**Software Requirements Specification**

For

**Heart Disease Detection Using Machine Learning**

07/12/2021

Prepared by

|  |  |  |
| --- | --- | --- |
| **Specialization** | **SAP ID** | **Name** |
| B.Tech CSE CCVT | 500075200 | Tanisha Goyal |
| B.Tech CSE CCVT | 500075998 | Vishakha Sharma |
| B.Tech CSE CCVT | 500076548 | Vishal Gupta |



Department of Systematics

School Of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,

DEHRADUN- 248007. Uttarakhand

Table of Contents

|  |  |  |
| --- | --- | --- |
| **Topic** | | **Page No** |
| Table of Content | | 2 |
| Revision History | | 3 |
| 1 | Introduction | 4 |
|  | 1.1 Purpose of the Project | 4 |
|  | 1.2 Target Beneficiary | 4 |
|  | 1.3 Project Scope | 4 |
|  | 1.4 References | 4 |
| 2 | Project Description | 4 |
|  | 2.1 Reference Algorithm | 5 |
|  | 2.2 Data/ Data structure | 6 |
|  | 2.3 SWOT Analysis | 6 |
|  | 2.4 Project Features | 6 |
|  | 2.5 User Classes and Characteristics | 8 |
|  | 2.6 Design and Implementation Constraints | 8 |
|  | 2.7 Design diagrams | 9 |
|  | 2.8 Assumption and Dependencies | 10 |
| 3 | System Requirements | 10 |
|  | 3.1 User Interface | 10 |
|  | 3.2 Software Interface | 10 |
|  | 3.3 Database Interface | 10 |
|  | 3.4 Protocols | 10 |
| 4 | Non-functional Requirements | 10 |
|  | 4.1 Performance requirements | 10 |
|  | 4.2 Security requirements | 10 |
|  | 4.3 Software Quality Attributes | 11 |
| 5 | Other Requirements | 11 |
| Appendix A: Glossary | | 11 |
| Appendix B: Analysis Model | | 11 |
| Appendix C: Issues List | | 11 |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Change** | **Reason for Changes** | **Mentor Signature** |
| 22-11-2021 | Naïve Bayes Algorithm was selected | It deals better with big datasets |  |
| 30-11-2021 | For dataset we used XAMPP Control panel | This software is started/stopped and is used for testing the our project and can be modified offline, the most very important functionality provided by this software(XAMPP) is that it is the creation of the MySQL Database |  |

