VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELAGAVI -590 014

A
Final Project Report
on

"Machine Learning Based Heart Disease Prediction System"

Submitted for partial fulfillment of requirement for the award of Degree of

Bachelor of Engineering

in

Computer Science & Engineering

Submitted by

KEERTHANA B G	**[1RL21CS045]**
KEERTHI.C. R	**[1RL21CS046]**
PALLAVI. R	**[1RL21CS081]**

of

VIII Semester

Under the guidance of

Mrs. B C HemaPriya

Associate Professor

Department of Computer Science & Engineering

R. L. JALAPPA INSTITUTE OF TECHNOLOGY

Doddaballapur, Bangalore Rural Dist-561 203

2024-25

R L JALAPPA INSTITUTE OF TECHNOLOGY

DODDABALLAPUR -561203 (KARNATAKA)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is Certified that the Project work entitled "**Machine Learning Based Heart Disease Prediction System" carried out by Keerthana B G, USN: [1RL21CS045], Keerthi C R, USN: [1RL21CS046], Pallavi R, USN: [1RL21CS081]** are bonafide students of R.L. Jalapa Institute of Technology, in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visveswaraya Technological University, Belagavi during the year 2024-25. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the department library. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the said degree.

…….	…….	…….
Mrs. B C Hemapriya	**Dr. Sunilkumar R M**	**Dr. P Vijaykarthik**
Assistant Professor	**Head of Department**	**Principal**
Dept of CS&E, RLJIT	**Dept of CS&E, RLJIT**	

	Name	**Signature with date**
Examiner 1:	_____	_____
Examiner 2:	_____	_____

ACKNOWLEDGEMENT

With Great pride we would like to convey our gratitude and appreciation to our alma-mater **"R. L Jalapa Institute of Technology"** for giving us the required platform for the fulfilment of the final project on **Machine Learning Based Heart Disease Prediction System** as per the V.T.U requirements, for the Eight semester subject.

We express our sincere thanks to **Dr. P Vijaykarthik**, Principal of **R. L. Jalapa Institute of Technology,** Doddaballapura for providing me with excellent infrastructure to complete the final project.

We express wholehearted gratitude to **Dr. Sunilkumar R M,** who is the respectable **HOD of Computer Science Department**. We wish to acknowledge his help in making my task easy by providing me with his valuable help and encouragement.

It is our pleasure to thank our guide **Mrs. B C HemaPriya, Assistant Professor**, **Department of Computer Science and Engineering,** RLJIT for their guidance, encouragement and valuable suggestion from the beginning of the final project work till the completion without which this final project work would not have been accomplished. We are greatly indebted to them.

And last but not the least, we would be very pleased to express our heart full thanks to the Teaching and Non-teaching staff of the Department of Computer Science and Engineering, RLJIT for their motivation and support.

We also thank all those who extended their support and co-operation while bringing out this final project.

Keerthana B G(1RL21CS045)

Keerthi C R(1RL21CS046)

Pallavi R(1RL21CS081)

ABSTRACT

Heart disease is one of the leading causes of the death worldwide. In today's common modern life, deaths due to the heart disease had become one of major issues, that roughly one person lost his or her life per minute due to heart illness. Predicting the occurrence of disease at early stages is a major challenge nowadays. Machine learning when implemented in health care is capable of early and accurate detection of disease. In this work, the arising situations of heart disease illness are calculated. Datasets used have attributes of medical parameters. The datasets are been processed in python using ML Algorithm i.e., Decision Tree Classifier. This technique uses the past old patient records for getting prediction of new one at early stages preventing the loss of lives. In this work, reliable heart disease prediction system is implemented using strong Machine Learning algorithm which is the Random Forest algorithm. Which read patient record data set in the form of CSV file. After accessing dataset, the operation is performed and effective heart attack level is produced. Advantages of proposed system are High performance and accuracy rate and it is very flexible and high rates of success are achieved.

TABLE OF CONTENT

List Of Figures

CHAPTER 1

INTRODUCTION

INTRODUCTION

The heart's function is affected by cardiac disease. According to a survey conducted by the World Health Organization, 10 million people are affected with heart disease and have died as a result of it. The difficulty that the healthcare industry is facing today is early disease prediction after a person has been impacted. Medical history records or data are vast, and data in the real world may be incomplete or inconsistent. In the past, accurately forecasting disease and providing treatment to patients may not have been possible for every patient at an early stage due to these variables.

Many scientists have attempted to develop a model that can predict cardiac disease in its early stages, but they have not been successful. Every proposed system has its own set of drawbacks. The user must enter all of the symptoms that he is experiencing in order for the system to anticipate the outcome. This research is based on data from the SAQ analysis. Back propagation is used to train neural networks and evaluate the prediction system.

