HEART ATTACK PREDICTION USING MACHINE LEARNING TECHNIQUE

BY

Shomaiya Afrin
ID: 191-15-12962

Rashed Mahmud
ID: 191-15-13008

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

Supervised By

Dr. Md. Zahid Hasan

Associate Professor

Department of CSE

Daffodil International University

Co-Supervised By

Dr. Md. Tarek Habib

Associate Professor

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled "Heart attack prediction using machine learning technique", submitted by Shomaiya Afrin, ID No: 191-15-12962, Rashed Mahmud, ID No: 191-15-13008, 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 (BSc) and approved as to its style and contents. The presentation has been held on 26 January 2023.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

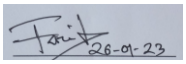
Internal Examiner

Subhenur Latif
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Mohammad Monirul Islam
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

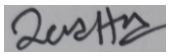


Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Md. Zahid Hasan, Associate Professor, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

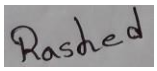


Dr. Md Zahid Hasan
Associate Professor
Department of CSE
Daffodil International University

Co Supervised by:

Dr. Md. Tarek Habib
Associate Professor
Department of CSE
Daffodil International University

Submitted by:



Rashed Mahmud
ID:191-15-13008
Department of CSE
Daffodil International University

Shomaiya Afrin
ID:191-15-12962
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to Dr. Md Zahid Hasan, Associate Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge& keen interest of our supervisor in the field of "*Data Mining & Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement,constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Head,** Department of CSE, for his kind help to finish our project and also to other faculty member 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 discuss while completing the course work.

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

ABSTRACT

The subject of AI known as machine learning has been at the forefront of several recent statistical and technological advances. It’s a branch of AI. Improved health outcomes can be achieved with the help of machine learning since it can increase patient participation in the treatment process. Machine learning methods can improve the diagnosis accuracy at every level by identifying the most likely reason for all similar patients' test findings. People over the age of 60 have a higher risk of experiencing a heart attack, and the prevalence of heart attacks increases with age. In order to foretell the onset of a heart attack, researchers are using a number of machine learning techniques. The goal of this study is to describe in depth the methods we use to predict cardiovascular disease, including Decision Tree, K-nearest neighbors, Logistic Regression, XGBoost, Support Vector Machine, and Random Forest. Predictive data mining techniques have been tested on the same dataset with varying degrees of success, and Random Forest methods have been shown to yield the highest accuracy of 87%.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Introduction

The medical conundrum that is a heart attack. As a result of a blockage in one or more coronary arteries, oxygen-rich blood stops flowing to the heart, and the resulting stocks in heart muscle cause an attack. All the oxygen it needs is being delivered. To the dying heart muscle that has insufficient oxygenated blood to function. Coronary artery disease is the leading cause of heart attacks, and atherosclerosis in the coronary arteries is the primary cause of this disease [1]. The coronary artery can also spasm or haggle in a casually exquisite fashion, which can lead to a heart attack. In this case, it serves to protect the blood supplies. Angina lasting between two and five minutes. A full-blown heart attack often lasts about 20 minutes. Cardiovascular crises are becoming commonplace all around the world. More than a million people have heart attacks every year in the United States. A cardiac attack poses several risks. Cigarette smoking, excessive alcohol consumption, leading a sedentary lifestyle, and the list goes on and on are all risk factors for cardiovascular disease and heart attacks. High blood pressure is the leading risk factor for heart attacks. The risk of having a heart attack increases dramatically after age 45 for males and after age 55 for women. WHO estimates that 108,530 people in Bangladesh would die from coronary heart disease in 2022. This represents 16% of all deaths in the country [2].

Information can be mined from massive data sets using a computer technique known as "data mining." In an exploratory investigation, data mining shines due to the fact that it can glean nontrivial information from massive amounts of data. There is a lot of untapped potential in medical data mining, which might be used to decipher hidden patterns within clinical data. These regularities have medical use in the context of diagnosis [3]. On the other hand, the raw medical data now at hand are dispersed, massive, and inherently diverse. It is important to collect these statistics in a systematic manner. Once all the information is gathered, it may be combined to produce a comprehensive health database. With data mining, users may take advantage of a pattern-discovery method that focuses on surfacing previously unseen relationships in the data Business questions may be answered

and new medical conditions can be predicted with the use of data mining technologies. The ability to foretell the onset of disease is an important component of data mining.

It's possible to interpret the difficulty as a tight, crushing weight, or as a disease that's triggering a heart attack. They could also rapture out in a wet place if they are feeling ill, lightheaded, or have measles [4]. Diagnosis of a heart attack involves checking blood pressure, temperature, and pulse rate. In order to see that the tests are being carried out and to cherub on good heart hygiene, one must picture the heart as being licked. In 2022, cardiovascular disease was responsible for the deaths of 18 million people worldwide, accounting for 35% of all deaths. A recent study found that only 65% of coronary heart disease could be actively anticipated by doctors [5]. In this study, we employ machine learning methods to determine the most precise results. When talking about heart attacks, ML (Machine Learning) is a great tool for making diagnoses, managing vasculature, and other elements of clinical practice. The primary goal of this research is to determine which machine learning algorithms provide the highest levels of accuracy.

1.2 Motivation of research

Machine learning is an approach to data analysis that uses computational algorithms to automatically examine big datasets in search of previously unknown patterns and insights. There are many different subfields within the larger field of machine learning. Machine learning encompasses a variety of techniques for improving prediction and accuracy, including supervised learning, unsupervised learning, and ensemble learning. To determine if a person has had a heart attack, we'll develop a machine learning model. Prior knowledge of machine learning and data analysis fundamentals is required for this assignment. For this task, we need to be conversant with a wide range of Machine Learning methods, such as Decision Tree, K-nearest neighbors, Logistic Regression, XGBoost, Support Vector Machine, and Random Forest. As part of our study, we can predict heart attack by machine learning algorithms, with the latter being 86% accuracy.

1.3 Rationale of the study

Machine learning algorithms are used often in our daily lives. Thus, machine learning algorithms are utilized in the medical field for prognostic purposes. We developed a data set and put forth a model for identifying cardiac events using machine learning techniques.

When an artery supplying blood and oxygen to the heart becomes clogged, a heart attack occurs. The heart's blood supply was carelessly cut off. We can't always predict when our hearts will give out. That's a logical conclusion to draw from this thesis.

1.4 Research Questions

- How can we collect data for the experiment?
- Can we provide the machine learning repertory dataset a better specialist viewpoint to forecast heart attack?
- Can we discover better predicted machine learning algorithms to minimize statistically cardiac attacks?
- What are the performances of used machine learning algorithms?
- What are the approaches of predicting heart attack?
- What is the approach of preprocessing data?
- What is the accuracy of proposed model?

