**HEART FAILURE PREDICTION**

# By

**Group No. 24**

**Rishabh Mundhada, TE B, 3**

**Dev Rathod, TE-B, 11**

**Raj Shah, TE-B, 17**

**Akshat Singh, TE-B, 28**

**Yogesh Yadav, TE-B, 58**

*Under the Guidance of*

**Mrs. Shiwani Gupta Assistant Professor**

*for the subject*

**Machine Learning**

*In*

**T.E. COMPUTER ENGINEERING**

# (Academic Year: 2020-21)

# \\162.162.15.215\admin dept\Harshali\print\project\NEW\uom.jpg

CERTIFICATE

*This is to certify that*

**Authors**

**Rishabh mundhada , Dev rathod ,**

**Raj shah , Akshat singh,**

**Yogesh Yadav**

*Have satisfactorily completed the requirements of the T.E Mini Project Report*

On

**“HEART FAILURE PREDICTION**

**”**

Shiwani Gupta Dr. Harshali Patil

**Subject In-charge HOD COMP**

**Examiners**

1. Signature: …………………. 2. Signature: ………………….

Name: Name:

Date:

Place: Mumbai

# Table of content

[List of Figures 1](#_TOC_250011)

Chapter 1. Introduction 2

* 1. [*Motivation 2*](#_TOC_250010)
  2. [Application 4](#_TOC_250009)

Chapter 2. Problem Definition 6

Chapter 3. Technology Used 7

* 1. Hardware and Software Requirement 7
  2. [Description of libraries used 7](#_TOC_250008)

Chapter 4. Implementation 9

4.1 Algorithms

|  |  |  |  |
| --- | --- | --- | --- |
|  | 4.1.1 | Decision Tree(DT) . . . . . . . . . . . . . . . . . . . . . . | . |
| 4.1.2 | Naïve Bayes (NB) . . . . . . . . . . . . . . . . . . . . . . . |  |
| 4.1.3 | Support Vector Machine (SVM) . . . . . . . . . . . . . . . . . . | . |
| 4.1.4 | Random forest classifier (RF) . . . . . . . . . . . . . | . |
| 4.2 | Design | of system . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | . |
|  | 4.2.1 | Dataset . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | . |
|  | 4.2.2 | Preprocessing . . . . . . . . . . . . . . . . . . . . . . . . . . . . | . |
|  | 4.2.3 | Load data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | . |
|  | 4.2.4 | Analyze features . . . . . . . . . . . . . . . . . . . . . . . . . . | . |
|  | 4.2.5 | Modeling and predicting with machine learning . . . . . . . . . . | . |
|  | 4.2.6 | Finding the result . . . . . . . . . . . . . . . . . . . . . . . . . . | . |

Chapter 5. Result and Analysis 17

Chapter 6. Conclusion and Future Scope 18

Chapter 7. Case Study 19

* 1. Problem Definition 19
  2. [Introduction 19](#_TOC_250004)
  3. [Executive Summary 19](#_TOC_250003)
  4. [Conclusion 20](#_TOC_250002)
  5. [Future Scope 20](#_TOC_250001)

[List of References 21](#_TOC_250000)

# Chapter 1 :

**INTRODUCTION**

Health is one among the world challenges for humanity. World health organization (WHO) has mentioned that for an Individual proper health is the fundamental right. So to keep people fit and healthy proper health care services should be provided. 31 percentage of all deaths worldwide are because of heart related problems . Diagnosis and treatment of heart disease is very complex, particularly in developing countries, due to the lack of diagnostic devices and a shortage of physicians and other resources affecting proper prediction and treatment of cardiac patients. With this concern in the recent times computer technology and machine learning techniques are being used to develop software to assist doctors in making decision of heart disease in the preliminary stage. Early stage detection of the disease and predicting the probability of a person to be at risk of heart disease can reduce the death rate. Medical data mining techniques are used in medical data to extract meaningful patterns and knowledge. Medical information has redundancy, multi-attribution, incompleteness and a close relationship with time. The problem of using the massive volumes of data effectively becomes a major problem for the health sector. Data mining provides the methodology and technology to convert these data mounds into useful decision-making information. This predication system for heart disease would facilitate Cardiologists in taking quicker decisions so that more patients can receive treatments within a shorter period of time, resulting in saving millions of life .