What is Machine Learning?

Machine Learning is a collection of computer algorithms that can learn from their mistakes and improve themselves without needing to be explicitly programmed by a programmer. Machine learning is a branch of AI that integrates data with statistical methods to predict an output that may be used to provide meaningful insights.

The breakthrough is based on the premise that a computer may learn from data (for example) to create correct results on its own. Data mining and Bayesian predictive modeling are both strongly related to machine learning. The computer takes data as input and generates answers using an algorithm.

Providing a recommendation is a common machine learning problem. All recommendations of movies or series are based on the user's past data for those who have a Netflix account. Unsupervised learning is being used by tech companies to improve the customer experience by personalizing recommendations.

Machine learning is also utilized to do activities such as fraud detection, predictive maintenance, portfolio optimization, task automation, and so on.

Machine Learning vs. Traditional Programming

Machine learning differs greatly from traditional programming. A programmer codes all the rules in traditional programming in conjunction with an expert in the industry for which software is being produced. Each rule is built on a logical foundation, and the machine will follow the logical statement and produce an output. More rules must be written as the system becomes more complicated. Maintaining it can quickly become untenable.

Machine learning differs greatly from traditional programming. A programmer codes all the rules in traditional programming in conjunction with an expert in the industry for which software is being produced. Each rule is built on a logical foundation, and the machine will follow the logical statement and produce an output. More rules must be written as the system becomes more complicated.

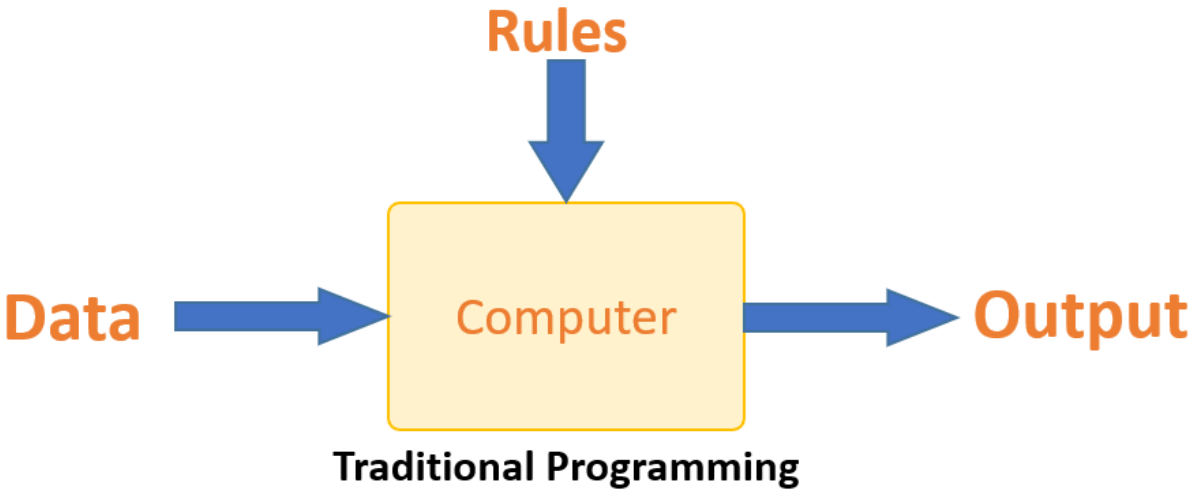

Fig 1.1 Traditional Programming

This problem is meant to be solved via machine learning. The machine deduces the relationship between the input and output data and creates a rule. Each time there is fresh data, the programmers do not need to design new rules. To improve efficacy over time, the algorithms adjust in reaction to new data and experiences.

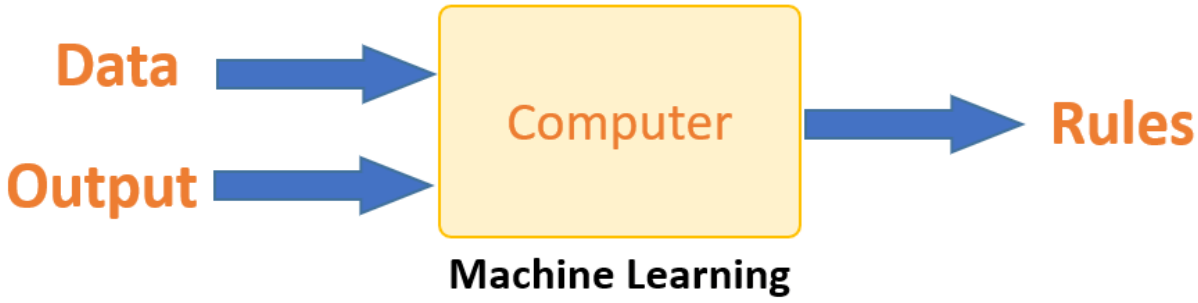

Fig 1.2 Machine Learning

How does Machine Learning Work?