1.5 Expected Output

In this research, we used several machine learning classifiers to try to predict the occurrence of heart attacks. It is anticipated that the completion of this research will result in the development of a model that is accurate to a high degree in its ability to forecast primary heart attacks.

1.6 Report Layout

The report has total 6 Chapters which will be followed given by instructions:

In Chapter 1, we have discussed about the introduction of this research In this part of this research, we discussed in details about the importance of heart disease prediction. Research question, motivation of this research, rationale and expected output has also discussed in this part.

In Chapter 2, we reviewed existing work on heart disease prediction in this chapter. We discussed about other authors approaches, limitations, results, methods in this part. Research scope and challenges also have discussed here.

In Chapter 3, research methodology have mainly discussed here. This chapter shows the data collection procedure, statistics of data, classifiers, figures of stats. Implementation requirements also have discussed in this chapter.

In Chapter 4, results of this research have mainly showed. In this chapter we have showed the result of all the classifiers we have used in this research.

In Chapter 5, shows how this research can impact in our society. Why this work is important and how it will sustain in this arena, this chapter shows mainly.

In Chapter 6, discussion and conclusion of this research have shown. In this part we have also discussed about the future work, and the limitations of this work.

CHAPTER 2
BACKGROUND

2.1 Introduction

Every day, the medical community finds new applications for machine learning. The development is helpful for scientific study. There is a plethora of study being done on this question. Some research publications that are relevant to our work were discovered. The goal of this research is to develop a method for foreseeing cardiac arrest. The majority of the dataset comes from the Dhaka Medical College Hospital, but we also pulled some information from Kaggle. Some new techniques were introduced to us throughout the course of our study. Not a simple effort to execute on our end. We shall go into further depth on this concept in the following chapter. We studied some earlier work that relates to heart attack prediction to fully apply this work and to learn about this new word.

2.2 Related Work

Authors Sharma et al. [6] noted the limited analytical ability and models currently available for predicting cardiovascular illnesses. Significant data on the status of artificial models has been gathered for this study. Learning at a deep level and artificial intelligence go hand in hand. They need 67% precision or better to trade a heart attack. The probabilities of heart illness were predicted using the Naïve Bayesian Decision Tree, the K-Nearest Neighbor, and the Neural Network methods. Application of deep learning to the problem of cardiac illness prognosis. Support vector machines (SVMs) are being utilized to detect the hyper plane, which is necessary for the execution of a ramification. Taking a two-pronged strategy. To continue the good job of defunct medical data sets in the future.

Recently, Hannan et al. [7] recommended condensing a number of studies that used data mining to foretell cardiac issues. Association Rule Construction Data mining is being applied to machine learning systems that are particularly vulnerable. In this investigation, we employed a variety of data mining technologies, including WEKA, RAPID MINER, and MATLAB. In the future, perhaps more informed approaches will emerge that will allow patients to select from among more precise service models.

Avis et al. [8] hoped to use epidemiologically-established techniques to evaluate the link between actual risk and goal. The results of the various responses to potential danger are provided below. Two logistic regressions were performed: one to foretell the likelihood of success and another to foretell the likelihood of failure. Both the real risk and the health risk assessment (HRA) are crucial factors.

Elhoseny et al. hypothesized that binary flections in heart disease may be carried out automatically. The highest accuracy achieved by Automated Heart Diseases Diagnosis (AHDD) is 90.1%. K-Nearest Neighbor, Artificial Neural Network, AdaBoost, and Support Vector Machine are some of the popular ML models used for classification. In our subsequent efforts, let's make an effort to incorporate a data set for both training and testing [9].

Mohan et al. used machine learning techniques to create fiction models with increasing accuracy, with the goal of satisfying the need for crucial form. Hybrid random forest linear models have an accuracy of 88.7 percent (HRFLM). Data classification strategies such as decision trees, K-nearest neighbors, and naive bayes are being applied. Applying different combinations of machine learning and prediction expertise in future attempts is key to achieving the best possible results [10].

The researchers Srinivas et al. advocated doing a review of the anticipated infliction of categorization revealed data mining systems to a significant degree in the medical field. The usage of Naive Bayes, Decision Tree, and ANN may be seen here. They have devoted themselves to developing an insightful and effective prediction method for heart attacks using data mining. The train techniques are used to do an analysis of the purpose. In the future, this may evolve and provide explanation [11].

This study was inspired by the extensive research into the use of Machine Learning algorithms for the detection of cardiovascular heart disease. An overview of the relevant literature is provided in this study. Several methods, such as logistic regression, k-nearest neighbors, random forest classifier, etc., have been used effectively to predict the onset of cardiovascular disease. In Results, you can see that different algorithms excel in different ways when it comes to capturing the specified goals [12].

Marjia Sultana et al. provided an example of how the raw nature of the datasets available for heart disease makes them highly redundant and inconsistent. There must be some sort of pre-processing done on these data sets, during which the high-dimensional data is transformed into a lower-dimensional one [13].

Author demonstrates that, since every conceivable feature is present in the data set, it is possible to extract the features that are most relevant to the problem at hand. By narrowing in on crucial features, we can cut down on the time and effort spent training the algorithm. When comparing algorithms, it's important to look at more than just execution time. Accuracy, for example, is a crucial metric. Bayes Net and SMO classifiers' performance is superior to that of MLP, J48, and KStar, according to research showing that the method proposed in works to improve accuracy. The predictive accuracy, ROC curve, and ROC value are utilized to evaluate the performance of the implemented algorithms (Bayes Net and SMO) on the data set obtained from the WEKA software [14].

Researchers at the University of California, Irvine uses neural networks (NN), decision trees (DT), support vector machines (SVM), and naive bayes to analyze data on people with heart disease. Comparisons are made between the results and these algorithms in terms of speed and precision. The proposed hybrid approach outperforms state-of-the-art methods, with F-1 measure results of 86.8 percent. Convolutional Neural Networks (CNN) are introduced for classification without segmentation. In the learning phase, this technique takes into account Electrocardiogram (ECG) signals from a variety of beginning positions within heart cycles [15].

2.3 Comparative Analysis and Research Summary

These days, heart attacks are a leading cause of death across the world. Data mining had received a lot of attention and had been employed in numerous sorts of diseases including in medical in current research. We are attempting to foresee cardiac arrest. That's why we read certain articles which are linked to my research aim. We're limiting the best machine learning techniques that work effectively with this data set. In this part we are going to compare one some earlier work with other work. Here's a handy table to help you keep track of what papers you've read and what ones are referenced in the last part.

Table 2.1: Comparison between previous works

References	Work	Algorithms	Best Accuracy
[16]	Machine learning for cardiac illness prognosis	NB, DT, KNN, NN	67%
[17]	Method for identifying heart disease using a machine learning classifier	SVM, ANN, KNN, DT	92.37%
[18]	Effective Prediction of Heart Disease Using Hybrid Machine Learning	RF, LR, NB, DT	88.7%
[19]	Machine learning method for cardiac illness detection using multi-agent feature wrapping	KNN, ANN, AdaBoost, SVM	90%

2.4 Research Scope

Classification revealed machine learning algorithms can forecast heart attack utility or not. We find better accuracy using machine learning accession. We 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 we want to work on deep learning algorithms that are based on artificial neural networks that work as a human brain.

