AN EFFECTUATION ANALYSIS OF HEART ATTACK PREDICTION USING MACHINE LEARNING ALGORITHMS

BY

HIMU AKTER
ID: 213-25-040

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

Supervised By

Md. Sadekur Rahman

Assistant Professor

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2022

APPROVAL

This Project/Thesis titled **"An Effectuation Analysis of Heart Attack Prediction Using Machine Learning Algorithms"** submitted by **Himu Akter** ID No: **213-25-040** 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 M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 21-09-2022 .

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori, PhD   **Chairman**
Professor and Associate Head
Department of Computer Science and Engineering
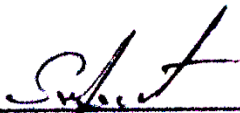Faculty of Science & Information Technology
Daffodil International University

Dr. Moushumi Zaman Bonny   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Sazzadur Ahamed   **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Safael Hossain   **External Examiner**
Associate Professor & Head
Department of Computer Science and Engineering
City University

i

DECLARATION

I hereby declare that this thesis has been done by me under the supervision of **Md. Sadekur Rahman, Assistant Professor, Department of CSE Daffodil International University**. I 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:

Sadek 20.9.22

Md. Sadekur Rahman
Assistant Professor
Department of CSE
Faculty of Science and Information Technology
Daffodil International University

Submitted by:

Himu Akter

Himu Akter
ID: 213 -25-040
Department of CSE
Faculty of Science and Information Technology
Daffodil International University

ACKNOWLEDGEMENT

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

I am really very grateful and wish profound indebtedness to **Md. Sadekur Rahman**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep knowledge and keen interest of my supervisor in the field of "Natural Language Processing and Machine Learning" to carry out this thesis. His scholarly guidance, patience, constructive criticism, motivation, constant and energetic supervision, continual encouragement, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this thesis.

I would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor & Head,** Department of CSE, for his motivation and appreciation. I am also very thankful to other faculty members and the staff of CSE department of Daffodil International University.

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

Finally, I am very thankful to our parents and friends who were always motivated and criticized my work in a manner to improve my work. At least I thank all of them from the core of my heart.

ABSTRACT

The Machine Learning field turns out consecutive statistics and artificial intelligence denomination. A sub sector of artificial intelligence is machine learning. Machine Learning can promote the treatment procedure by growing patient involvement and so fetching excellent health outcomes. Machine Learning models can qualify the motive explanation of all receivable results for the similar patient and it can raise the diagnostic accuracy of every gradation. Heart attack is more common in people over the age 65 and are more likely to occur as people get older. For predicting heart attack various machine learning algorithms are being used. Such as Logistic Regression, Naive Bayes, Decision Tree, Random Forest and SVM KNN. For all of these algorithms I found the best accuracy in Naive Bayes. Confusion matrix is being used for all classification and shows better outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Heart Attack is a medical quandary. The flux of oxygen-rich blood in one or more coronary arteries that creates heart attack and those stocks in heart muscle, casually becomes enclosed and a partition of heart muscle. It receives sufficient oxygen. To the heart muscle famishes that lack blood stocks of oxygen. The most general reason for heart attack is coronary artery affliction, contrariwise the coronary arteries deserve obdurate bound to atherosclerosis. Another reason for heart attack is casually exquisite spasm or haggling of the coronary artery. It encloses the blood stocks. Heart attack lasting two or five minutes. 20minutes lasting for a full heart attack. Nowadays, global heart attacks are very general. In the USA over 1 million masses have heart attacks. Several hazards occur with heart attack. The hazards are smoking, unlimited alcohol destruction, sedentary livelihood and so many other things that occurs heart attack. The most significant venture for heart attack is high blood pressure. The majority of men 45 and older, the majority of women 55 and older are exceedingly reliable to have a heart attack. According to the latest WHO in Bangladesh extended 108,528 or 15.16% of total departure data exposed in 2020 Coronary Heart Disease Deaths. The trouble can be perceived as unquiet pressure crushing completeness or affliction for the cause of heart attack. Feeling sick mild headed or measles they can also rapture out within a flux damp. Checking blood pressure, temperature, pulse rate is the heart attack diagnosis. To view the heart is licking that the tests are being committed and to cherub thorough heart hygiene. In purchases 17.9 million masses died from cardiovascular disease in 2019 recounting 32 percent of universal departure of those demises 85 percent that due to stroke and heart attack. The medical profession actively predicted heart disease only 67 percent following the latest observation. In this paper machine learning algorithms are being used for finding best accuracy. The infliction of ML (Machine Learning) in the discussion of heart attack is the excellent contrivance to diagnose vatic nation management and other aspects of clinical rule. The main objective of this paper is to find the best accuracy by applying different types of machine learning algorithms and which algorithms give the best accuracy.

1.2 Motivation

I will create a machine learning model that predicts if a patient can be diagnosed with a heart Attack or not. We should be acquainted with the basics of machine learning and data analysis to work on this project. This project requires me to be familiar with multiple Machine Learning algorithms, including Random Forest, Decision Tree, Logistic Regression and many others.

1.3 Rationale of Study

In our everyday life we use machine learning algorithms. So, machine learning algorithms are used in medical science to predict various health diagnoses. I created a dataset and proposed a model to detect Heart Attack using machine learning algorithms. A heart attack is created when the heart is blocked and an artery that sends blood and oxygen. The blood stocks to the heart casually interrupted. Many times, we do not know when we are likely to have a heart attack. For this thesis I can predict that.