|  |  |  |
| --- | --- | --- |
| 1 | INTRODUCTION | |
|  | 1.1 Purpose of the Project | The goal of the Heart Disease Detection is to detect the disease earlier then it can be treated properly and can be kept under control. Here, the earlier detection is the main key. The main purpose behind this project is to present a heart disease model for the earlier prediction of heart disease. Further our project is aimed towards identifying the best classification of machine learning algorithms for identifying the possibility of heart disease in a patient. |
|  | 1.2 Target Beneficiary | Heart is the next major organ comparing to brain which has more priority in human body It is one of the most critical human diseases in the world and affects human life very badly. In this work, we present a real -time patient’s data where we could develop a heart disease detection system with accurate few tests and attributes present in heart disease. The main target is to predict the risk of having the heart disease with attributes and faster efficiency in the form of dataset in the patient’s report. |
|  | 1.3 Project Scope | The scope of this project is to create a system to recognize data collection, pattern matching to calculate the accuracy of the datasets. Symptoms of the heart attack can be used as an alternative prediction more accurate diagnosis which also reduces the cost of treatment using machine learning algorithms Naïve Baye’s (NB) Algorithm (Baye’s Theorem) |
|  | 1.4 References | 1. Brownlee, J. (2016). Naive Bayes for Machine Learning. Retrieved March 4, 2019, from https://machinelearningmastery.com/naive-bayes-for-machine-learning/ 2. Animesh Hazra, Arkomita Mukherjee, Amit Gupta,  Asmita Mukherjee, “Heart Disease Diagnosis and  Prediction Using Machine Learning and Data Mining  Techniques: A Review”, Research Gate Publications,  July 2017, pp.2137-2159. 3. Research.ijcaonline.org. 2021. *A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)*. [online] Available at: <https://research.ijcaonline.org/volume68/number16/pxc3887250.pdf>. |
| 2 | PROJECT DESCRIPTION | |
|  | 2.1 Reference Algorithm | All the Algorithms mentioned in the introduction will be implemented in JAVA. This language is statically typed and is a system language and thus would increase the efficiency of our implementation.  To build the algorithm, a data flow diagram and activity diagram will be used as this project is more research oriented.  As this project is practically based, so we need to see whether the dataset of a person matches with all the details and cross checked with the previous data. For Example:-   1. **Data Collection and Pre-processing:-** The data set for the research was taken from the Kaggle data repository. 10 data accessed from the Kaggle Machine Learning Repository is freely available. 2. **Pattern Matching and Prediction:-** The overall objective of our work is to predict more accurately the presence of heart disease. In this project, Kaggle repository dataset are used to get more accurate results. Data mining classification technique which was applied is Naive Bayes.   Common attributes used for heart disease are Age, Sex, Fasting Blood Pressure, Chest Pain type, Resting ECG(test that measures the electrical activity of the heart), Number of major vessels coloured by fluoroscopy, Threst Blood Pressure (high blood pressure), Serum Cholesterol(determine the risk for developing heart disease), Thala Ch (maximum heart rate achieved), ST depression (finding on an electrocardiogram, trace in the ST segment is abnormally low below the baseline), pain lock (chest pain location (substernal=1, otherwise=0)), Fasting blood sugar, Exang(exercise included angina), smoke, hypertension, food habits, weight, height and obesity where the Baye’s Theorem algorithm will work. |
|  | 2.2 Characteristic of Data | Our project focuses on real time tests of the heart patients so we have a particular dataset. (Moreover we have also attached an Excel sheet)    In this Flowchart we can see how our project dataset of the patients can work in the heart disease model . |
