Forecasting Of Heart Disease Using Data Mining

SUBMITTED BY:
MD. SHAHRIAR RASHID (19-41372-3)

Section: N

SUPERVISED BY:
DR. AFZORA NAHAR

SURVEY PROPOSAL SUBMITTED IN THE PARTIAL FULFILMENT OF THE REQUREMENTS FOR THE COURSE OF RESEARCH METHODOLOGY

Date of Submission:
16-12-2022

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF SCIENCE AND TECHNOLOGY
AMERICAN INTERNATIONAL UNIVERSITY - BANGLADESH

2022

Preface

In this report we have forecasted of heart diseases. This is a major problem not only in our country but also all over the world. Also, we have analyzed the data of different country's people for this report. For our report, we made a survey where different ages and different professional people from all over the country participated. From that data we used our method which showed the overall scenario of the efficiency of virtual class and its future implementations. Also, it will give us an idea of how many people are facing particular problems and what the step should fix in the near future to reduce heart diseases

Acknowledgment

We are heartily thankful to our course instructor, whose encouragement, supervision and support from the preliminary to the concluding level enabled us to develop an understanding of the subject. We also thankful to all the person who participated in this study. We greatly appreciate their sincere cooperation. Also, we want to thank some of us friends and group members who helped us with collecting all analytical data from all over the country. Lastly, we offer our regards and blessings to all of those who supported us in any respect during the completion of the study.

Table of Contents

Chapter & Title **Page No.**

Preface i
Acknowledgement ii
Table of contents iii
Abstract iv

1.0 Chapter 1: Introduction
1.1 Motivation 05
1.2 Problem Statement 06
1.4 Research Objective 08

2.0 Chapter 2: Methodology
Research and Experiments of our applications 08
2.1 Data Set 09
2.2 Experimental Set up 09
2.3 Data Pre-processing 09

3.0 Chapter 3: Findings and Analysis
3.1 Findings 10
3.2 Analysis of the result 11

4.0 Chapter 4: Conclusion
4.1 Summery 11
4.2 Limitation 12
4.3 Future direction 12

Chapter 5: Working Schedule---------- 13

Reference ---------- 14

Abstract

The human body's most important and vital organ is the heart. The proper operation and efficient working of our hearts are absolutely necessary for life. It is a significant cause of death in the modern world. One of the most important health problems facing people today is heart disease. It is reportedly the main cause of death across the globe. Medical specialists frequently find it challenging to predict a cardiac illness early on. There are many valuable hidden facts and information in the health sector today that might be used to make predictions, particularly in the realm of medicine. This paper how classification techniques in data mining can be applied for heart disease prediction. To predict and alert about any future coronary ailment in the patient’s techniques like Naïve Bayes, and Decision tree and KNN are applied and efficiency of these algorithms is compared.

1.0 Chapter 1: Introduction

1.1 Motivation:

In modern Healthcare system there is a huge amount of information accessible heart disease is one of the most striking cause of death around this world[1]. heart disease can happen for many reasons like chain smoking, intake of high fat-based foods, pressure problems, pollution so on. heart disease can occurred specially in old patients because of diet non-steroidal ant-inflammatory drugs and it leads even two wards death[2]. all this date can be analyst using date mining technique such as predictive data mining Naive bayes, decision tree, KNN, k-means cluster. we went to use this tool to predict early heart disease problem of a person to lower the cause of death. our goals is to maintain the privacy and security of patient and health information. Most common types of heart diseases, are listed in Table I:

Table 1. Heart diseasestypes

Heart Disease Types	Description
Coronary heart disease	Block in the coronary blood vessels, which reduces blood and oxygen supply to heart.
Angina pectoris	Chest pain will occur due to the insufficient supply of blood toheart.
Congestive heart failure	Heart is not able to pump enough blood.
Cardiomyopathy	Weakening or a change in the heart muscle.
Congenital heart disease	Defect in the structure of the heart or great vessel present at time of birth
Arrhythmias	Problem with the rate or rhythm of the heartbeat
Myocarditis	Inflammation of heart muscle , can affect heart muscle or heart's electrical system