1.4 Research Questions

- During the research work some questions occur about this work. The main questions of my work in given below:
- Can I allot a better specialty viewpoint to predict heart attack more correctly to the dataset on machine learning repertory?
- Can I find better predictive machine learning models that can predict the appearance of heart attack to reduce from statistical experiment more correctly?
- Which classifiers perform better to predict heart attack.

1.5 Expected Output

To predict heart attack in this paper I executed several machine learning classifiers. The expected outcome of this research is a model which will be able to predict heart attack in primary level.

1.6 Research Layout

 6 chapters are being included in my report which are given below:

In Chapter 1, I mentioned my whole research work's outline and divided this chapter into multiple subchapters. Such as introduction, motivation, rationale of the study, research question and expected output of my project.

In Chapter 2, I have discussed the previous work of heart attack classification, the scope of the problem and challenges in this work.

In Chapter 3, I will talk about r my work procedure, methods and techniques.

In Chapter 4, I will discuss the Experimental Results and Discussion of my creating models.

In Chapter 5, I will talk about the Impact of Society and Sustainability plan of my work.

In Chapter 6, I have discussed the Summary, Conclusion and Further Study of the work.

CHAPTER 2

BACKGROUND

2.1 Terminologies

The use of machine learning in medical science is increasing day by day. It's a good thing for research. Many researchers are working on this topic. I found some papers related to our work. In my work I am trying to predict heart attack. Dataset is being collected from Kaggle.com. While doing this research, I got to know some new methods that I didn't know before. It was not an easy task for me to implement. In the upcoming chapter I will discuss this term elaborately. To implement my work completely and to know about this new term I reviewed some previous work which is related to predicting heart attack. Discuss about them shortly.

2.2 Related Works

Sharma et al. mentioned the shortened position of artifice skill and obtainable models for forecast heart diseases. The study gathers a lot of knowledge concerning the position of artifice models Deep learning and Machine learning. To exchange a heart attack 67% accuracy that they can get. Used Naive Bayes Decision Tree, KNN, and Neural Network algorithms to forecast chances of heart diseases. Deep learning in the ground of forecasting heart diseases by the profound solution on utilization. To execute a ramification support vector machine (SVM) is being used for noticing the hyper plane. That approach within two groups. In future to bear onwards the work of perished medical dataset [1].

Hazra et al. proposed to shorten a few recent explorations on forecasting heart diseases applying data mining skills. To build Association Rules Mining is being used to even susceptible machine learning. WEKA, RAPID MINER, MATLAB are data mining tools that were used in this study. Knowledgeable methods may appear which can conduct a choice of accurate provision models for patients in future [2].

Mohan et al. aimed at fiction models that target to demand important shape by employing machine learning methods resulting in progressing accuracy. Find accuracy 88.7% with Hybrid Random Forest Linear Models (HRFLM). Decision Tree, K-Nearest Neighbor, Naïve Byes is being used to categorize data. In the next try to implement various compositions of machine learning skill to excellent prognosis skill [3].

Hannan et al. proposed that elaborated knowledge within a patient and preprocessing and look after aptness competence of ANN models GRNN, RBF. Used General Regression Neural Network (GRNN), Radial Basis Function (RBF). Data is being gathered from the OPD committee. The study elaborated in future by applying Support Vector Machine (SVM) [4].

LI et al. aimed at an expert machine diagnosis process that has appeared of heart disease. Applied SVM, ANN, KNN, DT to showing allowance of data in graphical distribution which helps the public comprehend the pregnancy of data shortened. Found accuracy 92.37% from SVM. Next, they used other algorithms to find better results [5].

Elhoseny et al. proposed that automated heart diseases accomplished binary flections. Automated Heart Diseases Diagnosis (AHDD) gives best accuracy 90.1%. For classifying different types of machine learning models KNN, ANN, AdaBoost, SVM are being used. Try to add training and test data set in our future work [6].

Avis et al. aimed to bear epidemiological revealed methods to review the connection within realized hazard and purpose. This contains the outcome of the different hazard response. They accomplished two logistic regressions: forecast the feasibility and forecast detraction logistic model for real hazard and Health risk appraisal (HRA) are significant things [7].

Kukar et al. proposed to develop the indication perfecting of non-invasive indication models. Dataset is being used for 327 patients. K- Nearest neighbors, Decision tree, Naive Bayes algorithms are being used for this study. Predictive power of the indication methods is the most important result [8].

Firdous et. al. aimed the study to separate into two parts. Growing the number of confidential seams on neural networks is being proficiency to restraint. In data science Artificial intelligence is used to resolve extensive limits of matter. Finding accuracy 99%. Different machine learning algorithms used to predict the response on the database. In future deep learning algorithms applicant that novel dataset [9].

Ashrafuzzaman et al. proposed a model able of estimating the defeat of heart disease with a camera from a commercial smartphone and many other heart related diseases. The aim of this work is to ordain the mattre formerly to alleviate heart attack death rate. Fuzzy logic is being used. In next work they will attempt to apply the techniques of Microchip [10].

Shantakumar et al used proficient access for the expulsion of important samples of heart disease storehouses for the prediction of heart attack. K means clustering MAFIA algorithms are used for this work. The process is an appliance in java. They have designed to draft and develop an expert heart attack prediction process with the help of these elected,important models appling AI(Artificial Intelligence) in future[11].

Wang et al proposed to design algorithms leveraging clustering under and over sampling for heart attack prediction. Find the best accuracy from random forest 70.29% and 70.05% precision among Decision Tree, Adaboost KNN. 4049 medical records are being used as a dataset and 26 csv files. In the future research on different immitigable data for the mentioned UCO algorithms. To provide medical aids for receiving precaution with heart attack in stroke patients [12].

