A MINI PROJECT REPORT ON

Heart Diseases Prediction

Submitted by

Priyanshu Agrawal	**(204200156)**
Prajwal Negi	**(204200146)**
Rihans Jain	**(204200170)**
Sarthak Agarwal	**(204200185)**
Saksham Gupta	**(204200180)**
Prerak	**(204200148)**

Under the Guidance of

Mr.Atul Kumar Uttam (Assistant Professor)

Department of Computer Engineering & Applications

Institute of Engineering & Technology

GLA UNIVERSITY
MATHURA
Recognised by UGC Under Section 2(f)
Accredited with A Grade by NAAC
12-B Status from UGC

GLA University Mathura

INDEX

CERTIFICATE

This is to certify that the project entitled **Heart Diseases Prediction Using Ml** submitted by **Candidates** for the award of **Bachelor of Computer Application** is a record of the bonafide work carried out by them under my supervision and guidance at the **Department of Computer Science and Application, GLA University Mathura**.

Project Supervisor
(Mr. Atul Kumar Uttam)

Date: ______________

Head of Department
(Dr. Rohit Agrawal)

Program Coordinator
(Mr. Narendra Mohan)

Project in Charge
(Mr. DP Yadav)

DECLARATION

We hereby declare that the project entitled **Heart Diseases Prediction Using Machine Learning** submitted by us, for the award of **Bachelor of Computer Application** is a record of the bonafide work carried out by us under the guidance of Mr.Atul Kumar Uttam (Associate Professor) of the **Department of Computer Engineering & Application , GLA Mathura.**

Submitted By: -

Priyanshu Agrawal -	204200156
Prajwal Negi -	204200146
Rihans Jain -	204200170
Sarthak Agarwal -	204200185
Saksham Gupta -	204200180
Prerak -	204200148

ACKNOWLEDGEMENT

We sincerely thank **Mr.Atul Kumar Uttam** (Associate Professor) of Department of Computer Engineering & Application , **GLA Mathura** , for his kind and invaluable guidance in the preparation and completion of the project entitled **“Heart Diseases Prediction Using Machine Learning”.**

We feel highly esteemed of getting the privilege of his guidance without which it would have been a difficult task to accomplish the assigned task.

Priyanshu Agrawal:	(204200156)
Prajwal Negi:	(204200146)
Rihans Jain:	(204200170)
Sarthak Agarwal:	(204200185)
Saksham Gupta:	(204200180)
Prerak:	(204200148)

BACKGROUND

Heart attacks are the most common cause of death among all deadly disorders. Medical professionals perform various surveys on heart disorders in order to acquire information about heart patients, their symptoms, and the progression of their disease. Patients with prevalent diseases that exhibit typical symptoms are increasingly being reported. People in this fast-paced world want to live a very luxurious life, so they work like machines in order to earn a lot of money and live a comfortable life. As a result, they forget to take care of themselves, and as a result, their eating habits and overall lifestyle change. As a result, they are more tense, have high blood pressure and sugar at a young age, and they don't give themselves enough rest.

The term "heart disease" refers to a variety of disorders that affect the heart. The number of persons who have heart disease is increasing (health topics, 2010). According to the World Health Organization, a substantial number of people die each year as a result of heart disease all over the world. Heart disease is also included as one of Africa's leading causes of death. Marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining, and mobile computing are just a few of the applications where data mining has been employed. Data mining has recently been used to successfully discover healthcare fraud and abuse incidents.

Background of The Study

In the medical industry, data analysis is critical. It provides a solid foundation for making important judgments. It aids in the creation of a comprehensive study proposal. One of the most important applications of data analysis is that it aids in the removal of human bias from medical conclusions through statistical treatment. Because of nontrivial information in vast volumes of data, data mining is used for exploratory analysis. The health-care industry collects massive volumes of data that may contain hidden information that is useful for giving appropriate findings and making two effective judgments based on data. Some data mining techniques are utilized to improve the experience and conclusions that have been provided. The heart predicting system will be based on.

Problem Statement

Heart disease can be effectively managed with a combination of lifestyle changes, medication, and surgery in some circumstances. The symptoms of heart disease can be lessened and the heart's function enhanced with the correct treatment. The projected outcomes can be utilized to prevent and thereby minimize the cost of surgery and other costly treatments**.**

The overarching goal of my research will be to accurately predict the presence of heart disease using only a few tests and features. The attributes that are taken into account are the primary foundation for testing and, for the most part, provide accurate findings. Many more input attributes might be used, but our goal is to forecast the risk of heart disease with fewer and more efficient features. Rather than the knowledge-rich data hidden in the data set and databases, decisions are frequently made purely on doctors' intuition and expertise. This approach leads to unfavorable biases, errors, and high medical expenses, all of which have an impact on the quality of care offered to patients.

OBJECTIVE

Main Objectives

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set.

Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

Specific Objectives

- Provides a new approach to concealed patterns in the data.
- Helps avoid human biases.
- To implement Naïve Bayes Classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests

JUSTIFICATION

Clinical choices are frequently decided on the basis of a doctor's intuition and experience rather than the knowledge-rich data contained in the dataset. This practice results in unintended biases, errors, and exorbitant medical costs, all of which have an impact on the quality of care offered to patients. Clinical decision support will be integrated with computer-based patient records in the proposed system (Data Sets). Medical errors will be reduced, patient safety will be improved, undesirable practice variation will be reduced, and patient outcomes will improve. This approach holds promise since data modeling and analysis technologies, such as data mining, have the ability to create a knowledge-rich environment that can improve the quality of healthcare judgments dramatically.

Due to the large number of records in the medical data domain, it has become vital to apply data mining techniques to aid in decision support and prediction in the healthcare industry. As a result, medical data mining contributes to business intelligence, which is beneficial in disease diagnosis.

Scope

Integration of clinical decision support with computer-based patient records could reduce medical errors, increase patient safety, reduce undesirable practice variation, and improve patient outcomes, according to the project's scope. This approach holds promise since data modeling and analysis technologies, such as data mining, have the ability to create a knowledge-rich environment that can assist enhance the quality of healthcare judgments dramatically.

