A

Mini Project report on

**HEART DISEASE PREDICTION SYSTEM USING DATA SCIENCE**

Submitted in partial fulfillment of the requirements for the award of Degree of

**BACHELOR OF TECHNOLOGY**

**in**

**INFORMATION TECHNOLOGY**

**By**

**A.Jayavani (16881A1201)**

**S.Rakshitha (16881A1239)**

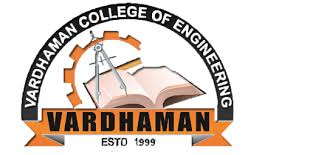
**D.Shresta (16881A1213)**

Under the Guidance of

**Mr. Srikanth Reddy Gopu**

Associate Professor

Department of Information Technology

****

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**VARDHAMAN COLLEGE OF ENGINEERING**

(AUTONOMOUS)

(Affiliated to JNTUH, Approved by AICTE and Accredited by NBA)

Shamshabad - 501 218, Hyderabad

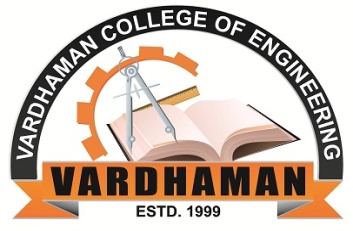
2016-2020

**VARDHAMAN COLLEGE OF ENGINEERING**

(AUTONOMOUS)

Shamshabad - 501 218, Hyderabad

**DEPARTMENT OF INFORMATION TECHNOLOGY**

****

**CERTIFICATE**

This is to certify that the project report entitled, **“Heart Disease Prediction System Using Data Science”**, done by **A.Jayavani(16881A1201), S.Rakshitha (1681A1239), D.Shresta (16881A1213),** Submitted to the department of Information Technology, **Vardhaman College of Engineering**, in partial fulfillment of the requirements for the Degree of **BACHELOR OF TECHNOLOGY** in **Information Technology**, during the year 2019. It is certified that he/she has completed the project satisfactorily.

**Signature of Supervisor: Head of the Department:**

**Mr. Srikanth Reddy Gopu****Dr. M. Gopi Chand****M.Tech, Ph.D.**

Associate Professor, Professor & Head,

Dept. of Information Technology, Dept. of Information Technology,

Vardhaman College of Engineering, Vardhaman College of Engineering,

Hyderabad. Hyderabad.

Viva-Voce held on……………………………………………**­­­­­­­­­­**

**SIGNATURE OF THE EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

The satisfaction that accompanies the successful completion of the task would be put in complete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to, **Mr. Srikanth Reddy Gopu,** Associate Professor, Department Information Technology, Vardhaman College of Engineering, for his able guidance and useful suggestions, which helped us in completing the project work, in time.

We are particularly thankful to **Dr. M. Gopi Chand**, Professor& Head of Department, Information Technology for his guidance, intense support and encouragement, which helped us to mold our project into a successful one.

We show gratitude to our honorable Principal **Dr. S. Sai Satyanarayana** **Reddy**, for providing all the facilities and support.

We avail this opportunity to express our deep sense of gratitude and heartfelt thanks to all the members of management, for providing congenial atmosphere to complete this project successfully.

We also thank all our teaching and non-teaching staff members of Information Technology department for their valuable support and generous advice. Finally thanks to all our friends and family members for their continuous support and enthusiastic help.

**A.JAYAVANI (16881A1201)**

**S.RAKSHITHA (16881A1239)**

**D.SHRESTA (16881A1213)**

**ABSTRACT**

Heart Disease is one of the major causes of increased mortality rate in India. Generally, a large most of data is generated from various sectors in the medical industry. For of diagnosis these data we require a good decision support systems and diagnosis techniques. Data mining techniques are helpful for early diagnosis of heart disease in a patient. These data mining techniques are very effective in designing a medical support system and enrich ability to determine the unseen patterns and associations in clinical data. In this project, we are designing a heart disease prediction system using the data mining techniques. The system analyzes the patient’s data and derives the patterns from it. These patterns are used to predict whether the patient is having the heart disease or not. The application is fed with various details and the heart disease associated with those details. The application allows user to share their heart related issues.

This project deals with implementing of heart disease prediction system (HDPS) with algorithms. Heart disease dataset is used for experimental analysis.

**CONTENTS**

Acknowledgement iii

Abstract iv

List of Abbreviations vi

**CHAPTER 1 INTRODUCTION 01**

1.1 Motivation 02

1.2 Problem Statement 02

1.3 Objective of the project 03

1.4 Scope of the project 03

**CHAPTER 2 LITERATURE SURVEY 04**

2.1 Existing System 05

2.2 Proposed System 05

**CHAPTER 3 SYSTEM ANALYSIS 06**

3.1 Software Requirements Specification 07

3.1.1 Hardware Requirements 07

3.1.2 Software Requirements 07

3.2 Algorithms and Flowcharts 07

**CHAPTER 4 DESIGN 15**

4.1 Introduction 16

4.2 UML Diagrams 17

4.3 Modules Design and Organization 21

**CHAPTER 5 IMPLEMENTATION AND RESULTS 22**

5.1 Introduction 23

5.2 Explanation of Key Functions 24

5.3 Methods of Implementation 24

5.3.1 Output Screens 25

**CHAPTER 6 TESTING AND VALIDATION 36**

6.1 Introduction 37

6.2 Design of Test Cases 38

**CHAPTER 7 CONCLUSION 41**

7.1 Project conclusion 42

7.2 Project Future Enhancement 42

**REFERENCES 43**

**LIST OF ABBREVIATIONS**

**1. HTML**: Hyper Text Markup Language is a markup language used to design static web pages.

**2. CSS:**  Style Sheets is the language for describing the presentation of Web pages, including colors, layout, and fonts.

**3. Bootstrap**: Bootstrap is the most popular HTML, CSS, and JavaScript framework for developing responsive, mobile-first websites.

**4. Python**: Python is a programming language. Python can be used on a server to create web applications.

**Chapter – 1**

**INTRODUCTION**

**1.1 Motivation:**

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: “How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?” This is the main motivation for this research.

**1.2 Problem Statement:**

Many hospital information systems are designed to support patient billing, inventory management and generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like “What is the average age of patients who have heart disease?”, “How many surgeries had resulted in hospital stays longer than 10 days?”, “Identify the female patients who are single, above 30 years old, and who have been treated for cancer.” However, they cannot answer complex queries like “Identify the important preoperative predictors that increase the length of hospital stay”, “Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?”, and “Given patient records, predict the probability of patients getting a heart disease.” Clinical decisions are often made based on doctors’ intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed 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 modelling 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.

**1.3 Objectives of the Project:**

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining modeling techniques, namely Decision tree. IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms.

**1.4 Scope of the project:**

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 .The main objective of this research is to develop a prototype Heart Disease Prediction System (HDPS) using data mining modeling techniques, namely Decision tree. So it provides effective treatments, it also helps to reduce treatment costs and also enhances visualization and ease of interpretation. With immense knowledge and accurate data in that field. Large corporations invest heavily in this kind of activity to help focus attention on possible events and risks that are involved. Such work brings together all available past and current data, as a basis on which to develop reasonable expectations about the future.

**Chapter - 2**

**LITERATURE SURVEY**

**2.1 EXISTING SYSTEM:**

**[1] Intelligent heart disease prediction system using data mining techniques:** In this paper heart disease prediction is done using data mining techniques such as decision trees, neural network and Naïve Bayes. This system answers “what if” query. It is implementing on .net platform. It is used for heart disease prediction.