CardioVascular disease (CVD) or also known as heart disease include blood and heart of the human body. myocardial infarction (as a heart attack) is also a part of the CVD. Another type of Heart Disease is called Coronary Heart Disease(CHD), in this type of disease, a substance called Plaque develop in the coronary arteries. The development of plaque can block the vessel completely through the course of time. The symptoms of the Heart Attack : 1. Chest Pain: The most common sign of a heart attack is chest pain. It mainly happens cause of the blockage of the coronary vessel of the body due to the plaque. 2. Arms pain: The pain normally starts in the chest and move towards the arm mainly left arm. 3. Low in oxygen: Because of the plaque the level of oxygen drops in the body and causes the dizziness and loss of balance. 4. tiredness: this cause for fatigues means simple chores become harder to do. 5. Excessive Sweating: Another common symptom is sweating. 6. Diabetics: In this, the patients have a heart rate of ~ 100 bpm and also occasionally having a heart rate of 130bpm. 7. Bradycardia: In this, the patient will have a slower heartbeat of 60 bpm. 8. Cerebrovascular Disease: The patient will have a high heart rate than normal usually of 200 bpm and higher than this can cause a Heart attack.

Some other reasons for the occurrence of a heart disease are lifestyle habits like smoking and certain eating habits. An estimated assumption is that more than 17.5 million deaths occur because of cardiovascular disease in the whole world. In India, there is more than 30 million heart disease patient right now. In India, more than 2 lake open heart surgeries are done per year. The patients affected by the heart attack is growing in India is 20% to 30 % every year. The development of the sensor network in the human monitoring system is more applicable from recent years. In this project, we show the use of sensor collected data which automatically generated by the sensor how it can be applied in Machine learning algorithms. Our project is based on the data of the sensor which collects the human heart rate. By using the sensor data and applying it in Machine learning algorithm we will predict the heart disease. The rest of the paper is described in the following manner. Section 2 describes the proposed system. Section 3 describes the Architecture. Section 4 includes experimental result and analysis..

## Methods:

Data mining provides the methodology and technology to convert data mounds into useful decision-making information. In this research the comparison of different machine learning techniques like- Support Vector Machine, Decision Tree, Random Forest, Naive Bayes are implemented to predict heart disease. Naïve mathematician used probability for predicating heart disease, SVM used on classification and regression technique, Random Forest works with varied decision Tree. These algorithms show different accuracy. We will try to tuning our techniques to obtain better accuracy which will be beneficial for more accurate prediction .

## 

## 1.2Objectives:

The main objective of this study is to predict weather a patient is affected with heart disease or not using different machine learning algorithms on a qualified dataset. Find out the co- relations between different attributes . Obtaining clear idea of our proposed data mining techniques and analyze the result and comparing between the results of different data mining techniques. We will analyze our techniques if there is any possibility to bring improvement for our results.

# 

# Chapter 2:

# PROBLEM DEFINITION

Health is one among the world challenges for humanity. World health organization (WHO) has mentioned that for an Individual proper health is the fundamental right. So to keep people fit and healthy proper health care services should be provided. 31 percentage of all deaths worldwide are because of heart related problems . Diagnosis and treatment of heart disease is very complex, particularly in developing countries, due to the lack of diagnostic devices and a shortage of physicians and other resources affecting proper prediction and treatment of cardiac patients. With this concern in the recent times computer technology and machine learning techniques are being used to develop software to assist doctors in making decision of heart disease in the preliminary stage. Early stage detection of the disease and predicting the probability of a person to be at risk of heart disease can reduce the death rate. Medical data mining techniques are used in medical data to extract meaningful patterns and knowledge. Medical information has redundancy, multi-attribution, incompleteness and a close relationship with time. The problem of using the massive volumes of data effectively becomes a major problem for the health sector. Data mining provides the methodology and technology to convert these data mounds into useful decision-making information. This predication system for heart disease would facilitate Cardiologists in taking quicker decisions so that more patients can receive treatments within a shorter period of time, resulting in saving millions of life .