The discomfort that a person experiences has a big impact on their heart disease symptoms. certain signs and symptoms typically recognized by the general public. However, common signs include heart palpitations, shortness of breath, and chest pain. Angina, also known as angina pectoris, is a type of chest discomfort that is present in many types of heart disease and is brought on when a portion of the heart does not receive enough oxygen. Stressful situations or physical exertion can cause angina, which often lasts less than 10 minutes. Heart attacks can also occur as a result of different types of heart disease. The signs of a heart attack are similar to angina except that they can occur during rest and tend to be more severe. The symptoms of a heart attack can sometimes resemble indigestion. Heartburn and a stomach ache can occur, as well as a heavy feeling in the chest. Other symptoms of a heart attack include pain that travels through the body, for example from the chest to the arms, neck, back, abdomen, or jaw, lightheadedness and dizzy sensations, profuse sweating, nausea and vomiting.

Data mining is the process of obtaining crucial decision-making data from a pool of historical records in order to analyze or anticipate the future. Without data mining, the information might be hidden and unidentifiable. One data mining technique is classification, which uses the historical data already in existence to predict future events or outcomes.

1.2 Problem Statement:

Forecasting the possibility of heart disease problem and to enable significant knowledge of heart disease. Heart diseases are now one of the major problems of health. Peoples dies because of lack of knowing the reason and proper treatment. We cannot determine if a person is going to have heart disease problem until the person is already affected or diagnosed.

1.3 Research Objective:

The main objective of the research are: Predict heart disease from an unknown dataset using data mining algorithms. To evaluate the performance of KNN, Naïve bayes in predicting heart disease.

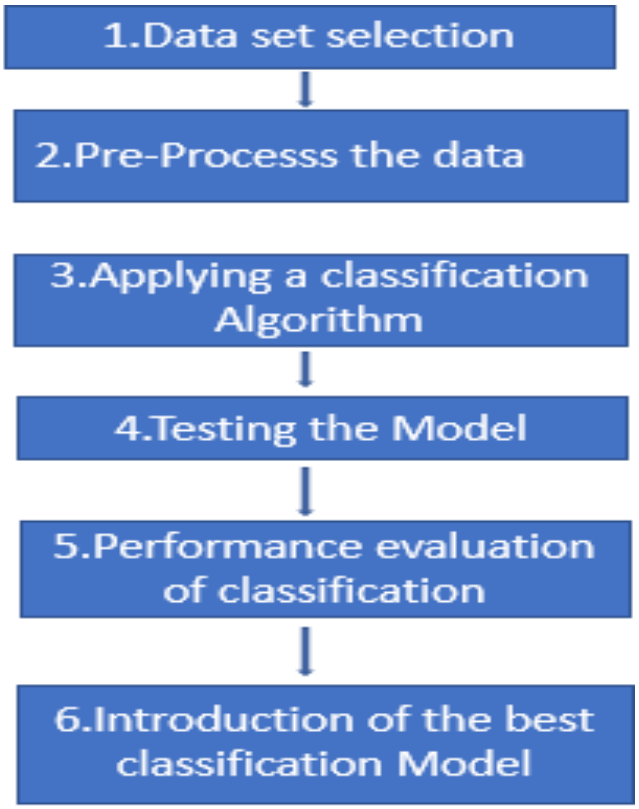


Fig:1 Sequential Overview

#	Features	Description	Data Type
1	age	age	Numeric
2	Sex	sex	Numeric
3	Cp	chest pain type (4 values)	Numeric
4	trestbps	resting blood pressure	Numeric
5	Chol	serum cholestoral in mg/dl	Numeric
6	Fbs	fasting blood sugar > 120 mg/dl	Numeric
7	Restecg	resting electrocardiographic results (values 0,1,2)	Numeric
8	Thalach	maximum heart rate achieved	Numeric
9	Exang	exercise induced angina	Numeric
10	oldpeak	ST depression induced by exercise relative to rest	Numeric
11	Slope	the slope of the peak exercise ST segment	Numeric
12	Ca	number of major vessels (0-3) colored by flourosopy	Numeric
13	thal	thal: 0 = normal; 1 = fixed defect; 2 = eversible defect The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.	Numeric
14	target	Class attributes	Numeric