**[2] An empirical study on applying data mining techniques for analysis and prediction of heart disease:**

It is found that health environment is poor in extracting knowledge so in this paper data mining techniques are applied. This paper deals with application of data mining.

**[3] Prediction system for heart disease using Decision tree:**

It is web-based classification. It takes input from the user and predicts the output. It compares the value with trained dataset. In this project it is mentioned that because of this system the treatment cost are reduced.

**[4] Decision support in heart disease prediction system using Decision tree mining:**

This research developed using data mining techniques mainly Decision tree. It takes input as the patients attributes. It helps trained nurses and medical students to treat patients.

**[5] Intelligent and effective heart attack prediction system using data mining:**

In this paper k-means clustering is used. This system is capable of predicting heart disease.

**2.2 PROPOSED SYSTEM:**

The practice of examining large pre existing databases in order to generate new information. It coverts raw data into useful information. It analyzes the data for relationships that have not previously been discovered. The steps of data mining are: Data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation. Medical data mining is a domain of lot of imprecision and uncertainty. The clinical decisions are usually based on the doctor’s intuition. Therefore this may lead to disastrous consequences. Due to this there are many errors in the clinical decisions and it results in excessive medical costs. Serialization is also used in this system. It converts the data objects into streams of bytes and stores it into database. The proposed system compares and contrasts different machine learning algorithms in predicting the best effective way to know the result for heart disease prediction system.

**Chapter – 3**

**SYSTEM**

**ANALYSIS**

**3.1 SOFTWARE REQUIREMENTS SPECIFICATION**:

**3.1.1 Hardware requirements:**

Processer : Any Update Processer

Ram : 1 GB

Hard Disk : 2 GB

**3.1.2 Software requirements:**

Operating System : Windows 7 / Windows 8 /Windows 10 / Ubuntu.

Technology : Python

Web Technologies : Html, Html-5, Bootstrap, CSS

UML : Star UML

**3.2 ALGORITHMS AND FLOWCHARTS:**

**1. Decision tree:**

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Assumptions while creating Decision Tree

The below are the some of the assumptions we make while using Decision tree:

* At the beginning, the whole training set is considered as the **root.**
* Feature values are preferred to be categorical. If the values are continuous then they are discretized prior to building the model.
* Records are **distributed recursively** on the basis of attribute values.
* Order to placing attributes as root or internal node of the tree is done by using some statistical approach.

**Decision Tree Algorithm Pseudocode**

1. Place the best attribute of the dataset at the **root** of the tree.
2. Split the training set into **subsets**. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
3. Repeat step 1 and step 2 on each subset until you find **leaf nodes** in all the branches of the tree.

**Advantages**

**Comprehensive**

A significant advantage of a decision tree is that it forces the consideration of all possible outcomes of a decision and traces each path to a conclusion. It creates a comprehensive analysis of the consequences along each branch and identifies decision nodes that need further analysis.

**Specific**

Decision trees assign specific values to each problem, decision path and outcome. Using monetary values makes costs and benefits explicit. This approach identifies the relevant decision paths, reduces uncertainty, clears up ambiguity and clarifies the financial consequences of various courses of action.

When factual information is not available, decision trees use probabilities for conditions to keep choices in perspective with each other for easy comparisons.

**Easy to Use**

Decision trees are easy to use and explain with simple math, no complex formulas. They present visually all of the decision alternatives for quick comparisons in a format that is easy to understand with only brief explanations.

They are intuitive and follow the same pattern of thinking that humans use when making decisions.

**Versatile**

A multitude of business problems can be analyzed and solved by decision trees. They are useful tools for business managers, technicians, engineers, medical staff and anyone else who has to make decisions under uncertain conditions.

The algorithm of a decision tree can be integrated with other management analysis tools such as Net Present Value and Project Evaluation Review Technique (PERT).

Simple decision trees can be manually constructed or used with computer programs for more complicated diagrams.

Decision trees are a common-sense technique to find the best solutions to problems with uncertainty. Should you take an umbrella to work today? To find out, construct a simple decision-tree diagram.

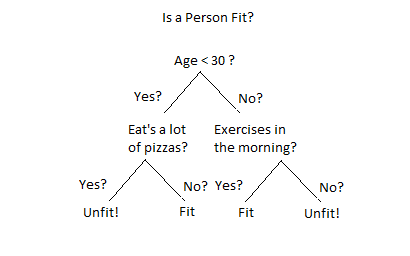
**Disadvantages**

* They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.
* They are often relatively inaccurate. Many other predictors perform better with similar data. This can be remedied by replacing a single decision tree with a random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree.
* For data including categorical variables with different number of levels, information gain in decision trees is biased in favor of those attributes with more levels.
* Calculations can get very complex, particularly if many values are uncertain and/or if many outcomes are linked.

**Common terms used with Decision trees:**

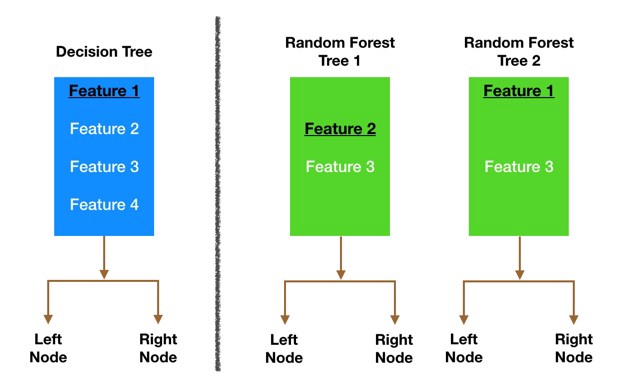
1. **Root Node:** It represents entire population or sample and this further gets divided into two or more homogeneous sets.
2. **Splitting:** It is a process of dividing a node into two or more sub-nodes.
3. **Decision Node:** When a sub-node splits into further sub-nodes, then it is called decision node.
4. **Leaf/ Terminal Node:** Nodes do not split is called Leaf or Terminal node.
5. **Pruning:** When we remove sub-nodes of a decision node, this process is called pruning. You can say opposite process of splitting.
6. **Branch / Sub-Tree:** A sub section of entire tree is called branch or sub-tree.
7. **Parent and Child Node:** A node, which is divided into sub-nodes is called parent node of sub-nodes whereas sub-nodes are the child of parent node.