2.5 Challenges

There are some challenges we have faced during this work -

- Collecting quality data.
- Data preprocessing.
- Data purification and source authentication.

- Preparation and publication of a data collection for research study.
- Getting accuracy above 80%.
- Making a decision based on the results of tests.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

It is in this chapter that we explore our research's theoretical foundations. Some methods and datasets will help us get the task done faster and more accurately. Even this is sufficient to get a clear picture of the work involved. There are a number of instruments and technologies that we'll explore in this chapter. Machine learning and data mining are used to assess the collected data and implement it in a way that is beneficial for everyone.

3.2 Research Subject and Instrumentation

This research aims to anticipate heart attacks and achieve the highest level of precision possible. To run the model, I must first build the dataset and choose which methods to employ. Supervised Learning and Unsupervised Learning are two techniques to develop a method in machine learning algorithms. We employed supervised machine learning methods in this study. In supervised learning, some classification methods are utilized to address classification issues. We employed the Logistic Regression, K-nearest neighbors, Decision Tree, Support Vector Machine, Random Forest, and XGBoost method in this study. In the forthcoming part on the suggested technique, we will explore all of the algorithms, how they function, and which one is the best, and we will attempt to determine which algorithm produces the best results based on the data set we gathered.

3.3 Data Collection Procedure

Machine Learning algorithms work exceptionally well when the data collected to train the machine is more dependable. Therefore, data collecting is crucial to our effort. The dataset is split into trains and tests. From this information, we determined which individuals suffer from heart attacks and which do not. We gathered data manually from the Dhaka Medical College Hospital and from the open-source dataset repository Kaggle. It has 4 thousand records and 15 columns. Table 3.1 shows the demography of dataset.

Table 3.1: Demography of Dataset

Sex	Age	Behavioral
Male	Continuous Value	Current Smoker
Female		Avg cigarette per day

3.4 Statistical Analysis

We have collected our data from the Dhaka Medical College Hospital and from the open-source dataset repository Kaggle. It has 4 thousand records and 15 columns.

Demographic

Sex: male or female

Age: continuous value.