Salman aimed to respond in this work with two major parts. For classification Bayesian network used dataset is being collected from 787 patients from 2 several characteristics and contains 24 variables. BN learning is capable of finding out arbitrate interrelations [13].

Takei applied feature selection and machine learning algorithms together. This work aimed at better feature selection and machine learning algorithms to find the heart attack. 84.81% accuracy in the support vector machine that finds the best accuracy. For feature selection reliefF, Fisher Filtering is used. In this research 2.22% develop the success rate as feature selection. in the future developing this rata for heart attack prediction [14].

Chitra et al used a recent classification model for the prediction of heart attack in the promptly period. Gather record from patients 270 the proficiency of the classifier is turned. Using ANN and Cascaded neural networks (CNN) to find out the risk of heart attack. CNN gives 85% accuracy which is 5%linked to ANN [15].

Srinivas et al proposed to review the expectation infliction of classification revealed data mining systems to a large extent in health care. Naive Bayes, Decision Tree, ANN are being used. They have dedicated an intellectual and efficient heart attack prediction applying data mining. The purpose is evaluated with the train methods. In future this can develop and explain [16].

2.3 Comparative Analysis and Summary

Today's heart attack has become one of the moving reasons for death worldwide. Data mining had obtained a lot of outlook and had been applied in various kinds of infliction including in medical in recent research. I am working with heart attack prediction. That's why I read some papers which are related to my research purpose. Basically, I am going to restrain which machine learning algorithms perform very well with this dataset. In this section I am going to compare one some previous work with other work. Now given a table which paper I read previous and which paper reference is given in the last chapter.

Table 2.1: Comparison between previous works

Work Title	Algorithm Usage	Accuracy
Prediction of heart disease using machine learning [1].	Naive Bayes Decision Tree, KNN, and Neural Network	67%(NB)
Effective heart disease prediction using hybrid machine learning [3].	Hybrid Random Forest Linear Naive Bayes Decision Tree.	88.7%(HRFM)
Heart disease identification method using machine learning classifier[5].	SVM, ANN, KNN, DT	92.37% (SVM)
A new multi-agent feature wrapper machine learning approach for heart disease diagnosis[6].	KNN, ANN, AdaBoost, SVM	90.1% (AHDD)

In this table I include 4 papers title name, best model and accuracy among all those papers which I read. So, from the previous work I got an initial and primary idea for my work which algorithms we use for my work.

2.4 Research Scope

Classification revealed machine learning algorithms can forecast heart attack utility or not. I find better accuracy using machine learning accession. I want to produce amazing prediction

models regarding accuracy, time and learning lack into selected methods by applying different machine learning algorithms and classifiers. In the future I want to work on deep learning algorithms that are based on artificial neural networks that work as a human brain.

2.5 Challenges

I face some problems while working on this paper. I face problems collecting dataset. I want to collect data from Bangladeshi hospitals to predict heart attack like a real dataset. But it is not too easy for me to go everywhere and collect data. Moreover, the patients did not show much interest in giving information to them. That's why I collect dataset from online platforms. So, for this limitation and challenges I cannot collect more real data. If I collect real data from reality then I find accurate accuracy.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

I want to try to predict heart attacks and try to find the best accuracy in my research. First, the dataset needs to be created to run the model and I need to know which algorithms to use. In machine learning algorithms Supervised Learning and Unsupervised Learning are two ways to build a method. In my research I used supervised machine learning algorithms. There are some classification algorithms in supervised learning that are being used to resolve the classification problems. In my research work I used Random Forest (RF), Decision Tree (DT), KNN algorithm. I will talk about all of the algorithms, how it works and which one is best in the next upcoming proposed methodology section and I will try to figure out which algorithm gives good results from my collecting data set.

3.2 Data Collection Procedure

Machine Learning algorithms perform very well when the collecting data is more balanced and reliable to train the machine perfectly. So, data collection is very significant for my work. Dataset is being divided into trains and tests. From this dataset I found which are affected by heart attack which are not. From this dataset I found that 508 people are affected and 410 people are not affected among 918 data.
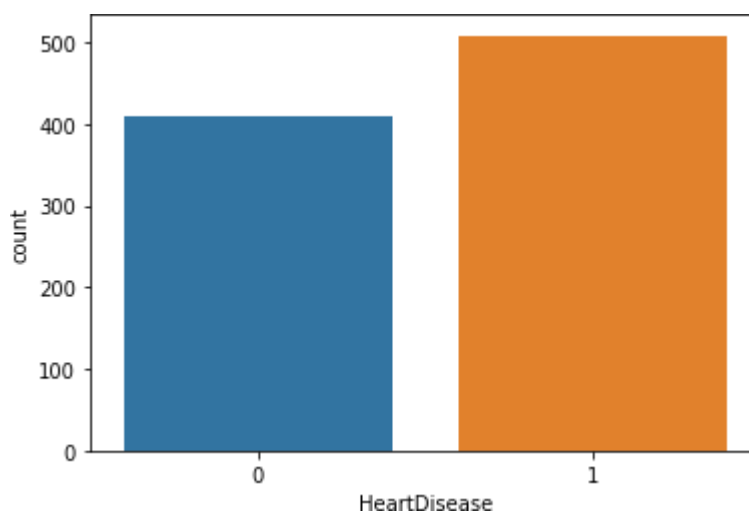

Figure 3.1: Histogram of target class

Hear the heart disease diagram that 0 indicates not affected and 1 indicates affected.