All learning takes place in the brain of machine learning. The machine learns in the same way that humans do. Humans gain knowledge from their experiences. The more information we have, the easier it is to make predictions. By analogy, the chances of success are lower in an unknown situation than in a known situation. Machines are taught in the same way. The machine examines an example in order to produce an accurate forecast. When we give the machine a similar scenario, it can predict what would happen. However, just like a human, if the machine is fed an example that has never been seen before, it has trouble predicting what will happen next.

Learning and inference are the two main goals of machine learning. First and foremost, the machine learns by recognizing patterns. The data helped to make this discovery. One of the most important tasks for a data scientist is to carefully select the data to provide to the computer. A feature vector is a collection of attributes used to solve an issue. A feature vector can be thought of as a subset of data that is utilized to solve a problem.

The computer employs some sophisticated algorithms to simplify reality and turn this finding into a model. As a result, the learning step is used to characterize and summarize the data in order to create a model.

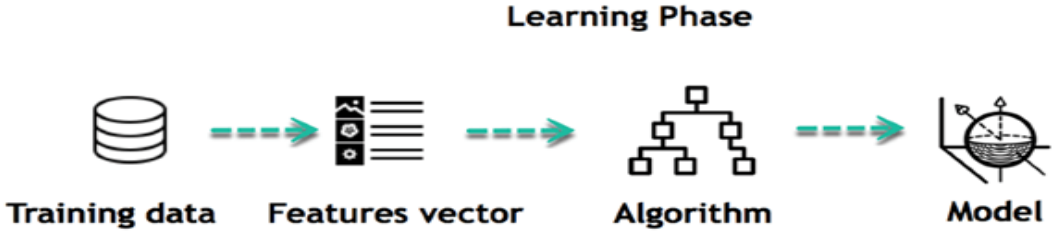

Fig 1.3 Learning Phase

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING AND PROPOSED SYSTEM

2.1.1EXISTING SYSTEM:

- Shen et al. developed a system based on self-administered questionnaires at first. The user must enter all of the symptoms that he is experiencing in order for the system to anticipate the outcome. This research is based on data from the SAQ analysis.
- Chen et al. devised a method for predicting cardiac disease. For categorization and prediction, he employed the Vector Quantization technique, which is an artificial intelligence technique. Back propagation is used to train neural networks and evaluate the prediction system. On the testing set, roughly 80% accuracy is attained during the testing phase. It takes time to put data from prior records to practical use.

DISADVANTAGES OF EXISTING SYSTEM:

- Low rate of accuracy.
- Process that takes a long time.
- Medical errors pose a severe threat to our healthcare profession. People will be afraid of going to the hospital for treatment if this continues. We can put an end to medical misdiagnosis by educating the public and initiating claims and lawsuits against the doctors who are responsible.
- The majority of these research are macro-level theoretical analyses, with quantitative investigations lacking.

2.1.2PROPOSED SYSTEM:

- To address this, we're using a Decision Tree Classifier to get more accurate results in less time. Machine learning is becoming increasingly important in modern life, particularly in the healthcare industry. Our issue is to predict heart disease by processing

- ❖ patient's dataset and data of patients, i.e., users for whom we need to predict the risks of occurrence of a heart illness.
- ❖ Our goal is to create a heart disease prediction system utilizing the Decision Tree Classifier, a powerful Machine Learning method. As input, a CSV file is provided. The outcome is anticipated and displayed after the surgery is completed successfully.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ The suggested model has the following advantages: high performance and accuracy rate.
- ❖ It is incredibly adaptable, and high success rates are achieved.
    - o When compared to alternative methods, the application with Decision Tree Classifier has a higher accuracy rate. We got roughly 96 percent in this system.

2.2 FEASIBILITY STUDY

In this step, the project's feasibility is assessed, and a business proposal is presented, complete with a general project plan and cost estimates. A feasibility study of the proposed system must be conducted during system analysis. This is to ensure that the planned system will not cause the organization any problems. A basic understanding of the system's primary requirements is required for feasibility analysis.

The feasibility analysis includes three main considerations.

- ♦ ECONOMICAL FEASIBILITY
- ♦ TECHNICAL FEASIBILITY
- ♦ SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is being conducted to determine the system's economic impact on the organization. The amount of money the corporation can invest in system research and development is limited. It is necessary to justify spending. As a result, the produced system came in under budget,

which was possible because the majority of the technologies used were freely available. Only the customized items have to be bought.