|  | 2.3 SWOT Analysis | **1. Strengths :**   * Earlier Detection of the heart problems * Predicts the risk of having Heart Disease with faster efficiency and attributes in the form of Kaggle Datasets.   **2. Weaknesses :**  This prediction involves several challenges, including different complex machine learning algorithms, complex background, processing time for the Heart Disease System and mainly the datasets which challenges to calculate the highest accuracy present in the heart of the patient.  **3.Opportunities :**   * Background Reports of the patients can be used as an alternative input device from which we can easily predict their heart problems by testing Electro Cardio Gram(ECG), Blood Pressure (BP), Diabetes, etc as accurate as possible. * All the Attributes in one system is calculated with faster efficiency then it can save many lives.   **4.Threats :**   * Control of High Blood Pressure and Diabetes. * Cholesterol to keep under control. * Stay at a healthy weight * Get a Regular Exercise with Healthy Diet * Do not Smoke or Drink Alcohol * Make sure of having at least 7 to 9 hours of sleep per night. |
|  | 2.4 Project Features | **CLASS DIAGRAM:**    **ACTIVITY DIAGRAM:**    CLASS and ACTIVITY use case diagram are used to describe the functionality of a particular system. The following shows the functionality of the user of the system. The role of system is always on and collects the data , analysis the test and then report the detected problems in the heart till it gives the highest accuracy of the datasets. |
|  |  |  |
|  | 2.5 User Classes and Characteristics | * Predicts the Heart rate Efficiency of the patients. * Develop a heart Disease Detection System with accurate tests. * Detect the risk of heart disease in patients. * System to provide the effective treatment report to avoid severe consequences. |
|  | 2.6 Design and Implementation Constraints | * System Performance: High Performance system and a good structure for the prediction of tests are required for analyzing the real time scenarios and perform the task assigned to them in pre-defined functions defined in the system. * Language Requirement: JAVA |
|  | 2.7 Design diagrams | **USE CASE DIAGRAM:**    **CLASS DIAGRAM** |
|  |  |  |
|  | 2.8 Assumption and Dependencies | Assumptions: we developed our model under the assumption that in the real time the accurate test will have faster efficiency in lesser time.  Dependencies: Our project is dependent on various external libraries such as Eclipse IDE, Kaggle Dataset, time, records, test cases, attributes etc. |
| 3 | SYSTEM REQUIREMENTS | |
|  | 3.1 User Interface | JAVA Runtime Environment (JRE) to use Eclipse IDE. |
|  | 3.2 Software Interface | This project uses Windows 10, Eclipse IDE for development purposes. |
|  | 3.3 Database Interface | Kaggle Dataset is used in this project. Kaggle is a machine learning practitioners platform where it allows users to find and publish data sets, explore it and then build the models in a database environment. |
|  | 3.4 Protocols | There is no as such defined protocols in the project. |
| 4 | NON-FUNCTIONAL REQUIREMENTS | |
|  | 4.1 Performance requirements | The performance shall minimize the number of calculations needed to perform pattern matching and rate of Heart Disease Detection. Each calculation shall be processed within 850 milliseconds to achieve 3 scans per second performance. It should minimize the use of Central Processing Unit(CPU) and software shall utilize less than 80% of the system’s CPU resource and less than 100 megabytes of system memory. |
|  | 4.2 Security requirements | The HDD should be made secure so that it should not breach the security of the patient’s records and also it should not collect the private information from the user system. |
|  | 4.3 Software Quality Attributes | The software shall be extensible to support future developments and add-ons to HDD software. The detection of heart disease shall be at least 94% accuracy in a dataset. |
| 5 | Other Requirements | HDD system does not require special requirements equipment except for a personal computer. The CPU of this computer should have at least two cores to handle the enormous number of calculations needed for the prediction of heart disease with faster and accurate efficiency. |
| Appendix A: Glossary | | HDD: Heart Disease Detection  Kaggle : Library used for datasets in the form of 0 and 1.  Eclipse IDE: To form a system  XAMPP Control Panel: Dataset is stored |
| Appendix B: Analysis Model | | Naïve Bayes Algorithm will be used in this project. |
| Appendix C: Issues List | | No issues |

