CARDIOVASCULAR DISEASE FORECASTING USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

CIBIRAJAN V (927621BAL008)

HARIKISHORE S (927621BAL013)

SAKTHIVEL S (927621BAL043)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

M. KUMARASAMY COLLEGE OF ENGINEERING, KARUR

ANNA UNIVERSITY:: CHENNAI 600 025

NOV 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"CARDIOVASCULAR DISEASE FORECASTING USING MACHINE LEARNING"** is the bonafide work of **"CIBIRAJAN V [927621BAL008], HARIKISHORE S [927621BAL013], SAKTHIVEL S [927621BAL043]"** who carried out the project work under my supervision.

SIGNATURE

Dr. R. RAJA GURU, M.Tech.,Ph.D.,
ASSOCIATE PROFESSOR AND HEAD
Department of Artificial Intelligence,
M.Kumarasamy College of Engineering,
(Autonomous)
Thalavapalayam,
Karur-639113.

SIGNATURE

Dr.T. SARAVANAN, Ph.D.,
Assistant Professor,
Department of Artificial Intelligence,
M.Kumarasamy College of Engineering,
(Autonomous)
Thalavapalayam,
Karur-639113.

TABLE OF CONTENTS

ABSTRACT

The healthcare industry has a large data repository thus, it is necessary to use a variety of strategies to process and extract useful insights from this richness of data. One such method is data mining, which is widely used to elucidate hidden correlations and patterns in healthcare data. Heart disease, the leading cause of death worldwide, is of utmost importance in the healthcare industry. A system has been created to forecast the risk of heart disease incidence and show its findings as a percentage likelihood in order to address this important issue. The system uses datasets that are organized according to medical parameters and analyzes these datasets using data mining classification techniques. It specifically makes use of the Decision Tree Algorithm the most important machine learning techniques. The workhorses of the system, these algorithms enable it to make meaningful inferences and predictions from the input data. Python programming is the method of choice for data processing and algorithm development due to its adaptability and robust machine learning library ecosystem. The decision tree algorithm is used by the heart disease prediction system to analyze the medical parameters included in the datasets and identify patterns and associations that affect the chance of developing heart disease. The precision of the forecasts is crucial in the context of these analyses. As a result, the goal of this study is to identify which of the machine learning algorithms-Decision Tree Algorithm produces the highest accuracy in heart disease prediction. The outcomes of this comparative analysis will offer priceless information about which algorithm is better suited to tackling this important healthcare challenge, furthering our understanding of heart disease prediction and assisting in the creation of more efficient diagnostic and preventive measures.

Keywords: Healthcare, Data mining, Heart disease, Decision Tree Algorithm, Machine learning

LIST OF FIGURES

LIST OF TABLES

ACRONYMS/LIST OF ABBREVIATIONS

Acronym	Abbreviations
CART	Classification and Regression Trees
CHAID	Chi Square Automatic Interaction Detection
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheets
JS	Javascript

CHAPTER-1

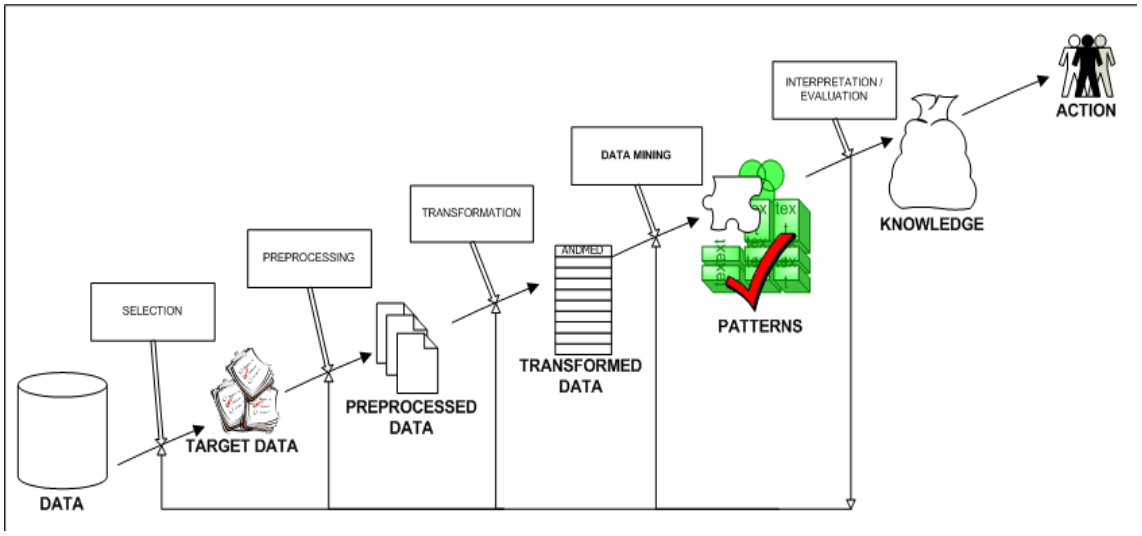
INTRODUCTION

INTRODUCTION

Data Mining

Structure of Data Mining

Generally, data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.



Data Mining Works

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyses relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. **Generally, any of four types of relationships are sought:**

- **Classes**: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

- **Clusters**: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

- **Associations**: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

- **Sequential patterns**: Data is mined to anticipate behaviour patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

1. Extract, transform, and load transaction data onto the data warehouse system.
2. Store and manage the data in a multidimensional database system.
3. Provide data access to business analysts and information technology professionals.
4. Analyse the data by application software.
5. Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure. Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

- **Decision trees**: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include **Classification and Regression Trees** (CART) and **Chi Square Automatic Interaction Detection** (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

- **Nearest neighbour method**: A technique that classifies each record in a dataset based on a combination of the classes of the *k* record(s) most similar to it in a historical dataset (where *k*=1). Sometimes called the *k*-nearest neighbour technique.

- **Rule induction**: The extraction of useful if-then rules from data based on statistical significance.

