A CAL Project Report

on

<< HEALTH STATUS (HEART) >>

to be submitted in partial fulfilling of the requirements for the course on

FUNDAMENTALS OF DATA ANALYTICS

by

<<Shivani G>> <<20BCS0121>>

<<Keerthana K>> <<20BCS0160>>

FALL Semester 2022-2023

ABSTRACT

The heart rate of a person is able to tell whether they are healthy. A heart-rate is able to measure or record the heart rate of a person in real time, whether it is an electrocardiogram (ECG) . In this work, a microprocessor system loaded with a heart-rate algorithm is implemented.

INTRODUCTION

A Heart-rate that measures or records the heart rate of a person in real time. A heart rate reveals the health condition of a person. This device performs non-invasive heart-rate monitoring by measuring physiological signals, using (ECG) . Electrocardiography measures the electrical activities of the heart by using electrodes placed on the person's body for a given period of time.

REQUIREMENT GATHERING:

Tools :

For application development, the following Software Requirements are:

Operating System: Windows 7

Language: Html and CSS

Software requirements:

Operating System : Any OS with clients to access the internet

Network : Wi-Fi Internet or cellular Network

Github : Versioning Control

Google Chrome: Medium to find reference to do system testing, display and run .

Hardware Requirements

For application development, the following Software Requirements are:

Processor : Intel or high

RAM: 1024 MB

Space on disk: minimum 100mb

For running the application: Device: Any device that can access the internet Minimum space to execute: 20 MB

The effectiveness of the proposal is evaluated by conducting experiments with a cluster formed by 3 nodes , configured with an Intel CORE™ i7-4770 processor (3.40GHZ, 4 Cores, 8GB RAM, running Ubuntu 18.04 LTS with 64-bit Linux 4.31.0 kernel)

Literature Review

Numerous studies have been done that have focus on diagnosis of heart disease. They have applied different data mining techniques for diagnosis & achieved different probabilities for different methods. (Polaraju, Durga Prasad, & Tech Scholar, 2017) proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. (Deepika & Seema, 2017) focuses on techniques that can predict chronic

disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine (SVM) and Artificial Neural Network (ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

(Science & Faculty, 2009) suggested heart disease prediction using data mining and machine learning algorithm. The goal of this study is to extract hidden patterns by applying data mining techniques. The best algorithm J48 based on UCI data has the highest accuracy rate compared to LMT. (Purushottam, Saxena, & Sharma, 2016) proposed an efficient heart disease prediction system using data mining. This system helps medical practitioner to make effective decision making based on the certain parameter. By testing and training phase a certain parameter, it provides 86.3% accuracy in testing phase and 87.3% in training phase.

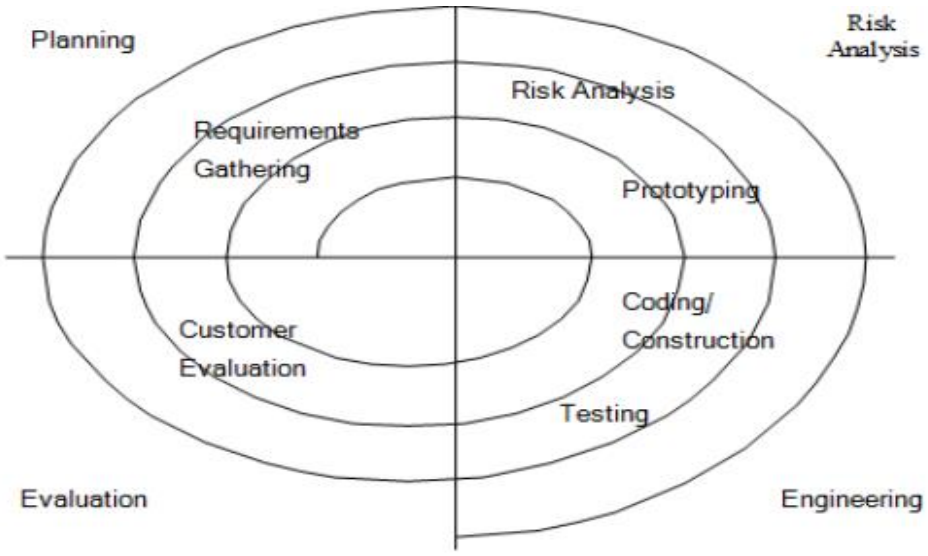
Problem Identification:

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in

the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

System Development Methodology :

The methodology of software development is the method in managing project development. There are many models of the methodology are available such as Waterfall model model, Incremental model, RAD model, Agile model, Iterative model and Spiral model. However, it still need to be considered by developer to decide which is will be used in the project. The methodology model is useful to manage the project efficiently and able to help developer from getting any problem during time of development. Also, it help to achieve the objective and scope of the projects. Methodology provides a framework for undertaking the proposed DM modeling. The methodology is a system comprising steps that transform raw data into recognized data patterns to extract knowledge for users.



