd.read_csv("Heart_Disease_Prediction.csv")

# creating a copy of dataset so that will not affect our original dataset.
heart_df = heart.copy()

# Renaming some of the columns
heart_df = heart_df.rename(columns={'Heart Disease':'target'})
print(heart_df.head())

x= heart_df.drop(columns= 'target')
y= heart_df.target

x_train, x_test, y_train, y_test= train_test_split(x, y, test_size= 0.25, random_state=42)

#feature scaling
scaler= StandardScaler()
x_train_scaler= scaler.fit_transform(x_train)
x_test_scaler= scaler.fit_transform(x_test)

# creating K-Nearest-Neighbor classifier
model=RandomForestClassifier(n_estimators=20)
model.fit(x_train_scaler, y_train)
y_pred= model.predict(x_test_scaler)
p = model.score(x_test_scaler,y_test)
print(p)

print('Classification Report\n', classification_report(y_test, y_pred))
```

```
print('Accuracy: {}%\n'.format(round((accuracy_score(y_test, y_pred)*100),2)))

cm = confusion_matrix(y_test, y_pred)
print(cm)

# Creating a pickle file for the classifier
filename = 'heart-disease-prediction-knn-model.pkl'
pickle.dump(model, open(filename, 'wb'))
```

SCREENSHOTS




Heart Disease Predictor
localhost:5000/predict
HEART DISEASE PREDICTION SYSTEM
Prediction: Presence of Heart Disease.


Heart Disease Predictor
localhost:5000/predict
HEART DISEASE PREDICTION SYSTEM
Prediction: Absence of Heart Disease.

CHAPTER-8
TESTING

8.1 TEST CASES

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and coding. In fact, testing is the one step in the software engineering process that could be viewed as destructive rather than constructive. A strategy for software testing integrates software test case design methods into a well-planned series of steps that result in the successful construction of software. Testing is the set of activities that can be planned in advance and conducted systematically. The underlying motivation of programtesting is to affirm software quality with methods that can economically and effectively apply to both strategic to both large and small-scale systems

CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

Performance Analysis is the one used for predicting the algorithm based on various metrics such as accuracy, precision, recall/F1-Score, etc. Performance Analysis aims at comparing the accuracy and performance of machine learning models. The metrics is evaluated with four measures.

TP [True Positive]

If the positive input (from dataset) is given to the classifier, it gives positive output (predicted value). It predicts the total true positive cases identified correctly. True positive values will be 1(True) for heart failured patients. The predicted value will be value 1 if the true positive value is 1.

TN [True Negative]

If the negative input (from dataset) is given to the classifier, it gives negative output (predicted value). It predicts the total true negative cases identified correctly. The true negative values will be 0(False) for heart failured patients. The predicted value will be 1(True) if the true negative value is 0(false).

FP [False Positive]

If the negative input (from dataset) is given to the classifier, it gives positive output (predicted value). It predicts the total false positive cases identified incorrectly. The false positive values will be 0 for heart failured patients. The predicted value will be 1 if the false positive value is 0.

FN [False Negative]

If the positive input (from dataset) is given to the classifier, it gives negative output (predicted value). It predicts the total false negative cases identified incorrectly. The false negative values will be 0 for heart failured patients. The predicted value will be 0 if the values predicted are false.The important metrics for performance analysis are accuracy, recall (F1-Score) and precision. These above measures are used to define the metric.

Accuracy(A)

Accuracy is a performance metric that has the correct predictions for the test data. It gives the percentage of correct predictions for testing the data. In machine learning, accuracy is calculated using the formula as shown in equation (1),

$$(TP+TN)/(TP+TN+FP+FN)$$

Recall (R) or F1-Score

Recall metric is used to predict the number of correct samples (all samples identified as positive). It is the fraction of values (results) returned to the total number of values that can be returned. It is the ratio between true positives and all the actual positives. It measures the correctly identified positive samples out of all the actual positive samples. The recall is calculated using the formula as shown in equation (2),

$$TP/(TP+FN)$$

Precision (P)

Precision which is also called as “positive predictive value”. It gives the percentage of true positives. It is the ratio between true positives and all the predicted positives. It measures the correctly identified positive samples out of all positively predicted samples. The precision is calculated using the formula as shown in equation (3),

$$TP/(TP+FP)$$

CHAPTER-10

ADVANTAGES AND DISADVANTAGES

ADVANTAGES

- Increased accuracy for effective heart disease diagnosis.
- Handles roughest(enormous) amount of data using random forest algorithm and feature selection.
- Reduce the time complexity of doctors.
- Cost effective for patients.

DISADVANTAGES

- Prediction of cardiovascular disease results is not accurate.
- Data mining techniques do not help to provide effective decision making.
- Cannot handle enormous datasets for patient records.

CHAPTER-11

CONCLUSION

we proposed a method for heart disease prediction using machine learning techniques, these results showed a great accuracy standard for producing a better estimation result. By introducing new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heartrate are to be taken at different stages of the same subjects, we find the information from the above input via ML Techniques. Firstly, we introduced a support vector classifier based on datasets.

CHAPTER-12

FUTURE SCOPE

The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks. Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions

CHAPTER 13

APPENDIX

13.1 SOURCE CODE

https://github.com/IBM-EPBL/IBM-Project-23859-1659932243/tree/main/Final%20Deliverables/Final%20Code

13.2 GITHUB AND PROJECT DEMO LINK

GITHUB

https://github.com/IBM-EPBL/IBM-Project-23859-1659932243

PROJECT DEMO LINK

https://github.com/IBM-EPBL/IBM-Project-23859-1659932243/tree/main/Final%20Deliverables/Demo%20Video