EXAMPLE:



**2. Random forest Algorithm:**

**Random forests** or **random decision forests** are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

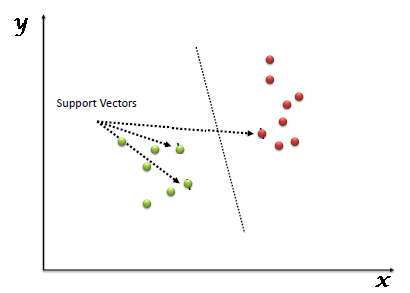


3. **Support-vector machine:**

In machine learning, **support-vector machines** (**SVMs**, also **support-vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on the side of the gap on which they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

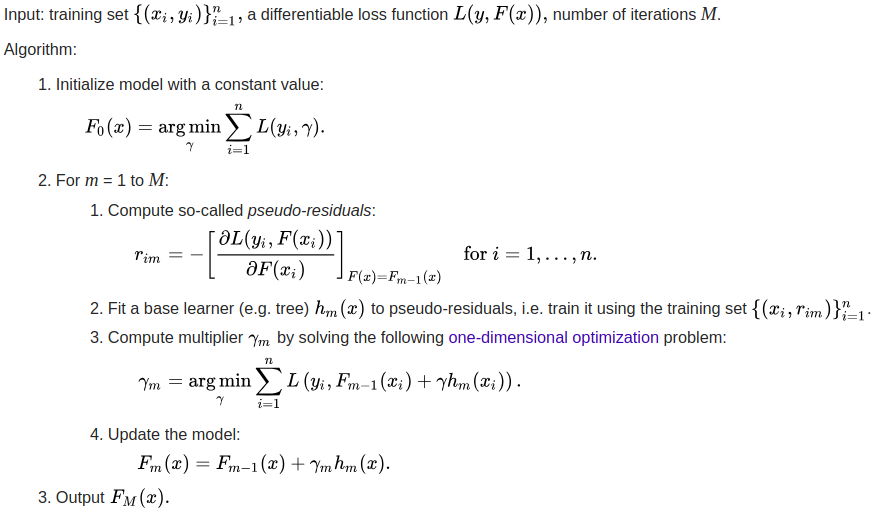
When data are unlabeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The **support-vector clustering** algorithm, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed in the support vector machines algorithm, to categorize unlabeled data, and is one of the most widely used clustering algorithms in industrial applications.

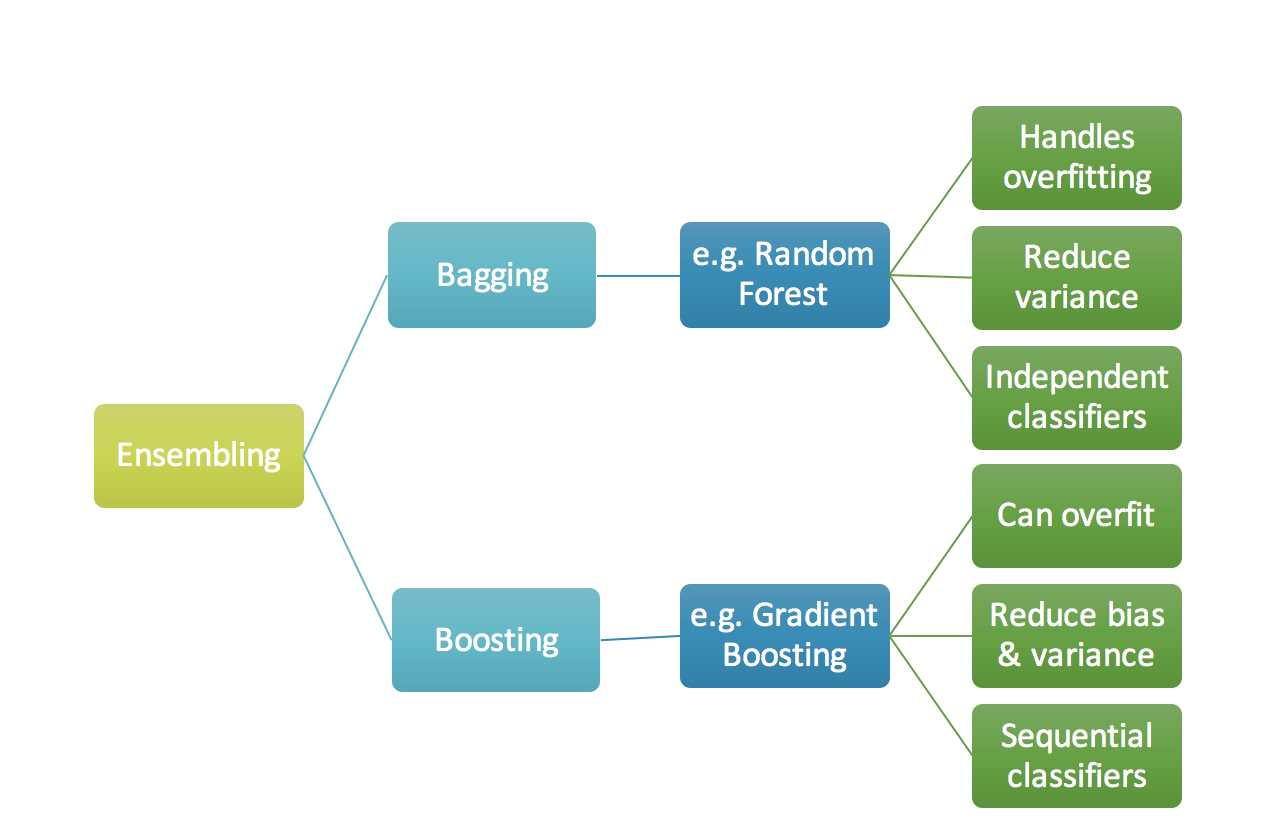


4. **Gradient boosting:**

**Gradient boosting** is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

**Bagging**is a simple ensembling technique in which we build many independent predictors/models/learners and combine them using some model averaging techniques. (E.g. weighted average, majority vote or normal average)





**ATTRIBUTES IN HEART DISEASE PREDICTION SYSTEM:**

#age

Age in years

#sex

(1 = male; 0 = female)

#cp

Chest pain type

#trestbps

Resting blood pressure (in mm Hg on admission to the hospital)

#chol

Serum cholesterol in mg/dl

#fbs

(Fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

#restecg

Resting electrocardiographic results

#thalach

Maximum heart rate achieved

#exang

Exercise induced angina (1 = yes; 0 = no)

#oldpeak

ST depression induced by exercise relative to rest

#slope

The slope of the peak exercise ST segment

#ca

Number of major vessels (0-3) colored by flourosopy

#thal

3 = normal; 6 = fixed defect; 7 = reversible defect

#target

1 or 0(1=true, 0=false)

**Chapter – 4**

**DESIGN**

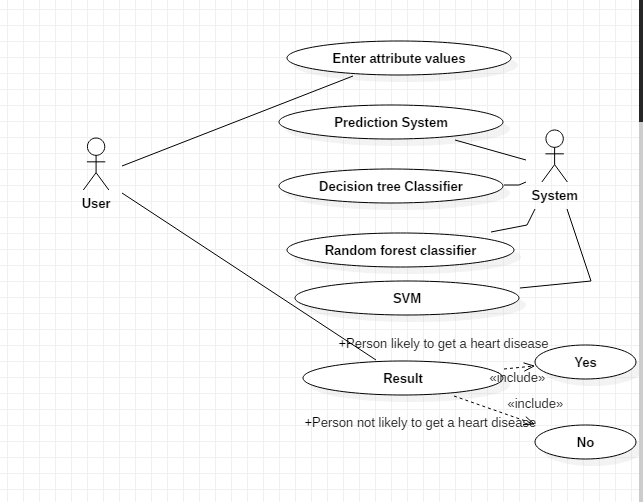
**4.1 Introduction:**

Heart Disease rate is increasing now-a-days in many countries. In today’s world with such higher heart attack rate there must be some protection against this disease. Here we introduced a system by which heart attacks can be reduced. Heart disease data must be fed into the system. We introduced data mining algorithm to predict attacks. K-means algorithm plays an important role in analyzing and predicting. K-means algorithm will cluster the attributes which are related to the heart disease occurrence and identify various relevant patterns by grouping the data. This system will prevent heart attack occurrences to some extent in the society. Heart data is analyzed which is stored in the database. Data mining algorithm will extract information and patterns from database. System will group the heart data. Clustering will be done based on attributes due to which heart attack occurs. This will help to predict the heart attack occurrence in future by analyzing the data. Admin will enter the heart dataset into the system which is required for prediction. Admin can view heart disease historical data. Mainly with the help of this system we can predict whether heart attack occurs or not based on some attributes related to heart dataset which is given and it is the historical dataset record from various hospitals where heart attacks occurred.