LIMITATION

Medical diagnosis is seen as an important yet complex task that must be completed accurately and efficiently. It would be really advantageous to automate the process. Clinical judgments are frequently made based on the doctor's intuition and experience rather than the database's knowledge-rich facts. This practice results in unintended biases, errors, and exorbitant medical costs, all of which have an impact on the quality of care offered to patients. Data mining has the ability to create a knowledge-rich environment that can improve the quality of therapeutic judgments dramatically.

SOFTWARE REQUIREMENT ANALYSIS

System Analysis is a detailed study of the various operations performed by a system and their relationship within and outside the system. It is a systematic technique that defines goals and objectives the goal of the development is to deliver the system in the line with the user's requirements, and analysis is this process. System study has been conducted with the following objectives in mind: -

- Identify the client's need.
- Evaluate the system concept for feasibility.
- Perform economical and technical analysis.
- Allocate functional to hardware, software, people, database and other system elements.
- Establish cost and schedule constraints
- Both hardware and software expertise is required to successfully attain the objectives.

Requirement Analysis

Information gathering is usually the first phase of the software development project. The purpose of this phase is to identify and document the exact requirements for the system. The user's request identifies the need for a new information system and on investigation re-defined the new problem to be based on MIS, which supports management. The objective is to determine whether the request is valid and feasible before a recommendation is made to build a new or existing manual system continues.

The major steps are:

- Defining the user requirements
- Studying the present system to verify the problem
- Defining the performance expected by the candidate to user requirement

Hardware Requirements

Processor	: Intel Dual Core or more
Processor Speed	:1.5 GHZ
RAM	: 4 GB
Hard Disk	: 20 GB of free space

Software Requirements

Operating System	: Window 7 and higher
Back End	: Python

Tools and Technologies

Tools

Python

Technologies:

Sklearn

Numpy

Pandas

Matplotlib

Abstract:

Heart disease causes a high mortality rate around the world and has become a very significant health threat for many people. Early prediction can save many lives; detecting cardiovascular disease early can help prevent it from getting worse .Heart diseases and other related diseases are growing at an alarming rate. Machine learning can help provide a healthier future by making accurate predictions about the development of said diseases. The medical industry should not reject machine learning because it will save people in the long run. In the proposed work, a novel machine learning approach is proposed to predict heart disease.The study proposed using the Cleveland heart disease dataset and applied machine learning techniques such as regression, classification, Random Forest and Decision Tree. This study presents a novel technique of machine learning which has not been performed previously. 3 Algorithms like 1. Random Forest, 2. Decision Tree and 3. Hybrid models (Hybrid of random forest and decision tree) are used in machine learning models to create accuracy levels of 88.7% through heart disease prediction models, for instance The interface is designed to take the user's input parameter, which is then analyzed using a hybrid model of Decision Tree and Random Forest.

Key Words: Kaggle Heart Disease Database, Decision Trees, Random forest, Hybrid algorithm, Machine learning

INTRODUCTION

Machine Learning is helpful for analyzing and understanding massive amounts of information. It is used to extract data and make decisions about subsequent applications. Clustering, association rule mining, and classifications are the most common data mining approaches. These data mining approaches can be implemented using a variety of algorithms.

Though simulation tools such as Weka are accessible, Python programming is gaining traction with these methods implemented using scikit-learn packages. As a result, the implementation of data mining principles in real time is more trustworthy than ever before.

Machine learning is becoming increasingly popular in the medical diagnosis business, where computer analysis may reduce manual error and enhance accuracy. Machine learning algorithms make disease diagnosis more accurate. Machine learning techniques are used to forecast diseases such as heart disease, liver disease, diabetes, and tumor. In the medical industry, classification techniques such as decision trees, naive Bayes, and SVM (Support Vector Machine) were employed; similarly, regression algorithms such as Random forest, lasso, and logistic regressions were used. Deep learning algorithms are widely employed in the medical diagnosis area for most tumor forecasts.

According to surveys, almost 17 million people die each year as a result of cardiovascular disorders (CVD). Many lives could be saved if sickness is detected early, and mortality can be minimized if patients receive therapy on time .Cardiovascular diseases cover a wide range of hazards, including heart disease, stroke, and other conditions. These disorders are becoming more widespread even in younger age groups as a result of a lack of physical exercise caused by lifestyle changes. The primary causes of heart disease are smoking, lack of physical activity, high cholesterol foods, junk food, and poor living practices.

Motivation

1) A major challenge facing healthcare organizations(hospitals,medical centers) is the priority of quality service at affordable cost.
2) Hospitals must also minimize the cost of clinical tests.they can achieve these results by employing appropriate computer-based information and/or decision support systems.

LITERATURE SURVEY

There are ten research papers that explore the computational methods to predict heart diseases. The summaries of them have been presented in a nutshell**.**

Ashish Chhabbi et al. [1] have investigated various data mining strategies for finding hidden patterns from a dataset in order to answer complicated questions in heart disease prediction. The data was gathered from the UCI repository. They updated the k-means algorithm and used Naive Bayes. The results suggest that modified k-means outperform simple k-means in terms of accuracy (where number of clusters were predefined)

Mai Shouman et al. [2] k-Nearest-Neighbors (k-NN) was used in the identification of heart disease.The accuracy of k-NN is higher than that of neural network ensembles, according to this article. However, unlike Decision tree classifiers, where voting improves accuracy, integrating voting did not improve k-NN accuracy in the identification of heart disease patients. Voting is an aggregation method for combining the results of many classifiers.Without voting, K-NN had the highest accuracy of 97.4%.With voting, however, the accuracy for k-NN dropped to 92.7 percent.

K. Sudhakar et al. [3] Data mining was used to research heart disease prediction.The healthcare business generates a large amount of data that is "information rich."As a result, it can't be deciphered by hand.From these datasets, data mining may be utilized to efficiently forecast diseases.Different data mining strategies are examined in the heart disease database in this paper.Decision trees, Naive Bayes, and neural networks are among the classification approaches used here.Associative classification is a new and efficient technique that combines association rule mining with classification to create a prediction model with the highest level of accuracy.Finally, this research examines and compares the performance of various classification methods on a heart disease database.