IMPLEMENTATION

**Connection**

**import** java.sql.\*;

**public** **class** MysqlConnection {

**public** **static** Connection getConnection()

{

Connection con=**null**;

**try**

{

con=DriverManager.*getConnection*("jdbc:mysql://localhost/rec","root","");

}

**catch**(Exception e) {

e.printStackTrace();

}

**return** con;}

**public** **static** **void** main(String args[]) {

Connection con=*getConnection*();

System.***out***.println(con);

**if**(con==**null**) {

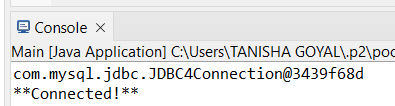
System.***out***.println("Not Connected!");

}

**else**

{

System.***out***.println("\*\*Connected!\*\*"); }



**DataFilter**

import java.util.\*;

import java.sql.\*;

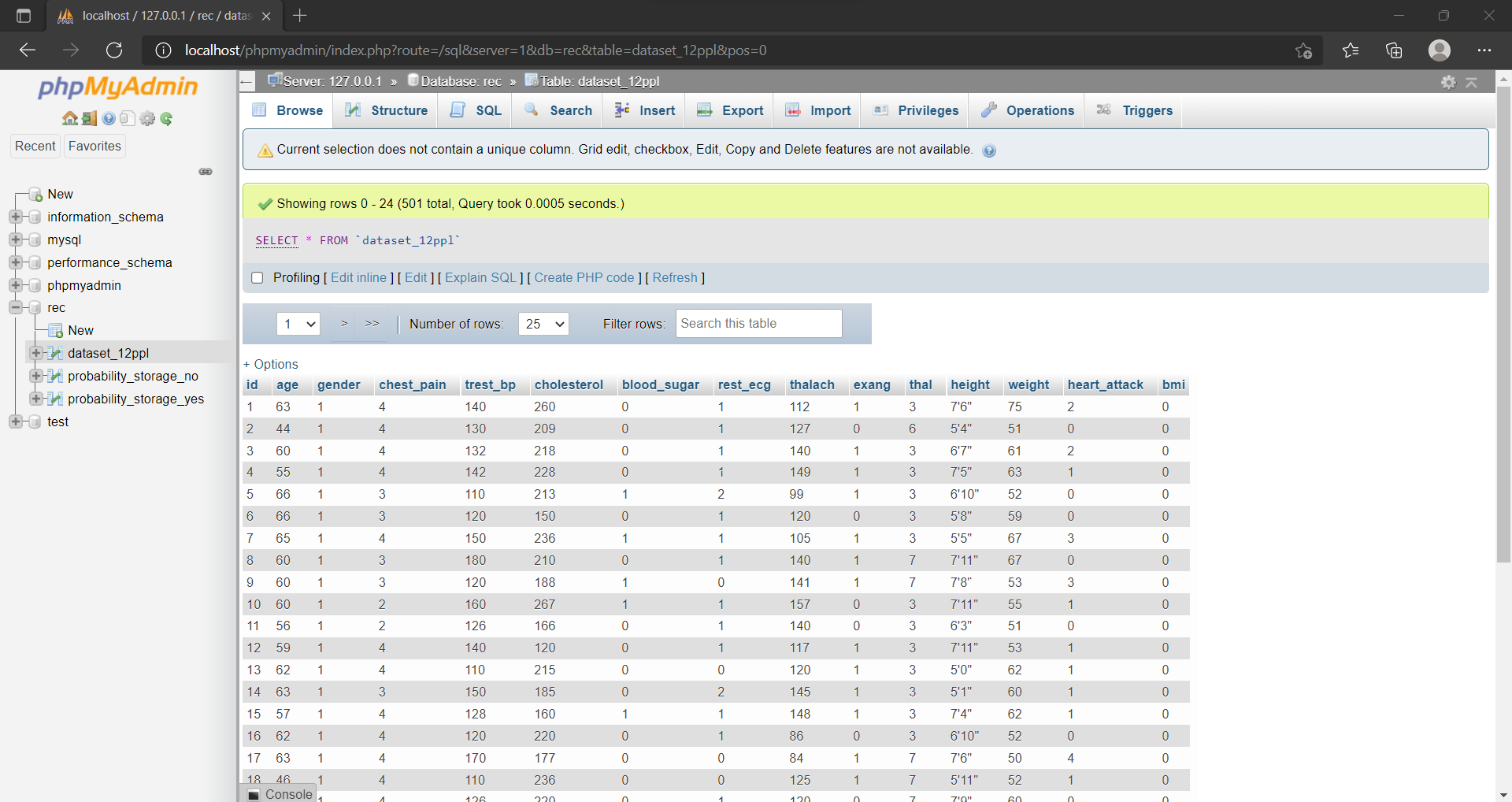
import java.util.HashMap;

class patient

{

int id,age, gender, cholesterol, trest\_bp, chest\_pain, blood\_sugar, rest\_ecg, thalach, ps,exang, thal, bmi=0, heart\_attack,value;

float height, weight;