3.3 Statistical Analysis

Data is being separated into two types who are affected and who are not affected by heart diseases. Among 918 data affected508 and not 410. In my data set 918 rows and 12 columns. My dataset is available in csv format. In the below section I provide the statistical analysis of my dataset. Now I am giving the percentage table for heart disease who are affected and who are not.

Table 3.1: Distribution dataset

Diseases	Data value	Percentage
Heart Disease	508	(508*100)/918=55.33%
No heart Diseases	410	(410*100)/918=44.66%

3.3.1 Confusion Matrix: The confusion matrix calculates the accuracy, specificity and impressibility measures [17]. For better data visualization confusion matrix is being used. It is used to discuss execution of classification models. All model confusion matrices are given .
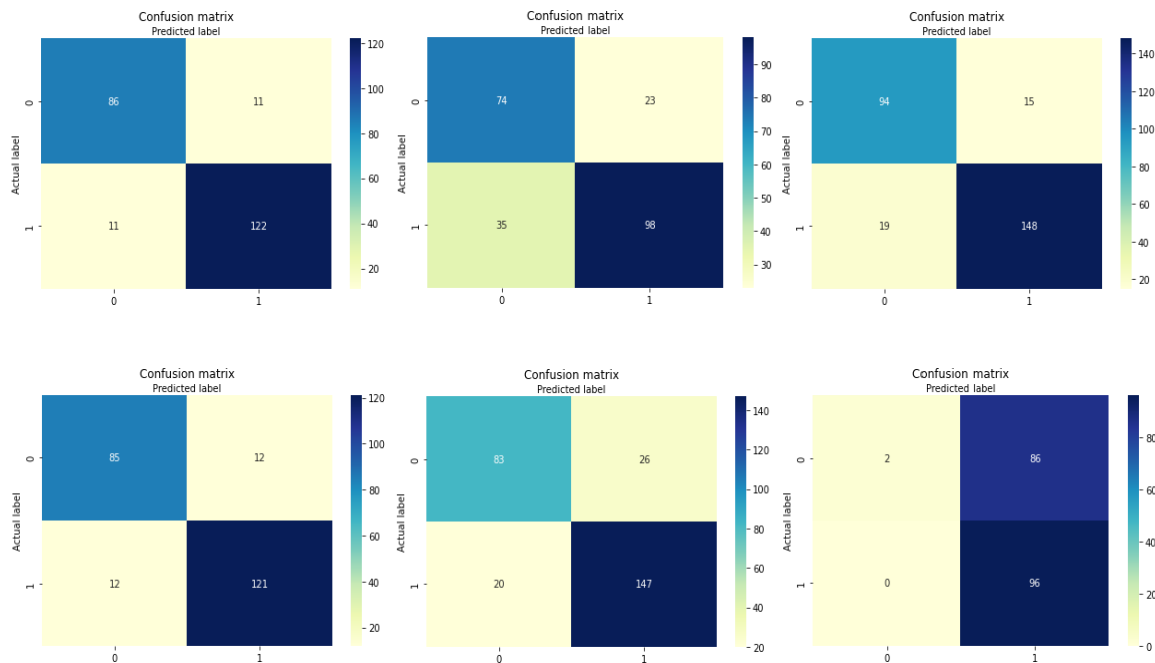
Figure 3.2: Confusion Matrix