Sairabi H. Mujawar et al. [4] Using modified k-means and Naive Bayes, researchers predicted cardiac disease.Heart disease diagnosis is a difficult undertaking that necessitates a high level of expertise.Cleveland Heart Disease Database provided the data for this study.A value of '1' for the property "Disease" indicates the presence of heart illness, while a value of '0' shows the absence of heart disease.Modified k-means can be used on categorical and combinatorial data, which is what we have here.We get the two farthest clusters by using two initial centroids.Finally, a suitable number of clusters is obtained.Naive Bayes produces a model that can predict the future.This predictor specifies which class a given tuple should belong to. This predictor has a 93 percent accuracy rate in predicting heart disease and an 89 percent accuracy rate in detecting whether or not a patient has heart disease.

K Cinetha et al. [5] Using fuzzy logic, a decision assistance system for preventing coronary heart disease was proposed.This approach forecasts a patient's risk of heart disease over the following ten years.The researchers gathered data from healthy people and coronary heart disease patients to see if a healthy individual may develop coronary heart disease and what circumstances might have caused it.Fuzzy logic and decision trees are used to investigate risk factor prevention.There are 1230 cases in the dataset.For the construction of fuzzy rules and the diagnosis of coronary heart disease, a decision tree is used. The clustered data is created using this way.The fuzzy rule is then extracted from the cluster using the Least Square Error method (LSE).During testing, the optimum cluster is determined using a fuzzy approach, and variant analysis is performed.Clustering works best with smaller variation boundaries.When using the TSK inference order-1 approach, the system's greatest accuracy for specified rules is 97.67 percent

Indira S. Fall Dessai [6] Using the Probabilistic Neural Network (PNN) technique, an efficient approach for heart disease prediction was proposed.The Cleveland Heart Disease Database provided the data set, which included 13 medical attributes.Using k-means, it is clustered.Probabilistic neural networks are a type of radial basis function (RBF) network that may be used for automatic pattern recognition, nonlinear mapping, and predicting class membership probabilities and likelihood ratios.For prediction, current algorithms such as decision trees, Nave Bayes, and BNN are compared to PBN.The Receiver Operating Characteristic Convex Hull (ROCCH) approach is used to do this.The proposed approach correctly predicts 94.6 percent of the time, according to the results

Serdar AYDIN et al. [7] have investigated and analyzed various data mining algorithms for identifying heart disease.Bagging, AdaBoostM1, Random Forest, Naive Bayes, RBF Network, IBK, and NN are some of the techniques employed.The information was gathered at the Long Beach VA Hospital.It has 200 samples, each with 14 characteristics.WEKA software is used to analyze the techniques.RBF Network has an accuracy of 88.20%, making it the most accurate classification technique in the detection of heart disease, according to the findings.

G Purusothaman et al. [8] have investigated and evaluated various categorization systems for predicting heart disease.Rather than using a single model, such as a Decision Tree, an Artificial Neural Network, or Naive Bayes, the authors concentrate on how hybrid models, which incorporate many categorization techniques, function.They looked at the work of researchers who looked into the efficacy of hybrid models.Single models such as Decision tree, artificial neural network, and Naive Bayes have respective performance of 76 percent, 85 percent, and 69 percent.Hybrid techniques, on the other hand, have a 96 percent accuracy rate.As a result, hybrid models produce trustworthy and promising classifiers for accurately predicting cardiac disorders.

Deepali Chandna [9] has combined a learning algorithm with a feature selection strategy using a hybrid approach.The data was gathered from the University of California, Irvine.Only 14 of the 76 traits in the set are chosen using the k-nearest neighbor techniques.Information gain and the Adaptive Neuro-Fuzzy Inference System are also used in this method (ANFIS).The combined effect of neural networks and fuzzy inference systems is known as ANFIS.The quality of qualities is chosen using information gained.The proposed method has a 98.24 percent accuracy.

Baharami et al. [10] J48 Decision Tree, k-Nearest Neighbors(k-NN), Naive Bayes(NB), and SMO Boshra are some of the classification algorithms that have been tested (SMO is widely used for training SVM).To extract the significant features from the dataset, a feature selection technique (gain ratio evaluation technique) is applied.The classification algorithms are implemented using WEKA software.The mining techniques are tested using a 10-fold cross-validation technique.With an accuracy of 83.732 percent, J48 is the most accurate.

PROPOSED SYSTEM

We used python and pandas operations to do heart disease classification using data collected from the UCI repository after reviewing the results from previous approaches. It gives a simple visual depiction of the dataset, working environment, and predictive analytics construction. The machine learning process begins with data preprocessing, followed by feature selection based on data cleaning, classification, and evaluation of modeling performance. To improve the accuracy of the outcome, the random forest technique is applied**.**