- **Data visualization**: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

Characteristics of Data Mining:

- **Large quantities of data**: The volume of data so great it has to be analysed by automated techniques e.g., satellite information, credit card transactions etc.
- **Noisy, incomplete data**: Imprecise data is the characteristic of all data collection.
- **Complex data structure**: conventional statistical analysis not possible
- Heterogeneous data stored in legacy systems

Benefits of Data Mining:

1. It's one of the most effective services that are available today. With the help of data mining, one can discover precious information about the customers and their behaviour for a specific set of products and evaluate and analyse, store, mine and load data related to them
2. An analytical CRM model and strategic business-related decisions can be made with the help of data mining as it helps in providing a complete synopsis of customers
3. An endless number of organizations have installed data mining projects and it has helped them see their own companies make an unprecedented improvement in their marketing strategies (Campaigns)
4. Data mining is generally used by organizations with a solid customer focus. For its flexible nature as far as applicability is concerned is being used vehemently in applications to foresee crucial data including industry analysis and consumer buying behaviours
5. Fast paced and prompt access to data along with economic processing techniques have made data mining one of the most suitable services that a company seek

Advantages of Data Mining:

1. Marketing / Retail:

Data mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaigns such as direct mail, online marketing campaign…etc. Through the results, marketers will have appropriate approach to sell profitable products to targeted customers.

2. Finance / Banking

Data mining gives financial institutions information about loan information and credit reporting. By building a model from historical customer's data, the bank and financial institution can determine good and bad loans. In addition, data mining helps banks detect fraudulent credit card transactions to protect credit card's owner.

3. Manufacturing

By applying data mining in operational engineering data, manufacturers can detect faulty equipment and determine optimal control parameters. For example, semi-conductor manufacturers have a challenge that even the conditions of manufacturing environments at different wafer production plants are similar, the quality of wafer are lot the same and some for unknown reasons even has defects. Data mining has been applying to determine the ranges of control parameters that lead to the production of golden wafer. Then those optimal control parameters are used to manufacture wafers with desired quality.

4. Researchers:

Data mining can assist researchers by speeding up their data analysing process; thus, allowing those more time to work on other projects.

1.1. BACKGROUND

The cardiovascular disease forecasting using machine learning is a web application using HTML, CSS, JAVASCRIPT are the main technical tools which are used to develop the web application. The main purpose of our project is the healthcare sector has a sizable data repository, so it is vital to analyses and extract insightful information from this wealth of data using a range of methodologies. Data mining is one approach that is frequently used to reveal hidden connections and trends in healthcare data. The largest cause of death in the world, heart disease, is of vital significance to the healthcare sector. In order to solve this crucial issue, a system has been developed to forecast the risk of heart disease incidence and display its findings as a % likelihood. Using data mining classification algorithms, the system examines datasets that are arranged according to medical parameters. Its uses of the most significant machine learning algorithms, the Decision Tree Algorithm, in particular. These algorithms are the backbone of the system, allowing it to draw conclusions and forecasts from the supplied data. Python programming is the preferred way for processing data and creating algorithms because of its adaptability and extensive ecosystem of machine learning libraries. The heart disease prediction system analyzes the medical parameters in the datasets and finds patterns and relationships that influence the likelihood of developing heart disease using the decision tree algorithm. In the context of these studies, the forecasts' accuracy is essential. Determining which of the machine learning algorithms—Decision Tree Algorithm produces the highest accuracy in heart disease prediction is the aim of this study. The results of this comparative analysis will provide invaluable insight into which algorithm is more capable of addressing this significant healthcare challenge, advancing our knowledge of heart disease prediction and assisting in the development of more effective diagnostic and preventive measures.

1.2. PROBLEM STATEMENT

The prevention and treatment of heart-related disorders could be revolutionized by using machine learning algorithms to predict cardiovascular illness based on an intensive examination of medical indicators and lifestyle factors. The importance of this strategy resides in its potential to significantly reduce the high number of fatalities attributable to cardiovascular illnesses, which remain a major worldwide health concern. The prevention and treatment of heart-related disorders could be revolutionized by using machine learning algorithms to predict cardiovascular illness based on an intensive examination of medical indicators and lifestyle factors. The importance of this strategy resides in its potential to significantly reduce the high number of fatalities attributable to cardiovascular illnesses, which remain a major worldwide health concern. The key to this tactic is its capacity to express prediction accuracy as a percentage, which is a key feature of machine learning models. High levels of accuracy and reliability give healthcare professionals a solid platform on which to base their decisions on the available evidence. Additionally, this technology can be seamlessly incorporated into clinical practice, supporting tailored medical approaches for patients and speeding up diagnosis and intervention. However, machine learning's influence on cardiovascular health goes beyond individual treatment. Public health professionals can more efficiently deploy resources and target interventions to target certain communities by identifying high-risk groups and geographic areas with a propensity for heart disease. This all-encompassing strategy to prevention has the potential to lessen the mortality rates and worldwide burden of heart disease. In conclusion, using machine learning algorithms to anticipate cardiovascular illness is a huge step forward for improving heart health. For better healthcare outcomes and a reduction in heart disease-related deaths, higher accuracy levels in these prediction models are essential. By saving lives, lowering healthcare costs, and eventually producing a healthier society, this ground-breaking combination of medical science and technology has the potential to revolutionize public health.

1.3. OBJECTIVES

The goal of this project is to create a reliable system for predicting heart disease using machine learning and data mining methods in the healthcare industry. The healthcare sector is flooded with enormous amounts of data, and our main goal is to efficiently utilize these data resources to forecast the chance of developing heart disease, a global health concern and the main cause of mortality. Our goal is to design a thorough and precise predictive model that can determine a person's risk of acquiring heart disease and display the results as a % likelihood. To accomplish this goal, we propose to use well-known machine learning techniques, the Decision Tree Algorithm, to organize and evaluate medical datasets containing pertinent information. Our predictive model is built on these algorithms, which allow us to find important patterns and relationships in the data. Because of its adaptability and the availability of strong machine learning packages, Python is our programming language of choice for data processing. This makes it the perfect platform for algorithm creation. In this study, we compare how well the Decision Tree Algorithm and the predict heart illness, with the goal of identifying the algorithm that predicts heart disease with the highest degree of accuracy. The accuracy of our forecasts is crucial since successful healthcare intervention and prevention methods depend on precise risk assessments. We compare these machine learning algorithms in an effort to determine which one is best for predicting heart disease, improving diagnostic and preventive procedures in the healthcare industry. The ultimate goal of this initiative is to give researchers and medical professionals a useful tool for a more precise and earlier prognosis of heart disease. We aim to advance our understanding of this crucial healthcare challenge and support the development of more effective and targeted healthcare interventions, ultimately saving lives and lowering the global burden of heart disease, by identifying the superior algorithm and evaluating its performance in predicting heart disease.