# 

# Chapter 3 TECHNOLOGY USED

## Hardware & Software Requirement

* + - Google Colabaratory
    - Notepad++
    - MySQL
    - Web Browser – Chrome
    - Kaggle
    - Personal Computer with Minimum 256 MB HDD, 2 GB RAM
    - Python 3.9

## Description of Libraries Used

* + 1. **Numpy**

NumPy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High-end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

## Pandas

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high-level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

## SciPy

SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

## SciKit-Learn

Scikit-learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit-learn supports most of the supervised and unsupervised learning algorithms. Scikit-learn can also be used for data-mining and data-analysis, which makes it a great tool who is starting out with ML.

## Matplotlib

Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides

features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc.

## Surprise

Surprise is a Python scikit for building and analyzing recommender systems that deal with explicit rating data. Give users perfect control over their experiments. To this end, a strong emphasis is laid on documentation, which we have tried to make as clear and precise as possible by pointing out every detail of the algorithms. Alleviate the pain of Dataset handling. Users can use both built-in datasets (Movielens, Jester), and their own custom datasets. Provide tools to evaluate, analyse and compare the algorithms’ performance. Cross-validation procedures can be run very easily using powerful CV iterators (inspired by scikit-learn excellent tools), as well as exhaustive search over a set of parameters.

## OS

This module provides a portable way of using operating system dependent functionality. If you just want to read or write a file see open(), if you want to manipulate paths, see the os.path module, and if you want to read all the lines in all the files on the command line see the fileinput module. For creating temporary files and directories see the tempfile module, and for high- level file and directory handling see the shutil module.

## Random

This module implements pseudo-random number generators for various distributions. For integers, there is uniform selection from a range. For sequences, there is uniform selection of a random element, a function to generate a random permutation of a list in-place, and a function for random sampling without replacement. On the real line, there are functions to compute uniform, normal (Gaussian), lognormal, negative exponential, gamma, and beta distributions. For generating distributions of angles, the von Mises distribution is available.

## Seaborn

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics. For a brief introduction to the ideas behind the library, you can read the introductory notes. Visit the installation page to see how you can download the package and get started with it. You can browse the example gallery to see what you can do with seaborn, and then check out the tutorial and API reference to find out how.

## 

## Chapter 4

## IMPLEMENTATION

## 4.1 Algorithms

In machine learning we can use different algorithms otherwise known as classifiers to help us predict for our project. Here in our project we are looking forward to predict the number of patient that have heart disease and the number of patient that do not have heart disease running four algorithm to our data set. The reason we are going to use four is that it will allow us to get better and more reliable prediction . Because if we are using one algorithm or classifier and do not have anything else to compare it with then we cannot say that it a reliable prediction because it might be giving us a very good accuracy but this algorithm might not be the best or more appropriate one to use for our scenario. Whereas if we use more than one algorithm or classifier in our case four of them, we can compare them with one another and if we find one classifier is giving us accuracy that is not even in the ball park of the other algorithm provided accuracy we can understand that something is going wrong. It can be that the algorithm itself is not suitable for the job or we made a mistake in our coding So using more than one algorithm is essential for any prediction based system. Now the algorithms that we have chosen to use in our project are: 1. Decision tree,

2. Naïve Bayes, 3.SVM (support vector machine) and lastly 4. Random Forest. We will be discussing each of those algorithms below.