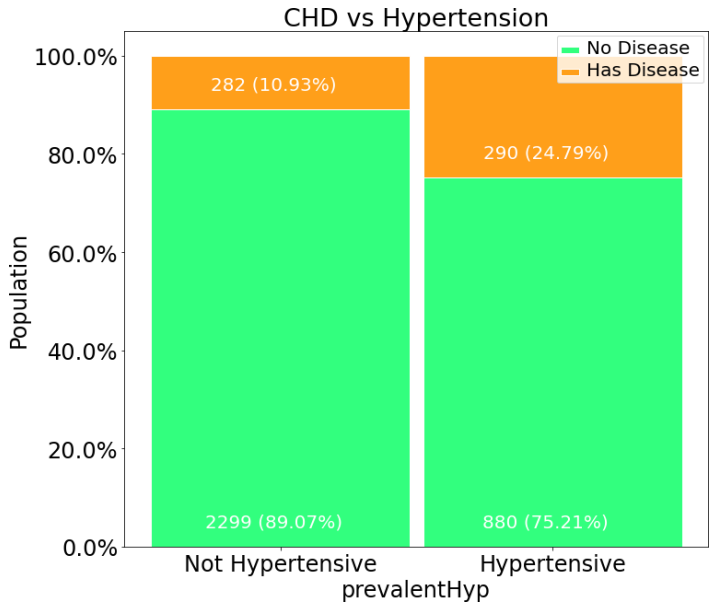


Figure 3.4.1: CHD vs Hypertension

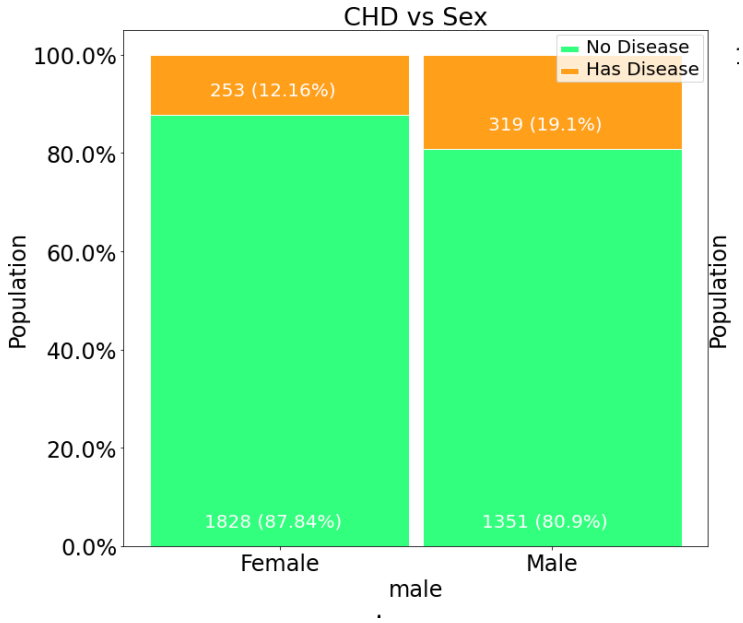


Figure 3.4.2: CHD vs Gender

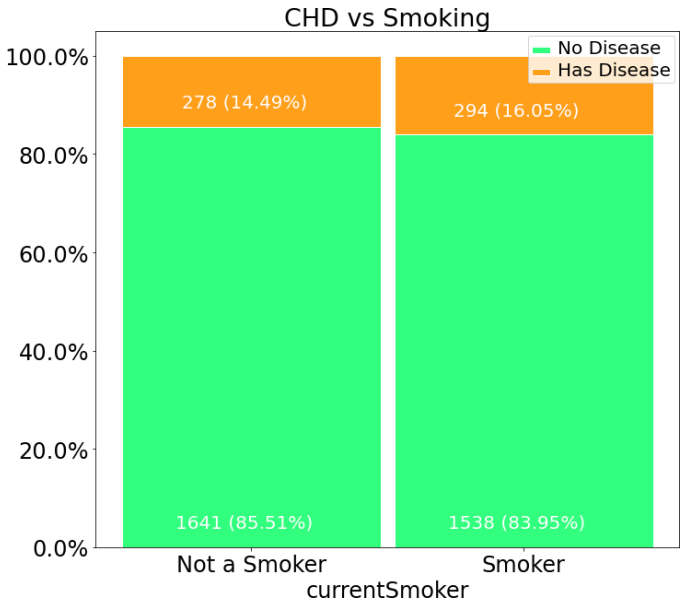


Figure 3.4.3: CHD vs Smoking

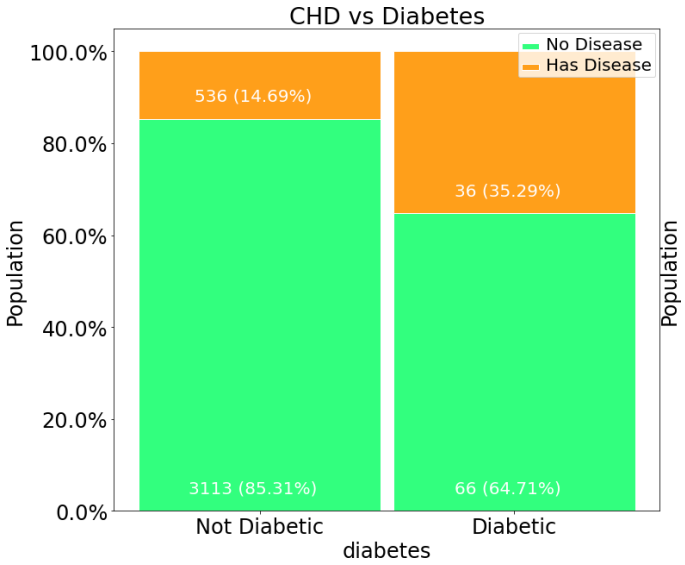


Figure 3.4.4: CHD vs Diabetes

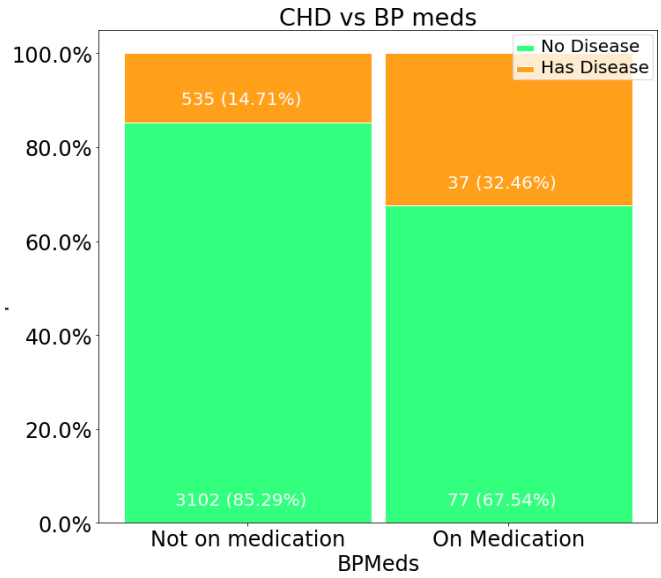


Figure 3.4.5: CHD vs Blood Pressure

3.5 Data Pre-Processing

When doing research, the initial step is to preprocess data. The first phase of data mining is data processing. This is why we obtain news from a variety of sources. To remedy this, we will first prepare the data. Each news organization produces a range of articles, which

comprise an Initial Data Set. These data sets are divided into several numbers. This data is handled on an individual basis.

3.6 Proposed Model Workflow

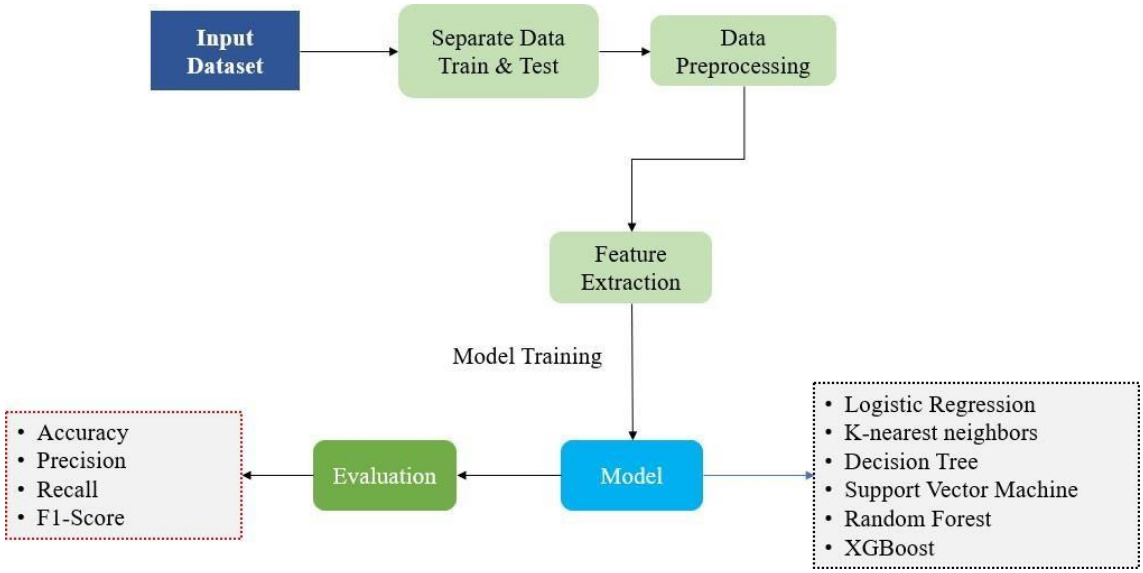


Figure 3.6.1: Proposed model workflow

3.7 Insert Data

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

3.8 Separate Data

To complete this task we 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.

3.9 Machine Learning Model

Machine learning algorithms used in medical science every day. For that's why we used machine learning algorithms to predict heart attack. We used different types of machine learning classifiers like Decision Tree, K-nearest neighbors, Logistic Regression, XGBoost, Support Vector Machine, and Random Forest. We tried to find out the best accuracy which models give.

3.9.1 Logistic Regression

Predicting a conditional unconditional object variable is a common use of logistic regression, a prominent supervised machine learning approach. Logistic regression of a binary variable is being used to determine the likelihood. It is used to make predictions about a categorical dependent variable. Due to the high size of the data set, logistic regression will not be able to properly classify the information. The association between a dependent variable and one or more i-uncommitted variables may also be calculated using logistic regression. There are four characteristics that may be used in a logistic regression analysis: sepal length, petal length, sepal width, and petal width. Logistic regression is used in machine learning applications because it provides space for classification algorithms based on past data. Discrete results are being delivered via logistic regression [20].

3.9.2 K-Nearest Neighbors

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 [21].

3.9.3 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 [22].

3.9.4 Random Forest

Supervised machine learning algorithms like Random Forest exist. The random forest technique is now being utilized in regression and classification. It is the composition of decision trees. It sends out favorable reckoning to understand spontaneously. Largedatasets 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 madewith 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 forthat 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 [23].

3.9.5 Support Vector Machine

A supervised machine learning algorithm is Support Vector Machine (SVM) 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 [24].