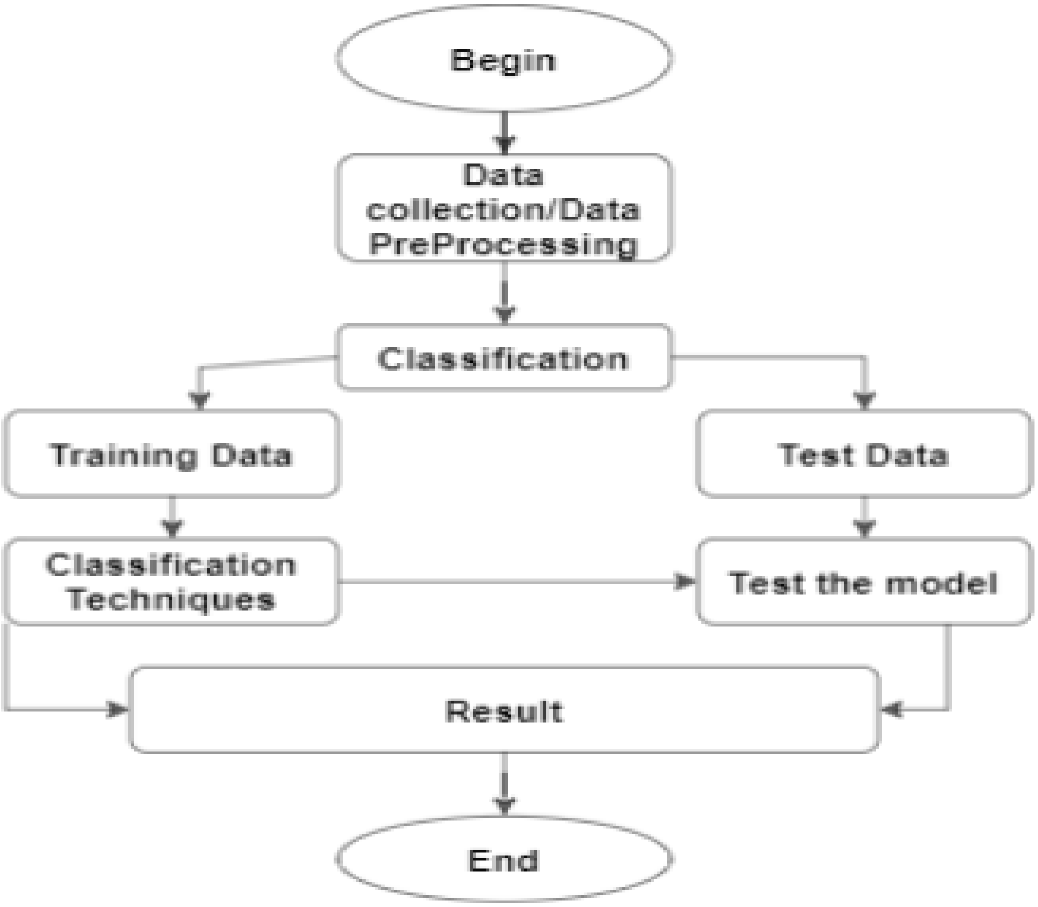


FIGURE 1: Experiment workflow with Dataset

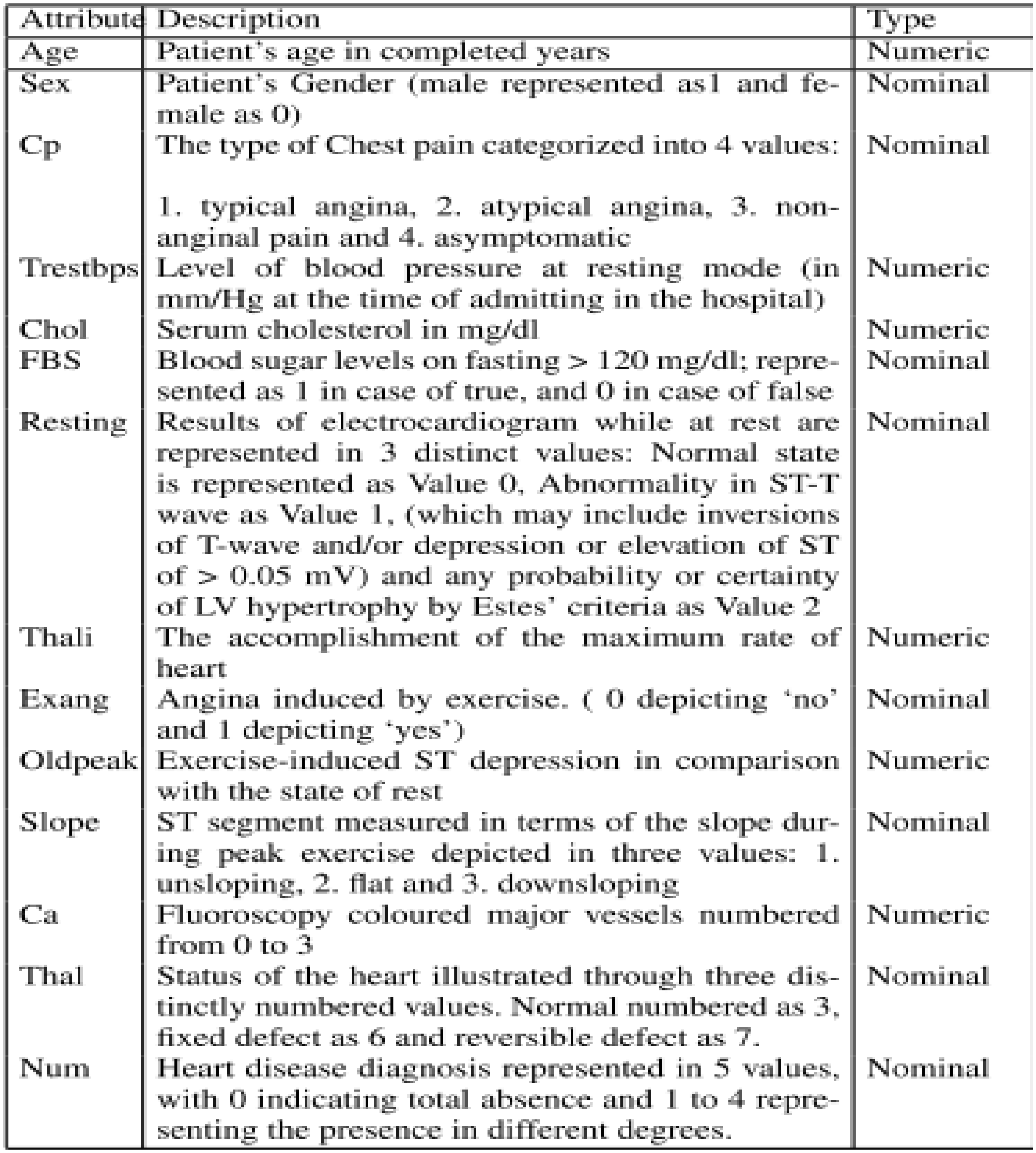
Advantages

1. Improved accuracy in the diagnosis of cardiac disease
2. Random forest technique and feature selection are used to handle the largest amount of data
3. Doctors' time complexity should be reduced
4. Patients will save money.