### DECISION TREE (DT)

For our first algorithm we will be using Decision Tree classifier. It is one of the most popular machine learning algorithms to this date . They are used for both classification and regression problems. Now a question might arise why we are willing to use Decision tree classifier over other classifiers. To answer that question we can bring about two reasons.

being, Decision trees often tries to mimic the same way human brain thinks so it is quite simple to understand the data and come

to some good conclusions or interpretations. Second reason can be, Decision trees actually allows us to see the logic for the data to interpret rather than being a black box algorithm like SVM, NN and others. It has the specialty of being simple and clear easily becoming one of the favorite among programmers of this generation. Now that we have discussed why Decision tree is good let us look further into what actually is Decision tree classifier. To start a decision tree is a tree where there are a bunch of nodes and each node represent a feature (attribute), each link (branch) represent a decision otherwise known as rule and each leaf of the tree represent an outcome otherwise known as categorical or continues value. The idea is to create a tree for the entire data and get an outcome at every leaf .

### NAIVE BAYES (NB)

We have already talked about our first machine learning algorithm, the decision tree classifier. Now we are going to talk about our second machine algorithm which we are going to use for our prediction purposes which is named Naive Bayes classifier. To give a brief introduction of naïve bayes algorithm we can say that it is a collection of classification algorithms based on Bayes Theorem [[7](#_bookmark42)]. We cannot call it a single algorithm because it is a family of algorithms where all of them share a common principle [[7](#_bookmark42)]. Here every pair of features that are being classified is independent from each other. Now in order to understand it better we are going to use our previous data set of weather and if we can play or not. We are going to analyze the features – outlook, temperature, humidity, windy and predict Play golf with yes or

### 4.1.3 Support Vector Machine (SVM):

Before we start discussing about SVM (support vector machine) we need to be accustomed with linear regression and

Logistic regression algorithms. If not it is suggested to look at them before moving on to support vector machine. Support vector machine is another simple algorithm that every machine learning expert should have in his or her arsenal. In this scenario Support vector machine is highly preferred by many as it produces significant accuracy with less computation power. Moreover support vector machine, abbreviated as SVM can be used for both regression and classification tasks. But it is widely used in classification objectives.

**4.1.4 RANDOM FOREST CLASSIFIER (RF):**

It is also known to be one of the most used algorithms, because its simplicity and the fact that it can be used for both classification and regression tasks. Here down below we are going to show how the random forest algorithm works and several other important things about it. First of all let us discuss about how it actually works. For starters random forest is a supervised learning algorithm. Like you will already see from its name, it creates a forest and makes it somehow random. The ‘forest’ it builds, is an ensemble of decision trees, most of the time trained with the ‘bagging’ method. The general plan of the textile technique is that a mixture of learning models will increase the result. If you want to say it in simple words: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction .

**4.2 Design of system**

In this portion of our report we are going to discuss how we prepared or designed the whole system. In terms of how we executed the system it will be discussed later in the book.

### 4.2.1 Dataset

We found our data set that has been used in our book from kaggle [(https://www.kaggle.com/](http://www.kaggle.com/) ronitf/heart-disease- uci/version/1) [[22](#_bookmark57)]. The dataset that we used in our thesis has in total 14 columns and 303 rows. First 13 of those columns are the features that we will be using later on in order to predict the final column ‘diagnosis’ which will tell us if the patient is going to be affected by heart disease or not. The 303 rows represents data of 303 patients that we found from the dataset.

### 4.2.2 Preprocessing

Before we start let us give a brief information about what data preprocessing actually is. Data preprocessing may be a data processing technique that involves remodeling data into a lucid format. Real-world data is often incomplete, inconsistent and lacking in certain behaviors or trends and is likely to contain many errors [[22](#_bookmark57)]. Data preprocessing may be a tried technique of partitioning such problems. Data preprocessing prepares raw data for further processing. Data preprocessing is used in database-driven applications such as customer relationship management and rule-based applications. For our thesis we are using standard scaler from the sklearn library for preprocessing our data. We choose this one over the many other ones because it suits very well with our system.