CHAPTER-2

LITERATURE REVIEW

LITERATURE REVIEW

1) Heart disease prediction using data mining techniques

AUTHORS: Rairikar and etc... (2022)

The healthcare industry collects large amounts of healthcare data, but unfortunately not all the data are mined which is required for discovering hidden patterns and effective decision making. We propose efficient genetic algorithm with the back propagation technique approach for heart disease prediction. This paper has analysed prediction systems for Heart disease using a greater number of input attributes. The System uses medical terms such as Gender, blood pressure, cholesterol like13 attributes to predict the likelihood of patient getting a heart disease.

2) Predictions in heart disease using techniques of data mining

AUTHORS: Gandhi and etc… (2015)

As huge amount of information is produced in medical associations (healing facilities, therapeutic focuses) yet this information is not properly utilized. The health care system is "data rich" however "knowledge poor ". There is an absence of successful analysis methods to find connections and patterns in health care data. Data mining methods can help as remedy in this circumstance. For this reason, different data mining techniques can be utilized. The paper intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analysed on medical data sets using algorithms.

3) HDPS: Heart disease prediction system

AUTHORS: Chen, A. H., and etc… (2011)

The diagnosis of heart disease in most cases depends on a complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels coloured, and that. Secondly, we develop

an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS). The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

4) Using Data Mining Techniques to Predict Diabetes and Heart Diseases

AUTHORS: Aldallal, A and etc… (2018)

Modernization and commercialization of life lead to an unhealthy Lifestyle that results in increasing non-communicable diseases such as heart diseases and diabetes. Non-communicable diseases have direct impact on inaction, inactivity, and idleness of people. Heart diseases and diabetes are two of the most dangerous killers affecting the society. This research aims to produce application software to be used by doctors and other medical practitioners to predict the occurrence or recurrence of non-communicable diseases (NCDs). The predictive data-mining model was applied in this project. Patients' records obtained from Bahrain Défense Force Hospital were used to examine the proposed software application. This application was executed and tested by the actual practitioner in the mentioned hospital. The results showed that the prediction system is capable of predicting NCDs' diseases effectively, efficiently and most importantly, instantly. This application is capable of helping a physician in making proper decisions towards patient health risks.

5) Analysis of data mining techniques for heart disease prediction

AUTHORS: Sultana and etc… (2016)

Heart disease is considered as one of the major causes of death throughout the world. It cannot be easily predicted by the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. This paper addresses the issue of prediction of heart disease according to input attributes on the basis of data mining techniques. We have investigated the heart disease prediction using K-Star, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a

standard data set as well as a collected data set. Based on performance factor SMO and Bayes Net techniques show optimum performances than the performances of K-Star, Multilayer Perceptron and J48 techniques.

6) A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)

AUTHORS: T. Mythili and etc… (2013)

The paper "A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)" by T. Mythili, Dev Mukherji, Nikita Padaila, and Abhiram Naidu, published in the International Journal of Computer Applications in April 2013, presents a predictive model for heart disease. The authors utilize a combination of machine learning techniques, including Support Vector Machines (SVM), Decision Trees, and Logistic Regression, referred to as SDL.Their approach aims to enhance the accuracy of heart disease prediction by integrating these diverse algorithms. This approach utilizes various patient data, such as medical history and diagnostic parameters, to generate predictions. SVM is known for its strong classification capabilities, while Decision Trees help identify crucial decision points in the prediction process. Logistic Regression contributes by modeling the probability of heart disease occurrence. The paper highlights the importance of leveraging data mining and machine learning in the healthcare sector, given the abundance of medical data. This model showcases the potential for combining multiple algorithms to improve predictive accuracy, which can have a significant impact on early diagnosis and preventive measures for heart disease. It signifies the continuous evolution of medical data analysis techniques to enhance healthcare decision-making.

7) Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques

AUTHORS: Mr. Chala Beyene and etc… (2018)

The paper titled "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques" was authored by Mr. Chala Beyene and Prof. Pooja Kamat. It was published in the "International Journal of Pure and Applied Mathematics" in the year 2018. The paper likely explores the use of data mining techniques for predicting and analysing the occurrence of heart disease. Data mining is a process of uncovering valuable insights and patterns from large datasets, and in the context of healthcare, it can help in making predictions related to heart disease. The authors might have reviewed and discussed various data mining

techniques, such as Naive Bayes, Neural Networks, and Decision Tree Algorithms, and their applications to medical datasets. This research is significant as it addresses the importance of leveraging data-driven methods to improve our understanding of heart disease and potentially enhance diagnostic and preventive measures in healthcare.

8) Heart disease Prediction using Machine Learning and Data mining Technique

AUTHORS: Jayami Patel and etc… (2017)

The paper titled "Heart Disease Prediction using Machine Learning and Data Mining Techniques," authored by Jayami Patel, Prof. Tejal Upadhyay, and Dr. Samir Patel, was published in March 2017. The authors aimed to investigate and develop predictive models to assist in identifying individuals at risk of heart disease based on medical data. Machine learning algorithms, such as decision trees, neural networks, and support vector machines, might have been employed to analyse medical datasets and make predictions regarding heart disease incidence. Data mining techniques would have been used to extract valuable insights and patterns from the healthcare data. This research is significant as it could contribute to more effective diagnostic and preventive measures for heart disease, a leading cause of global health concerns. The findings from this study might offer insights into the accuracy and performance of different algorithms in heart disease prediction, thereby improving our understanding of this critical healthcare issue.

9) Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique

AUTHORS: Mamatha Alex P and etc… (2019)

The paper titled "Prediction and Diagnosis of Heart Disease Patients using Data Mining Technique" was authored by Mamatha Alex P and Shaicy P Shaji and presented at the International Conference on Communication and Signal Processing in 2019. The research in this paper focuses on the application of data mining techniques to predict and diagnose heart disease in patients. Data mining is a method used to uncover hidden patterns and relationships within large datasets, and in this context, it's applied to medical data for heart disease analysis. The study aims to provide insights and solutions for improving heart disease prediction and diagnosis. By using data mining, the authors likely explored various algorithms and methods

such as decision trees, neural networks, or regression models to analyse medical data. Their findings and methodologies can contribute to more accurate and efficient diagnostic tools in the field of cardiology, ultimately benefiting patient healthcare and outcomes. The International Conference on Communication and Signal Processing is the forum where this research was presented and shared with the academic and scientific community.