APPROACH

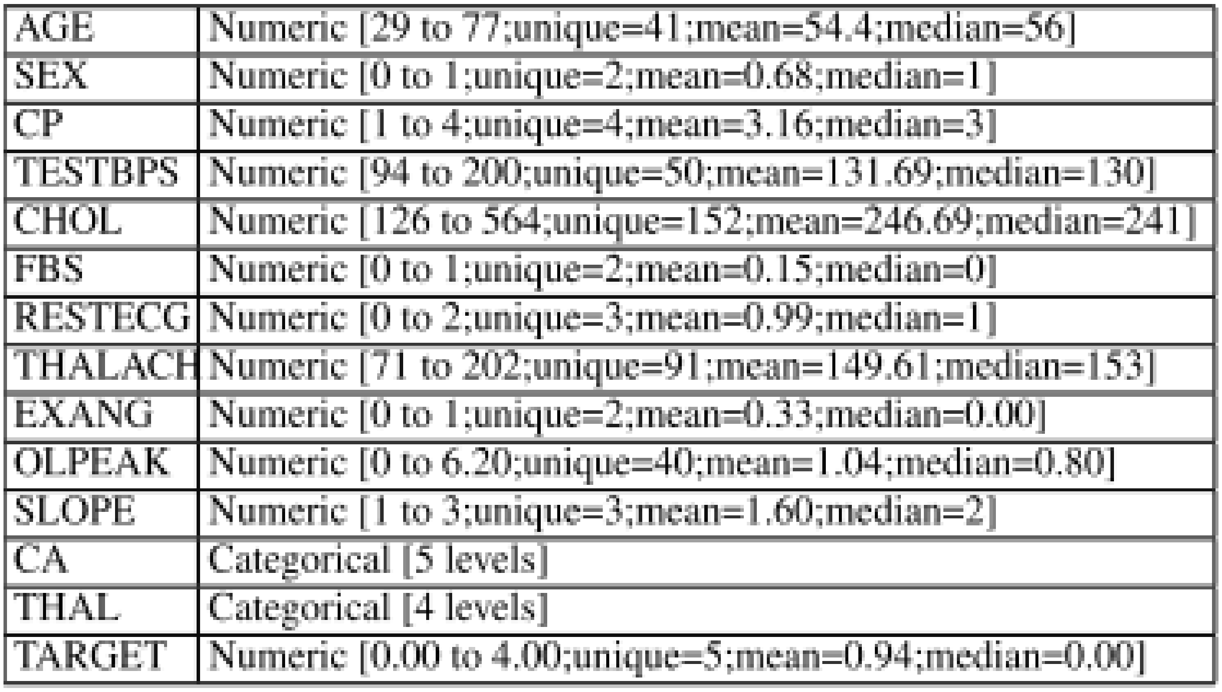
Data Pre-Processing

Dataset containing attributes sex denotes the patient's gender, age denotes the patient's age, trestbps denotes the patient's resting blood pressure, cp denotes chest pain, fbs denotes fast blood sugar, chol denotes cholesterol, thalach denotes the maximum heart rate achieved, restecg denotes the resting electroc. result (1 anomaly), oldpeak denotes the ST depression induced. ex, exang denotes exercise The slope of peak exercise is indicated by the word slope. The thalassemia is indicated by ST, pred attribute, thal**.**



Attribute	Description	Type
Age	Patient's age in completed years	Numeric
Sex	Patient's Gender (male represented as1 and female as 0)	Nominal
Cp	The type of Chest pain categorized into 4 values: 1. typical angina, 2. atypical angina, 3. non-anginal pain and 4. asymptomatic	Nominal
Trestbps	Level of blood pressure at resting mode (in mm/Hg at the time of admitting in the hospital)	Numeric
Chol	Serum cholesterol in mg/dl	Numeric
FBS	Blood sugar levels on fasting > 120 mg/dl; represented as 1 in case of true, and 0 in case of false	Nominal
Resting	Results of electrocardiogram while at rest are represented in 3 distinct values: Normal state is represented as Value 0, Abnormality in ST-T wave as Value 1, (which may include inversions of T-wave and/or depression or elevation of ST of > 0.05 mV) and any probability or certainty of LV hypertrophy by Estes' criteria as Value 2	Nominal
Thali	The accomplishment of the maximum rate of heart	Numeric
Exang	Angina induced by exercise. (0 depicting 'no' and 1 depicting 'yes')	Nominal
Oldpeak	Exercise-induced ST depression in comparison with the state of rest	Numeric
Slope	ST segment measured in terms of the slope during peak exercise depicted in three values: 1. unsloping, 2. flat and 3. downsloping	Nominal
Ca	Fluoroscopy coloured major vessels numbered from 0 to 3	Numeric
Thal	Status of the heart illustrated through three distinctly numbered values. Normal numbered as 3, fixed defect as 6 and reversible defect as 7.	Nominal
Num	Heart disease diagnosis represented in 5 values, with 0 indicating total absence and 1 to 4 representing the presence in different degrees.	Nominal

Table1. UCI Dataset attributes detailed information



AGE	Numeric [29 to 77;unique=41;mean=54.4;median=56]
SEX	Numeric [0 to 1;unique=2;mean=0.68;median=1]
CP	Numeric [1 to 4;unique=4;mean=3.16;median=3]
TESTBPS	Numeric [94 to 200;unique=50;mean=131.69;median=130]
CHOL	Numeric [126 to 564;unique=152;mean=246.69;median=241]
FBS	Numeric [0 to 1;unique=2;mean=0.15;median=0]
RESTECG	Numeric [0 to 2;unique=3;mean=0.99;median=1]
THALACH	Numeric [71 to 202;unique=91;mean=149.61;median=153]
EXANG	Numeric [0 to 1;unique=2;mean=0.33;median=0.00]
OLPEAK	Numeric [0 to 6.20;unique=40;mean=1.04;median=0.80]
SLOPE	Numeric [1 to 3;unique=3;mean=1.60;median=2]
CA	Categorical [5 levels]
THAL	Categorical [4 levels]
TARGET	Numeric [0.00 to 4.00;unique=5;mean=0.94;median=0.00]

Table 2. UCI dataset range and data types

Feature Selection and Reduction

Two qualities related to age and sex are utilized to identify the patient's personal information among the 13 attributes in the data set. The remaining qualities are significant because they provide critical clinical information. Clinical data are essential for determining the degree of cardiac disease and diagnosing it.