### 4.2.3 Load data

We created an array called col names and put down all our columns on that array. Then we read the csv file also known as the dataset find.

### 4.2.4 Analyze features

In this section we are going to distribute the target value is vital for choosing appropriate accuracy metrics and consequently properly assess different machine learning models. First of all we are going to count values of explained variable otherwise known as the determining variable which is going to give us the prediction of a patient being affected by heart disease or not [[22](#_bookmark57)]. Second of all we are going to separate numeric features from categorical features. Then we are going to show the relation between the categorical features in various plots and try to figure out or rather observe the influence of those categorical features in the actual determining variable “diagnosis”.

### 4.2.5 Modeling and predicting with machine learning

The main goal of the entire project is to predict heart disease occurrence with the highest accuracy. In order to achieve this we will test several classification algorithms. This section includes all results obtained from the study and introduces the best performer according to accuracy metric. I have chosen several algorithms typical for solving supervised learning problems throughout classification methods. First of all, let’s equip ourselves with a handy tool that benefits from the cohesion of SciKit Learn library and formulate a general function for training our models. The reason for displaying accuracy on both, train and test sets, is to allow us to evaluate whether the model over fits or under fits the data (so-called bias/variance tradeoff). Then we are going to split the data then test and train them in the ratio of 70:30. Then we are going to create a model where we are going to run all our algorithms [[24](#_bookmark59)].

### 4.2.6 Finding the result

At the end we are going to create a summery table where we are going to show the different accuracy percentage of different algorithms [[25](#_bookmark60)]. Where we are going to find out that it does not come as a surprise that the more complex algorithms like SVM and Random Forests generated better results compared to the basic ones. It is worth to emphasize that in most cases hyper parameter tuning is essential to achieve robust results out of these techniques. By producing decent results, simpler methods proved to be useful as well. Machine learning has absolutely bright future in medical field. Just imagine a place where heart disease experts are not available. With just basic information about a certain patient’s medical history, we may quite accurately predict whether a disease will occur or not [[25](#_bookmark60)]. We are going to discuss them more in details in the later section.

**Chapter 5:**

**RESULT AND ANALYSIS**

In this what we found is during small datasets in some other cases most of time decision trees direct us to a solution which is not accurate, but when we look at Naïve Bayes results we are getting more accurate results with probabilities of all other possibilities but due to guidance to only one solution decision trees may miss lead. Finally we can say by this experiment that Naïve Bayes is more accurate if the input data is cleaned and well maintained even though ID3 can clean it self it cannot give accurate results every time, and in this same way Naïve Bayes also will not give

accurate results every time we need to consider results of different algorithms and by all its results if a prediction is made it will be accurate. But we can use Naïve Bayes consider variables as indi-vidual we can use combination of algorithms like Naïve Bayes and K-means to get accuracy

## Chapter 6

## CONCLUSION & FUTURE SCOPE

## Conclusion:

In this reasearch we have tried to compare different machine learning algorithms and predict if a certain person, given various personal characteristics and symptoms, will get heart disease or not. The main motive of our report was to comparing the accuracy and analyzing the reasons behind the variation of different algorithms. We have used Cleveland dataset for heart diseases which contains 303 instances and used 10-fold Cross Validation to divide the data into two sections which are training and testing datasets [[5](#_bookmark40)]. We have considered 13 attributes and implemented four different algorithms to analyze the accuracy. By the end of the implementation part, we have found Gaussian Naive Bayes and Random Forest are giving the maximum accuracy level in our dataset which is 91.21 percent and Decision Tree is performing the lowest level of accuracy which is 84.62 percent. Probably for other instances and other datasets other algorithm may work in better way but in our case we have found this result [[6](#_bookmark41)]. Moreover, if we increase the attributes, maybe we can found more accurate result but it will take more time to process and the system will be slower than now as it will be little more complex and will be handling more data’s. So considering these possible things we took a decision which is better for us to work with.

