

# **Prediction of Heart Diseases Using Machine Learning**

**BY**

**Md. Mubasshir Rahman**  
**ID: 181-15-11085**

**Mahmud Islam Sykat**  
**ID: 181-15-11025**

**Md. Golam Rafi Khan**  
**ID: 181-15-11070**

This Report Presented in Partial Fulfillment of the Requirements for  
The Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Abdus Sattar**  
Assistant Professor  
Department of CSE  
Daffodil International University

**Aniruddha Rakshit**  
Senior Lecturer  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**  
**DHAKA, BANGLADESH**  
**DECEMBER, 2021**

## APPROVAL

This Project/internship titled “**Prediction of heart diseases using machine learning**”, submitted by Md.Mubasshir Rahman , Mahmud Islam Sykat , Golam Rafi Khan, ID No: 181-15-11085 , 181-15-11025 , 181-15-11070 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 03-01-2022.

### BOARD OF EXAMINERS



**Chairman**

---

**Dr. S.M Aminul Haque**  
**Associate Professor and Associate Head**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**Internal Examiner**

---

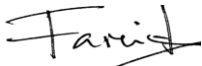
**Naznin Sultana**  
**Assistant Professor**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**Internal Examiner**

---

**Raja Tariqul Hasan Tusher**  
**Senior Lecturer**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



**External Examiner**

---

**Dr. Dewan Md. Farid**  
**Professor**  
Department of Computer Science and Engineering  
United International University

## DECLARATION

We hereby declare that, this thesis has been done by us under the supervision of **Abdus Sattar, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

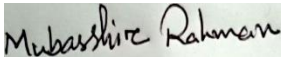
### Supervised by:



---

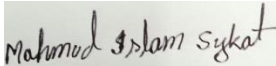
**Abdus Sattar**  
Assistant Professor  
Department of CSE  
Daffodil International University

### Submitted by:



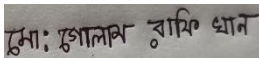
---

**Md. Mubasshir Rahman**  
ID: 181-15-11085  
Department of CSE  
Daffodil International University



---

**Mahmud Islam Sykat**  
ID: 181-15-11025  
Department of CSE  
Daffodil International University



---

**Golam Rafi Khan**  
ID: 181-15-11070  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible to complete the final thesis successfully.

We are really grateful and wish our profound indebtedness to **Abdus Sattar**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Human Computer Interaction in Education” to carry out this thesis. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior drafts and correcting them at all stages have made it possible to complete this thesis.

We would like to express our heartiest gratitude to **Pro. Dr. Touhid Bhuiyan, Professor, and Head**, Department of CSE, for his kind help to finish our thesis and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and passion of our parents.

## ABSTRACT

Now-a-days, we can see the number of heart disease cases increasing highly. Especially old people affected by this . It is so concerning for the world. We thought about this kind of disease and how we could predict this in advance. Though it's difficult to diagnose, it should be done correctly and quickly too. We made a prediction system named heart diseases prediction system , which uses a patient's medical data to predict whether or not they will be diagnosed with heart disease. The fundamental recognition of the studies paper is on which sufferers are extra likely to expand coronary heart sickness primarily based totally on numerous clinical characteristics. We used four machine learning algorithms to predict and classify heart disease patients such as Decision Tree, Logistic Regression, Random Forest Classifier and k-N Neighbor . To adjust how the version may be used to enhance the accuracy of prediction of Heart Attack in any individual , a very helpful approach was used . The proposed model's power changed into pretty satisfying, because it changed into capable of are expecting proof of getting a coronary heart sickness in a particular person the use of Logistic Regression , Random forest and Decision Tree which showed a high level of accuracy when compared to k-N Neighbor.

# TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-5</b>
1.1 Introduction	02-03
1.2 Motivation	03
1.3 Problem Definition	03
1.4 Research Methodology	04
1.5 Research Question	04
1.6 Research Objective	04
1.7 Report Layout	05
<b>CHAPTER 2: BACKGROUND</b>	<b>6-8</b>
2.1 Introduction	06
2.2 Related work	06-07
2.3 Bangladesh Perspective	08
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>9-13</b>
3.1 Introduction	09
3.2 Data Pre-Processing	10
3.3 Architecture of The Model	10-12
3.4 Modeling and Predicting with Machine Learning	12-13
<b>CHAPTER 4: PERFORMANCE OF THE PROPOSED MODEL</b>	<b>14-18</b>
4.1 Training, Testing and the Validation of the model	14
4.2 Accuracy of Models with All Features	14-15
4.3 Feature Engineering	15
4.4 Feature Importance	15
4.5 Accuracy of Models with Selected Features	16-18
<b>CHAPTER 5: RESULT AND ANALYSIS</b>	<b>19</b>
<b>CHAPTER 6: CONCLUSION AND FUTURE WORK</b>	<b>20</b>
<b>REFERENCES</b>	<b>21-22</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO.</b>
Figure 3.1 Steps of Data Collecting and Processing	09
Figure 3.2 Architecture of KNN	11
Figure 3.3 Architecture of DT	11
Figure 3.4 Architecture of Random Forest	12
Figure 3.5 Architecture of Logistic Regression	12
Figure 4.1 KNN Importance Checker	16
Figure 4.2 Decision Tree Importance Checker	17
Figure 4.3 Random Forest Importance Checker	17
Figure 4.4 Logistic Regression Importance Checker	18