3.4 Proposed Methodology:
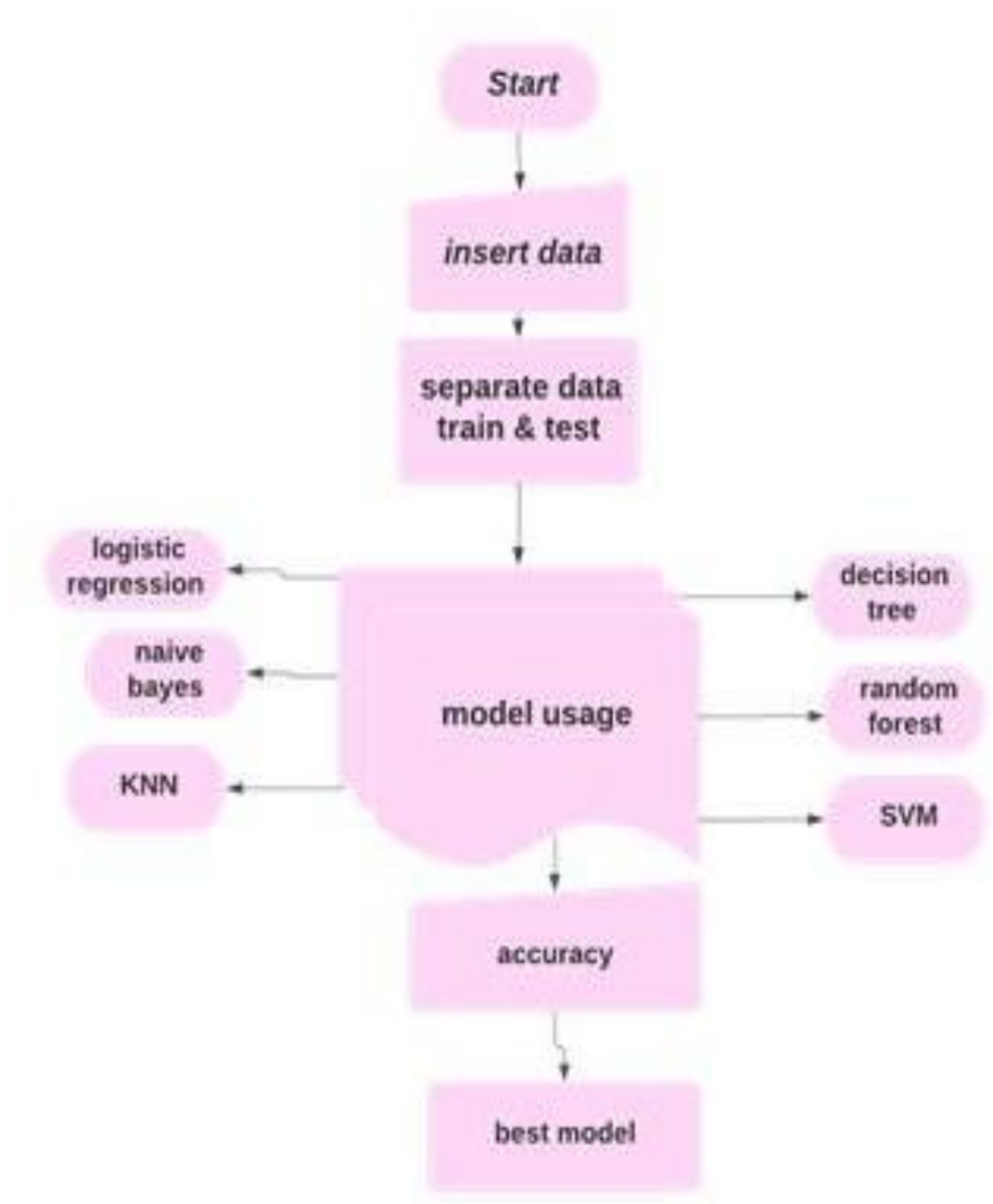
3 .3: Proposed Methodology

3.3.2 Insert Data: After plucking data I am trying to appliance the proposed method. The process has been accomplished with input data in this gradation. For this work goal I have used 12 attributes and among them one attribute used for output result.

3.3.3 Separate Data: To complete this task I split the dataset into two parts one is training dataset and another is test dataset. Training dataset is the primary data which is being used to teach machine learning algorithms that can calculate the values. Test dataset is the minor dataset that helps to test machine learning algorithms.

3.3.4 Model Usage: Machine learning algorithms used in medical science every day. For that's why I used machine learning algorithms to predict heart attack. I used different types of machine learning classifiers like Random Forest, Decision Tree, KNN and others to discuss next. I tried to find out the best accuracy which models give.

3.3.5 Accuracy & Best model: The best model refers to which model is best among all the models. Best models performed anything from the research related. In this step I found which model is best to predict accuracy. I accomplished the research work to help people.

3.3.6 Model Selection: Supervised and Unsupervised learning are two types of machine learning. In my research work I used some supervised learning algorithms. The names of these algorithms are Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (DT) and Support Vector Machine (SVM). Now I am describing the algorithms which I used for prediction.

3.3.6.1 Logistic Regression: Logistic regression is a popular supervised machine learning algorithm that is used to predict a contingent unconditional object variable. The probability is being calculated by logistic regression of a binary. It predicts a dependent variable which is categorical. The data set is being much larger than the logistic regression can help to categorize the data. Logistic regression is also applied to calculate the relationship within dependent variables and one or more I uncommitted variables. Sepal length, petal length, sepal width, petal width are the features of logistic regression. Logistic regression accommodates algorithms to classify oncoming data revealed on historical data that applied in machine learning applications. Discreet outputs are being provided by logistic regression [18].

3.3.6.2 Naive Bayes: To classify an object's Naïve Bayes algorithms used. It is also known as simple Bayes. To classify data Naïve Bayes applied probability theory. The probability of a case can be fixed as new data is initiated. It is an outfit of machine learning algorithms which create wants of statistical distinction. For scoring and dwelling it allows fast and high scalable methods. Discrete and continuous data is being handled by naïve Bayes. To real time prediction it is used. It doesn't look at much training data. It predicts the class of the test data set which is simple and fast. It is also used for multiple class prediction and also used for collecting data in the medical sector. Text classification, email spam detection is becoming high dimensional data that works on naïve Bayes algorithms. The process is very helpful for difficult problems on a large range. The excellent accomplishment for this position would be to apply the Naive

© Daffodil International University

Bayes classifier that is perfectly faster in parallelism to another classification algorithm [1,3,21].

3.3.6.3   KNN: A non-parametric model is k-nearest neighbors which are applied for classification. In KNN data set out in a shape. KNN applied as a recommendation system. It depots all the obtainable data and classifies a recent data point. It applied for regression. Both classification and regression KNN is being used. KNN is used for text mining, medical sector, Finance, Agriculture and so many others. It can gather high accuracy. It is impossible to gain labeled data that is so costly. KNN algorithms intend a number k w that is the nearest Neighbor to the data item is to be classified. If the value of K=4 then it searches for about 4 data items [8].

3.3.6.4   Decision Tree: A non-parametric algorithm name is Decision Tree that is a supervised machine learning algorithm. It takes advantage of classification and also regression. According to a particular parameter data is being intersected. To build a method that forecasts the measure of an object variable it's the destination of the decision tree. It helps to appreciate one's choice. Its tools are very excellent that helps one too alike among different courses of activities. Decision tree is very easy to understand, explain and recognize. Any types of data are managed by decision trees like categorical, numerical. In decision tree normalization is not obligate. It acts by recursively partitioning the data revealed input ground measure. It has three main parts. In every class value measures the number of items [3,16 ,19].