10) Cardiovascular disease detection using a new ensemble

AUTHORS: Hamidreza Ashrafi Esfahani and etc.… (2017)

The paper titled "Cardiovascular Disease Detection using a New Ensemble" was authored by Hamidreza Ashrafi Esfahani and Morteza Ghazanfari and was published on December 22, 2017, in the IEEE (Institute of Electrical and Electronics Engineers) domain. This research paper focuses on the detection of cardiovascular diseases, a significant health concern. The authors propose a novel ensemble-based approach to improve the accuracy and reliability of disease prediction. Ensembles in machine learning combine multiple models to enhance predictive performance. The paper likely presents a detailed investigation into the methodology and techniques used to develop this ensemble model. It may discuss the dataset, feature selection, and the specific algorithms employed to predict cardiovascular disease. This research contributes to the field of healthcare analytics and provides a potential tool for more effective and early detection of cardiovascular diseases, which can be critical for timely medical interventions and patient care.

CHAPTER-3

FEASIBILITY STUDY

FEASIBILITY STUDY

A feasibility study includes such vital information and data as the funding needed to complete the project, the market opportunity, government regulations, risk factors, strength and weaknesses, the management team and the financials of the project.

3.1 Technical Feasibility: -

- **Hardware Requirements:** Assess the availability and compatibility of the hardware required for data processing and running machine learning algorithms in the healthcare facility or research environment. Ensure that the system's hardware requirements are within reach, including computational resources and storage capacity.
- **Software Requirements:** Evaluate the availability of the necessary software, including Python programming environment, machine learning libraries, and data mining tools. Ensure that the software stack can be readily implemented.

3.2 Data Feasibility: -

- **Data Availability:** Verify the availability of comprehensive and high-quality medical datasets required for the heart disease prediction system. Ensure that the datasets include relevant medical parameters.
- **Data Quality:** Assess the quality and reliability of the datasets, checking for data completeness, accuracy, and consistency. Address any data quality issues if identified.

3.3 Economic Feasibility: -

- **Budgeting:** Develop a budget for the project, including costs for hardware, software, data acquisition, personnel, and ongoing maintenance. Ensure that the project is financially viable and that the necessary funding can be secured.
- **Cost-Benefit Analysis:** Evaluate the potential benefits of implementing the heart disease prediction system, including reduced healthcare costs, improved patient outcomes, and research advancements.

3.4 Operational Feasibility: -

- **Resource Availability:** Determine if the necessary human resources, including data scientists, medical experts, and IT professionals, are available or can be readily sourced for the project.
- **Integration:** Assess the feasibility of integrating the system into the existing healthcare infrastructure, ensuring minimal disruption to existing operations.

3.5 Legal and Ethical Feasibility: -

- **Data Privacy and Compliance**: Ensure that the system complies with data privacy regulations (e.g., HIPAA) and ethical guidelines in healthcare. Address any legal and ethical concerns related to patient data usage.
- **Regulatory Approval:** Determine if regulatory approvals are required for the system's deployment and ensure that the necessary steps are taken to obtain them.

3.6 Schedule Feasibility: -

- **Timeline:** Develop a realistic project timeline that includes milestones, testing phases, and deployment schedules. Ensure that the project can be completed within the desired timeframe.

3.7 Risk Assessment: -

- **Identify Risks:** Identify potential risks and challenges, such as data security breaches, technical issues, or regulatory hurdles. Develop risk mitigation strategies to address these challenges.