## LIST OF TABLES

<b>TABLE NO.</b>	<b>PAGE NO.</b>
Table 4.1 Attributes List	14
Table 5.1 Accuracy Rate	19



# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

Machine learning is a methodology of determining and extracting constant, antecedently unknown/known, and probably relevant information knowledge". Machine Learning may be a broad and complicated range, and its breadth and application are developing all the time. Machine learning includes supervised, unattended, and ensemble learning classifiers that are won't to anticipate and notice the accuracy of a dataset. We will use that data to our HDPS initiative, because it can profit an oversized variety of people. vas diseases are quite common currently, they describe a spread of conditions that would have an effect on your heart. World health management appraisal that seventeen.9 million world deaths from (Cardiovascular-diseases ) CVDs. it's the first reason for deaths in adults. Our project will facilitate predicting the people that are doubtless to diagnose with a heart condition by facilitating their medical record. It acknowledges WHO all are having any symptoms of sickness are (heart condition /cardiomyopathy /cardiovascular disease) like hurting or high vital sign and may facilitate in identification disease with less medical tests and effective treatments, in order that they will be cured consequently. This project focuses on principally 3 data processing techniques namely: (1) Logistic regression, (2) KNN (3) Random Forest Classifier (4) Decision Tree. The accuracy of our project is 81% that is healthier than the previous system wherever just one data processing technique is employed. So, mistreatment of additional data processing techniques exaggerated the HDPS accuracy and potency. provision regression falls underneath the class of administered learning. solely separate values are utilized in provision regression. the target of this project is to see whether or not the patient is probably going to be diagnosed with any vas heart diseases supporting their medical attributes like gender, age, chest pain, abstinence sugar level, etc. A dataset is chosen from the UCI repository with the patient's medical record and attributes. By mistreating this dataset, we tend to predict whether or not the patient will have a heart condition or not. To predict this, we tend to use fourteen medical attributes of a patient and classify him if the patient is probably going to possess a heart condition. These medical attributes are competent underneath 4 algorithms: Logistic regression, K-N N ,

Random Forest Classifier and Decision Tree. Most effective of those algorithms is Logistic regression which provides the nation with an accuracy of 81%. And, finally we tend to classify patients that are in danger of obtaining a heart condition or not and conjointly this methodology is completely price economical.

## **1.2 Motivation**

The main goal of this study is to gift a cardiovascular disease prediction model that may be wont to predict the incidence of cardiovascular disease. what is more, the goal of this analysis is to search out the simplest classification algorithmic rule for detection the presence of cardiovascular disease in an exceedingly patient. This work is supported by a comparative study and analysis victimization 3 classification algorithms, namely KNN, supply Regression, and Random Forest Classifier, at numerous levels of analysis.

Although these square measure are unremarkably used machine learning algorithms, the guts illness prediction may be an important task involving highest attainable accuracy. Hence, the 3 algorithms square measure evaluated at various levels and kinds of analysis methods. this can give researchers and medical practitioners to determine a far better understanding and facilitate them determine an answer to spot the simplest technique for predicting the guts diseases.

## **1.3 Problem Definition**

In today's world, artificial intelligence (AI) is a very important term in the terrain of ICT[11]. AI will aid in the development of our health-care sector. It is essential to identify the problems and associated necessity in this field in order to provide an appropriate solution. To implement AI in the health sector, it's also needed to understand government protocol or etiquette, software corporation demands, and course methodologies. Conduct a brief survey of patients and developers to learn about health-related issues[11].

## **1.4 Research Methodology**

The Experiment Data Set, Data Pre-processing, Model Architecture, Learning Rate and Optimizer, and Model Training are all discussed in this section of our research paper. The proposed model's performance will be described at the end of this chapter.

## **1.5 Research Questions**

The main questions of our research is given below:

1. What is the current situation of heart diseases in Bangladesh?
2. What process should be helpful in this kind of research?
3. Which method is useful for this type of research?

## **1.6 Research Objectives**

There are some advantages to using AI to predict heart disease. Artificial intelligence has some technical and medical goals.

The following are some of the technical goals:

- Create an effective model for detecting heart diseases.
- To encourage software developers to use the model to work with AI.
- Embedding the model in mobile apps and websites is a must.

The following are some of the medical goals:

- Assist patients in detecting heart diseases.
- Assist the patients in becoming self-sufficient.
- Reducing the cost of treatment is a good idea.

## **1.7 Research Layout**

Chapter 1: will cover the following topics: introduction, motivation, problem definition, research question, research methodology, and our project's expected outcome.

Chapter 2: will discuss the background of this research and related work, as well as the current state of affairs from the perspective of Bangladesh, as well as government goals and regulations.

Chapter 3: will describe the state of AI in Bangladesh's Cardiology sector of medical science.

Chapter 4: will look at the future of AI in medical science from a development standpoint.

Chapter 5: focuses on the outcomes and benefits of AI in medical science.

Chapter 6: It describes the conclusion of this research.