3.3.6.5   Random Forest: Random Forest is a machine learning algorithm which is supervised. In regression & classification a random forest algorithm is being used. It is the composition of decision trees. It sends out favorable reckoning to understand spontaneously. Large datasets can be managed by random forest. It measures higher alignment of accuracy to forecast production of decision trees. It is eminent with high dimensional data that is made with subsets of data. In random forest training data is faster. Several decision trees are being manufactured for every exhibit. An outcome will be created in every decision tree. Over-fitting matter can be minimized by random forest and also minimized the conflict for that reason promotes accuracy. Decision trees are composited in random forest and give the final results which are being multitude. It gives a higher True positive rate [19,20,21].

3.3.6.6   SVM: A supervised machine learning algorithm is Support Vector Machine which is used for both regression and classification. The motive of the Support Vector Machine algorithm is to search a hyper plane in an N-dimensional place which distinctly distributes the

data items. Dimension depends on the number of features. SVM is one of the most strong and accurate algorithms within the various classification algorithms. It is useful in high dimensional incidents. Several kernel activities can be given off the decision actions and it's feasible to mark more kernels. It is used in handwriting recognition, face detection, email classification and so many other things. Analyzed the data for both regression and classification [5,6].

3.5 Implementation Requirements

An Effectuation Analysis of Heart Attack Prediction Using Machine Learning Algorithms is my research title. I gathered data from online platforms. To fully evaluate the work, I need useful hardware and need high computer configuration with GPU. Useful tools, libraries and hardware, software is used to process the work. Below I have mentioned all the used hardware, software and tools that I have used in my research purposes.

Advance Libraries and Tools:

- Windows 10
- Pandas
- Python 3.8
- Matplotlib
- NumPy
- Scikit-Learn

Hardware and Software:

- High Speed Net Connection
- Google Collab
- 1TB Hard Disk
- 8GB ram in Intel core i5 7th gen integrated.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

In my work I described the algorithms which are applied in my research work. I applied six classification methods to predict heart attack. The six classification methods are being described in the section of methodology. The algorithms are Logistic Regression, Naive Bayes, K-Nearest Neighbor, Decision Tree, Random Forest and Support Vector Machine. In text data set all the classifiers execute very well. Table 4.2 displays the accuracy of all classifiers. By knowing the accuracy rate we can say which model is best for these dataset. From table 4.2 the algorithms are Logistic Regression, Naive Bayes, K-Nearest Neighbor, Decision Tree, Random Forest and Support Vector Machine and the accuracy is gradually for all classifiers 89%, 90%, 74%, 76 %, 87%, 53%. Comparing all classifiers accuracy and selecting the best model among six models. For the better result, I found out the confusion matrices for all classifiers that are shown in chapter 3 and table number 3.2. Confusion matrix builds it simple to view if the process baffling 2 classes. Precision, Recall, F1-Score found for all classifiers and the score is split into heart disease and no heart disease. Now in this section I describe the accuracy, precision, recall and so many other things. The table are given below:

4.1.1 Precision: Most functions are defined in the math module for precision testing. Without any false positive machine learning methods are challenged to recognize all positive things by using precision. Precision is being used when the vaccination of 1 needs to be as appropriate as possible [23,24,25].

4.1.2 Recall: The work of recall is to measure or determine the number of appropriate positive predictions among all positive predictions. Recall is a metric which measures the positive exhibit. Recall gives the percentage of total conversant outcomes c appropriately classified by our algorithms. Recall is extremely applied when we have to appropriately classify several instances which have already happened [22,23,25].

4.1.3 F1-Score: In machine learning f1 score is the highest significant rating metrics. A set of prediction labels f1 score is being calculated. It is initially applied to assimilate the perfection of 2 classifiers. For regression the f1 score is used. In each individual dataset f1 scores are almost analogous [22,23,24].

Table 4.1: Precision, Recall, F1-score and Support

Algorithms	Diseases	Precision	Recall	F1-Score	Support
Logistic Regression	No heart disease	0.88	0.88	0.88	97
	Heart disease	0.91	0.91	0.91	133
Naive Bayes	No heart disease	0.89	0.89	0.89	97
	Heart disease	0.92	0.92	0.92	133
KNN	No heart disease	0.68	0.76	0.72	97
	Heart disease	0.81	0.74	0.77	133
Decision Tree	No heart disease	0.66	0.83	0.73	109
	Heart disease	0.86	0.72	0.79	167
Random Forest	No heart disease	0.82	0.88	0.85	109
	Heart disease	0.92	0.87	0.90	167
SVM	No heart disease	0.73	0.76	0.74	88
	Heart disease	0.77	0.74	0.76	96

4.2 Experimental Results and Analysis:

Machine learning is the model of the Artificial Intelligence process that operates the task usually forecasting output norm from bestwed input data. No model can provide complete accuracy in machine learning algorithms. I am trying to find a satisfactory solution to predict heart attack. But it is not possible to give 100 percentage accuracy to all models.I am trying to find out the best accuracy for my dataset and which is best. From table 4.2 saw the accuracy of all models. Seeing these we found that the accuracy is not so much. Among all the models Logistic Regression accuracy 89% and Naive Bayes accuracy 90% give better accuracy. But KNN accuracy 74%, Decision Tree accuracy76%, Random forest accuracy 87% give good results but not so much expectation for this dataset. Support Vector Machine accuracy 53%

gives average results. In my research models I appreciate precision, Recall, F1-Score for the dataset. For all calcification in the applied confusion matrix. After evaluating the confusion matrix I recognized the dataset. In 3.3.1 I gave the confusion matrix diagrams of all models. Table 4.2 gives the accuracy of all models. In applying all of these models I found the best accuracy from the Naive Bayes algorithms. The accuracy of Naive Bayes algorithm is 90%. So at the end of the research work I found out 90% accuracy to predict heart attack. That is the best accuracy in my work.