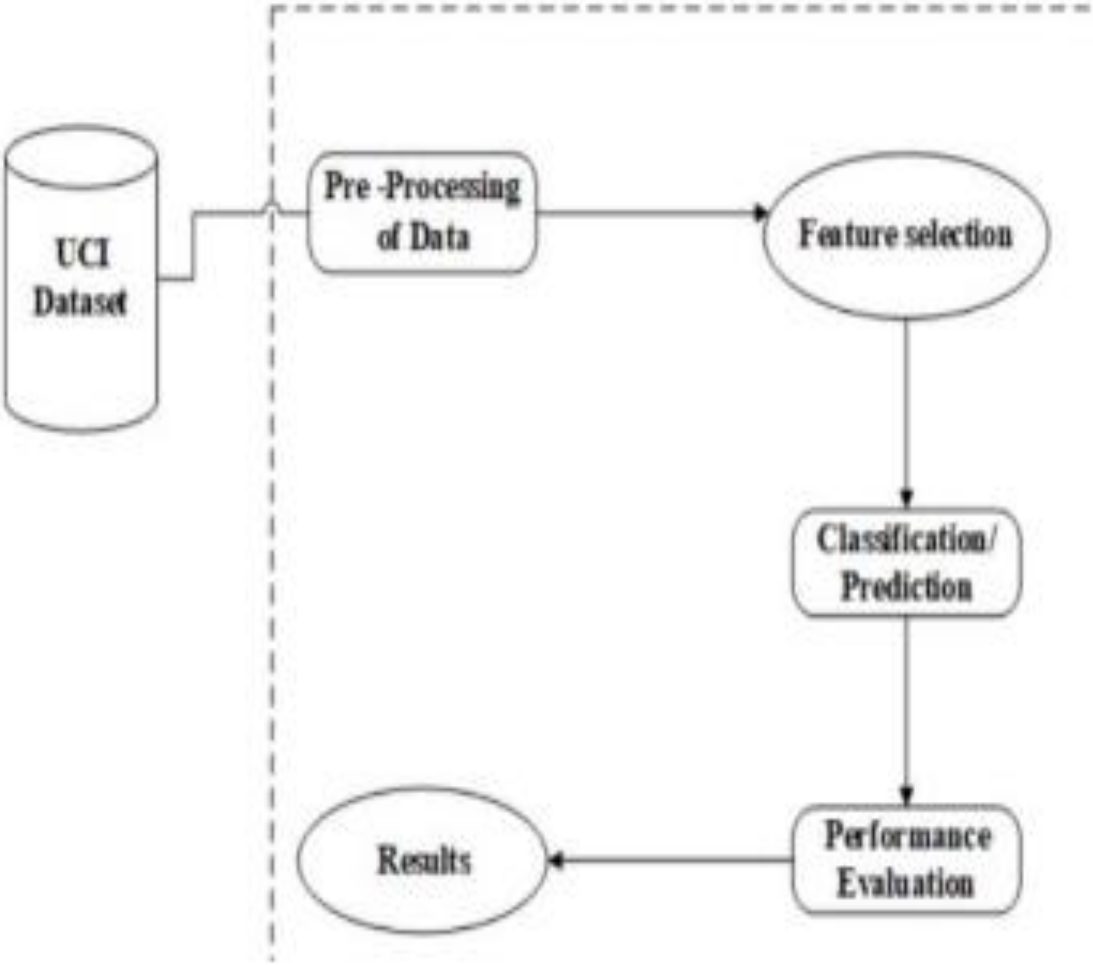


Diagram Representing overall flow of the data

MACHINE LEARNING

Classification Modeling

The attributes listed in Table 1 are fed into several machine learning methods such Random Forest, Decision Tree, Support Vector Classifiers, and K Neighbor Classifiers.The input dataset is divided into two parts: 67% is used for training, while the remaining 33% is used for testing.A training dataset is a collection of data that is used to train a model.The testing dataset is used to evaluate the trained model's performance.The performance of each method is computed and analysed using several metrics such as accuracy, precision, recall, and F-measure scores, as discussed below.

1) Random Forest classifier

For both classification and regression, Random Forest methods are used.It builds a tree out of the data and generates predictions from it.Even when substantial sets of record values are missing, the Random Forest technique can yield the same result on large datasets.The decision tree's generated samples can be kept and used on additional data in the future.There are two stages in random forest: first, generate a random forest, and then make a prediction using the random forest classifier that was created in the first stage.

To achieve the optimum result, this ensemble classifier constructs many decision trees and combines them.It primarily uses bootstrap aggregation or bagging for tree learning

2) Decision Tree Classifier

The Decision Tree algorithm is represented as a flowchart, with the inner node representing the dataset properties and the outer branches representing the result.Decision Trees were chosen because they are quick, dependable, and simple to read, and they require very little data preparation.The prediction of class label comes from the root of the tree in a Decision Tree.The root attribute's value is compared to the record's attribute.The matching branch is followed to that value and a jump to the next node is performed based on the result of the comparison.

3) Support Vector Classifier

SVM (Support Vector Machine) is a supervised machine learning technique that can be used to solve classification and regression problems.It is, however, mostly employed to solve categorization difficulties.Each data item is plotted as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm.Then we accomplish classification by locating the hyperplane that best distinguishes the two classes.
Simply put, support vectors are the coordinates of each individual observation.
The SVM classifier is a frontier that separates the two classes (hyper-plane/line) the most effectively.

4) K neighbors Classifier

The k-nearest neighbours (KNN) algorithm is a data classification that estimates the likelihood that a data point will belong to one of two groupsbased on the data points clos

The supervised machine learning algorithm k-nearest neighbour is used to address classification and regression problems. It is, however, mostly employed to solve categorization difficulties.est to it.