**4.2 UML Diagrams**

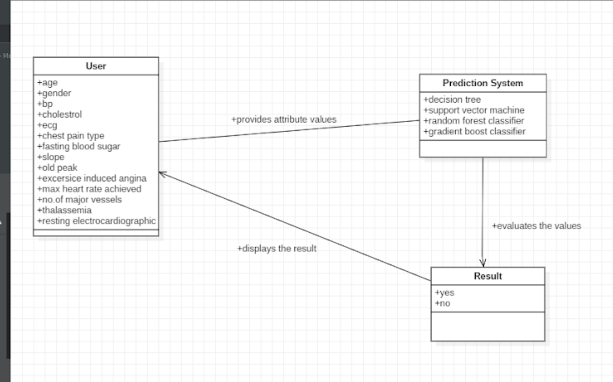
**1. Use Case Diagram:**

A use case diagram is a graphic depiction of the interactions among the elements of a system. A use case is a methodology used in system analysis to identify, clarity and organize system requirements. A use case diagram contains four components. The boundary, which defines system of interest in relation to the world around it.The actors usually individuals involved with the system defined according to their roles. The use cases in which the specific roles are played by the actors within and around the system. The relationships between and among the actors and the use cases.

****

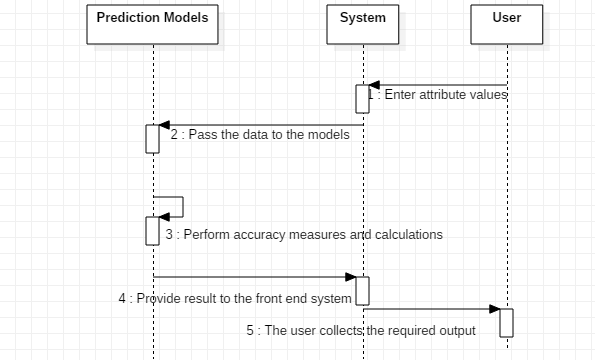
**2. Class Diagram:**

A class diagram is an illustration of the relationships and source code dependencies among classes in the Unified Modeling Language (UML). In this context, a class defines the methods and variables in an object, which is a specific entity in a program or the unit of code representing that entity.



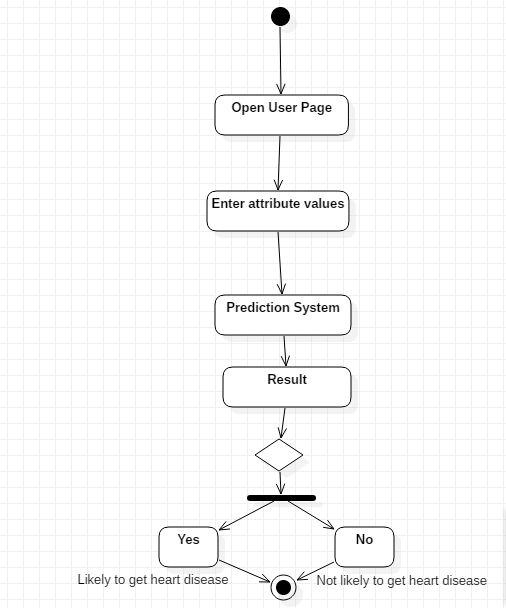
**3. Sequence Diagram**

A **sequence diagram** simply depicts interaction between objects in a **sequential** order i.e. the order in which these interactions take place. We can also use the terms event **diagrams** or event scenarios to refer to a **sequence diagram**. **Sequence diagrams** describe how and in what order the objects in a system function.

****

**4. Activity Diagram**

**Activity diagram** is another important **diagram** in UML to describe the dynamic aspects of the system. **Activity diagram** is basically a flowchart to represent the flow from one **activity** to another **activity**. The **activity** can be described as an operation of the system. The control flow is drawn from one operation to another.

****

**4.3 MODULE DESIGN AND ORGANIZATION:**

**Here is the snippets of code working at the backend to predict the existence of heart disease in a person. We compared and contrasted between different algorithms such as Decision tree, Gradient boost classifier, Random forest classifier and SVM.The idea behind comparison is to visualize which algorithm can give best output with high accuracy rate and precision. The comparison lets us know that the decision tree classifier can produce the output with more precision and accuracy among the other classifier algorithms.**

****

Middleware used:

To combine the back end (Python) with front end (Bootstrap), we use middleware as flask framework of Python.

Flask:

Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.



Front end:

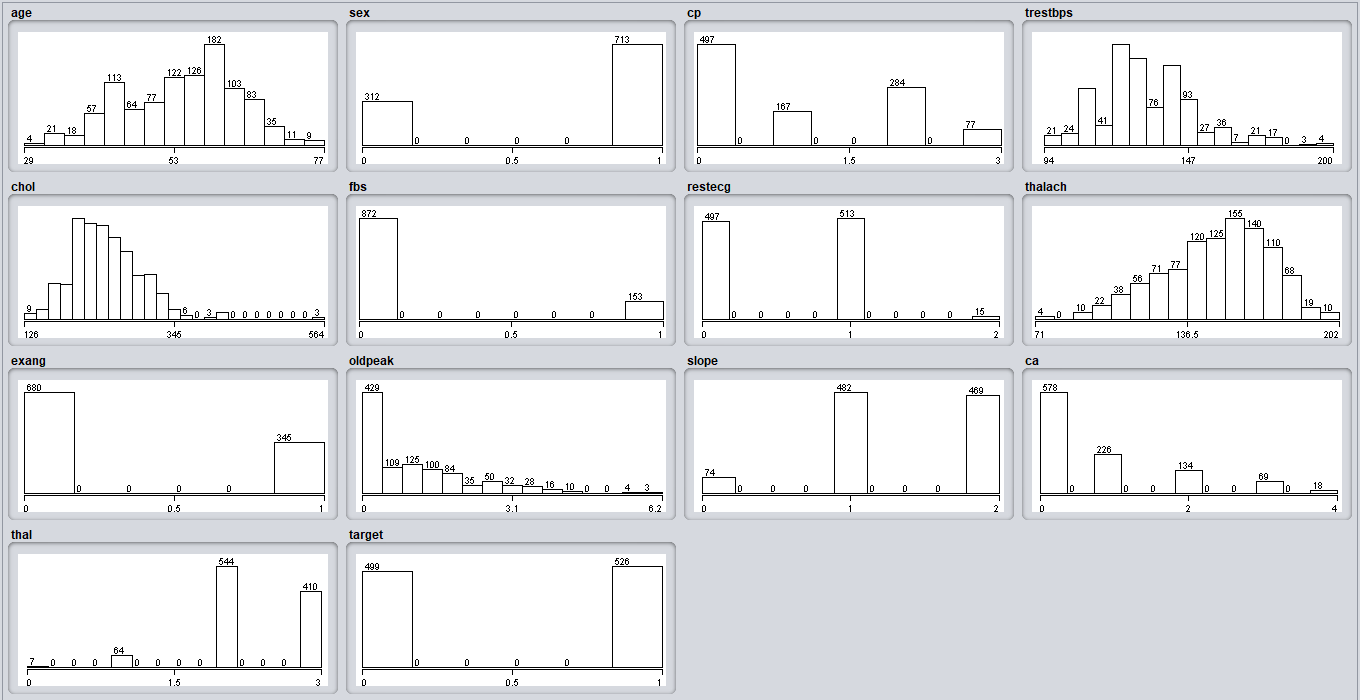
We use HTML, CSS and Bootstrap technologies to design front end where in the user can give the input to predict the output.