CHAPTER-4

PROJECT METHODOLOGY

4.1DESCRIPTION OF THE WORKING FLOW OF PROPOSAL SYSTEM

SYSTEM DESIGN

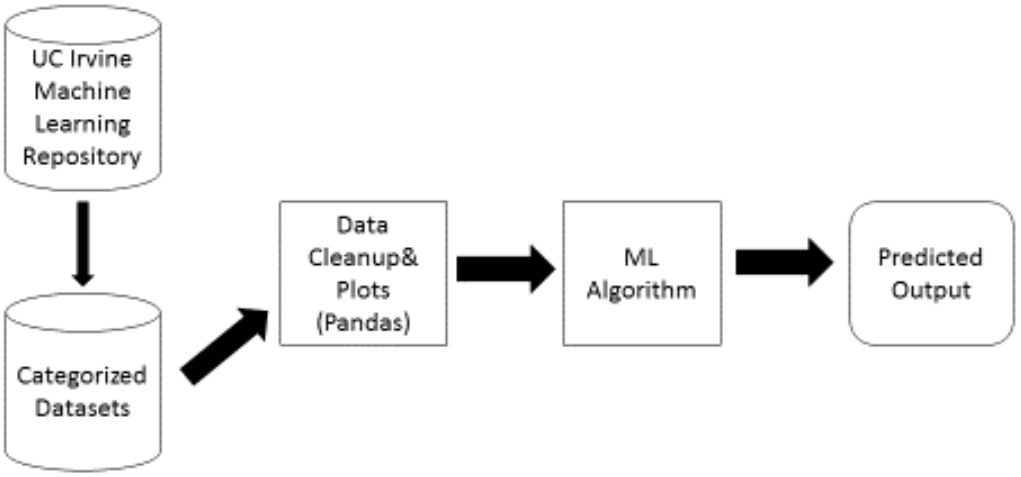


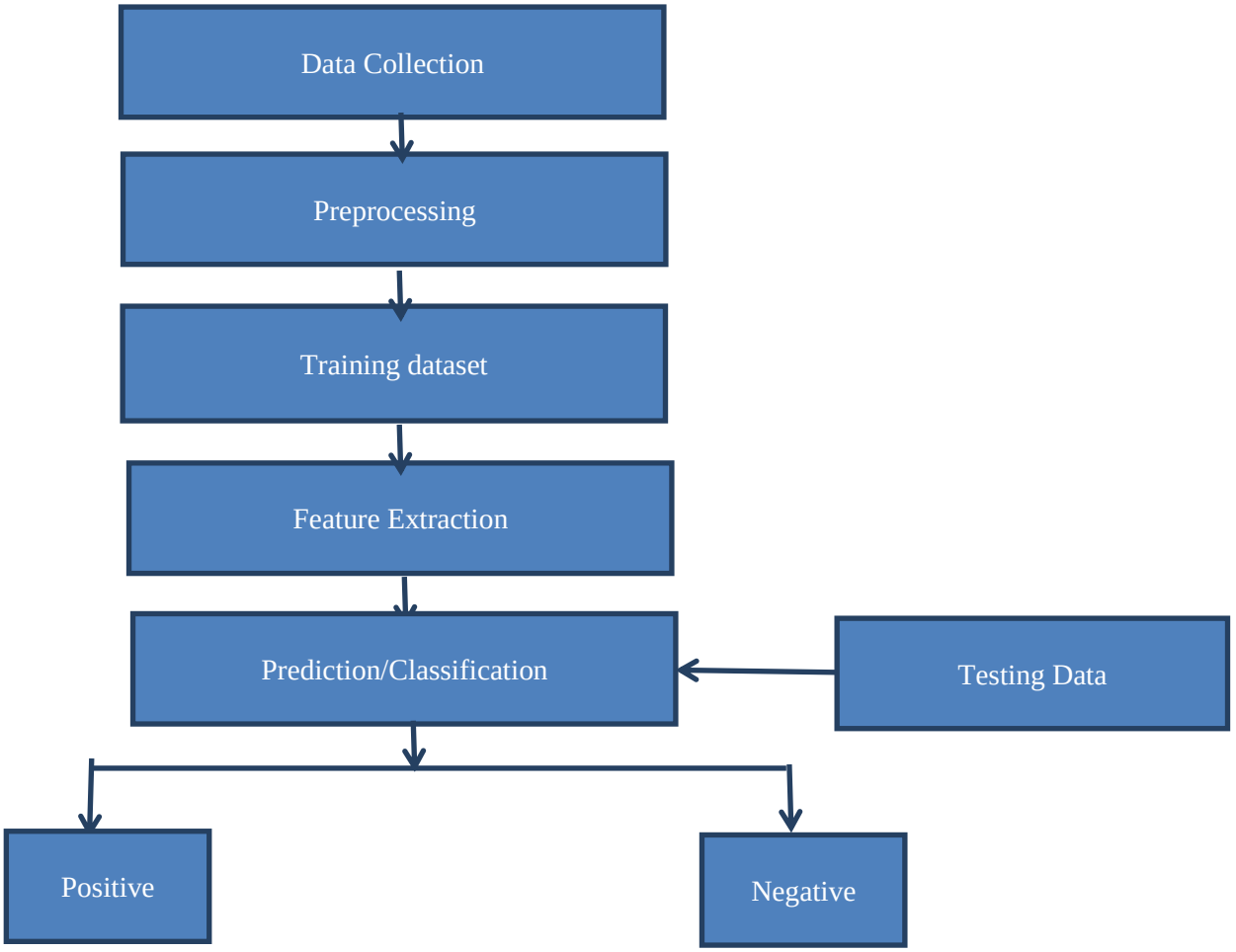
Figure 4.1: Proposed Diagram

PROJECT METHODOLOGY

The goal of this project is to develop a trustworthy system for the healthcare sector that uses machine learning and data mining techniques to forecast cardiac disease. Our primary objective is to effectively use these data resources to anticipate the likelihood of acquiring heart disease, a worldwide health problem and the leading cause of mortality. The healthcare sector is inundated with massive volumes of data. Our objective is to create a complete and accurate prediction model that can calculate an individual's risk of developing heart disease and show the findings as a percentage likelihood. We suggest using the Decision Tree Algorithm, well-known machine learning methods, to arrange and assess relevant medical datasets in order to achieve this aim. These algorithms serve as the foundation for our prediction model and enable us to identify significant trends and links in the data. Python is our preferred programming language for data processing due to its versatility and the availability of powerful machine learning tools. It is therefore the ideal environment for the development of algorithms. The objective of this study is to determine which algorithm is most accurate in predicting heart disease by contrasting how well the Decision Tree Algorithm and the predict heart disease. Our estimates must be exact since effective healthcare intervention and preventive strategies rely on accurate risk assessments. To improve diagnostic and preventative practices in the healthcare sector, we evaluate these machine learning algorithms to see which one is more effective in predicting heart disease. The end result of this endeavor aims to provide scientists and medical experts with a helpful tool for a more accurate and earlier prognosis of cardiac disease. By determining the best algorithm and assessing its ability to predict heart disease, we hope to improve our understanding of this important healthcare challenge and support the development of more effective and targeted healthcare interventions, ultimately saving lives and reducing the global burden of heart disease.

4.2 DATA FLOW DIAGRAM: -

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.



UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

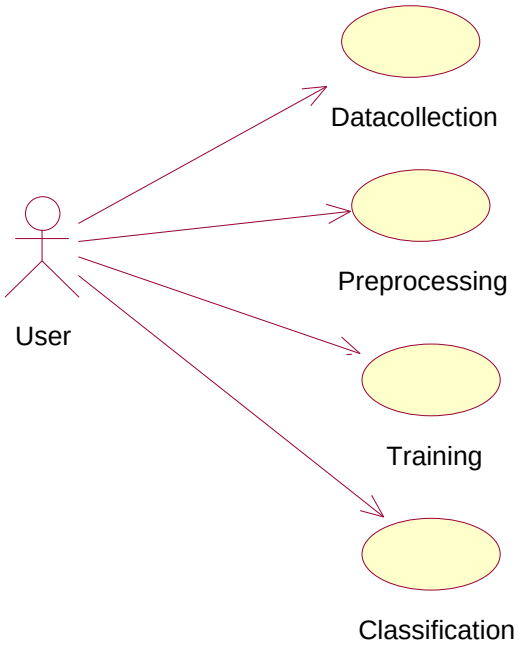
GOALS:

The Primary goals in the design of the UML are as follows:

1. Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extendibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of OO tools market.
6. Support higher level development concepts such as collaborations, frameworks, patterns and components.
7. Integrate best practices