Table 4.2: Accuracy of different classifiers

Model Name	Accuracy
Logistic Regression (LR)	89%
Naive Bayes (NB)	90%
K-Nearest Neighbor (KNN)	74%
Decision Tree (DT)	76%
Random Forest (RF)	87%
Support Vector Machine (SVM)	53%

4.3 Discussion

At the end of my work I am attempting to subscribe to heart attack prediction. There is much research work on heart attack prediction. I am attempting to add a new dimension of heart attack prediction in my work. So I am happy to have fulfilled this work. In this work I am able to create a machine learning model which gives better results of heart attack prediction. For finding the best model I used six classifiers. Among all the classifiers Naive Bayes gives the best results in my dataset. All model accuracy is given in table 4.2. To find better accuracy I find out the score of precision ,Recall and F1 score . Also find out the confusion matrix. All of this helps me to find better accuracy of heart attack prediction. .

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Heart Attack creates a great impact on society. Heart plays an especially significant role in our impulsive wisdom. The wisdom of emotion results from the heart ,body and brain performance in adjuncts. A variety of emotional and behavioral problems, like anxiety and depression can usually be attributed to these heart attacks. These concerns may prevent adopting and maintaining health promoting practices. Patients with heart attacks suffer from various mental and physical weakness including chest pain, sleep problems and depression. Patients with these symptoms curtail their physical and social activities resulting in impaired QOL. High mortality rates are related to poor QOL. The undiluted enumeration of social determinants impacting a several can improve the hazard dying the disease of heart attack.

5.2 Impact on Environment

Environment can play a vital role in a heart attack. Particulate matter has been linked to death from other diseases, including heart attack in air pollution. Heart attack can increase due to high temperatures in efficient patients for the reason of blood density. The number of environmental elements may increase the risk of heart disease including lead, arsenic and cadmium which have been studied extensively. In heart failure carbon monoxide is linked. The hazard may be grown in those with a family history of the disease and those who share other factors in the common environment. Eating unhealthy foods and smoking cigarettes may increase the risk of hay fever in people with a family history of the disease. The improvement of inherent heart disease is versatile with environmental and genetic domination. Environmental disclosure that contributes to the improvement and cruelty of heart attack are important but largely unfounded hazard factors.

5.3 Ethical Aspects

The making of hazard heart disease is complicated with the query of partitive equity in the accountable targeting of these hazards. A discussion that has become more apparent in the composition of medical but less in the composition of ethics. Careful discretion must be given to legal consignment, ethical concerns and cultural regulation when medicating a patient with a heart attack. This helps to see the prediction of heart attack and get the results affected or not

© Daffodil International University

affected. Ethical consignment in the attempt of patients with heart attack involve measurement of decision-making retention when understanding or indication is ruined, potent, evaluation of influence of life.

5.4 Sustainability Plan

My research work will help new researchers to predict heart diseases and they can get ideas about who to predict and which models give the best results. By reading this work anyone can know when heart disease and when no heart disease.This is the reason why I produced the model and why I want to decide this completely so that the public can freely demand the appropriate solution. I should have recognized this idea. For this reason the models are being sustained and that is needed for me. I created an appropriate model that provides the appropriate knowledge about this heart attack prediction. I also need to take action on how to sustain the model incessantly.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH

6.1 Summary

I used machine learning algorithms in my research to make a heart attack prediction. In my research I am trying to build a proper model using machine learning algorithms. So, for making this model I have several problems and so more aligns in appearance during this research work. My research summary is given below

- Thinking about work
- Preparative Problem
- Data Collection
- Data labeling
- Clean up Data
- Separate Train & Test
- Choose Model
- Execution & Evaluation Model

After following these steps, I am able to create my model for predicting heart attack.

6.2 Conclusion

In machine learning the experimental analysis suggests that variant intercourse to the prediction of heart attack have been tested. The work demands the purification of the algorithms used and then identifies the efficient contrivance. In python software the research subscribes to the compatible solicitation and exploration of exclusive machine learning algorithms. Pre-processing is a drastic gradation in machine learning which supports obtaining exact results to sum up. I restrained the explanation of different machine learning algorithms on the database and found the best accuracy. Now at the end of work Naive Bayes Gives best accuracy 90%.

6.3 Recommendation

For my research I have recommendations. Machine learning classification is being used in my work for developing accuracy of the model. For a large number of datasets there are many algorithms and methods, datasets. So, that model will predict more exactly for heart attack. Recommendations are given below:

- Heart Attack Prediction Dataset.
- Try to create better classification models.
- Try to improve datasets.
- Understand conversion.
- Try to find better execution of accuracy.

6.4 Implication for Further Research

In my research some limitations and drawbacks. In my research one of the most important limitations is collecting data set online platforms. I used only a machine learning algorithms classifier. In future research to collect real dataset from various hospitals and collect data physically. Deep learning algorithms are being used in my future research which works as a human brain.