TECHNICAL FEASIBILITY

This research is being carried out to determine the system's technological feasibility, or technical requirements. Any system that is created must not place an excessive burden on the available technical resources. This will put a lot of strain on the technical resources available. As a result, the client will face high expectations. Because very minor or no changes are necessary to implement this system, the designed system must have a low requirement.

SOCIAL FEASIBILITY

The purpose of the study is to determine the user's level of acceptance of the system. This includes teaching the user how to utilize the technology effectively. The user must accept the system as a need rather than feeling threatened by it. The methods used to educate and familiarize the user with the system are totally responsible for the level of acceptance by the users. His self-esteem must be boosted so that he may offer constructive criticism, which is encouraged because he is the system's final user.

2.3 TOOLS AND TECHNOLOGIES USED

2.3.1 Flask Framework:

Flask is a Python-based web application framework. It is developed by Armin Ronacher, who leads the Pocco international network of Python aficionados. The Flask template engine is based on the Werkzeug WSGI toolkit and the Jinja2 template engine. Both are Pocco initiatives. The HTTP protocol is the backbone of the internet's data communication. This protocol specifies many techniques for retrieving data from a given URL.

2.3.2 MySQL

MySQL is a free database management system that may be used to store data and retrieve information from the system. The acronym SQL stands for "structured query language," and it is also known as a relational database management system. Sun Microsystems was the company that created it. Because MySQL is a query language, we write in the form of queries, which are basic English terms that are easier to comprehend. When we write a query, the result or output that matches the query's conditions is generated. We must utilize the "AND" clause

when we need to combine two or more searches. We utilize the "OR" clause when either of the queries must be executed. There are various options provisions for many purposes SELECT, DELETE, INSERT, ADD, DROP, and UPDATE are some basic clauses. The SELECT clause is used to select specific table entries. The DELETE command is used to remove records from a table. To delete the entire table, use the DROP command. The INSERT command is used to add records to the table. UPDATE is used to change the values of the Inserted tuple. To add new columns to an existing table, use ADD.

2.4 HARDWARE AND SOFTWARE REQUIREMENTS

HARDWARE REQUIREMENTS:

- System : Pentium i3 Processor.
- Hard Disk : 500 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 4 GB

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10.
- Coding Language : Python
- Web Framework : Flask

CHAPTER 3

SOFTWARE REQUIREMENTS SPECIFICATION

We must define a user and their capabilities as part of this process. We'll need to identify which activities will be employed in this offering. And it tells what all the actions are, as well as the functions of each process.

3.1 USERS

3.1.1 User function

We can try to design a device data protection system for sending and receiving data in our project. It is a simple and effective programmer to use. As a result, any user with a basic understanding of storage and basic user can use it. In our project, we organize a new idea called device authority, which is used to decrypt files. Then devising a method of revoking the device. In this project, you will build a method for transmitting an encrypted file and then decrypting it using specific keys. After that, implement a search for a file and send a key for that file.

3.1.2 Admin Functions

We gathered a requirement for developing an admin login page based on the survey. Before logging in, the owner can register the login into the sites. We try to plan the admin login into that page on that page. Using a username and password Upload a file based on the user's requirements. This plan will be used to collect and design the information.

3.2 Functional Requirements

A steady need is employed in encoding to illustrate the consistency of a thing structure or its part. A significant portion of the reason for existing is to appear as a technique of information sources, leads, and yields. Practical requirements imply that the principal action must be completed. This specifies what the framework must accomplish and which types of movement will be anticipated. The framework's useful prerequisites are divided into three main categories: Receiver nuances, admin nuances, cloud nuances, and their capacities

- Clustering Server
- Customer

- Marketing

3.2.1 Clustering Server:

(a) Given the appropriate fragments, we can determine the high homogeneity of individual groups and the high level of uniqueness among persons.

(b) a two-layer grouping model based on client attributes, client commitments, and bunch division analysis. We group the estimations of various clients and execute the qualities of clients in a precise method all of the time.

(c). Using these cards, businesses can track and distinguish client usage, as well as use our proposed bunching model for business analysis. Inclination research can help a company spot changes in customer value and behavior and, if necessary, adjust its product process to retain top customers.

3.2.2 Customer:

(a). Apart from effectively managing the homogeneity of quality across a large customer base, this research can also be applied to a small number of clients—for example, to develop board processes and marketing standards to help executives with client relationships (CRM).

(b). Our methodology outlines a strategy for businesses to prepare for long-term CRM and retain loyal customers. Additionally, temporary advertisers can use this displaying method to specifically deal with advance items or administrations.

(c). Business applications include targeted or direct advertising based on client segmentation and grouping, customization services, excellent CRM, and client behavior, properties, and preferences.