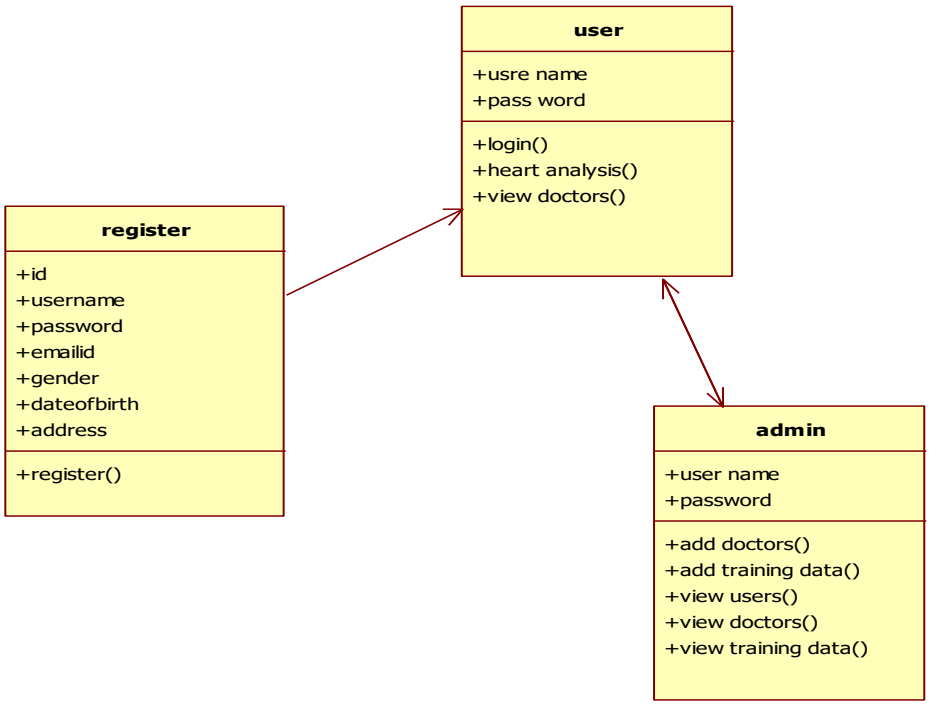
4.3 USE CASE DIAGRAM:

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



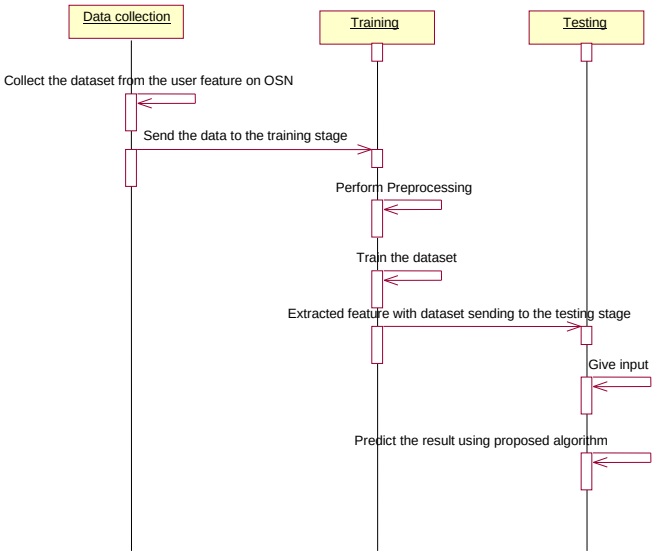
4.4 CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



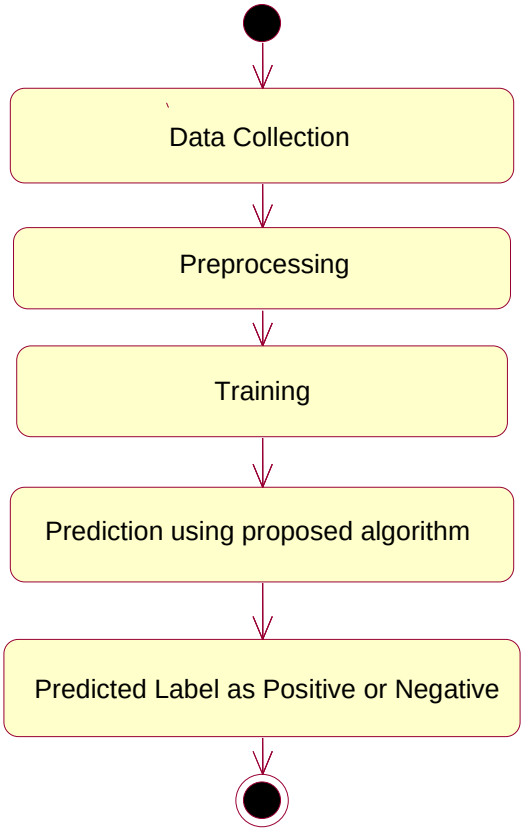
4.5 SEQUENCE DIAGRAM:

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



4.6 ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



CHAPTER-5

RESULTS AND DISCUSSION

RESULTS: -

In this project, the likelihood that a patient will have heart disease was predicted by analyzing a dataset using a supervised data mining algorithm. The Decision Tree Classification was the main classification model applied in this study. To find the most accurate algorithm for heart disease prediction, these algorithms were methodically applied to the same dataset. The study's findings showed that the Decision Tree Classification model had an amazing degree of accuracy, successfully predicting heart disease patients with 91% accuracy. This result emphasizes how well the Decision Tree Classification method works and how appropriate it is for handling medical datasets, particularly heart disease-related ones. Given these results, it is safe to say that the Decision Tree Classification algorithm is the most effective option for predictive modeling related to heart disease. This finding underscores the algorithm's efficacy and usability in healthcare applications, in addition to highlighting its potential. In the future, the system that has been built using the machine learning classification algorithm has great potential for the diagnosis and prognosis of additional diseases. Subsequent investigations may delve into the elaboration and refinement of this framework, conceivably integrating supplementary machine learning methods to expand its range and relevance. This growth may lead to a more thorough and automated approach to the analysis of cardiac disease, expediting the diagnosis procedure and opening the door to better patient care. In conclusion, the project's remarkable 91% accuracy in predicting patients with heart disease highlights the usefulness and efficacy of the Decision Tree Classification algorithm in healthcare data analysis. Given its potential for use in the diagnosis of a number of illnesses, it appears that this research has set the stage for further developments in the fields of predictive modeling and medical data analysis.