References

1. Sharma, H., & Rizvi, M. A. (2017). Prediction of heart disease using machine learning algorithms: A survey. *International Journal on Recent and Innovation Trends in Computing and Communication*, *5*(8), 99-104.
2. Hannan, S. A., Manza, R. R., & Ramteke, R. J. (2010). Generalized regression neural network and radial basis function for heart disease diagnosis. *International Journal of Computer Applications*, *7*(13), 7-13.
3. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
4. Hannan, S. A., Manza, R. R., & Ramteke, R. J. (2010). Generalized regression neural network and radial basis function for heart disease diagnosis. *International Journal of Computer Applications*, *7*(13), 7-13

5. Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*, *8*, 107562-107582.

6. Elhoseny, M., Mohammed, M. A., Mostafa, S. A., Abdulkareem, K. H., Maashi, M. S., Garcia-Zapirain, B., ... & Maashi, M. S. (2021). A new multi-agent feature wrapper mach**7.** Avis, N. E., Smith, K. W., & McKinlay, J. B. (1989). Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed?. American Journal of Public Health, 79(12), 1608-1612.

7. Avis, N. E., Smith, K. W., & McKinlay, J. B. (1989). Accuracy of perceptions of heart attack risk: what influences perceptions and can they be changed?. American Journal of Public Health, 79(12), 1608-1612.

8. Kukar, M., Kononenko, I., Grošelj, C., Kralj, K., & Fettich, J. (1999). Analysing and improving the diagnosis of ischaemic heart disease with machine learning. Artificial intelligence in medicine, 16(1), 25-50.

9. Firdous, N., Bhardwaj, S., & Bhat, A. H. (2021). Prevention of autopsy by establishing a cause-effect relationship between pulmonary embolism and heart-failure using machine learning. International Journal of Computer Applications in Technology, 66(2), 165-171.

10. Ashrafuzzaman, M., Huq, M. M., Chakraborty, C., Khan, M. R. M., Tabassum, T., & Hasan, R. (2013). Heart attack detection using smart phone. *International journal of technology enhancements and emerging engineering research*, *1*(3), 23-27.

11. Patil, S. B., & Kumaraswamy, Y. S. (2009). Extraction of significant patterns from heart disease warehouses for heart attack prediction. *IJCSNS*, *9*(2), 228-235.

12. Wang, M., Yao, X., & Chen, Y. (2021). An imbalanced-data processing algorithm for the prediction of heart attack in stroke patients. *IEEE Access*, *9*, 25394-25404.

13. Salman, I. (2019). Heart attack mortality prediction: an application of machine learning methods. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(6), 4378-4389.

14. Takci, H. (2018). Improvement of heart attack prediction by the feature selection methods. *Turkish Journal of Electrical Engineering and Computer Sciences*, *26*(1), 1-10.

15. Chitra, R., & Seenivasagam, V. (2013). Heart attack prediction system using cascaded neural network. In *Proceedings of the International Conference on Applied Mathematics and Theoretical Computer Science-2013* (p. 223).

16. Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. *International Journal on Computer Science and Engineering (IJCSE)*, *2*(02), 250-255.

17. Y. Xing, J. Wang, Z. Zhao, and A. Gao, "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease," in 2007 International Conference on Convergence Information Technology (ICCIT 2007), 2007, pp. 868–872

18. .Keya, M. S., Emon, M. U., Akter, H., Imran, M. A. M., Hassan, M. K., & Mojumdar, M. U. (2021, January). Predicting performance analysis of garments women working status in Bangladesh using machine learning approaches. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (pp. 602-608). IEEE.

19. Keya, M. S., Akter, H., Rahman, M. A., Rahman, M. M., Emon, M. U., & Zulfiker, M. S. (2021, January). Comparison of Different Machine Learning Algorithms for Detecting Bankruptcy. In *2021 6th International Conference on Inventive Computation Technologies (ICICT)* (pp. 705-712). IEEE.

20. [1]"Naive Bayes classifier - Wikipedia", *En.wikipedia.org*, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Naive_Bayes_classifier. [Accessed: 16- Sep- 2022].

21. 2]S. Dube, "What is Logistic Regression in Machine Learning? - Scaler Topics", *Scaler Topics*, 2022. [Online]. Available: https://www.scaler.com/topics/logistic-regression-machine-learning/. [Accessed: 16- Sep- 2022].

22. [12]"All Mathematical Functions Defined under Math Module in Python 3", *Programiz.com*, 2022. [Online]. Available: https://www.programiz.com/python-programming/modules/math. [Accessed: 17- Sep- 2022].

23. M. Rahman, M. M. Zahin and L. Islam, "Effective Prediction On Heart Disease: Anticipating Heart Disease Using Data Mining Techniques," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2019, pp. 536-541, doi: 10.1109/ICSSIT46314.2019.8987776.

24. Khourdifi, Y., & Bahaj, M. (2019). Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. *International Journal of Intelligent Engineering and Systems*, *12*(1), 242-252.

25. Padmanaban, K. A., & Parthiban, G. (2016). Applying machine learning techniques for predicting the risk of chronic kidney disease. *Indian Journal of Science and Technology*, *9*(29), 1-6.

AN EFFECTUATION ANALYSIS OF HEART ATTACK PREDICTION USING MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

10% SIMILARITY INDEX	10% INTERNET SOURCES	3% PUBLICATIONS	2% STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	www.upgrad.com Internet Source	1%
3	www.jatit.org Internet Source	1%
4	ijsrcseit.com Internet Source	<1%
5	Submitted to Daffodil International University Student Paper	<1%
6	academic-accelerator.com Internet Source	<1%
7	Augusto Junio Guimarães, Paulo Vitor de Campos Souza, Edwin Lughofer. "Hybrid Model for Parkinson's Disease Prediction", Information Processing and Management of Uncertainty in Knowledge-Based Systems Internet Source	<1%