# **CHAPTER 2**

## **BACKGROUND**

### **2.1 Introduction**

One of humanity's greatest challenges is health. consistent with the globe Health Organization (WHO), a human right to physiological condition is key. As a result, correct health care services ought to be provided to stay individuals working and healthy. cardiovascular disease is answerable for 31% of all deaths worldwide. The diagnosing and medication of cardiovascular disease is very troublesome, especially in growing countries, because of a scarcity of diagnostic devices, physicians, and different resources that have an effect on correct internal organ patient prediction and treatment. In response to the present concern, engineering and machine learning techniques have recently been wont to develop software systems to help doctors in creating preliminary selections concerning cardiovascular disease. The death rate may be reduced by early detection of the malady and predicting the chance of an individual being in danger of cardiovascular disease. Medical data processing techniques are n't to extract important patterns and information from medical knowledge. Redundancy, multi-attribution, unity, and an in depth relationship with time characterize medical knowledge. The health sector faces a serious challenge in effectively utilizing large amounts of knowledge. The technique and automation for changing these knowledge mounds into helpful managerial data is provided by data processing. This cardiovascular disease prediction system would enable cardiologists to form quicker selections, permitting additional patients to receive treatment in a very shorter quantity of time, doubtless saving immeasurable lives.

### **2.2 Related Works**

At the Survey of Techniques for Data Mining on Medical Data for Finding Frequent Diseases in the Local Environment, Abdul Khaleel delivered a paper. This research delves into the information mining processes needed for medicinal data mining, especially to locate locally visited disorders including heart disease, lung cancer, and bosom disease, among others. For employed information mining to assess and diagnose cardiac illness. Information mining is a process of gathering data for the goal of discovering inactive examples. In this situation, the Naive Bayes

method was applied. As a result, Naive Bayes offers a lot of flexibility when it comes to making assumptions [1].

There was a paper called Remote. Based on the first month's data and information from the intervention's baseline, predicting the efficiency of a health monitoring system. RHM systems are both expensive and efficient when it comes to decreasing disease. Wanda- CVD is an enhanced RHM framework based on mobile phones that was developed to give members remote training and social assistance. CVD prophylaxis is recognized as a must-have. This is a topic that social insurance groups all across the world are focusing on [2].

In a study titled Prediction for illness similarities by utilizing ID3 algorithm in television and mobile phone, Writers reported their findings. This research describes methods to detect designs that are masked by cardiovascular disease in a planned and hidden way. Data mining methods such as the ID3 algorithm are used in the proposed framework. This proposed strategy has the potential to lower both the death rate and the number of illness sufferers [3].

The researcher proposed a disease prediction system based on data mining techniques. This paper discusses MAFIA (Maximum Frequent Item Set Algorithm) and K-Means classification. Because illness prognosis necessitates categorization. The classification accuracy is attained using MAFIA and K-Means [4].

A study titled Intelligence System for Cardiovascular - Related Diagnosis Level using K-Star Algorithm. In this research, they provide a paradigm for predicting heart infection. Studying The neural system's calculation vectors are measured. This framework's brain system detects 13 clinical variables as data and forecasts the patient's existence or lack of coronary disease, as well as several allot [5].

Heart Disease Predictions the researcher spoke about using the Learning Vector Quantization Algorithm. In this study, they provide a heart disease expectation architecture learning - based vector quantization and the neural system calculations The neural system in this framework detects 13 clinical variables as information and predicts the existence or absence of coronary disease in the patient, as well as numerous execution metrics [6].

## **2.3 Bangladesh Perspective**

Cardiovascular disease (CVD) is turning into AN more and more vital reason for morbidity and mortality in Bangladesh, because it is elsewhere. Thanks to medical specialty transition, the predominant malady pattern during this country has shifted from preponderantly communicable to preponderantly non-communicable malady, with CVD enjoying a big role within the latter. In fact, disorder (CVD), notably arteria malady (CAD), is on the increase. Acute infectious disease is on the decline, whereas cardiovascular disease and heart disease area unit on the increase. Despite some efforts, reliable information on numerous aspects of CVD is presently lean. cardiovascular disease, arteria malady, infectious disease, rheumatic heart condition, and stroke have current prevalence rates of 20-25 %, 4-6 %, 1/1000, and 0.3-1.0 percent, severally. Aside from ancient CVD risk factors, genetic predisposition and novel problems like high salt intake, arsenicosis, malnutrition D, and pollution might all play a role within the Etiopathogenesis of CVD during this population. In Bangladesh, developing applicable policies and putting a larger stress on bar ways might aid within the fight against CVD.

# CHAPTER 3

## RESEARCH METHODOLOGY

### 3.1 Introduction

The suggested machine technique's intention is to use as a group processes to boost the accuracy of coronary heart sickness prediction. The recommended machine's structure is depicted in Figure 1. It includes six stages: statistics collection, statistics preprocessing, characteristic selection, statistics splitting, version training, and version evaluation.