3.9.10 XGBoost

XGBoost, an ensemble decision-tree method, is used by gradient-boosting frameworks. While decision trees are typically easy to grasp visually and conceptually, it may take some time to gain an intuitive understanding of older tree-based algorithms [25].

3.10 Implementation Requirements

There are some requirements to implemented this project.

Hardware or Software Requirements

- Windows 7 or above operating system
- Portable or in-built Hard Disk above 300 GB
- Minimum 2 GB RAM
- Google Chrome or Mozilla Firefox

Developing Tools

- Python with Tensor flow
- Jupiter
- Notepad++

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

Specifically, the descriptive analysis of the data utilized in the research, as well as the experimental outcomes of our project, are the primary emphasis of this chapter 4.

4.2 Experimental Results

In this section, we described the outcomes of the algorithms used in this research. We utilized six categorization techniques to forecast heart attack. The methodology section describes the six categorization methodologies. Logistic Regression, K-nearest neighbors, Decision Tree, Support Vector Machine, Random Forest, and XGBoost are the algorithms. In text data sets, all classifiers perform exceptionally well. By understanding the accuracy rate, we can determine the optimal model for these data. From table 4.1 the algorithms are Logistic Regression, K-nearest neighbors, Decision Tree, Support Vector Machine, Random Forest, XGBoost and the accuracy is gradually for all classifiers 66%, 78%, 76%, 70 %, 87%, 80%. Comparing the accuracy of each classifier and picking the optimal model from among six options. For a superior outcome, we determined the confusion matrices for all classifiers presented in Chapter 3. Confusion matrix makes it easy to determine if a procedure is confounding two classifications. Precision, Recall, and F1-Score were identified for all classifiers, and the resulting score is divided between heart disease and no heart disease. This section describes the accuracy, precision, recall, and several other aspects. Table 4.1 shows the classification report of six classifiers on this model.

TABLE 4.1: CLASSIFICATION REPORT OF THIS MODEL

Algorithms	Accuracy	Precision	Recall	F1-Score
LR	66%	0.65	0.67	0.65
KNN	78%	0.79	0.77	0.77
DT	76%	0.76	0.75	0.76

SVM	70%	0.69	0.69	0.71
RF	87%	0.86	0.85	0.88
XGBoost	80%	0.79	0.81	0.79

4.3 Descriptive Analysis

In this research work, we found the best result on Random Forest 87% accuracy rather than other classifiers. In this part we will show the break down result of each and every classifier.

Logistic Regression (LR)

```
Train Result:
================================================
Accuracy Score: 65.95%

_______________________________________________
CLASSIFICATION REPORT:
                    0            1  accuracy    macro avg  weighted avg
precision    0.677361     0.632842  0.659528     0.655101      0.657466
recall       0.734097     0.567237  0.659528     0.650667      0.659528
f1-score     0.704589     0.598247  0.659528     0.651418      0.657065
support    2531.000000  2045.000000  0.659528  4576.000000   4576.000000
_______________________________________________
Confusion Matrix:
 [[1858  673]
 [ 885 1160]]
```

Figure 4.3.1: Train Result of LR

```
Test Result:
================================================
Accuracy Score: 65.85%

_____________________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.685507    0.617582  0.658515     0.651545      0.655964
recall       0.731066    0.564257  0.658515     0.647662      0.658515
f1-score     0.707554    0.589717  0.658515     0.648635      0.656303
support    647.000000  498.000000  0.658515  1145.000000   1145.000000

_____________________________________________________
Confusion Matrix:
 [[473 174]
 [217 281]]
```

Figure 4.3.2: Test Result of LR

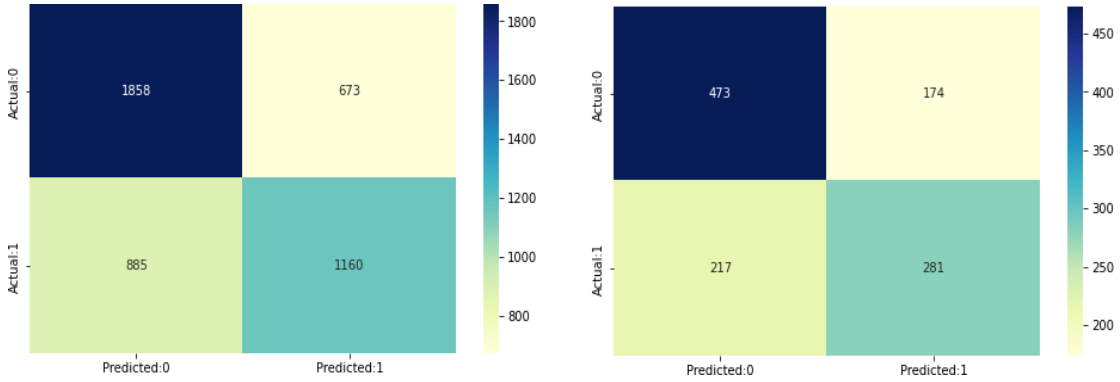


Figure 4.3.3: Confusion Matrix of train (left) & test (right) result of LR

K-Nearest Neighbors (KNN)

```
Train Result:
================================================
Accuracy Score: 85.64%

_________________________________________________
CLASSIFICATION REPORT:
                    0            1  accuracy    macro avg  weighted avg
precision    0.953095     0.776715  0.856425     0.864905      0.874271
recall       0.778744     0.952567  0.856425     0.865655      0.856425
f1-score     0.857143     0.855700  0.856425     0.856421      0.856498
support    2531.000000  2045.000000  0.856425  4576.000000   4576.000000
_________________________________________________
Confusion Matrix:
 [[1971  560]
 [  97 1948]]
```

Figure 4.3.4: Train Result of KNN

```
Test Result:
================================================
Accuracy Score: 77.21%

_________________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.886000    0.683721  0.772052     0.784860      0.798022
recall       0.684699    0.885542  0.772052     0.785120      0.772052
f1-score     0.772450    0.771654  0.772052     0.772052      0.772104
support    647.000000  498.000000  0.772052  1145.000000   1145.000000
_________________________________________________
Confusion Matrix:
 [[443 204]
 [ 57 441]]
```