KNN is a non-parametric, slow learning method.

Because it doesn't perform any training when you submit the training data, it's known as a lazy learning algorithm or lazy learner.Instead, it just saves the data and does not execute any calculations throughout the training period.It doesn't start building a model until the dataset is queried

RESEARCH METHODOLOGY

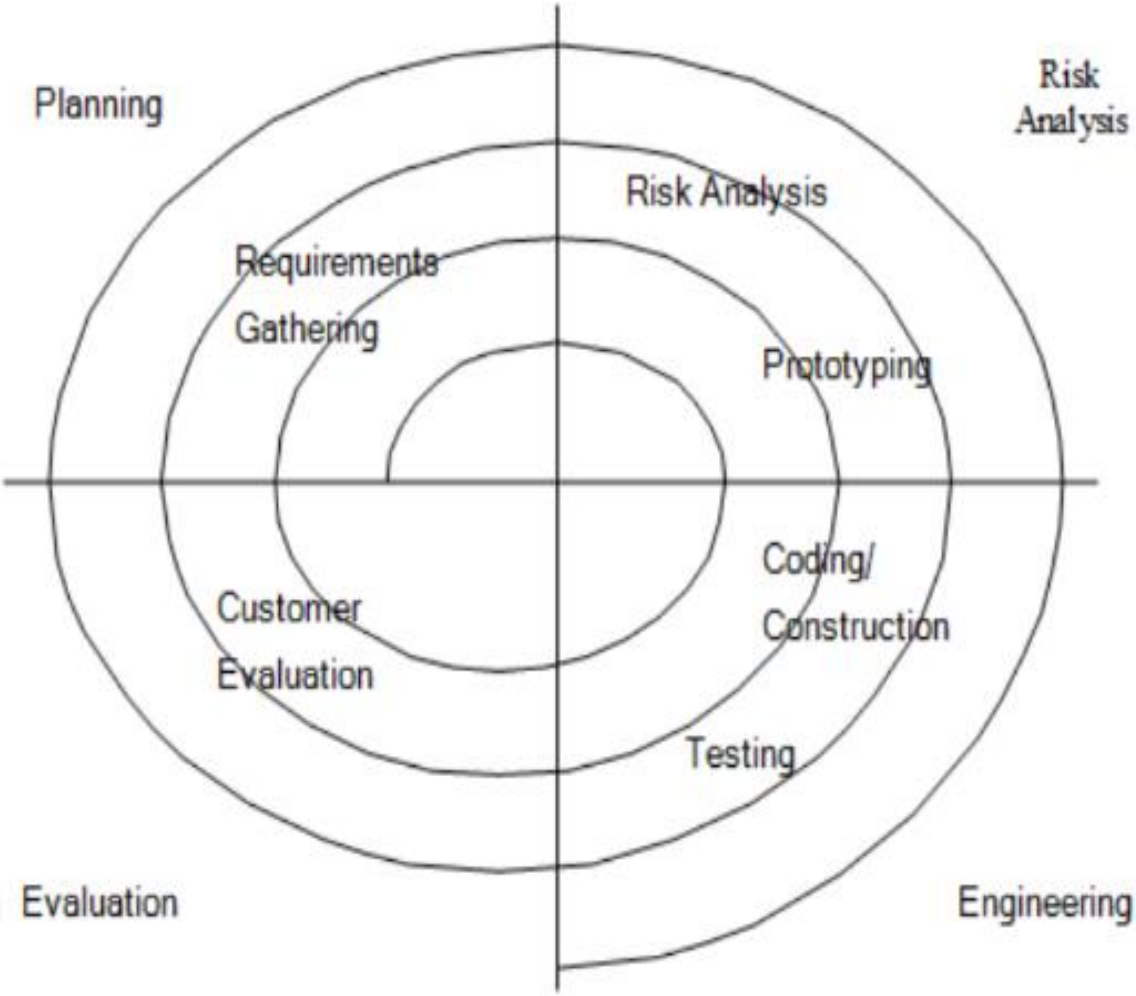
Research Design:

I'm going to use an experimental research design. It is a method of quantitative research. Essentially, it is a scientifically conducted study in which one set of variables is kept constant while another set of variables is measured as the experiment's topic. This is more practical when conducting face recognition and detection since it observes a subject's behaviours and patterns to see if the subject fits all of the details supplied and cross-checked against previous data. It is a time-bound approach of effect research that focuses on the relationship between the variables that produce actual results**.**

System Development Methodology:-

The strategy for controlling project development is known as software development methodology. Waterfall model, incremental model, RAD model, Agile model, Iterative model, and Spiral model are just a few of the methodologies accessible. However, the developer must still consider it when deciding which will be used in the project. The approach model is beneficial for efficiently managing projects and for preventing developers from encountering problems during development. It also aids in the achievement of the project's goal and scope. It is necessary to comprehend the stakeholder requirements in order to construct the project.

The proposed DM modeling can be carried out using Methodology as a framework. The approach is a set of actions that transform raw data into recognizable data patterns so that users can extract knowledge



There are four phases that involve in the spiral model:

1) Planning phase

Phase where the requirements are collected and risk is assessed. This phase where the title of the project has been discussed with the project supervisor. From that discussion, a Heart Prediction System has been proposed. The requirement and risk was assessed after doing study on the existing system and doing literature review about another existing research.

2) Risk analysis Phase

Phase where the risk and alternative solution are identified. A prototype is created at the end of this phase. If there is any risk during this phase, there will be suggestions about alternate solutions.