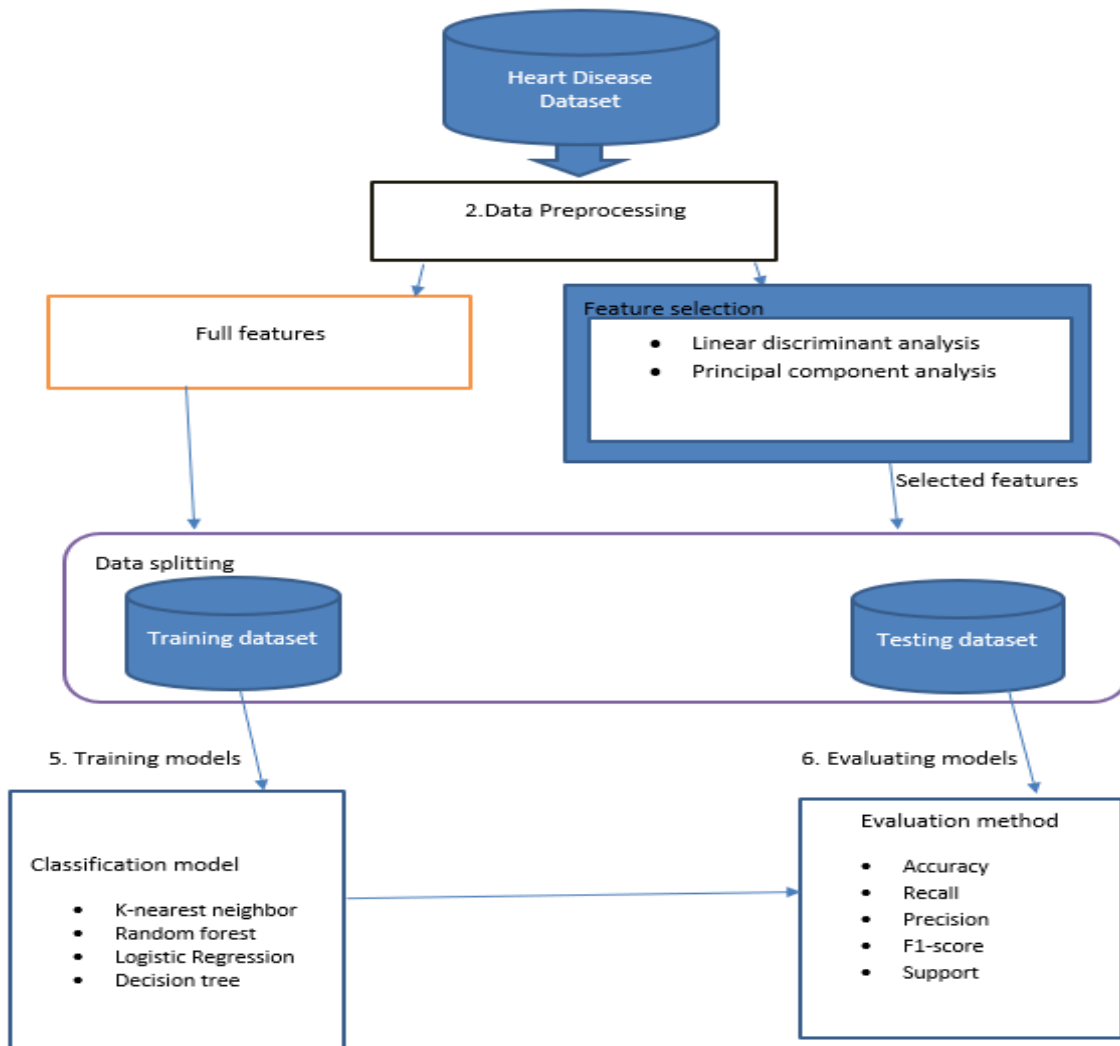


Figure 3.1: Research Methodology Steps



## 3.2 Data Pre-Processing

**Ablution:** Data that we need to method will now no longer be elegant, which means it is able to have noise or values which are lacking. If we method it, we can now no longer get first rate results, consequently we need to cast off all of this. The approach to cast off all of that is statistics cleaning. We'll fill in lacking statistics and decrease noise with the aid of using tactics inclusive of filling in lacking values with the maximum not unusual place value[15].

**Transformation:** Data transformation is changing statistics from one layout to every other so as to make it extra understandable with the aid of using the use of normalization, smoothing, and generalization procedures, in addition to combination techniques[15].

**Integration:** Data that we do now no longer want to method might also additionally come from more than one sources, and if we do now no longer integrate them, it is able to motivate a hassle in the course of processing. As a result, integration is a vital segment in statistics pre-processing, and diverse problems are mentioned right here to integrate[15].

**Reduction:** When operating with statistics, it is able to be complicated and tough to understand at times, as a consequence so as to make it comprehensible to the system, we can lessen it to an appropriate layout so as to achieve top results[15].

## 3.3 Architecture of the Model

Various types of machine learning algorithms: KNN, DT, RF, and LR are applied to allocate heart diseases.

(1). KNN is a nonparametric lazy studying set of rules that permits the prediction of the latest pattern category. It's utilized by lots of people. It may be utilized in each regression and category forecasting situations. It is, nevertheless, regularly utilized in category while implemented to commercial issues because it plays properly throughout all standards reviewed while comparing a technique's functioning, even though it is essentially used due to its ease of expertise and shorter calculation time [8–25, 27–30].

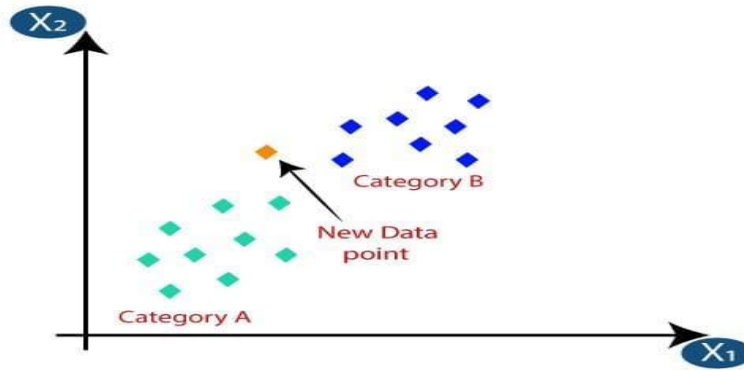


Figure 3.2

(2). DT is a tree form that works on the notion of custom. It is particular and consists of modern algorithms which may be applied in predictive modeling. Internal nodes, branches, and a terminal node have all been assigned to encompass them. Every internal node has a feature "take a look at," and branches have the take a look at conclusion, further to a class label for each leaf node. It can be used for class further to regression.