3.2.3 Marketing:

(a) Clients construct appropriate segments, allowing the company to focus on its target customers and develop CRM, marketing strategies, and limited-time activities.

(b). The extensive data stage cross-examination effort maintains each customer grouping and, in conjunction with effective marketing strategies, frames the second layer of client grouping study.

(c) provide continuous, diverse, and rich client data through pre-planned pre-investigation to strengthen the target customer base and relieve advertising personnel of the burdensome task.

(d) employ data mining technology to identify possible target consumers, increase the reach of advertising products and services, and improve the precision of accuracy advertising.

3.3 Non-Functional Requirements

3.3.1 Performance requirements

• The term "performance requirements" refers to the time it takes to respond to a request for system capability.

• Our project should be able to support whatever type of requirements the user desires. In addition, it meets the needs of the end consumers.

• Within a few seconds, the system will check the login details. It will be in the user, cloud, and administrator information.

• With the help of a device, our product should provide security. The device that has half of a password on it. Another section of their system will be saved.

3.3.2 Safety Requirements

Our product should meet safety standards. The devices should be revoked by the system. When the devices are misplaced or taken by someone else. Our product then disables the devices and assigns a new password.

3.3.3 Security Requirements

Our system should include security criteria and a password for device authentication. Exception handling mechanisms can assist you avoid making a mistake.

3.3.4 Software Quality Attributes

Availability:

Our programmed will not hang, will open swiftly, and will access data promptly. With these issues, the use will suffice.

Reliability

Our system should detect erroneous inputs and reject them. It should also be checked. If there is an error, the system will display error messages. Furthermore, our system will not fail. When the operating system within the method for the system is in interface, our system behaves in a user-acceptable manner.

Usability:

Our system effortlessly communicates with other users and distributes data.

Maintainability

This system will be utilized indefinitely. The system should be well-maintained, and the device password should be kept safe.

Portability

With minor or no modifications, our system can run on any web browser on any platform.

CHAPTER 4

SYSTEM DESIGN AND ANALYSIS

4.1 SYSTEM ARCHITECTURE:

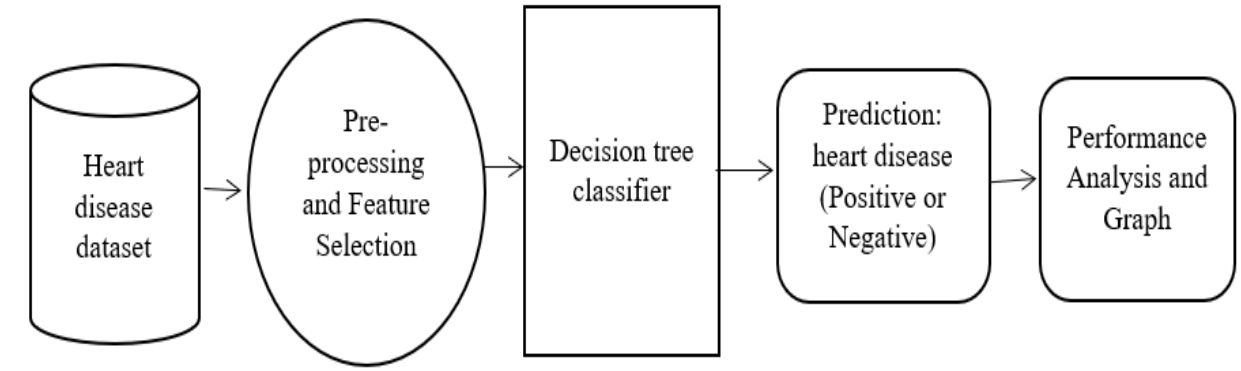

Fig 4.1 System Architecture

This structured workflow ensures a logical approach to predicting heart disease, combining data science techniques with medical applications. The detailed breakdown of the process illustrated in the diagram:

1. **Heart Disease Dataset** – The journey begins with a dataset containing various patient health attributes, such as age, blood pressure, cholesterol levels, and other medical indicators. This dataset forms the basis for training the predictive model.
2. **Pre-processing and Feature Selection** – Before feeding data into the model, it's crucial to clean it by handling missing values, removing irrelevant features, and normalizing numerical values. Feature selection ensures that only the most relevant variables (like blood pressure or ECG readings) are considered, improving accuracy and reducing unnecessary complexity.
3. **Decision Tree Classifier** – This is the core predictive model. A decision tree works by breaking down the data into smaller subsets based on feature conditions, forming a tree-like structure where each branch represents a decision rule. The final nodes (leaves) determine the classification—whether the patient is likely to have heart disease or not.