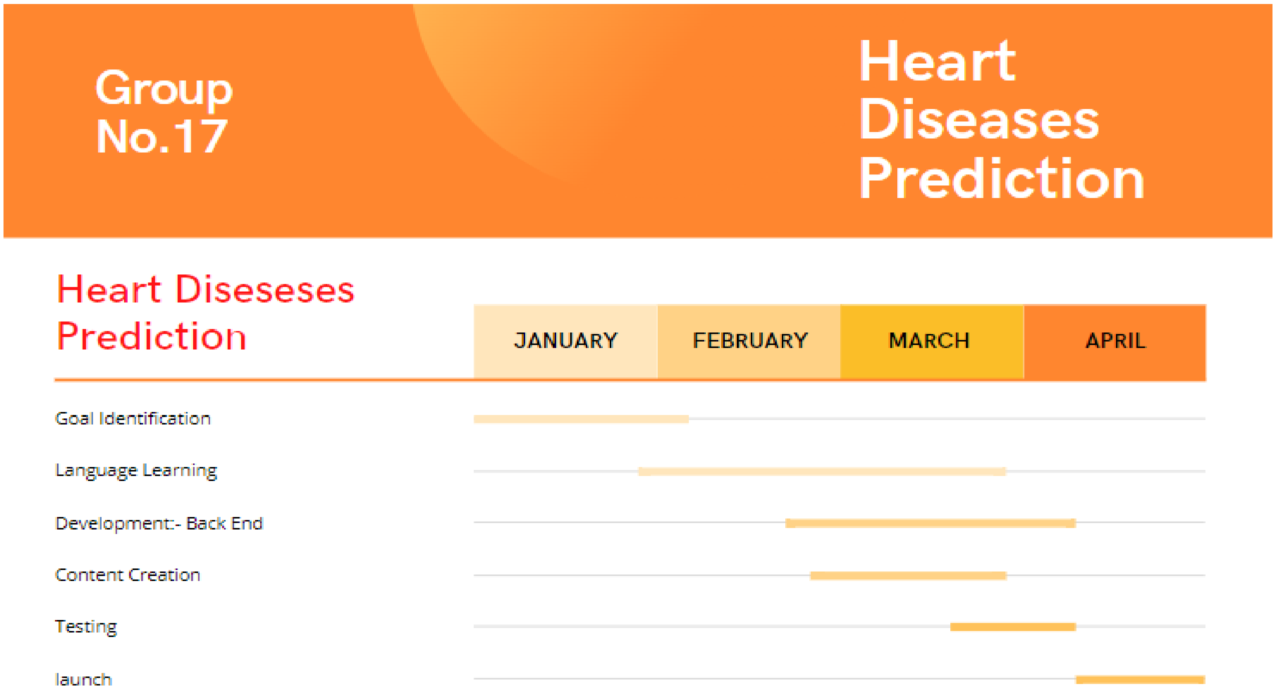
3) Engineering phase

At this phase, software is created and testing is done at the end of this phase.

4) Evaluation phase

At this phase, the user does evaluation toward the software. It will be done after the system are presented and the user will test whether the system meets with their expectation and requirement or not. If there is any error, user can tell the problem about system

GANTT CHART


Group
No.17
Heart
Diseases
Prediction
Heart Diseseses
Prediction
JANUARY
FEBRUARY
MARCH
APRIL
Goal Identification
Language Learning
Development:- Back End
Content Creation
Testing
launch

CONCLUSION

With the rising number of deaths due to heart disease, it is becoming increasingly important to build a system that can effectively and accurately forecast heart disease. The goal of the research was to find the most effective machine learning algorithm for detecting heart diseases. Using the UCI machine learning repository dataset, this study analyses the accuracy scores of Decision Tree,Support Vector Classifier, Random Forest, and K Neighbors algorithms for predicting heart disease. According to the findings of this study, the Random Forest algorithm is the most efficient algorithm for predicting heart disease, with an accuracy score of 99.00 percent. In the future, the work could be improved by creating a web application based on the Random Forest method and using a larger dataset than previously used.

REFERENCE

[1] Ashish Chhabbi,Lakhan Ahuja,Sahil Ahir, and Y. K. Sharma,19 March 2016,“Heart Disease Prediction Using Data Mining Techniques”, International Journal of Research in Advent Technology,E-ISSN:2321-9637,Special Issue National Conference “NCPC-2016”, pp. 104-106.

[2] Mai Shouman, Tim Turner, and Rob Stocker, June 2012,“Applying k-Nearest Neighbors in Diagnosing HeartDiseasePatients”,International Journal of Information and Education Technology, Vol. 2, No. 3,pp. 220-223.

[3] K.Sudhakar, and Dr. M. Manimekalai, January 2014, “Study of Heart Disease Prediction using Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 1157-1160

[4] Sairabi H. Mujawar, and P. R. Devale, October 2015,“Prediction of Heart Disease using Modified k-means and by using Naive Bayes”, International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.

[5] K Cinetha, and Dr. P. Uma Maheswari, Mar.-Apr. 2014,“Decision Support System for Precluding Coronary Heart Disease using Fuzzy Logic.”, International Journal of Computer Science Trends and Technology (IJCST), Vol. 2, Issue 2, pp. 102-107

[6] Indira S. Fal Dessai,2013,“Intelligent Heart Disease Prediction System Using Probabilistic Neural Network”, International Journal on Advanced Computer Theory and Engineering (IJACTE),Vol. 2, Issue 3, pp. 38-44.

[7] Serdar AYDIN, Meysam Ahanpanjeh,and Sogol Mohabbatiyan,February 2016, “Comparison And Evaluation of Data Mining Techniques in the Diagnosis of Heart Disease”,International Journal on Computational Science & Applications (IJCSA), Vol. 6,No.1, pp. 1-15

[8] G. Purusothaman, and P. Krishnakumari, June 2015,“A Survey of Data Mining Techniques on Risk Prediction: Heart Disease”, Indian Journal of Science and Technology, Vol. 8(12), DOI:10.17485/ijst/2015/v8i12/58385, pp. 1-5.

[9] Deepali Chandna, 2014,“Diagnosis of Heart Disease Using Data Mining Algorithm”, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5

(2), pp. 1678-1680.

[10] Boshra Bahrami, and Mirsaeid Hosseini Shirvani,February 2015,“Prediction and Diagnosis of Heart Disease by Data Mining Techniques”,Journal of Multidisciplinary Engineering Science and Technology(JMEST), ISSN:3159- 0040, Vol. 2, Issue 2, pp. 164-168.

[11] https://www.kaggle.com/datasets

[12] Github