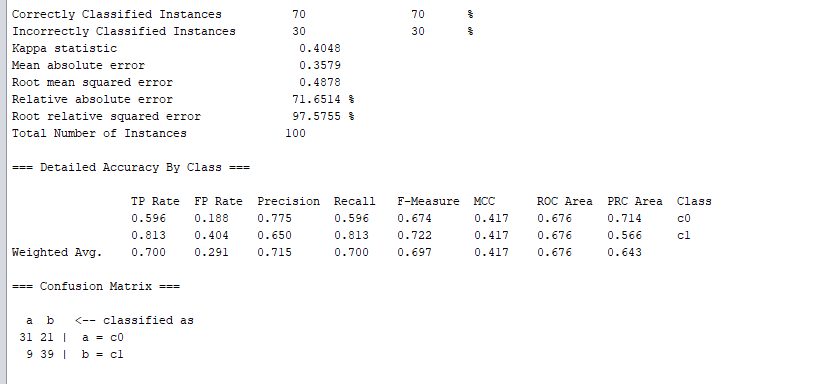
WEKA Tool Usage:

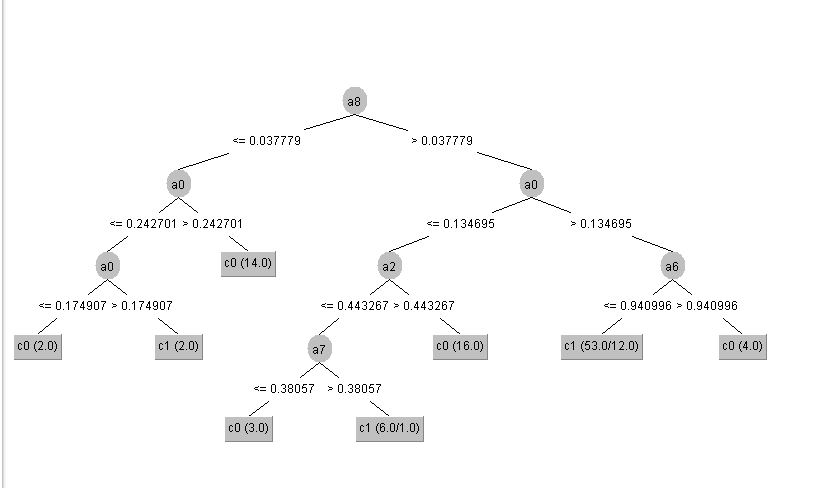
In order to predict the accuracy and precision rates, we can use weka tool wherein the data set in the form of csv file can be imported and the output can be generated.

**Attributes distribution:**

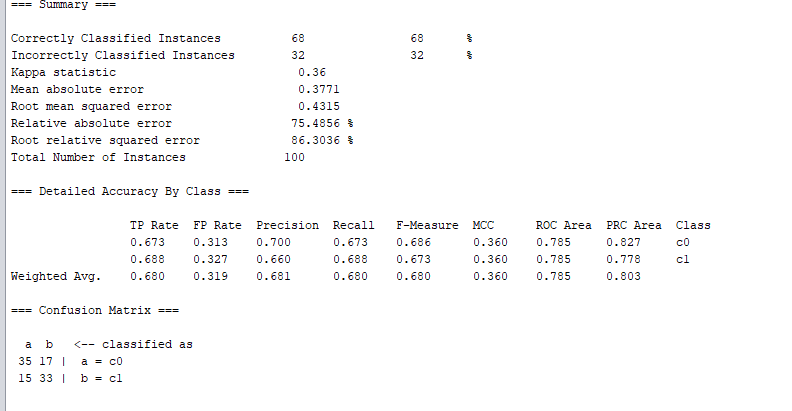
****

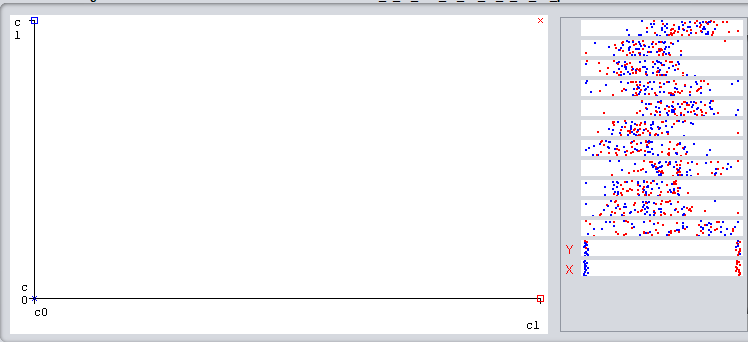
**Decision Tree:**

****

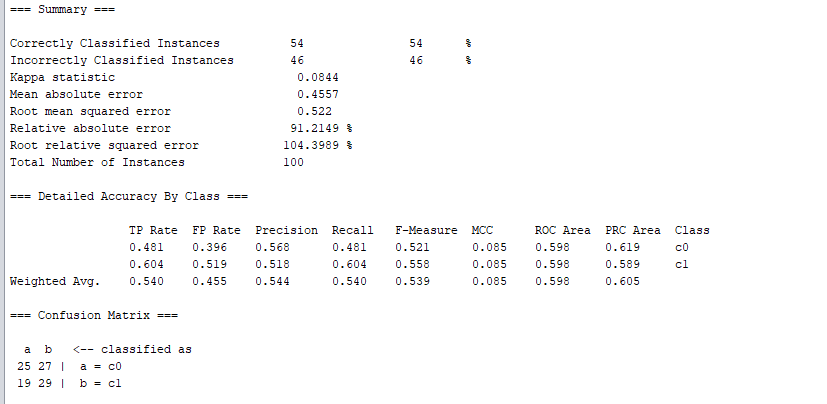
****

**Random forest:**

****

****

**Naïve Bayes:**

****

****

**Chapter – 5**

**IMPLEMENTATION**

**5.1 Introduction:**

The practice of examining large preexisting data bases in order to generate new information. It coverts raw data into useful information. It analyze the data for relationships that have not previously been discovered.

The steps of data mining are:

Data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge representation.

Medical data mining is a domain of lot of imprecision and uncertainty. The clinical decisions are usually based on the doctor’s intuition. Therefore this may lead to disastrous consequences. Due to this there are many errors in the clinical decisions and it results in excessive medical costs. Serialization is also used in this system. It converts the data objects into streams of bytes and stores it into database.

**Why Data Mining is used in Business?**

Data mining is used in business to make better managerial decisions by:

Automatic summarization of data

Extracting essence of information stored.

Discovering patterns in raw data.

* KDD is an iterativeprocess where evaluation measures can be enhanced, mining can be refined, new data can be integrated and transformed in order to get different and more appropriate results.
* Preprocessing of databases consists of Data cleaning and Data Integration.