There are four phases that involve in the spiral model:

1**) Planning phase**: Phase where the requirement are collected and risk is assessed. This phase where the title of the project has been discussed with project supervisor. From that discussion, Heart Prediction System has been proposed. The requirement and risk was assessed after doing study on existing system and do literature review about another existing research.

2**) Risk analysis Phase**: Phase where the risk and alternative solution are identified. A prototype are created at the end this phase. If there is any risk during this phase, there will be suggestion about alternate solution.

3) **Engineering phase**: At this phase, a software are created and testing are done at the end this phase.

4**) Evaluation phase** :At this phase, the user do evaluation toward the software. It will be done after the system are presented and the user do test whether the system meet with their expectation and requirement or not. If there is any error, user can tell the problem about system.

Dataset Explanation

The Heart Dataset selected for this project comes from the UCI Machine Learning Repository. The dataset consists of patient's data, which describe the individual's health factors and diagnosis of heart disease. The 12 health factors in the dataset used in this project are outlined below.

1. Age — age of the patient in years

2. Sex — sex of the patient

0 indicating Female

1 indicating Male

3. CP — chest pain type of the patient

1 indicating typical angina

2 indicating atypical angina

3 indicating non-anginal pain

4 indicating an asymptomatic patient

4. TrestBps — resting blood pressure in mmHg

5. Chol — serum cholesterol in mg/dl

6. Fbs — fasting blood sugar

7. RestEcg — resting electrocardiographic results

0 indicating normal

1 indicating having ST-T wave abnormality

2 indicating probable or definite left ventricular hypertrophy

8. Thalach — maximum heart rate achieved

9. Exang — exercised induced angina

0 indicating no

1 indicating yes

10. Oldpeak — ST depression induced by exercise relative to rest

11. Slope — the slope of the peak exercise ST segment

1 indicating upsloping

2 indicating flat

3 indicating downsloping

12. Cardio — diagnosis of heart disease

0 indicating absence

1 indicating presence

For the analysis and predictive modeling, the data was processed such that groupings were used for the age, resting blood pressure, and serum cholesterol factors.

CODING:

```
        html lang="en">
 <head>
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <meta http-equiv="X-UA-Compatible" content="ie=edge">
 <!--  <link rel="stylesheet" href="styles.css">-->
   <title>Heart Disease Prediction</title>
 </head>
 <style>
   @import
url('https://fonts.googleapis.com/css2?family=Montserrat:wght@500&display=sw
ap');
 *{
   margin: 0;
   padding: 0;
   box-sizing: border-box;
   font-family: 'Montserrat', sans-serif;
```

```
  color:black;
 }
 body{
  background: url("https://cdn.pixabay.com/photo/2016/08/10/20/26/stethoscope-1584223_1280.jpg") no-repeat top center;
  padding: 0 10px;
  background-size: cover;
 }
 .wrapper{
  max-width: 500px;
  width: 100%;
  background: rgba(0, 0, 0, 0.7);
  margin: 20px auto;
  padding: 30px;
  box-shadow: 1px 1px 2px rgba(0, 0, 0, 1.25);
 }
 .wrapper .title{
  font-size: 24px;
  font-weight: 700;
  margin-bottom: 25px;
```

```
  color: #fec107;
  text-transform: uppercase;
  text-align: center;
}
.wrapper .form{
  width: 100%;
}
.wrapper .form .input_field{
  margin-bottom: 15px;
  display: flex;
  align-items: center;
}
.wrapper .form .input_field label{
  width:80px;
  font: bold;
  color: wheat;
  margin-right: 10px;
  font-size: 14px;
}
.wrapper .form .input_field input{
```

```
  width: 100px;
 }
 .wrapper .form .input_field .textarea{
  resize: none;
  height: 250px;
  width: 500px;
 }
 </style>
 <body>
 <form action="{{ url_for('predict')}}" method="post">
  <div class="wrapper">
    <div class="title">
      <h1 style="color: white;">Heart Disease Prediction</h1>
    </div>
    <div class="form">
      <div class="input_field">
        <textarea style="color: white; background: rgba(0, 0, 0, 0.3);"
class="textarea" readonly>
 It's a clean, easy to understand set of data. However, the meaning of some of the column headers are not obvious. Here's what they mean,
 Age: displays the age of the individual.
```