**CalculateProbability**

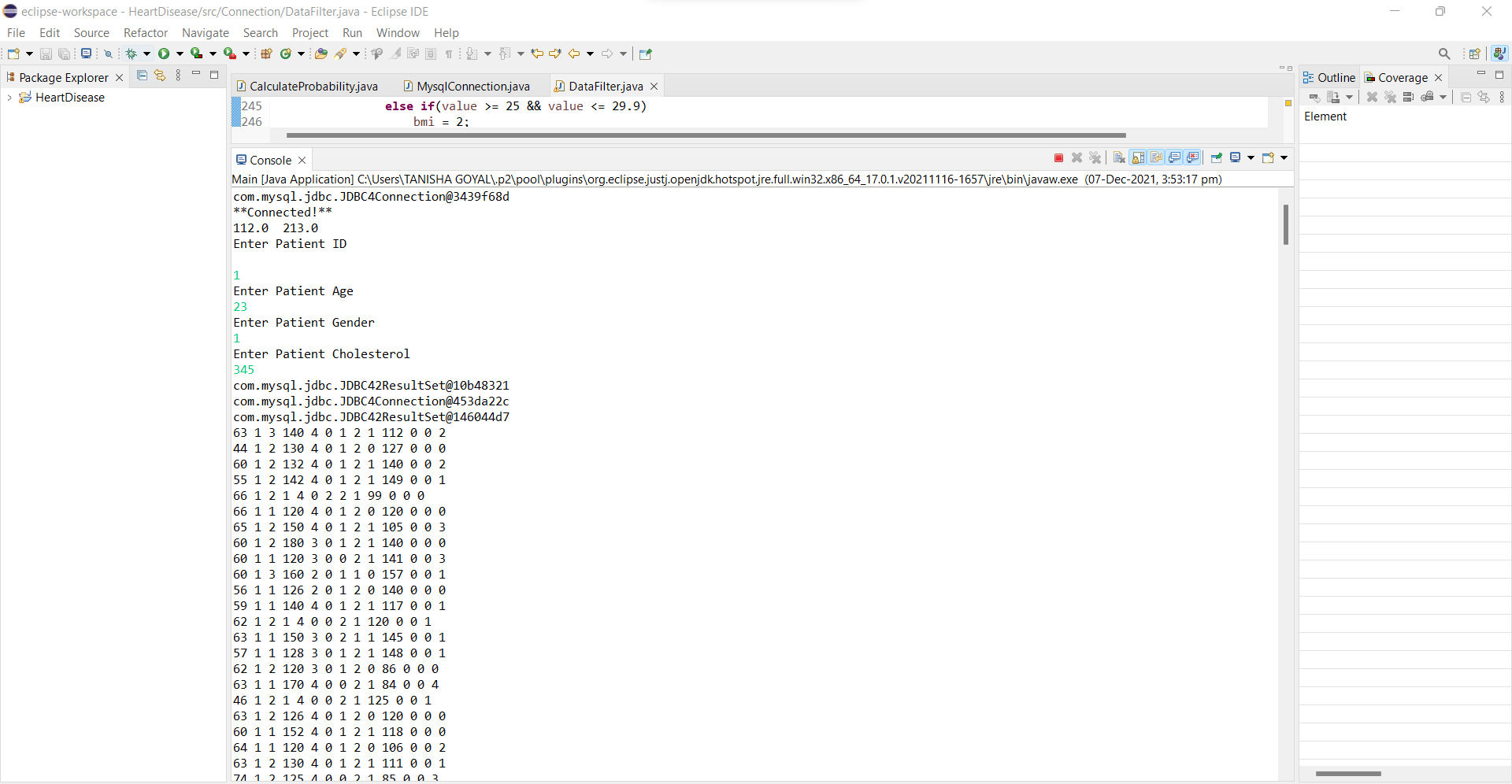
*EXTERNAL\_HA\_YES* = HA\_YES/ (**float**) total;