**5.2 Explanation of Key functions:**

The proposed project mainly contains one module:

1. User module

**1. User Module:**

In this module, user opens the web page and enters the required details such as values of 14 attributes and click the submit button. If any of the values are missing, a message appears as “missing values”. Once every values is entered, the prediction system in the back end predicts the result and produces the output to the user. There is a refresh button where the values can be refreshed by the user if required.

**5.3 Method of Implementation:**

**User:**

In this module user can open the web page. User can enter patient details, and system will do prediction for output.

**1. Decision tree:**

A **decision tree** is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

**2. Random forest Algorithm**:

**Random forests** or **random decision forests** are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

3**. Support-vector machine:**

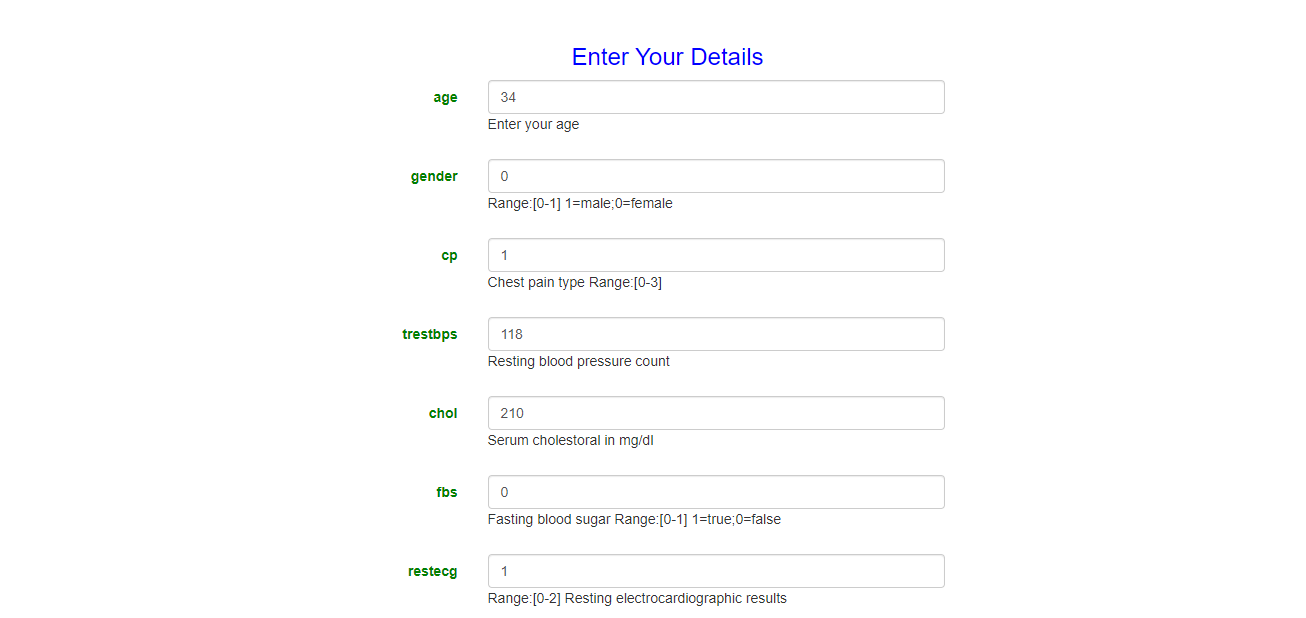
In machine learning, **support-vector machines** (**SVMs**, also **support-vector networks**) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

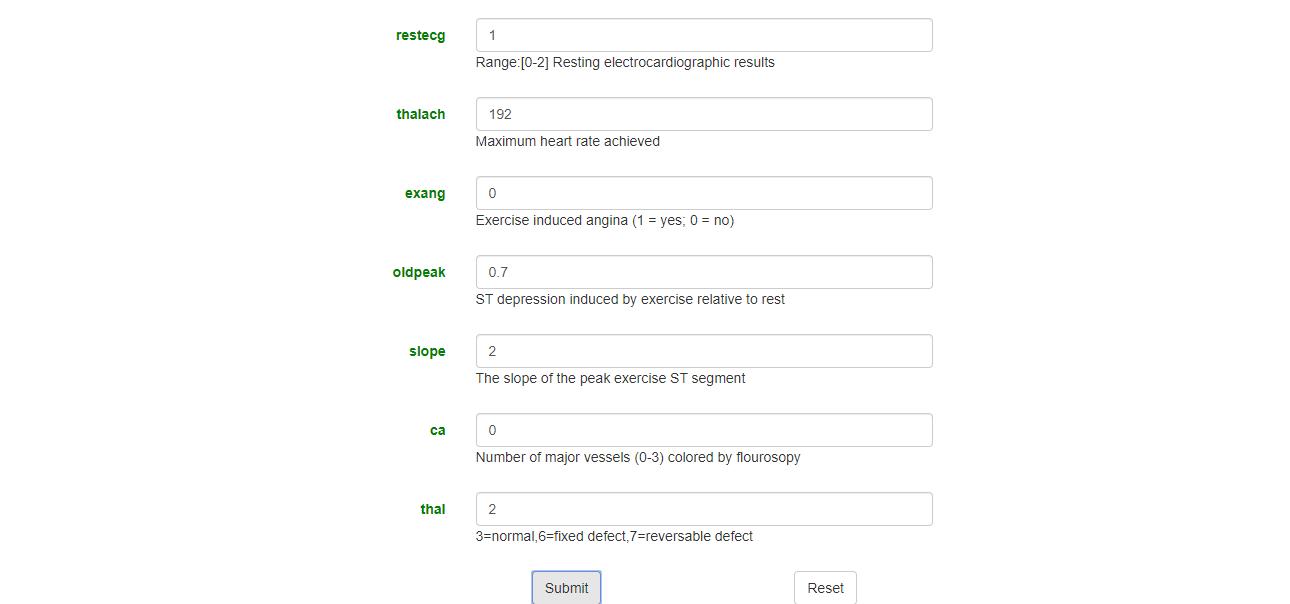
4. **Gradient boosting:**

**Gradient boosting** is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

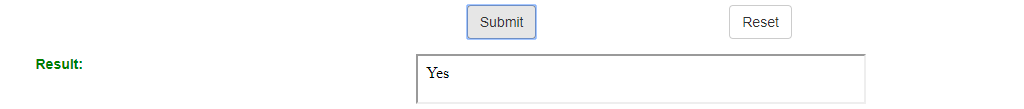
**5.3.1 Outputs:**

**1. The web page is the front end which takes input from the user to predict the result.**

****

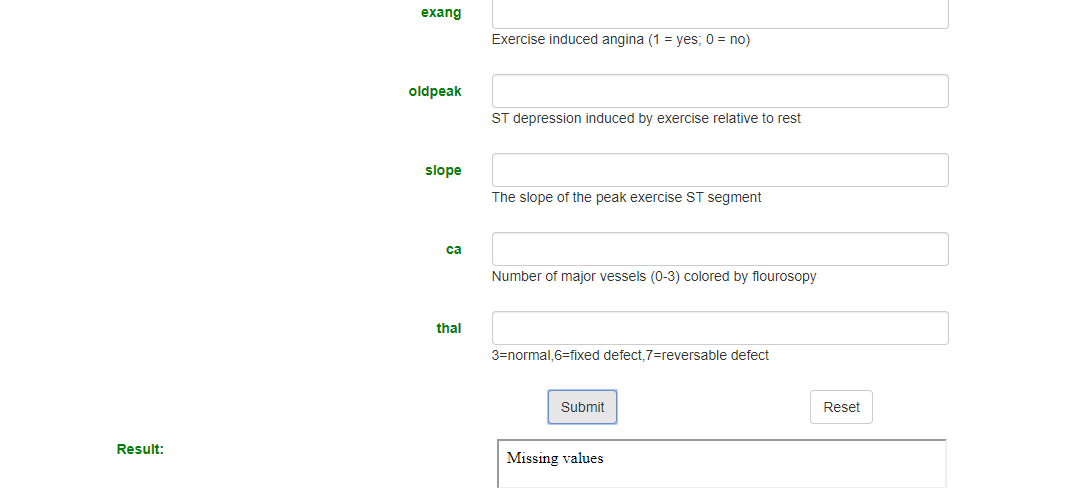
****

**2. Here is the output. Yes/No is displayed in the result box.**

****

If the user do not enter any single value, then a message “Missing Values” is appeared on the screen. Once the user enters all the values and clicks the submit button, the system predicts the result and produces the output to the user. Reset button is also placed for the convenience of the user where a user can refresh the values as per the requirement.