Figure 4.3.5: Test Result of KNN

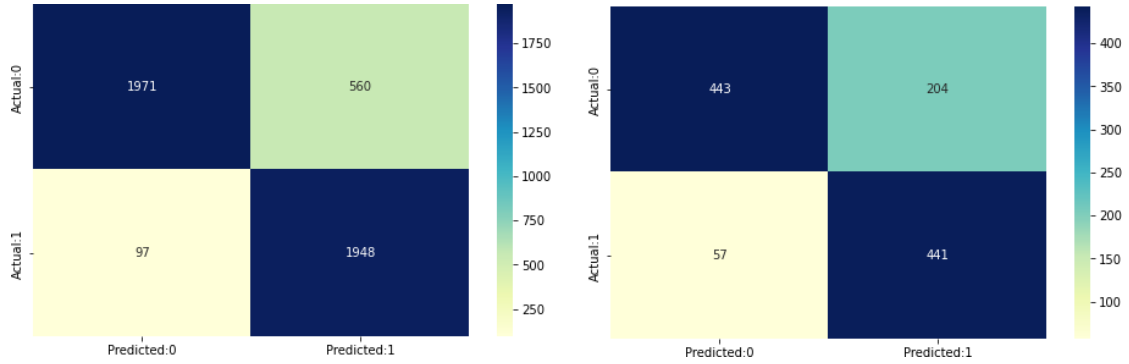


Figure 4.3.6: Confusion Matrix of train (left) & test (right) result of KNN

Decision Tree (DT)

```
Train Result:
================================================
Accuracy Score: 100.00%

_________________________________________________
CLASSIFICATION REPORT:
                0       1  accuracy  macro avg  weighted avg
precision     1.0     1.0       1.0        1.0           1.0
recall        1.0     1.0       1.0        1.0           1.0
f1-score      1.0     1.0       1.0        1.0           1.0
support    2531.0  2045.0       1.0     4576.0        4576.0

_________________________________________________
Confusion Matrix:
 [[2531    0]
 [   0 2045]]
```

Figure 4.3.7: Train Result of DT

```
Test Result:
================================================
Accuracy Score: 76.16%

________________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.807566    0.709497  0.761572     0.758531      0.764912
recall       0.758887    0.765060  0.761572     0.761974      0.761572
f1-score     0.782470    0.736232  0.761572     0.759351      0.762360
support    647.000000  498.000000  0.761572  1145.000000   1145.000000

________________________________________________
Confusion Matrix:
 [[491 156]
 [117 381]]
```

Figure 4.3.8: Test Result of DT

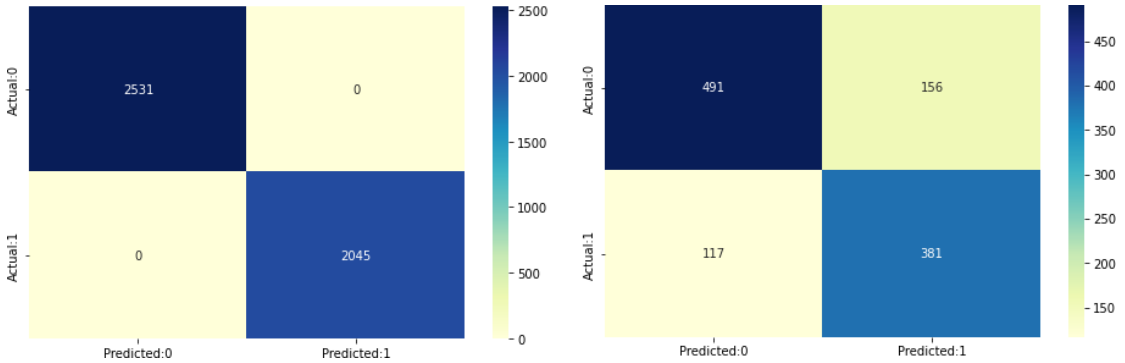


Figure 4.3.9: Confusion Matrix of train (left) & test (right) result of DT

Support Vector Machine (SVM)

```
Train Result:
================================================
Accuracy Score: 72.62%

_________________________________________________
CLASSIFICATION REPORT:
                    0            1  accuracy    macro avg  weighted avg
precision    0.751179     0.694882   0.72618     0.723031      0.726020
recall       0.755038     0.690465   0.72618     0.722751      0.726180
f1-score     0.753103     0.692666   0.72618     0.722885      0.726094
support    2531.000000  2045.000000   0.72618  4576.000000   4576.000000

_________________________________________________
Confusion Matrix:
 [[1911  620]
 [ 633 1412]]
```

Figure 4.3.10: Train Result of SVM

```
Test Result:
================================================
Accuracy Score: 69.34%

_________________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.736422    0.641618   0.69345     0.689020      0.695189
recall       0.712519    0.668675   0.69345     0.690597      0.693450
f1-score     0.724273    0.654867   0.69345     0.689570      0.694086
support    647.000000  498.000000   0.69345  1145.000000   1145.000000

_________________________________________________
Confusion Matrix:
 [[461 186]
 [165 333]]
```

Figure 4.3.11: Test Result of SVM

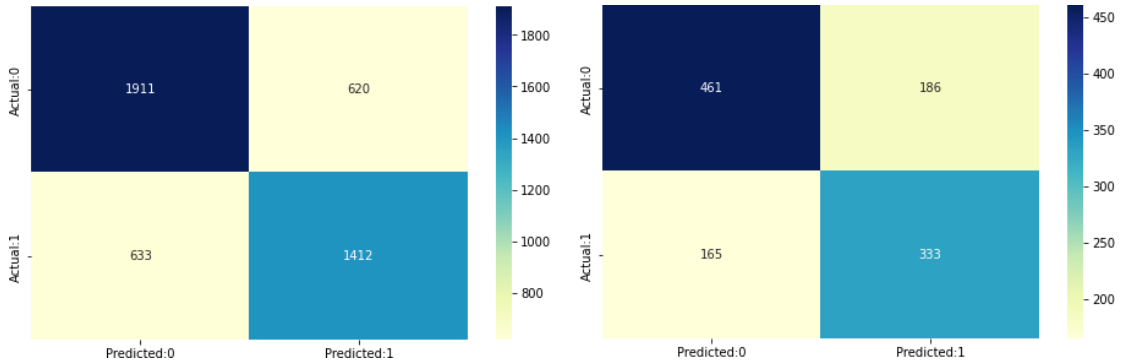


Figure 4.3.12: Confusion Matrix of train (left) & test (right) result of SVM

Random Forest (RF)

```
Train Result:
================================================
Accuracy Score: 100.00%

_______________________________________________
CLASSIFICATION REPORT:
                    0       1  accuracy  macro avg  weighted avg
precision         1.0     1.0       1.0        1.0           1.0
recall            1.0     1.0       1.0        1.0           1.0
f1-score          1.0     1.0       1.0        1.0           1.0
support        2531.0  2045.0       1.0     4576.0        4576.0

_______________________________________________
Confusion Matrix:
 [[2531    0]
 [   0 2045]]
```

Figure 4.3.13: Train Result of RF

```
Test Result:
================================================
Accuracy Score: 86.55%

_____________________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.880989    0.845382  0.865502     0.863185      0.865502
recall       0.880989    0.845382  0.865502     0.863185      0.865502
f1-score     0.880989    0.845382  0.865502     0.863185      0.865502
support    647.000000  498.000000  0.865502  1145.000000   1145.000000

_____________________________________________________
Confusion Matrix:
 [[570  77]
 [ 77 421]]
```