The aim of this project is to know whether the patient has heart disease or not. The records in the datasets are divided into training set and test sets. After preprocessing the data, data mining classification technique namely decision tree and naïve Bayes were applied. This section shows the results of those classification model done using Python Programming. The results are generated for both training datasets and test data sets

. Figure 5.1: The Output shows Decision tree

Train set which classifies the 70% instances of dataset and displays Possibilities of having heart disease where Red denotes patients who not having heart disease (NO) Green denotes patients who\had heart disease (YES)

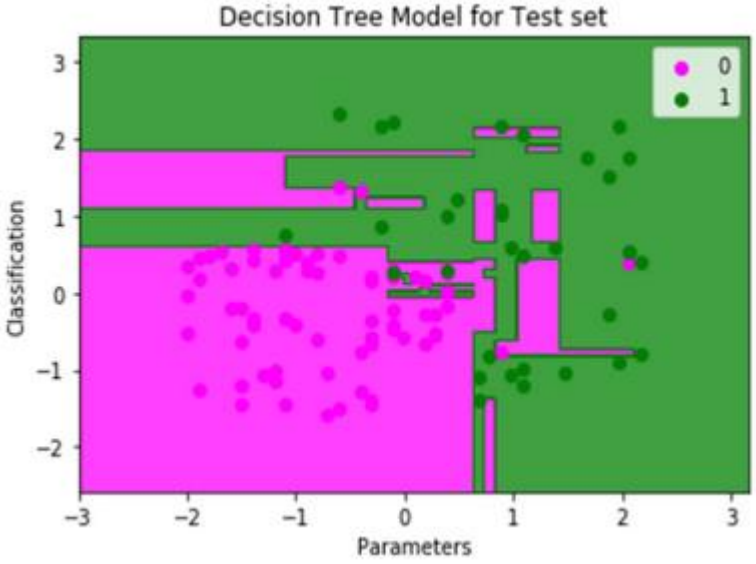


Figure 5.2: The Output shows Decision tree

Test set which classifies the 30% instances of dataset and displays Possibilities of having heart disease where magenta denotes patients who not having heart disease (NO), and Green denotes patients who had heart disease (YES)

5.3 LOGIN PROCESS

Once the user has registered, they can directly login into the web application by entering their email id and their password. Does the email and the password have verified the page will redirect to the next page.

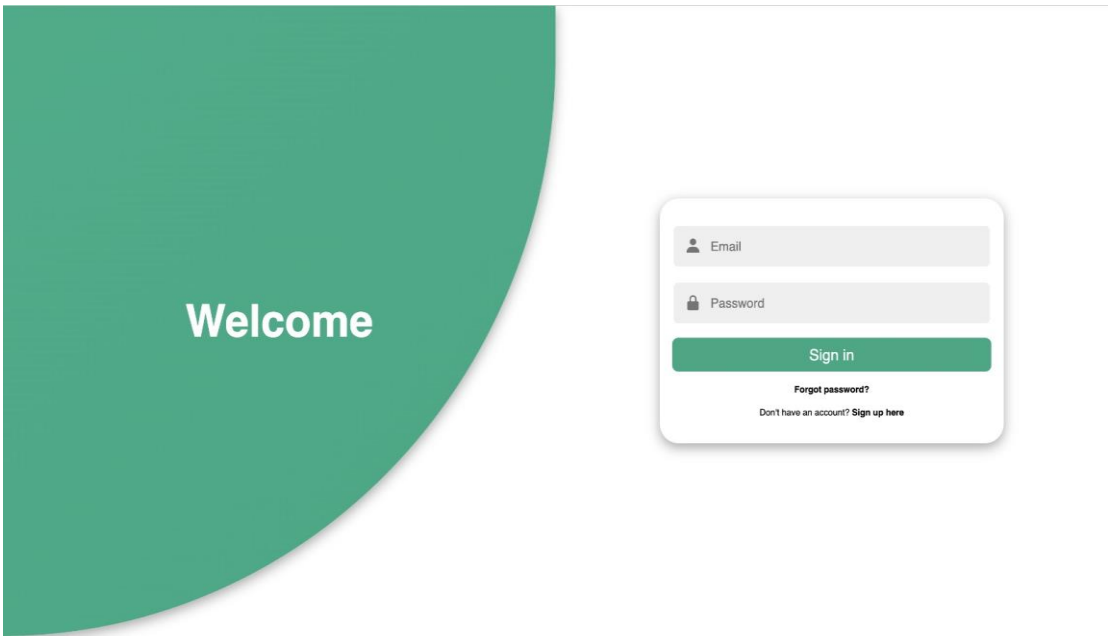


Figure 5.3: Login Process

5.4 EXECUTION PAGE

The user can make execution by entering patient report into the web application. As a result, the this is identified which of the machine learning algorithms-Decision Tree Algorithm produces the highest accuracy in heart disease prediction.