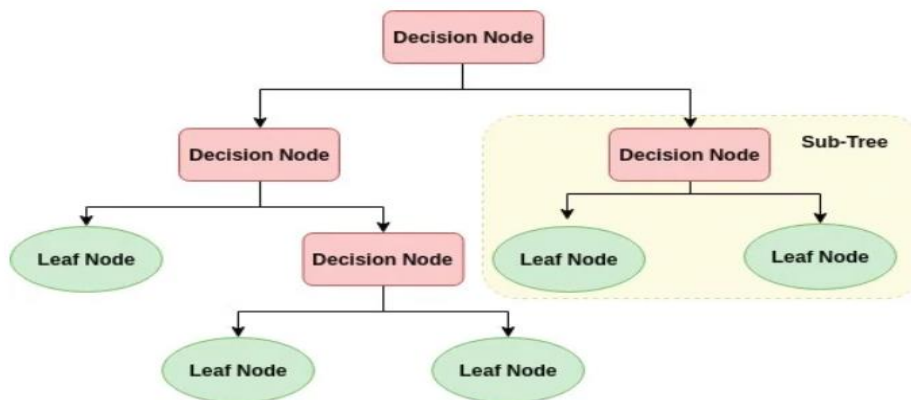


Figure 3.3

(3). Random selection forests are a sort of gadget gaining knowledge of a set of rules that may be used to resolve class and regression issues, in keeping with RF. The output is received with the aid of using improving all DT's output to work out a unmarried result, and the feature is accomplished with the aid of using building a distinctive variety of DT classifiers or regressors, and the feature is accomplished with the aid of using building a distinctive variety of DT classifiers or regressors, and the output is received with the aid of using improving all DT's result to settle a unmarried result.

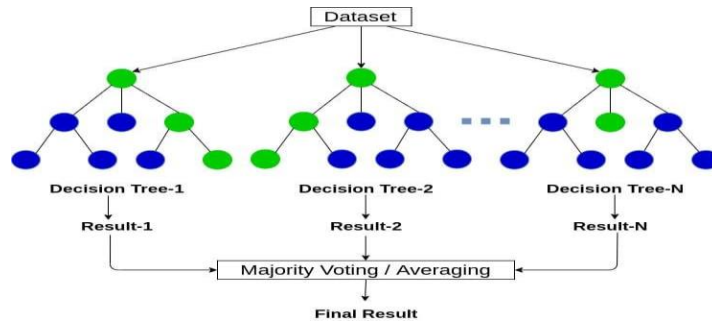


Figure 3.4

(4). Logistic regression is the most suitable regression method to make use of whilst the established fickle is angled (binary). Like different regression investigations, logistic regression is an auguring analysis. Logistic regression is used to explain records and provide an explanation for the connection among one established binary changeable and one or greater nominal, ordinal, interval, or ratio-degree impartial variables. Logistic regressions are probably hard to recognize at times; however, the Intellectus Statistics application makes the system easy after which explains the effects in simple English.

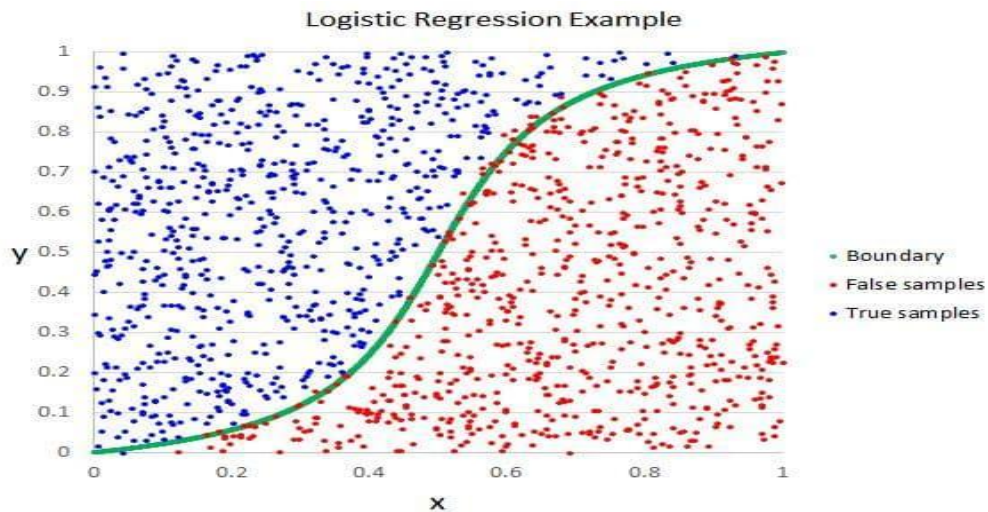


Figure 3.5

### 3.4 Modeling and predicting with machine learning

This phase compiles the study's findings and introduces the first-class acrobat based absolutely on accuracy metric. Throughout class access, I've decided on a few algorithms which might be

generally used to treat supervised learning challenges. First and foremost, permit us to create a useful tool that takes advantage of the Sci-Kit Learn library's harmony and enlarges a preferred operation for education of our models. The cause of presenting accuracy on every train and test gadgets is to look if the model over fits or beneath neath fits the data. After that, we can take the data and test and train them at 75:25 ratio. After that, we can enlarge a model in which we can execute all of our algorithms.