Figure 4.3.14: Test Result of RF

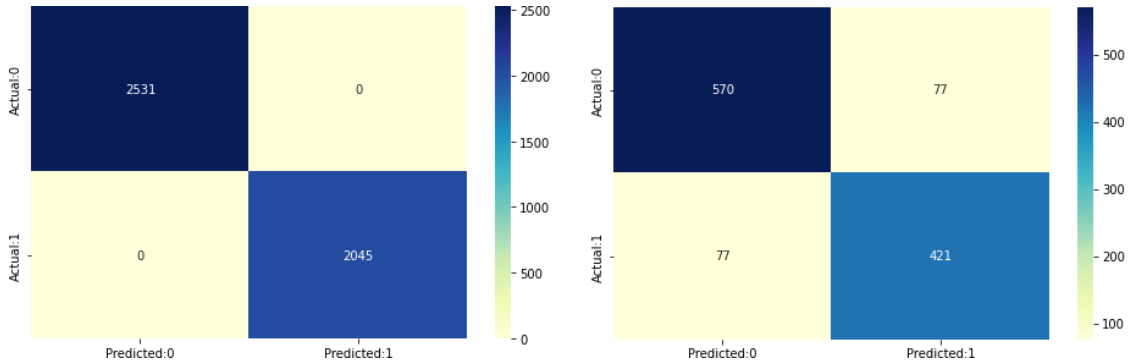


Figure 4.3.15: Confusion Matrix of train (left) & test (right) result of RF

XGBoost

```
Train Result:
================================================
Accuracy Score: 81.45%

_______________________________________________
CLASSIFICATION REPORT:
                    0            1  accuracy    macro avg  weighted avg
precision    0.822965     0.803245  0.814467     0.813105      0.814152
recall       0.846701     0.774572  0.814467     0.810637      0.814467
f1-score     0.834664     0.788648  0.814467     0.811656      0.814100
support    2531.000000  2045.000000  0.814467  4576.000000   4576.000000

_______________________________________________
Confusion Matrix:
 [[2143  388]
 [ 461 1584]]
```

Figure 4.3.16: Train Result of XGBoost

```
Test Result:
================================================
Accuracy Score: 79.13%

_______________________________________________
CLASSIFICATION REPORT:
                    0           1  accuracy    macro avg  weighted avg
precision    0.808157    0.768116  0.791266     0.788137      0.790742
recall       0.826893    0.744980  0.791266     0.785937      0.791266
f1-score     0.817418    0.756371  0.791266     0.786894      0.790867
support    647.000000  498.000000  0.791266  1145.000000   1145.000000

_______________________________________________
Confusion Matrix:
 [[535 112]
 [127 371]]
```

Figure 4.3.17: Test Result of XGBoost

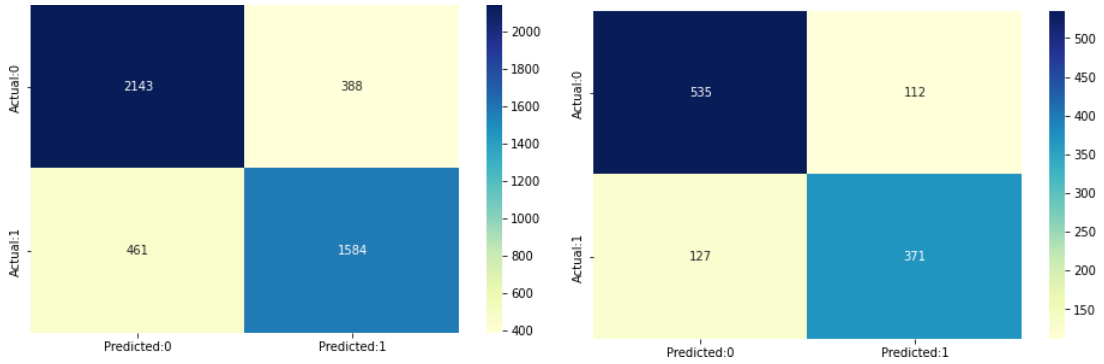


Figure 4.3.18: Confusion Matrix of train (left) & test (right) result of XGBoost

4.4 Comparison based on accuracy

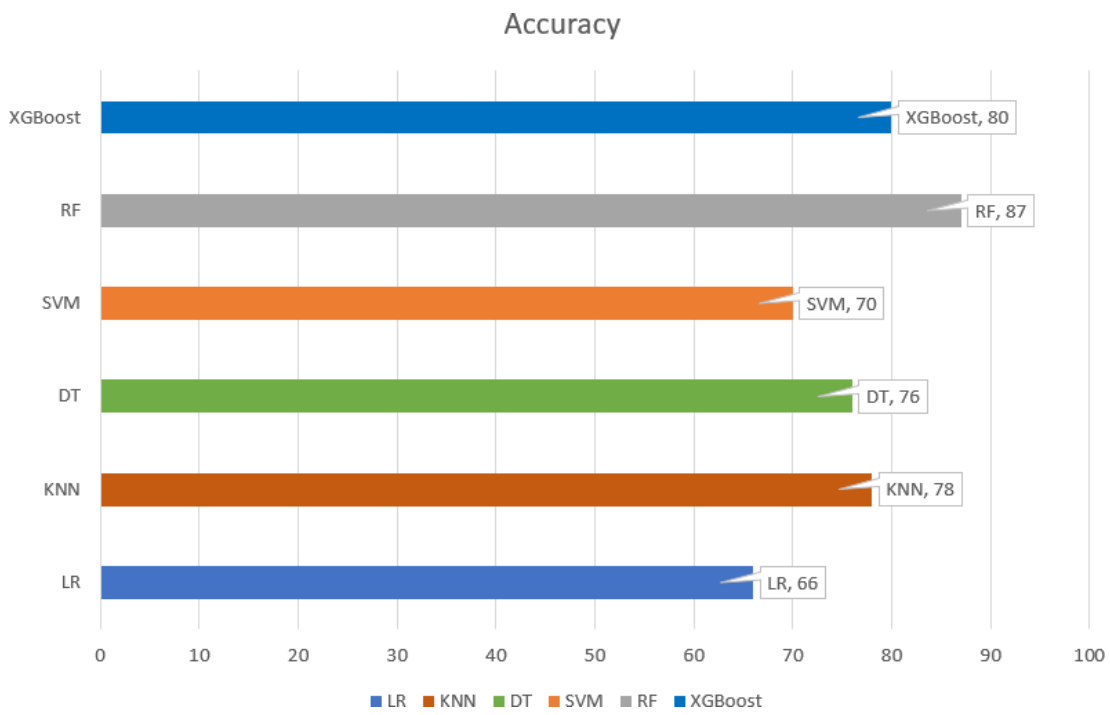


Figure 4.4.1: Comparative analysis of six classifiers based on test accuracy

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Heart Attack has a significant effect on society. The heart plays a crucial part in our impulsive intelligence. The performance of the heart, body, and brain as adjuncts contributes to the sagacity of emotion. Typically, a number of mental and behavioral issues, such as anxiety and despair, might be linked to these heart attacks. These worries may inhibit the adoption and maintenance of health-promoting habits. Patients suffering heart attacks experience a variety of mental and physical weaknesses, including chest pain, insomnia, and sadness. Patients suffering with these symptoms limit their physical and social activities, resulting in a diminished quality of life. High mortality rates are associated with a low quality of life. The exhaustive listing of social variables influencing a number of individuals can increase the risk of dying from 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 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