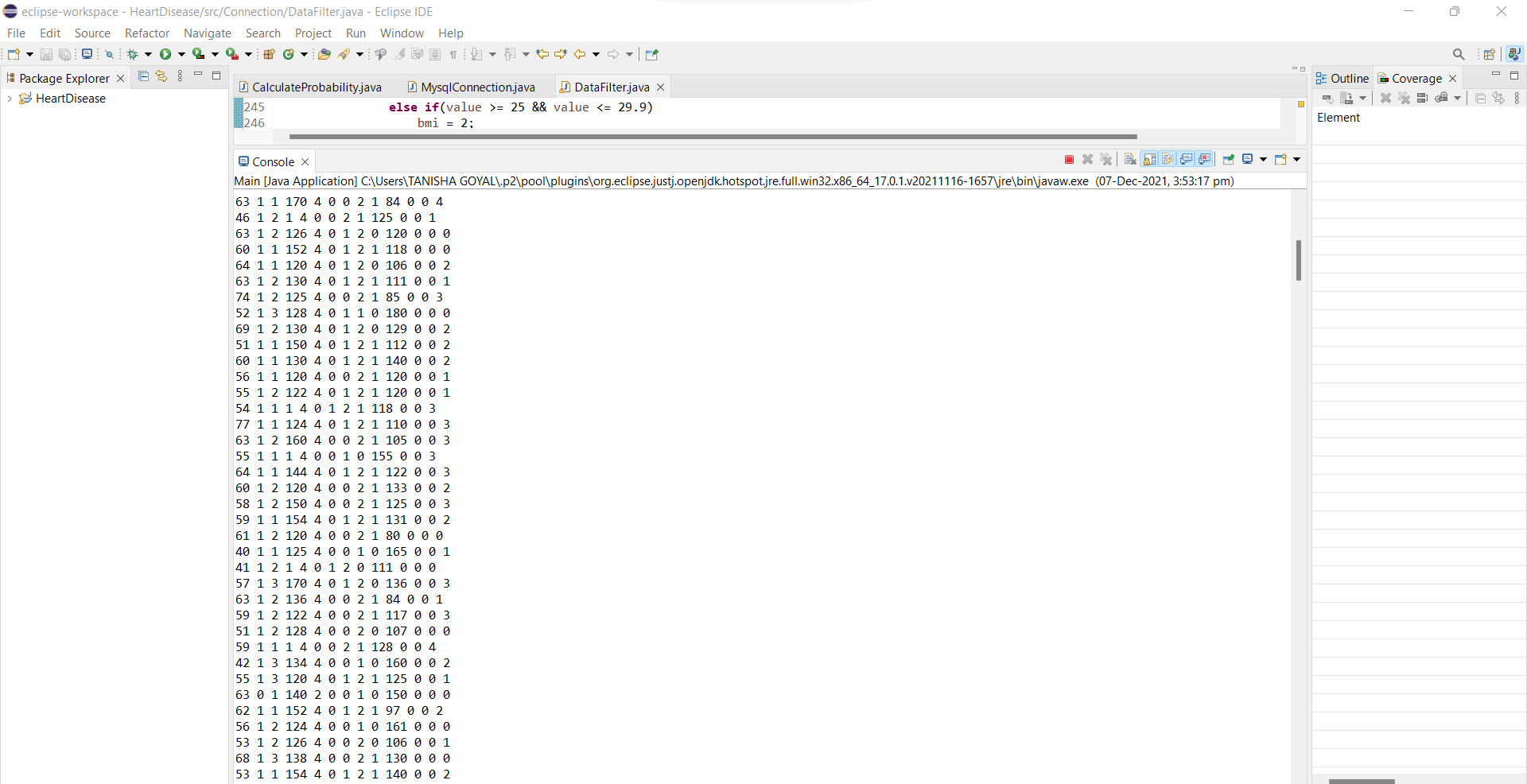
*EXTERNAL\_HA\_NO* = HA\_NO / (**float**) total;