4. **Prediction Output (Positive or Negative)** – After the classifier processes the input data, it generates a prediction: either the patient is at risk for heart disease (**Positive**) or not (**Negative**). These predictions can then be used for further medical assessments or preventive measures.
5. **Performance Analysis & Graph Visualization** – To assess how well the model is performing, evaluation metrics such as accuracy, precision, recall, and F1-score are calculated. Visualizations like confusion matrices or decision boundaries help interpret the model's effectiveness and identify areas for improvement.

4.2 SYSTEM DESIGN

DATA FLOW DIAGRAM:

1. A bubble chart is another name for a DFD. It's a basic graphical formalism that can be used to depict a system in terms of data intake, processing, and output data generated by the system.
2. One of the most essential modelling tools is the data flow diagram (DFD). It's used to represent the various system components. The system process, the data used by the process, an external entity that interacts with the system, and the information flows in the system are all examples of these components.
3. DFD depicts how data flows through the system and is transformed by a sequence of transformations. It's a graphical representation of information flow and the transformations that occur when data goes from input to output.
4. DFD is sometimes referred to as a bubble chart. At any level of abstraction, a DFD can be used to depict a system. DFD can be divided into levels, each representing a different level of information flow and functional detail.

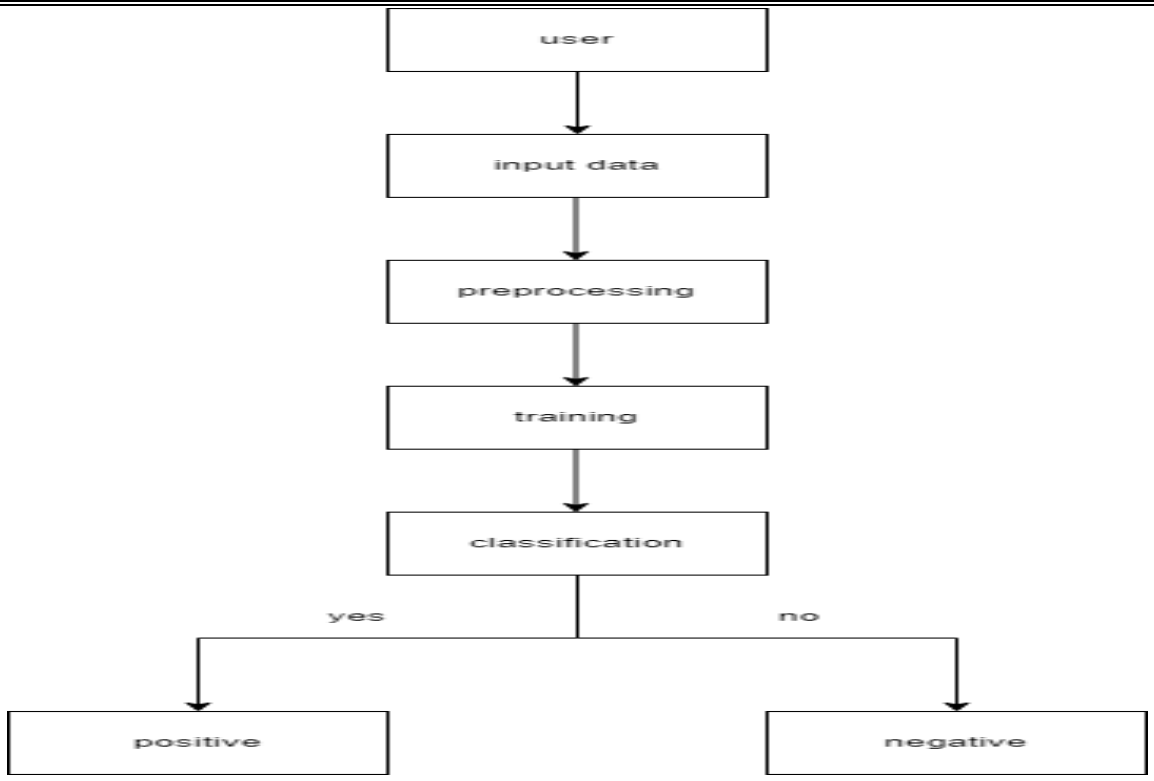

Fig 4.2 0-level DFD

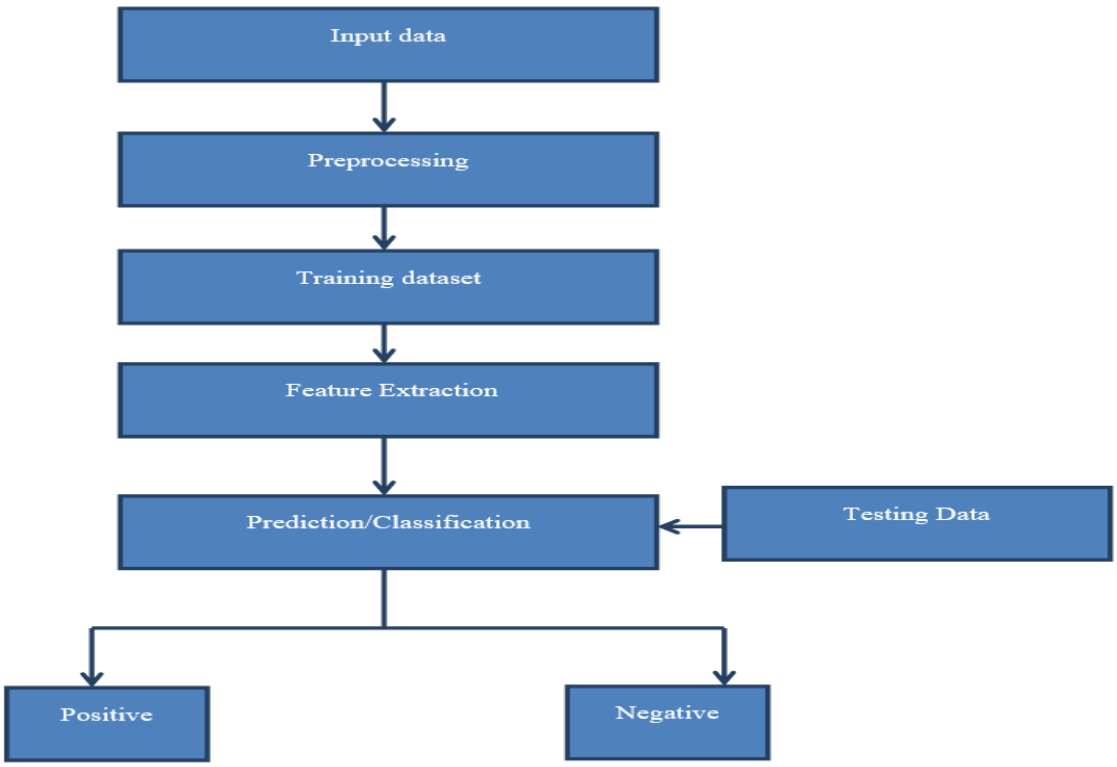

Fig 4.3 1-level DFD

ER-diagram:

The ERD (Entity Relationship Diagram) is a diagram that shows the relationship between entity sets recorded in a database. In other words, ER diagrams aid in the explanation of database logical structure. Entities, attributes, and relationships are the three core ideas used to generate ER diagrams.

Rectangles are used to represent entities, ovals are used to describe characteristics, and diamond shapes are used to show relationships in ER Diagrams.

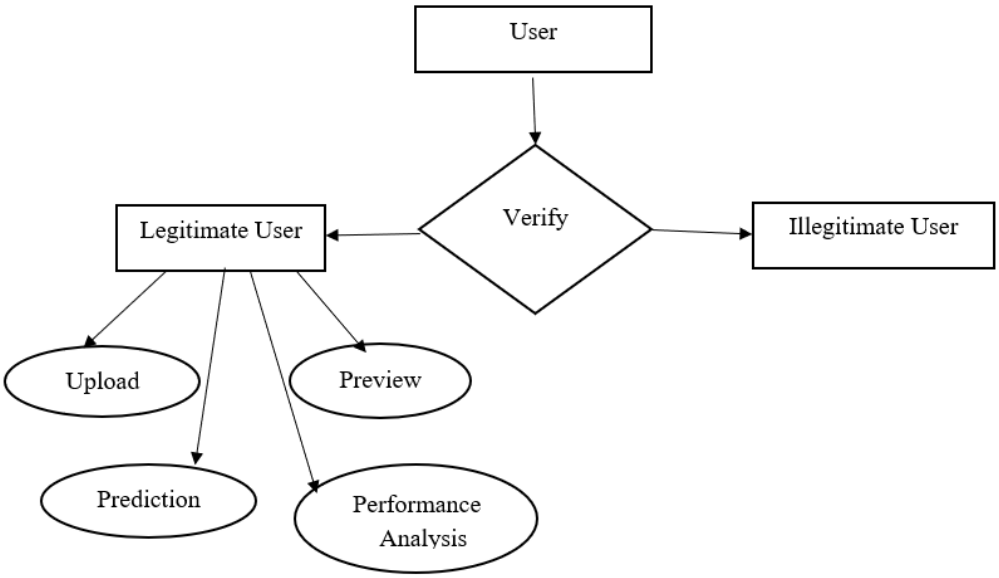

Fig 4.4 ER-Diagram

USE CASE DIAGRAM

A use case diagram is a form of behavioral diagram specified by and derived from a Use-case analysis in the Unified Modeling Language (UML). Its goal is to offer a graphical representation of a system's functionality in terms of actors, goals (expressed as use cases), and any dependencies between those use cases. A use case diagram's principal aim is to indicate which system functions are performed for which actor. The roles of the system's actors can be shown.