## Future Scope:

The dataset that is used in our thesis is very small and old. Moreover, no new dataset regarding heart disease has been introduced so far. There is a need of new dataset and we can collect that from various hospitals of Bangladesh . We can also evaluate the efficiency of each individual classifier and also such classifiers in combination, by employing the bagging, boosting and stacking technique.

**Chapter 7**

**CASE STUDY**

**7.1 Problem statement: Diabetes prediction**

Diabetes is one of the frequent diseases that targets the elderly population worldwide.

According to the International Diabetes Federation, 451 million people across the world were diabetic

in 2017. The expectations are that this number will increase to affect 693 million people in the coming

26 years [1]. Diabetes is considered as a chronic disease associated with an abnormal state of the

human body where the level of blood glucose is inconsistent due to some pancreas dysfunction that

leads to the production of little or no insulin at all, causing diabetes of type 1 or cells to become

resistant to insulin, causing diabetes of type 2 [2,3]. The main cause of diabetes remains unknown,

yet scientists believe that both genetic factors and environmental lifestyle play a major role in

diabetes. Even though it’s incurable, it can be managed by treatment and medication. Individuals

with diabetes face a risk of developing some secondary health issues such as heart diseases and nerve

damage. Thus, early detection and treatment of diabetes can prevent complications and assist in

reducing the risk of severe health problems. Many researchers in the bioinformatics field have

attempted to address this disease and tried to create systems and tools that will help in diabetes

prediction. They either built prediction models using different types of machine learning algorithms

such as classification or association algorithms. Decision Trees, Support Vector Machine (SVM), and

Linear Regression were the most common algorithms .

**7.2 Introduction :**

Diabetes is one of the frequent diseases that targets the elderly population worldwide.

According to the International Diabetes Federation, 451 million people across the world were diabetic

in 2017. The expectations are that this number will increase to affect 693 million people in the coming

26 years [1]. Diabetes is considered as a chronic disease associated with an abnormal state of the

human body where the level of blood glucose is inconsistent due to some pancreas dysfunction that

leads to the production of little or no insulin at all, causing diabetes of type 1 or cells to become

resistant to insulin, causing diabetes of type.

7.3 Executive summary :

Diabetes is one of the most common diseases worldwide. Many Machine Learning (ML)

techniques have been utilized in predicting diabetes in the last couple of years. The increasing

complexity of this problem has inspired researchers to explore the robust set of Deep Learning (DL)

algorithms. The highest accuracy achieved so far was 95.1% by a combined model CNN-LSTM.

Even though numerous ML algorithms were used in solving this problem, there are a set of

classifiers that are rarely used or even not used at all in this problem, so it is of interest to determine

the performance of these classifiers in predicting diabetes. Moreover, there is no recent survey that

has reviewed and compared the performance of all the proposed ML and DL techniques in addition

to combined models. This article surveyed all the ML and DL techniques-based diabetes predictions

published in the last six years. In addition, one study was developed that aimed to implement those

rarely and not used ML classifiers on the Pima Indian Dataset to analyze their performance. The

classifiers obtained an accuracy of 68%–74%. The recommendation is to use these classifiers in

diabetes prediction and enhance them by developing combined models.

**7.4 Conclusion:** Researchers are passionate to try different types of classifiers and build new models with an

effort to enhance the accuracy of diabetes prediction. In this paper, the same vision was followed to

reach high prediction accuracy. All the Machine Learning (ML) and Deep Learning (DL) classifiers

that have been used in the last six years were reviewed regarding their frequency of use and accuracy.

ML classifiers with one or zero frequency have been implemented on the PID dataset to set

recommendations regarding their usage. The obtained accuracy by these ML techniques was 68%–

74%. For the DL algorithms, the highest accuracy achieved by researchers was 95%. As a future work,

the non-used classifiers can be applied to other datasets in a combined model to enhance further the

accuracy of predicting the Diabetes disease.