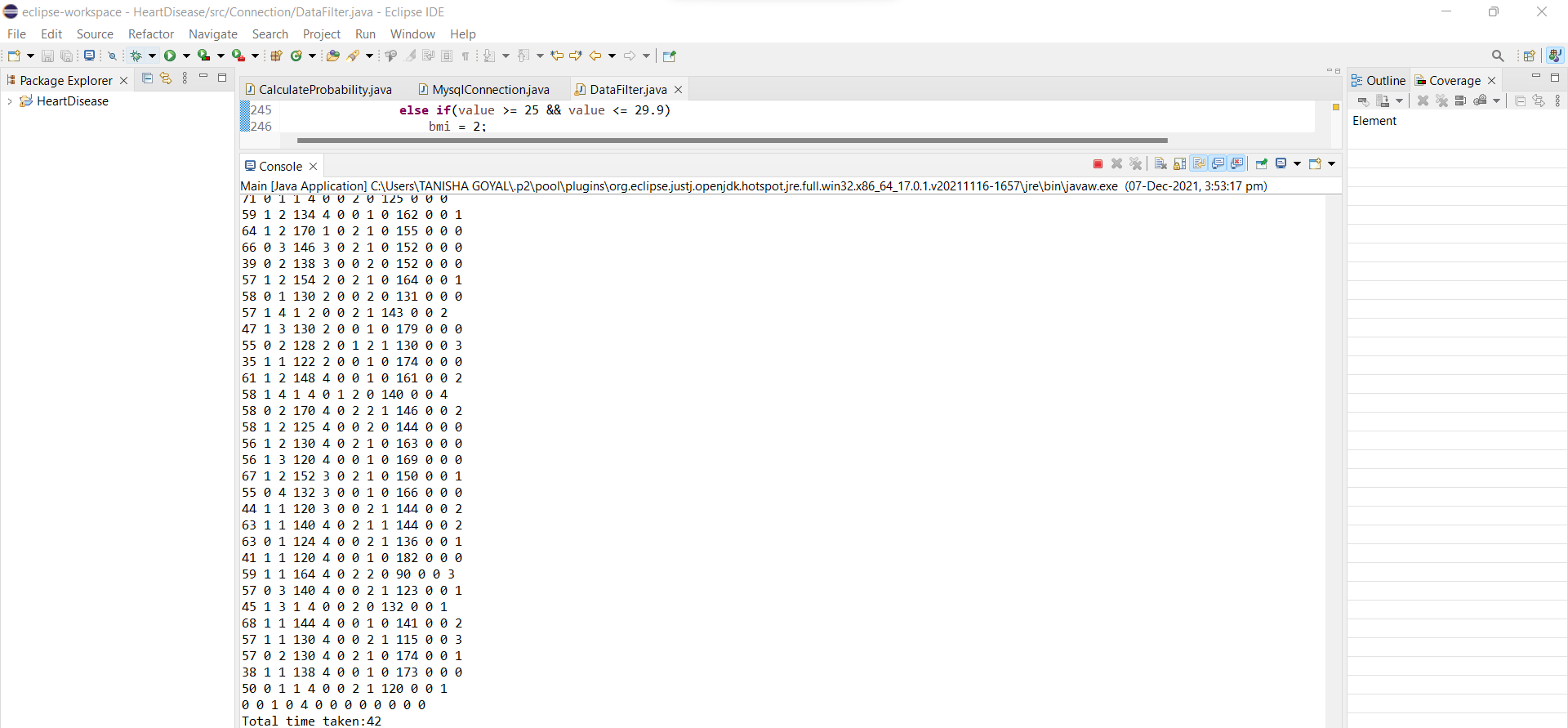
System.***out***.println(HA\_YES+" "+HA\_NO);

System.***out***.println(age + " " + gender + " " + cholesterol + " "+ trest\_bp + " " + chest\_pain + " " + blood\_sugar + " "+ rest\_ecg + " " + thalach + " " + exang + " " + thal + " " + height + " " + weight + " " + heart\_attack);

***RESULT***

******

******

******