Sex: displays the gender of the individual using the following format :

1 = male

0 = female

Chest-pain type: displays the type of chest-pain experienced by the individual using the following format :

0 = typical angina

1 = atypical angina

2 = non — anginal pain

3 = asymptotic

Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)

Serum Cholestrol: displays the serum cholesterol in mg/dl (unit)

Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl.

If fasting blood sugar > 120mg/dl then : 1 (true) else : 0 (false)

Resting ECG : displays resting electrocardiographic results

0 = normal

1 = having ST-T wave abnormality

2 = left ventricular hyperthrophy

Max heart rate achieved : displays the max heart rate achieved by an individual.

Exercise induced angina :

1 = yes

0 = no

ST depression induced by exercise relative to rest: displays the value which is an integer or float.

Peak exercise ST segment :

1 = upsloping

2 = flat

3 = downsloping

Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.

Thal : displays the thalassemia :

0 = normal

1 = fixed defect

2 = reversible defect

Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not :

0 = absence

1, 2, 3, 4 = present.

</textarea>

```
        <label>AGE</label>
        <input type="number" id="age" min="0" max="150" name="age"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(Age: 1 - 150)</i>
      </div>
      <div class="input_field">
        <label>SEX</label>
        <input type="number" min="0" max="1" name="sex" id="sex"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(Male: 1 & Female: 0)</i>
      </div>
      <div class="input_field">
        <label>CP</label>
        <input type="number" name="cp" id="cp" min="0" max="3"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(Enter Single Value From
Range 0-3)</i>
      </div>
      <div class="input_field">
        <label>TRESTBPS</label>
```

```
        <input name="trestbps" id="trestbps" type="number" min="0" step="1"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(Enter Non-Decimal
Value)</i>
      </div>
      <div class="input_field">
       <label>CHOL</label>
        <input name="chol" id="chol" type="number" min="0" step="1"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(Enter Non-Decimal
Value)</i>
      </div>
      <div class="input_field">
       <label>FBS</label>
        <input type="number" min="0" max="1" name="fbs" id="fbs"
class="input_text">
        <br/>  <i style="font-size: 10px; color: white;">(1 = True; 0 = False)</i>
      </div>
      <div class="input_field">
       <label>RESTECG</label>
        <input type="number" name="restecg" min="0" max="2" id="restecg"
class="input_text">
```


 (Enter Single Value From Range 0-2)

<label>THALACH</label>

<input name="thalach" id="thalach" type="number" min="0" step="1" class="input_text">

 (Enter Non-Decimal Value)

<label>EXANG</label>

<input type="number" min="0" max="1" name="exang" id="exang" class="input_text">

 (Exercise: 1 = YES; 0 = NO)