## CHAPTER 4

### PERFORMANCE OF THE PROPOSED MODEL

#### 4.1 Training, Testing and the Validation of the model

Attribute	Code given	Note	Values
1. age	Age	in years	Numeric
2. sex	Sex	1 = male; 0 = female	Binary
3. chest pain type	level of pain	0,1,2,3 4	Values
4. resting blood pressure	trestbps	in mm Hg	Numeric
5. serum cholesterol	cholesterol	in mg/dl	Numeric
6. fasting blood sugar	fbf	> 120 mg/dl	Numeric
7. resting electrocardiographic results	restecg	0,1,2 3	Values
8. maximum heart rate achieved	thalach	71–202	Numeric
9. exercise induced angina	exang	0,1	Binary
10. oldpeak = ST	oldpeak	depression	Numeric
11. the slope of the peak exercise ST segment	slope	0,1,2 3	Values
12. number of major vessels fluoroscopy	ca	0,1,2,3 4	Values
13 defect:normal;fixed;reversible; non-reversible	thal	0,1,2,3 4	Values
14. class	target	0,1	Binary

Table 4.1

#### 4.2 Models Accuracy with All Features

Different kinds of algorithms were used to gather the findings. We used the whole dataset with all abilities in our first take a look at and used K-Nearest Neighbor , Decision Trees, Random Forests, and Logistic Regression. Table 4.1 shows the accuracy of t several algorithms we used in our dataset.

		Predicted Class	
		High Risk (1)	Low Risk (0)
Actual Class	High risk (1)	True Positive (TP)	False Negative (FN)
	Low risk (0)	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} *100\%$$

### 4.3 Feature Engineering

The algorithm's accuracy can be influenced by a variety of factors. As a result, dealing with the features is critical. Some people may desire to work with certain aspects for a variety of reasons. We can train faster by selecting fewer characteristics. We can leverage interactions between the most significant aspects as new features by focusing on the most crucial characteristics. This can sometimes result in an unexpected improvement. Some characteristics are linked in a linear fashion to others. This could put the model under stress. Feature Selection refers to the process of selecting only the most necessary appearance in order to increase the algorithm's accuracy. It cuts down on training time and eliminates overfitting.

### 4.4 Feature Importance

What functions have the maximum effect on predictions, for example, is a quite essential inquiry we might also additionally ask of a model. This is known as function significance. A model's accuracy may be decreased with the aid of using some attributes in a few instances. As a result, it is vital to paint with precise qualities. So far, we have labored with all the dataset's attributes and in comparison the accuracy of numerous models. Now we need to analyze how the accuracies of various classifiers range after a subset of the attributes is selected. The relevance of a characteristic can be determined using K-Nearest Neighbor[Table 4.2] , Decision Tree [Table 4.2] ,Random Forest [Table 4.2], Logistic Regression[Table 4.2].

The K-Nearest Neighbor importance checker is shown in Figure 4.1.

The Decision Tree Importance Checker is shown in Figure 4.2.

The Random Forests Importance Checker is shown in Figure 4.3.

The Logistic Regression Importance Checker is shown in Figure 4.4.

## 4.5 Models Accuracy with Selected Features

We choose the following top features to see the difference in prediction after seeing the relevance of Table 4.2 features .