**7.5 Future scope:**

The highest accuracy obtained by implementing the above classifiers was 74.48% attained by

REPTree. REP stands for Reduces Error Pruning. It is a fast decision tree learner that generates

multiple trees. Since it is a fast learner, its running time is short. It follows the principle of computing

the information gain with entropy. Also, this principle is applied to minimize the error results

between the actual and the desired output, which contributes in reaching a high accuracy [59].

Another high-performance model that belongs to the tree category is M5P. It stands for Multivariate

Regression prediction model. It has been used in the prediction of student performance and revealed

an accuracy of 97.17%. Also, when it is applied in this study, it performed very well with a Root Mean

Squared Error reaching a value of 0.43.

**REFERENCE**

1. Arnold, C. (1990). *Heart disease*. Franklin Watts.
2. Canfield, J., Hansen, M. V., and Rackner, V. (2005). *Heart disease*. Health Communica- tions.
3. Carney, R. M. and Freedland, K. E. (2010). Psychotherapies for depression in people with heart disease. *Depression and Heart Disease*, page 145–168.
4. Dittmer, L. (2012). *Heart disease*. Creative Education. [5]Goetz, T. (2010). *The decision tree*. Rodale.

[6]Gold, J. C. and Cutler, D. J. (2000). *Heart disease*. Enslow Publishers. [7]Healey, J. (2005). *Heart disease*. Spinney Press.

1. Hook, S. V. (2001). *Heart disease*. Smart Apple Media.
2. Jiang, W. and Xiong, G. L. (2010). Epidemiology of the comorbidity between depression and heart disease. *Depression and Heart Disease*, page 1–37.
3. Johansson, P. (1998). *Heart disease*. Enslow Publishers.
4. Klapholz, M. (2003). Heart failure in the elderly. *Heart Disease*, 5(4):241–243.
5. Mensah, G. A. (2009). The burden of valvular heart disease. *Valvular Heart Disease: A Companion to Braunwalds Heart Disease*, page 1–18.
6. Mervyn, L. and Mervyn, L. (1990). *Heart disease*. Thorsons.
7. Moodie, D. (2016). The journal congenital heart disease-2016. *Congenital Heart Disease*, 11(1):5–6.
8. Morris, P., Warriner, D., Morton, A., and Mayhew, P. (2016). *Heart disease*. JP Medical Ltd.
9. Naff, C. F. (2008). *Heart disease*. Greenhaven Press.
10. Otto, C. M. (2009). Evaluation of valvular heart disease by echocardiography. *ValvularHeart Disease: A Companion to Braunwalds Heart Disease*,
11. Rajamannan, N. M. (2009). Cellular, molecular, and genetic mechanisms of valvular heart disease. *Valvular Heart Disease: A Companion to Braunwalds Heart Disease*, page 39–54.
12. Roberts, W. C. and Ko, J. M. (2009). Clinical pathology of valvular heart disease.

*Valvular Heart Disease: A Companion to Braunwalds Heart Disease*, page 19–38.

1. Shavelle, D. M. (2009). Evaluation of valvular heart disease by cardiac catheterization and angiocardiography. *Valvular Heart Disease: A Companion to Braunwalds Heart Disease*, page 85–100.
2. Sheen, B. (2004). *Heart disease*. Thomson/Gale.
3. Silverstein, A., Silverstein, V. B., and Nunn, L. S. (2006). *Heart disease*. Lerner.
4. Stout, K. (2009). Valvular heart disease in pregnancy. *Valvular Heart Disease: A Companion to Braunwalds Heart Disease*, page 424–436.
5. Thomas, R. (2002). *Heart disease*. Vega.
6. Tiger, S. and Reingold, M. (1986). *Heart disease*. J. Messner.