Missing Values case:



**Chapter – 6**

**TESTING**

**AND**

**VALIDATION**

**6.1 Introduction:**

**TESTING**

Testing is the debugging program is one of the most critical aspects of the computer programming triggers, without programming that works, the system would never produce an output of which it was designed. Testing is best performed when user development is asked to assist in identifying all errors and bugs. The sample data are used for testing. It is not quantity but quality of the data used the matters of testing. Testing is aimed at ensuring that the system was accurately an efficiently before live operation commands.

**Testing objectives:**

The main objective of testing is to uncover a host of errors, systematically and with minimum effort and time. Stating formally, we can say, testing is a process of executing a program with intent of finding an error.

A successful test is one that uncovers an as yet undiscovered error.

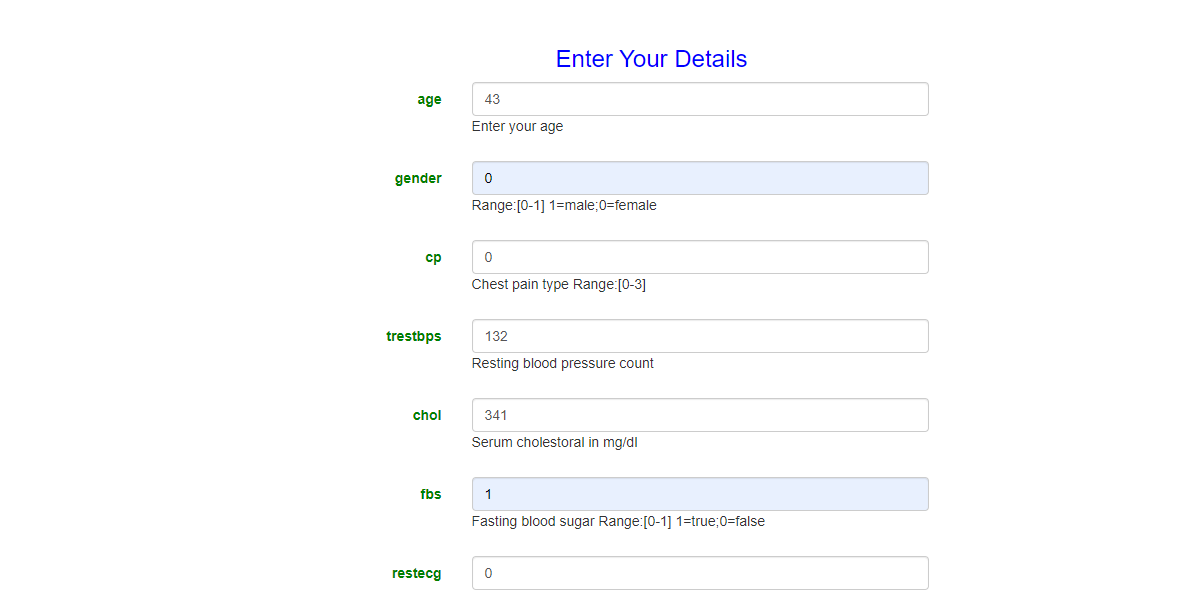
A good test case is one that has probability of finding an error, if it exists.

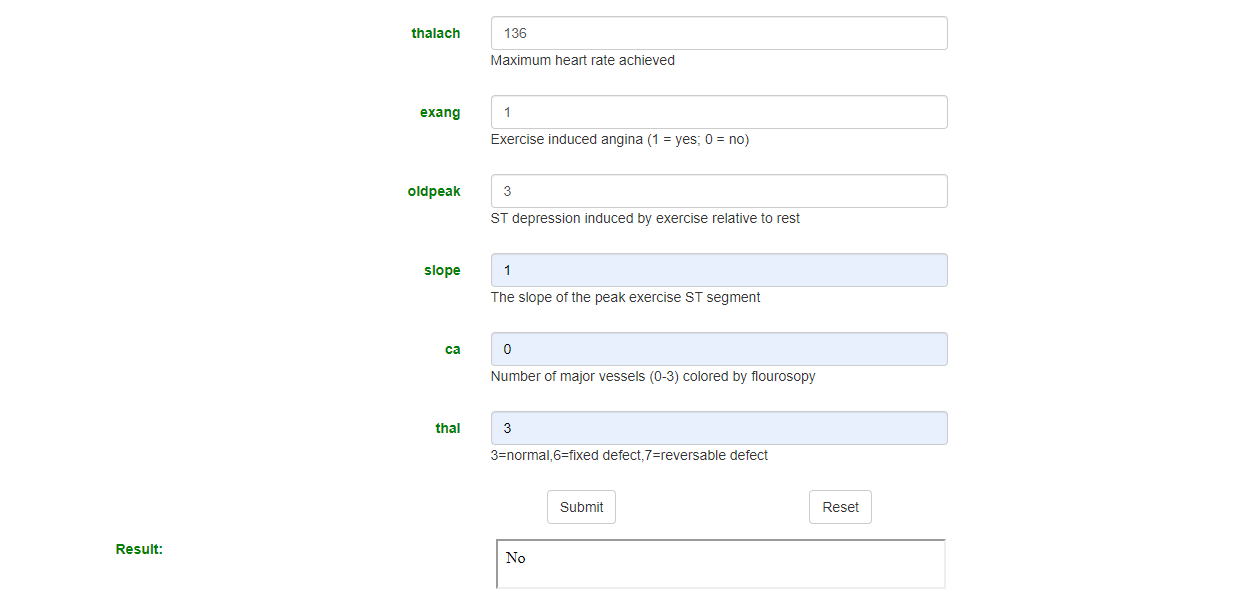
The test is inadequate to detect possibly present errors.

The software more or less confirms to the quality and reliable standards.