They are level of pain, thal, ca, age, cholesterol, thalach

### K-Nearest Neighbor

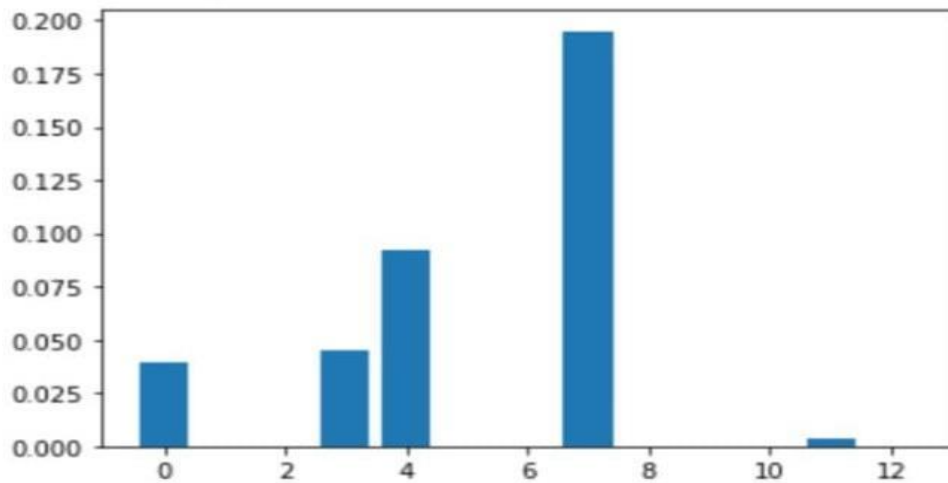


Figure 4.1

## Decision Trees

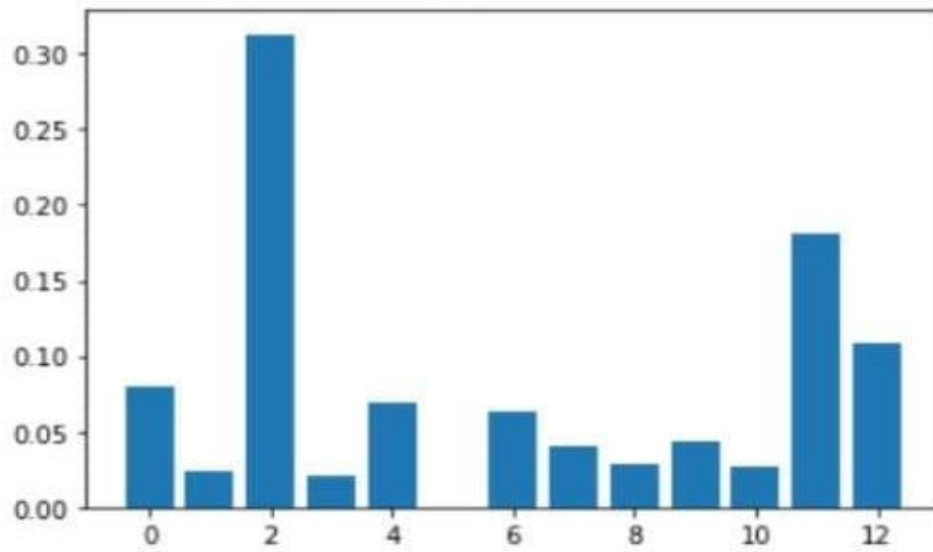


Figure 4.2

## Random Forests

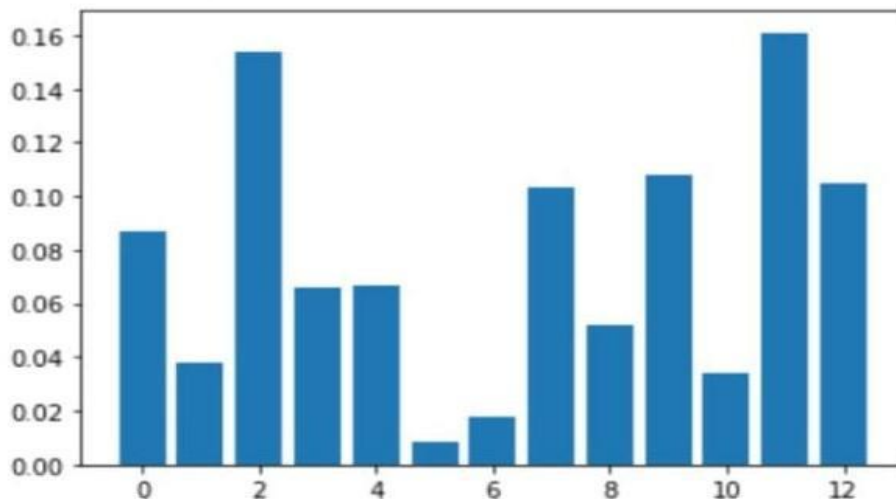


Figure 4.3



# Logistic Regression

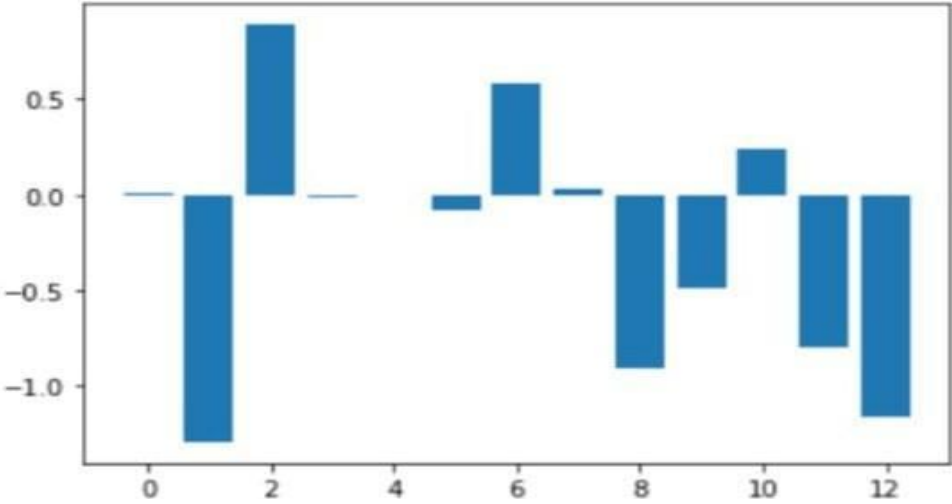


Figure 4.4

## CHAPTER 5

### RESULT AND ANALYSIS

We used four types of methods to justify which is more suitable for this kind of work . After implementing the methods we can see Logistic Regression is the more suitable with 81% accuracy.

No.	Classifier	Accuracy (%)
1	K-Nearest Neighbor	62%
2	Decision Trees	75%
3	Random Forests	77%
4	Logistic Regression	81%

Table 5.1

Different strategies carried out higher relying at the situation, whether or not move validation and characteristic choice have been applied or not, as visible withinside the tables above. Depending on the situation, any set of rules has the inherent cap-potential to outperform different algorithms. Random Forest, for example, plays extensively higher with a big range of datasets than with small datasets [11]. Missing values play a critical function in selection trees. Even after imputing, it's miles not able to supply the identical consequences as a wonderful dataset. On this dataset, Logistic Regression is the quality classifier [11]. The presumption that everyone's features are impartial is the reason for this. There might have been much less accuracy if there has been a dating among the attributes withinside the dataset.