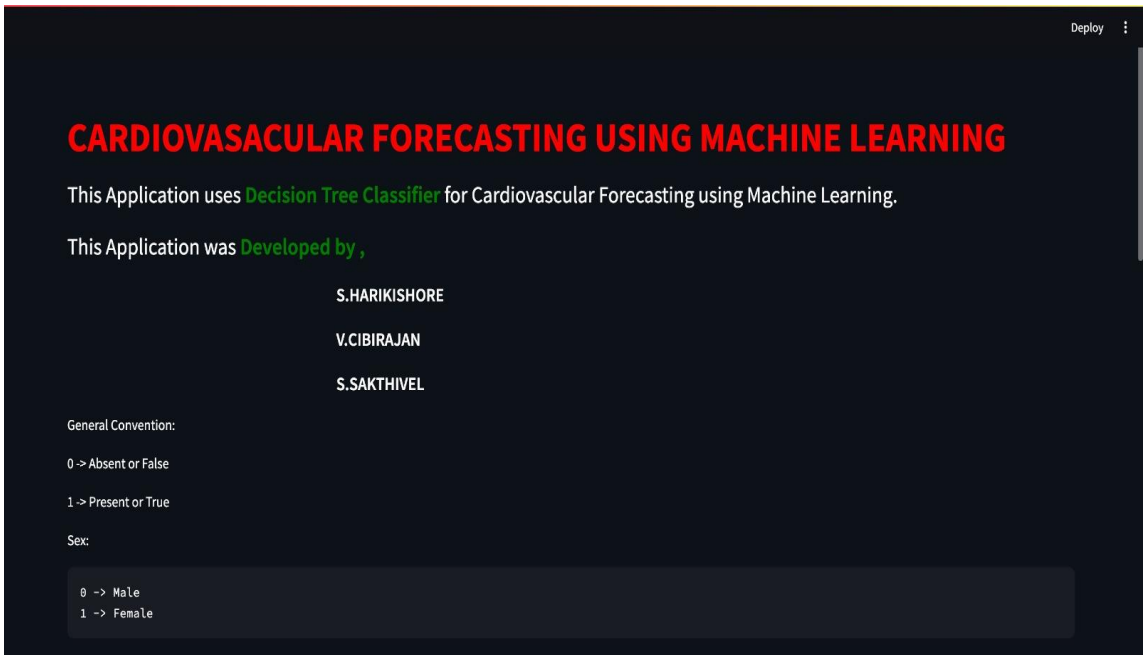


Figure 5.4: Execution Process

CHAPTER-6

CONCLUSION

CONCLUSION

Worldwide, heart disease is one of the main causes of death. Improving patient outcomes requires early detection and treatment. To help clinicians identify patients at high risk and deliver prompt therapies, machine learning algorithms can be utilized to construct predictive models for heart disease. In order to determine the likelihood that a patient has heart disease, a decision tree classification method was used in this study on a dataset of individuals with and without heart disease. The algorithm's 91% accuracy shows that it is a viable tool for predicting heart disease. The machine learning approach known as decision trees can be applied to both classification and regression applications. Recursively dividing the data into smaller and smaller subsets based on the features' values is how they operate. The algorithm chooses the attribute at each node of the tree that divides the data most effectively into two groups, one with a high likelihood of having heart disease and the other with a low probability. The procedure is repeated until a somewhat uniform collection of data points is present at each leaf node of the tree.

Decision trees provide a number of benefits for predicting heart disease. First, physicians and other healthcare professionals can grasp them since they are reasonably simple to interpret. Also, they are strong. The results of this study point to decision tree categorization as a potentially useful method for predicting heart disease. When applied to the dataset utilized in this investigation, the method had a high accuracy of 91%. The algorithm is also rather simple to understand and resilient to data noise and outliers.

Future Work:

The system that was created and the machine learning classification method that was utilized can be applied to forecast or identify different diseases. For the automation of heart disease analysis, incorporating some other machine learning techniques, the work can be enhanced or extended.

CHAPTER-7

REFERENCES

REFERENCES

[1]. Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.

[2]. Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.

[3]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[4]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150- 154). IEEE.

[5]. Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "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, 2016.

[6]. Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[7]. Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[8]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer Science & Information Technology Journal (2014): 53-59.

[9]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In 2015 2nd International Conference on Computing for Sustainable Global De

[10] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159.

[11] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.

[12] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2005.

[13] T.Mythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, vol. 68, 16 April 2013. International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 18, September 2018 25

[14] Nimai Chand Das Adhikari, Arpana Alka, and rajat Garg, "HPPS: Heart Problem Prediction System using Machine Learning".

[15] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.

[16] Marjia Sultana, Afrin Haider, "Heart Disease Prediction using WEKA tool and 10-Fold cross-validation", The Institute of Electrical and Electronics Engineers, March 2017.

[17] Dr.S.Seema Shedole, Kumari Deepika, "Predictive analytics to prevent and control chronic disease", https://www.researchgate.net/punlication/316530782, January 2016.

[18] Ashok kumar Dwivedi, "Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation", Springer, 17 September 2016.

[19] Megha Shahi, R. Kaur Gurm, "Heart Disease Prediction System using Data Mining Techniques", Orient J. Computer Science Technology, vol.6 2017, pp.457-466.

[20] Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.

[21] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", International Journal of Computer Science and Engineering, May 2018.

[22] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, "Heart disease Prediction using Machine Learning and Data mining Technique", March 2017.

[23] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, "Efficient Heart Disease Prediction System", 2016, pp.962-969.

[24] K.Gomathi, Dr.D.Shanmuga Priyaa, "Multi Disease Prediction using Data Mining Techniques", International Journal of System and Software Engineering, December 2016, pp.12-14.

[25] Mr.P.Sai Chandrasekhar Reddy, Mr.Puneet Palagi, S.Jaya, "Heart Disease Prediction using ANN Algorithm in Data Mining", International Journal of Computer Science and Mobile Computing, April 2017, pp.168-172.

[26] Ashwini Shetty A, Chandra Naik, "Different Data Mining Approaches for Predicting Heart Disease", International Journal of Innovative in Science Engineering and Technology, Vol.5, May 2016, pp.277-281.

[27] Jaymin Patel, Prof. Tejal Upadhyay, Dr.Samir Patel, "Heart Disease Prediction using Machine Learning and Data Mining Technique", International Journal of Computer Science and Communication, September 2015-March 2016, pp.129-137.

[28] Boshra Brahmi, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journals of Multidisciplinary Engineering Science and Technology, vol.2, 2 February 2015, pp.164-168.

[29] Noura Ajam, "Heart Disease Diagnoses using Artificial Neural Network", The International Institute of Science, Technology and Education, vol.5, No.4, 2015, pp.7-11.

[30] Prajakta Ghadge, Vrushali Girme, Kajal Kokane, Prajakta Deshmukh, "Intelligent Heart Disease Prediction System using Big Data", International Journal of Recent Research in Mathematics Computer Science and Information Technology, vol.2, October 2015 - March 2016, pp.73-77.

[31] Prbhavati, D.M.Chitra, "Analysis and Prediction of Various Heart Diseases using DNFS Techniques", International Journal of Innovations in Scientific and Engineering Research, vol.2, 1, January 2016, pp.1-7.

[32] Sairabi H.Mujawar, P.R.Devale, "Prediction of Heart Disease using Modified K-means and by using Naïve Bayes", International Journal of Innovative research in Computer and Communication Engineering, vol.3, October 2015, pp.10265-10273.

[33] Sharan Monica.L, Sathees Kumar, "Analysis of Cardiovascular Disease Prediction using Data Mining Techniques", International Journal of Modern Computer Science, vol.4, 1 February 2016, pp.55-58.

[34] Sharma Purushottam, Dr. Kanak Saxena, Richa Sharma, "Heart Disease Prediction System Evaluation using C4.5 Rules and Partial Tree", Springer, Computational Intelligence in Data Mining, vol.2, 2015, pp.285-294.

[35] K. Polaraju, D. Durga Prasad, "Prediction of Heart Disease using Multiple Linear Regression Model", International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017.