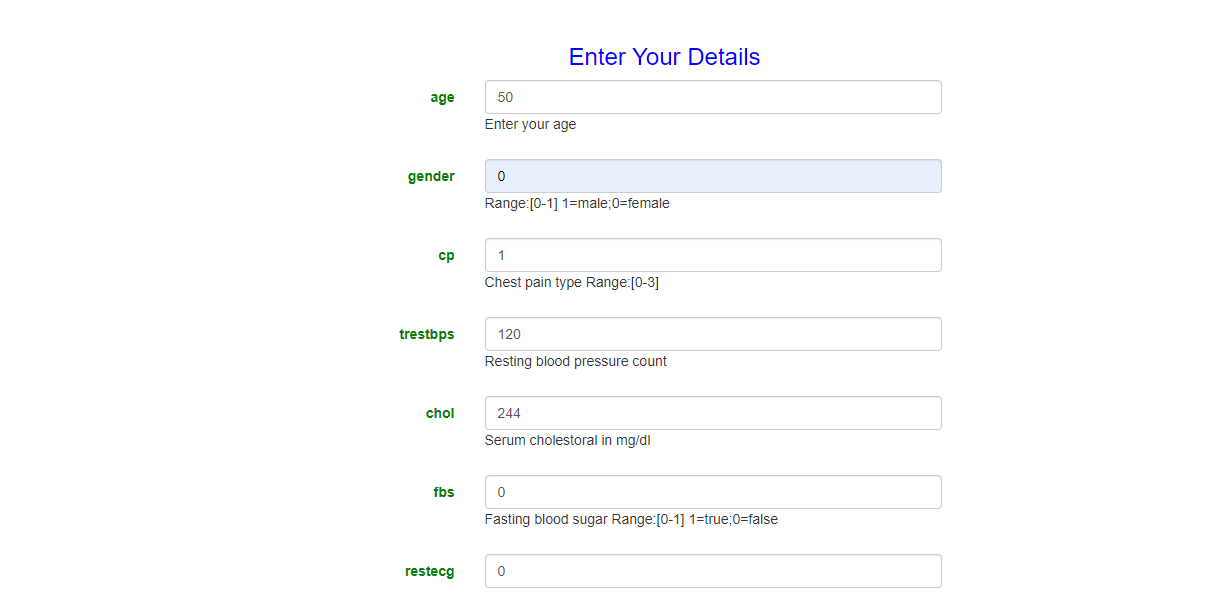
**6.2 Test Case Outputs:**

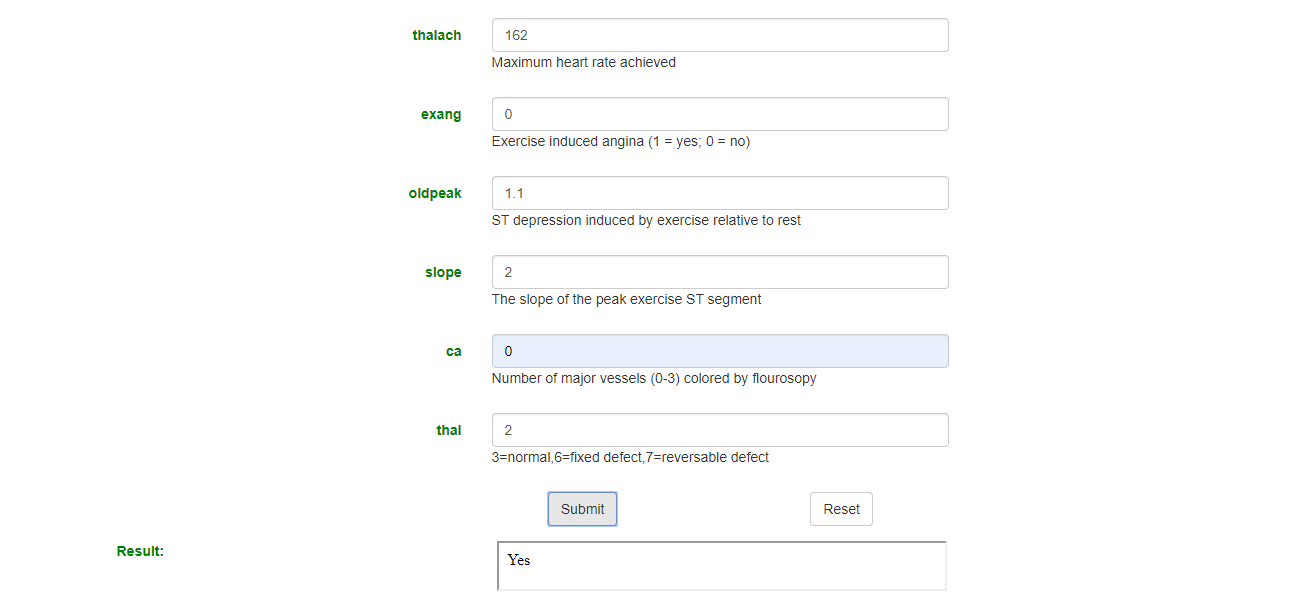
**i. The result evaluates to be false or ‘No’ heart disease in this case.**

****



ii. The result evaluates to be true or ‘Yes’ in this case.





**Chapter – 7**

**CONCLUSION**

**&**

**FUTURE ENHANCEMENT**

**7.1 CONCLUSION**

In this project we are proposing heart disease prediction system using decision tree algorithm. We are using decision tree clustering for increasing the efficiency of the output. This is the most effective model to predict patients with heart disease. Decision trees are commonly used in operations research, specifically in **decision** analysis, to help identify a strategy most likely to reach a goal

**7.2 PROJECT FUTURE ENHANCEMENT**

Medical related information’s are huge in nature and it can be derived from different birthplaces which are not entirely applicable in feature. In this work, heart disease prediction system was developed using clustering and classification algorithms to predict the effective risk level and accuracy of the patients. In future work, we have planned to propose an effective disease prediction system to predict the heart disease with better accuracy using different data mining techniques and compare the performance of algorithm with other related data mining algorithms.

**REFERENCES:**

[1] Sellappan Palaniappan, Rafiah Awang “Intelligent Heart Disease Prediction System Using Data Mining Techniques “Department of Information Technology Malaysia University of Science and Technology Block C, Kelana Square, Jalan SS7/26 Kelana Jaya, 47301 Petaling Jaya, Selangor, Malaysia .

[2] "CSV File Reading and Writing" (http:/ / docs. python. org/ library/ csv. html). . Retrieved July 24, 2011. "Is no ““CSV standard”.

[3] Y. Shafranovich. "Common Format and MIME Type for CommaSeparated Values (CSV) Files" (http:/ / tools. ietf. org/ html/ rfc4180) Retrieved September 12, 2011.

[4] home.deib.polimi.it/matteucc/Clustering/tutorial\_html/kmeans.html “A tutorial on clustering algorithms”.

[5] Shadab Adam Pattekari and Asma Parveen “Prediction System For Heart Disease Using Naïve Bayes” International Journal of Advanced Computer and Mathematical Sciences ISSN 2230-9624. Vol 3, Issue 3, 2012, pp 290-294.

[6] Mrs.G.Subbalakshmi (M.Tech), Mr. K. Ramesh M.Tech, Asst. Professor Mr. M. Chinna Rao M.Tech,(Ph.D.) Asst. Professor, “Decision Support in Heart Disease Prediction System using Naive Bayes” G.Subbalakshmi et al. / Indian Journal of Computer Science and Engineering (IJCSE)2011.

[7] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen “Association rule mining to detect factors which contribute to heart disease in males and females” Expert Systems with Applications 40 (2013) 1086–1093.

[8] Oleg Yu. Atkov (MD, PhD), Svetlana G. Gorokhova (MD, PhD), Alexandr G. Sboev (PhD), Eduard V. Generozov (PhD), Elena V. Muraseyeva (MD, PhD), Svetlana Y. Moroshkina,Nadezhda N. Cherniy “Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters” Journal of Cardiology (2012) 59, 190—194.