## **CHAPTER 6**

### **CONCLUSION AND FUTURE WORK**

This study, we in comparison distinct device studying algorithms are expecting whether or not someone will expand coronary heart sickness or now no longer primarily based totally on diverse private functions and symptoms. Our report's predominant intention became to evaluate the accuracy of various algorithms and examine the reasons for his or her differences. The Cleveland dataset for cardiac illnesses, which has 303 instances, became applied to partition the statistics into groups, education and checking out data -sets[12], use of 10-fold short Validation. To verify the accuracy, we took under consideration thirteen elements and used 4 awesome algorithms. By the end of the implementation stage, we've discovered that Logistic Regression and Random Forest have the highest degrees of accuracy in our dataset (81 % and 77 %), while Decision Tree and K-Nearest Neighbor have the lowest levels of accuracy (75 % and 62 % ).Other algorithms may match higher in different instances and with distinct datasets, but in our case, we were given this outcome. Furthermore, growing the characteristics can also additionally permit us to locate greater correct results, however it'll take longer to process, and the machine may be slower than it's far presently as it may be barely greater state-of-the-art and deal with greater statistics. So, after studying all the possibilities, we determined which alternative will be the best for us to pursue.

## REFERENCES

- [1] Kelley, Deeanna. "Heart disease: Causes, prevention, and current research." *JCCC Honors Journal* 5.2 (2014): 1.
- [2] Alshurafa, N., Sideris, C., Pourhomayoun, M., Kalantarian, H., Sarrafzadeh, M. and Eastwood, J.A., 2016. Remote health monitoring outcome success prediction using baseline and first month intervention data. *IEEE journal of biomedical and health informatics*, 21(2), pp.507-514.
- [3] Athilingam, P., Jenkins, B., Johansson, M. and Labrador, M., 2017. A mobile health intervention to improve self-care in patients with heart failure: Pilot randomized control trial. *JMIR cardio*, 1(2), p.e7848.
- [4] Abd, D., Alwan, J.K., Ibrahim, M. and Naeem, M.B., 2017, March. The utilisation of machine learning approaches for medical data classification and personal care system mangementfor sickle cell disease. In *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)* (pp. 213-218). IEEE.
- [5] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients Mai Shouman, Tim Turner, and Rob Stocker *International Journal of Information and Education Technology*, Vol. 2, No. 3, June 2012
- [6] Amudhavel, J., Padmapriya, S., Nandhini, R., Kavipriya, G., Dhavachelvan, P. and Venkatachalapathy, V.S.K., 2016. Recursive ant colony optimization routing in wireless mesh network. In *Proceedings of the Second International Conference on Computer and Communication Technologies* (pp. 341-351). Springer, New Delhi.
- [7]Xiao-Yan Gao,<sup>1</sup>Abdelmegeid Amin Ali,<sup>2</sup> Hassan Shaban Hassan,<sup>2</sup> and Eman M. Anwar ,” Improving the Accuracy for Analyzing Heart Diseases Prediction Based on the Ensemble Method”, Published 10 February 2021.
- [8] Almustafa, K.M., 2020. Prediction of heart disease and classifiers’ sensitivity analysis. *BMC bioinformatics*, 21(1), pp.1-18.
- [9] Rajesh N1 , T Maneesha<sup>2</sup> , Shaik Hafeez<sup>3</sup> , Hari Krishna<sup>4</sup> ,”Prediction of Heart Disease Using Machine Learning Algorithms”, published in 2018.
- [10] Musfiq Ali MD. Iftiyar Khan Masud Al Imran Musnath Siddiki ,” Heart Disease Prediction Using Machine Learning Algorithms”, published in April 2019.

- [11] Farhan Mashuk, Samsujjoha, Abdus Sattar, Nahida Sultana. "Machine Learning Approach for Bird Detection" , 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021.
- [12] Dittmer, L. (2012). Heart disease. Creative Education.
- [13] Goetz, T. (2010). The decision tree. Rodale.
- [14] <https://www.kaggle.com/ronitf/heart-disease-uci>
- [15] N Rajesh, Maneesha T, Shaik Hafeez, Hari Krishna. "Prediction of Heart Disease Using Machine Learning Algorithms" , International Journal of Engineering & Technology, 2018

## Heart Diseases Prediction

### ORIGINALITY REPORT

13%

SIMILARITY INDEX

0%

INTERNET SOURCES

12%

PUBLICATIONS

4%

STUDENT PAPERS

### PRIMARY SOURCES

1

Harshit Jindal, Sarthak Agrawal, Rishabh Khera, Rachna Jain, Preeti Nagrath. "Heart disease prediction using machine learning algorithms", IOP Conference Series: Materials Science and Engineering, 2021

Publication

9%

2

Nourin Islam, Nasrin Akter, Abdus Sattar. "Sentiment Analysis on Food Review using Machine Learning Approach", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021

Publication

2%

3

Submitted to Coventry University

Student Paper

1%

4

Submitted to University of Greenwich

Student Paper

1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off