Fig.2: Feature Description

Chapter 2: Methodology

2.0 Research and experiments of our application:

This section provides a detailed explanation of the approach employed, as displayed in Figures 1 and 2. Data pre-processing, feature selection, feature weight computation, using KNN and Naive Bayes, and model evaluation

2.1 Dataset:

This research uses the heart disease dataset that is obtained from Kaggle 7Machine learning Model [5]. Kaggle Machine Learning Repository is one of the largest available datasets. The Cleveland dataset from the Kaggle Machine Learning Repository is one of the heart disease datasets that researchers have utilized extensively up to this point[6]. This dataset, which has 1025 rows, will also be used in this study. The dataset has 14 attributes including class label are used

2.2 Experimental Setup:

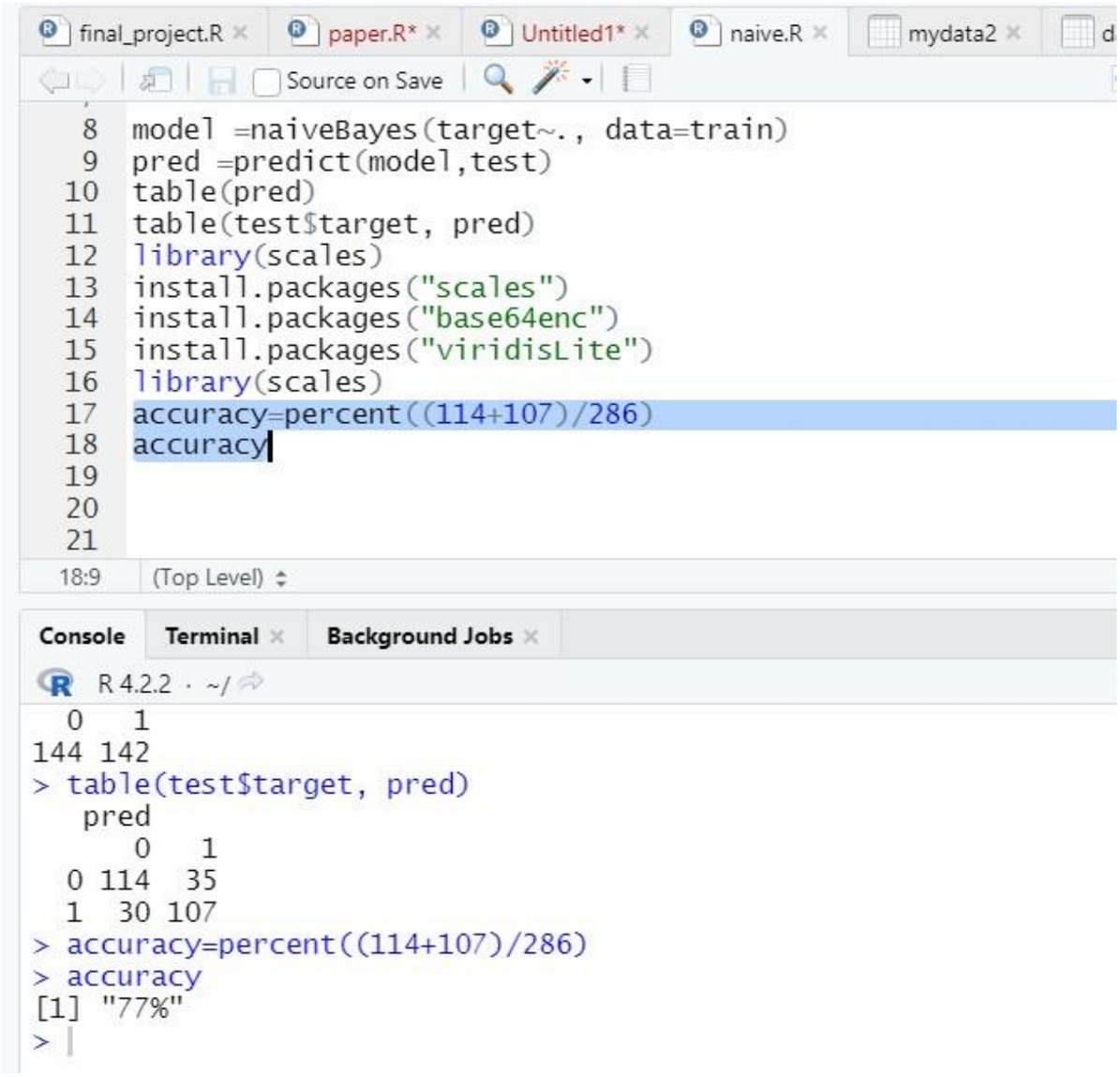
In this research, RStudio was used to conduct the experiments. The Cleveland dataset that was retrieved underwent pre-processing. The Cleveland dataset's 14 total attributes were used to find the significant features [5]. KNN and Naïve-Bayes was applied to the heart disease dataset to generate rules. Finally, accuracy was performed to obtain the prediction of heart disease using data mining technique based on significant features. The detailed explanation of each process is explained in the following sections.