This 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 we produced the model and why we want to decide this completely so that the public can freely demand the appropriate solution. We should have recognized this idea. For this reason, the models are being sustained and that is needed for me. We created an appropriate model that provides the appropriate knowledge about this heart attack prediction. We also need to take action on how to sustain the model incessantly.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the study

A branch of artificial intelligence called machine learning generates a steady stream of coding and statistical innovations. An area of AI is machine learning. Machine learning can help to improve health outcomes because it can increase patient involvement in the healing process. By determining the most likely explanation for all similar patients' test results, machine learning techniques can increase diagnosis accuracy at every level. Heart attacks are more common as people get older, and those over the age of 60 are at higher risk of having one. Researchers are using numerous machine learning techniques to predict the beginning of a heart attack. In-depth descriptions of the techniques we employ to forecast cardiovascular disease, such as Logistic Regression, K-nearest Neighbors, Decision Tree, Support Vector Machine, Random Forest, and XGBoost, were the main objective of this study. On the same dataset, various predictive data mining techniques have been tested with varying degrees of success, and Random Forest techniques have been found to produce results with the highest accuracy of 87%. The algorithms include XGBoost, K-nearest Neighbors, Decision Tree, Support Vector Machine, and Logistic Regression. For all classifiers, the accuracy increases gradually (66%, 78%, 76%, 70%, 87%, and 80%). All classifiers perform remarkably well in text data sets. We can choose the best model for these data by determining the accuracy rate.

6.2 Conclusions

In machine learning, the experimental analysis reveals that variation interactions with heart attack prediction have been examined. The job necessitates the purification of the employed algorithms before identifying the most effective mechanism. Python software subscribes to the compatible soliciting and investigation of unique machine learning algorithms. Pre-processing is a significant step in machine learning that facilitates the accumulation of precise findings. We restricted the explanation of many machine learning algorithms on the database and discovered the algorithm with the highest accuracy. After evaluating the model, this research found the best accuracy on Random Forest classifier which is 87%.

6.3 Implication for further study

For this research we 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.

REFERENCES

[1] Klarin, D., & Natarajan, P. (2022). Clinical utility of polygenic risk scores for coronary artery disease. *Nature Reviews Cardiology*, *19*(5), 291-301.

[2] Ahmed, A. U., & Sharmeen, T. (2004). Assessing the performance of conditional cash transfer programs for girls and boys in primary and secondary schools in Bangladesh. *International Food Policy Research Institute, Washington, DC*.

[3] Khan, F., Siva Prasad, B. V. V., Syed, S. A., Ashraf, I., & Ramasamy, L. K. (2022). An Efficient, Ensemble-Based Classification Framework for Big Medical Data. *Big Data*, *10*(2), 151-160.

[4] Wegener, S. (2022). Triggers of stroke: anger, emotional upset, and heavy physical exertion. New insights from the INTERSTROKE study. *European Heart Journal*.

[5] Möller-Leimkühler, A. M. (2022). Gender differences in cardiovascular disease and comorbid depression. *Dialogues in clinical neuroscience*.

[6] 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.

[7] 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.

[8] 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.

[9] 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 machine learning approach for heart disease diagnosis. Comput. Mater. Contin, 67, 51-71.

[10] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.

[11] Srinivas, K., Rao, G. R., & Govardhan, A. (2010, August). Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining

techniques. In *2010 5th International Conference on Computer Science & Education* (pp. 1344-1349). IEEE.

[12] Masih, N., & Ahuja, S. (2022). Application of data mining techniques for early detection of heart diseases using Framingham heart study dataset. *International Journal of Biomedical Engineering and Technology*, *38*(4), 334-344.

[13] Sultana, M., Haider, A., & Uddin, M. S. (2016, September). Analysis of data mining techniques for heart disease prediction. In *2016 3rd international conference on electrical engineering and information communication technology (ICEEICT)* (pp. 1-5). IEEE.

[14] Saleh, B., Saedi, A., Al-Aqbi, A., & Salman, L. (2020). Analysis of weka data mining techniques for heart disease prediction system. *International journal of medical reviews*, *7*(1), 15-24.

[15] Cheng, C. A., & Chiu, H. W. (2017, July). An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2566-2569). IEEE.

[16] Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018, March). Prediction of heart disease using machine learning. In *2018 second international conference on electronics, communication and aerospace technology (ICECA)* (pp. 1275-1278). IEEE.

[17] 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.

[18] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.

[19] 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 machine learning approach for heart disease diagnosis. *Comput. Mater. Contin*, *67*, 51-71.

[20] Huang, F. L. (2022). Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education*, *90*(1), 213-228.

[21] Lin, G., Lin, A., & Gu, D. (2022). Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient. *Information Sciences*, *608*, 517-531.

[22] Islam, T., Kundu, A., Islam Khan, N., Chandra Bonik, C., Akter, F., & Jihadul Islam, M. (2022). Machine Learning Approaches to Predict Breast Cancer: Bangladesh Perspective. In *International Conference on Ubiquitous Computing and Intelligent Information Systems* (pp. 291-305). Springer, Singapore.

[23] Balyan, A. K., Ahuja, S., Lilhore, U. K., Sharma, S. K., Manoharan, P., Algarni, A. D., ... & Raahemifar, K. (2022). A hybrid intrusion detection model using ega-pso and improved random forest method. *Sensors*, *22*(16), 5986.

[24] Noble, W. S. (2006). What is a support vector machine?. *Nature biotechnology*, *24*(12), 1565-1567.

[25] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, *1*(4), 1-4.

Heart Attack Prediction

ORIGINALITY REPORT

21%	18%	16%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
2	Ruby Hasan. "Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction", ITM Web of Conferences, 2021 Publication	1%
3	academic-accelerator.com Internet Source	1%
4	Akansha Singh, Anamika Jain. "Prediction of Heart Disease using Dense Neural Network", 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), 2022 Publication	1%
5	Submitted to Daffodil International University Student Paper	1%
6	Sunita Dhote, Rupesh Pais, Chandan Vichoray, S. Baskar. "Fuzzy Machine Learning Model in Real-World Physical Domains; A State-of-the-	1%