<label>OLDPEAK</label>

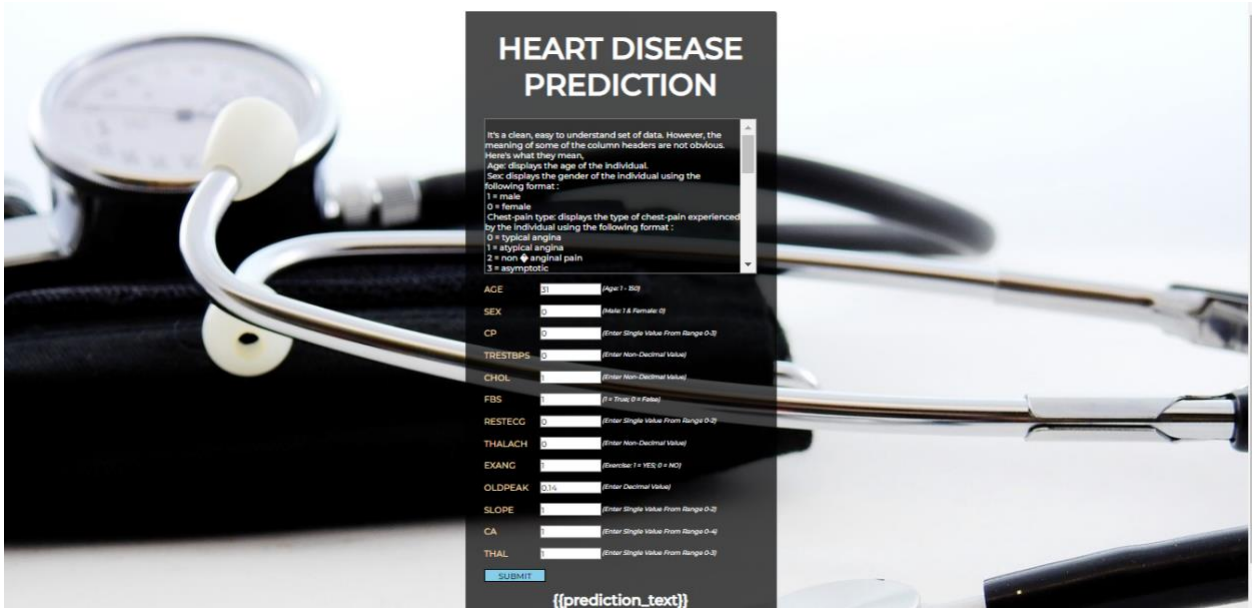
<input name="oldpeak" id="oldpeak" type="number" min="0" step="0.01" class="input_text">

 (Enter Decimal Value)

```
</div>
<div class="input_field">
 <label>SLOPE</label>
 <input type="number" min="0" max="2" name="slope" id="slope" class="input_text">
 <br/>  <i style="font-size: 10px; color: white;">(Enter Single Value From Range 0-2)</i>
</div>
<div class="input_field">
 <label>CA</label>
 <input type="number" name="ca" min="0" max="4" id="ca" class="input_text">
 <br/>  <i style="font-size: 10px; color: white;">(Enter Single Value From Range 0-4)</i>
</div>
<div class="input_field">
 <label>THAL</label>
 <input type="number" name="thal" min="0" max="3" id="thal" class="input_text">
 <br>  <i style="font-size: 10px; color: white;">(Enter Single Value From Range 0-3)</i>
</div>
```

```
        <div class="input_field">
         <input type="submit" style="color:black; background: skyblue;" id="submit" value="SUBMIT">
        </div>
        <div>
          <center>
            <h2 style="color: white; margin-bottom:10px;">
              {{prediction_text}}
            </h2>
          </center>
        </div>
      </div>
    </div>
  </form>
  </body>
  </html>
```

EXECUTION:



Data Analysis

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	cardio
age	1.000											
sex	-0.033	1.000										
cp	0.089	0.096	1.000									
trestbps	0.266	-0.003	0.029	1.000								
chol	0.130	-0.135	0.085	0.136	1.000							
fbs	0.176	0.081	-0.039	0.181	0.083	1.000						
restecg	0.149	-0.008	0.004	0.070	0.087	0.083	1.000					
thalach	-0.316	-0.122	-0.377	-0.132	-0.054	-0.036	0.084	1.000				
exang	0.092	0.214	0.468	0.159	0.082	0.021	-0.057	-0.462	1.000			
oldpeak	0.163	0.182	0.243	0.207	0.013	0.008	-0.013	-0.349	0.354	1.000		
slope	0.145	0.128	0.231	0.159	0.031	0.090	-0.002	-0.409	0.365	0.558	1.000	
cardio	0.176	0.326	0.436	0.199	0.080	0.055	0.033	-0.427	0.481	0.446	0.388	1.000

Figure 1

In the age column, trestbps (resting blood pressure) and thalach (maximum heart rate achieved) are the strongest correlated factors with age. As a patient's age increases, their resting blood pressure tends to increase and their maximum heart rate achieved

tends to decrease. The correlation between age and cardio is positive, however it is not a very strong correlation with a value.

CONCLUSION:

The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research. In DM warehouse, generally, the dimensionality of the heart database is high, so identification and selection of significant attributes for better diagnosis of heart disease are very challenging tasks for future research.