2.3 Data Pre-Processing:

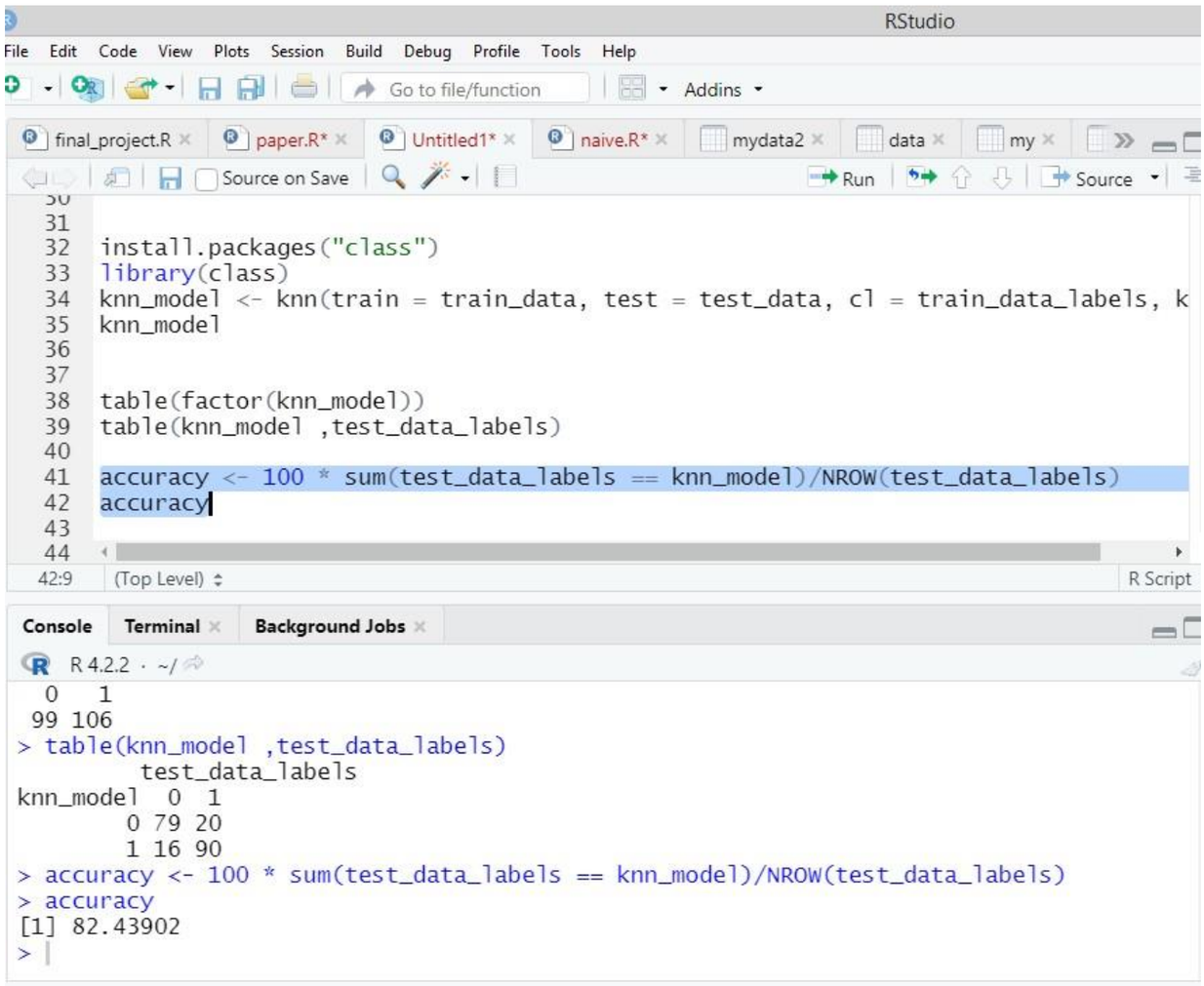
In the data pre-processing phase, first we normalize the data set so that the output remains unbiased. Because the scale of each variable is different, weights must be carried out to avoid bias. The KNN and Naïve-Bayes algorithm is applied to the training data set and the results are verified on the test data set. Splicing the data depends on the researcher, this time we used a 20% to test data and 80% data train. Then we build a machine learning Model. after building the model, then we can check the accuracy of forecasting using confusion matrix.

3.0 Chapter 3: Findings and Analysis

3.1 Findings:



Naïve bayes Method



KNN method

3.2 Analysis of the result:

Here two method used for the analysis. Naïve bayes method and KNN method. From both these method the accuracy is better in the KNN method. The accuracy of KNN method is almost 82% & the accuracy of naïve bayes method is 77%. KNN method gives more accurate result.

4.0 Chapter 4: Conclusion

4.1 Summery

Various classification techniques have been used in literature for prediction of heart disease. A significant accuracy has been achieved for such models. This paper discusses implementation of two classification methods; Naïve Bayes and KNN applied on clevarand and Statlog datasets with results showing higher accuracy using Naïve Bayes algorithm. Each classification algorithm has its own advantages and disadvantages, so a combination of such models may be used to achieve higher accuracy and to achieve scalability, such models may be applied on large data sets and can be validated.

4.2 Limitation

It is really tough to collect the data from all the people. Here limited amount of data have collected for the analysis. The symptoms can be changed time to time. Here recent symptoms taken to research.in future new symptoms can arise. It's just a prediction for the heart disease problem.

4.3 Future direction

The Future work of this research work can be made to produce an impact in the accuracy of the Decision Tree and KNN and Naive Bayes Classification for additional improvement after applying genetic algorithm in order to decrease the actual data for acquiring the optimal subset of attribute that is enough for heart disease prediction. Heart disease prediction automation using real-time data from healthcare institutions and agencies that can be developed using big data They can be provided as streaming data, and with the help of the data, a real-time study of the patients can be prepared.

Working Schedule

Week	Week (1-3)	Week (4-6)	Week (7-9)	Week (10-12)	Week (13-15)	Week (16-18)	Week (19-21)	Week (22-24)
Proposal Topic Selection	-							
Proposal Writing & Read Paper	-							
Data Collection		-	-					
Data Analysis			-	-				
Initial Report				-	-	-		
Final Report						-	-	
Typing							-	
Final Review								
Submission								

REFERENCE:

[1] Fadnavis, R., Dhore, K., Gupta, D., Waghmare, J. and Kosankar, D., 2021, May. Heart disease prediction using data mining. In Journal of physics: conference series (Vol. 1913, No. 1, p. 012099). IOP Publishing.

[2] Venkatalakshmi, B. and Shivsankar, M.V., 2014. Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology, 3(3), pp.1873-7.

[3] Yazdani, A., Varathan, K.D., Chiam, Y.K., Malik, A.W. and Wan Ahmad, W.A., 2021. A novel approach for heart disease prediction using strength scores with significant predictors. BMC medical informatics and decision making, 21(1), pp.1-16.

[4] Bahrami, B. and Shirvani, M.H., 2015. Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(2), pp.164-168.

[5] kaggle kernels output akashkotal/heart-disease-eda-with-7-machine-learning-model -p /path/to/dest

[6] Amin MS, Chiam YK, Varathan KD Identifcation of signifcant features and data mining techniques in predicting heart disease. Telem Inform. 2019;36;82–93.