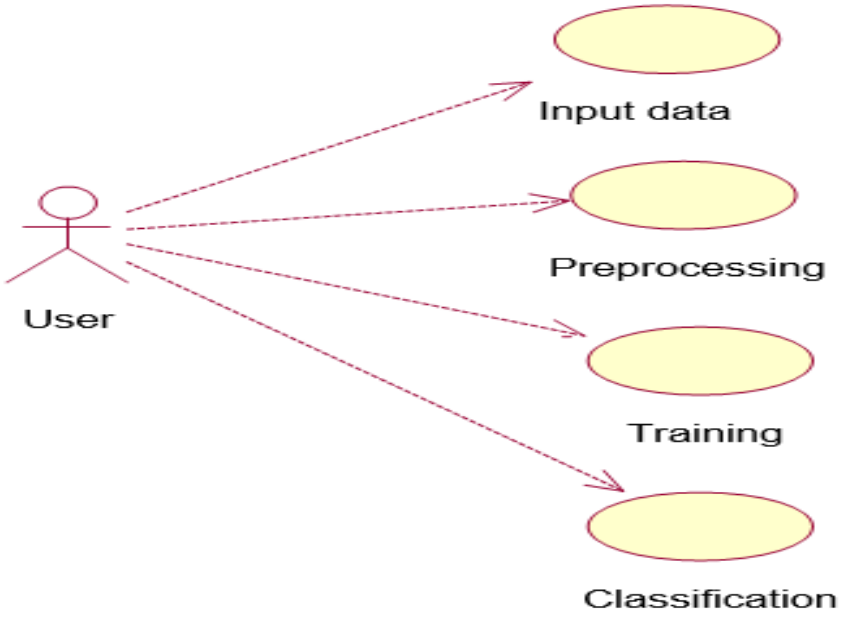

Fig 4.5 Use Case Diagram

SEQUENCE DIAGRAMS

In the Unified Modeling Language (UML), a sequence diagram is a type of interaction diagram that depicts how processes interact with one another and in what order. It's a Message Sequence Chart construct. Event diagrams, event scenarios, and timing diagrams are all terms that are used to describe sequence diagrams.

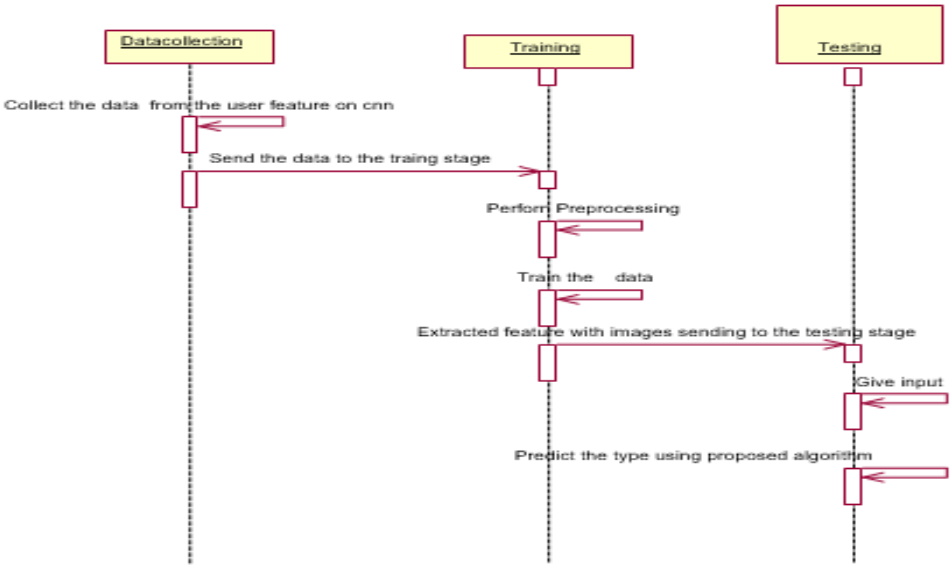

Fig 4.6 Sequence Diagram

CLASS DIAGRAM:

A class diagram in the Unified Modeling Language (UML) is a form of static structure diagram in software engineering that depicts the structure of a system by displaying the system's classes, attributes, operations (or methods), and relationships among the classes. It clarifies which class holds data.

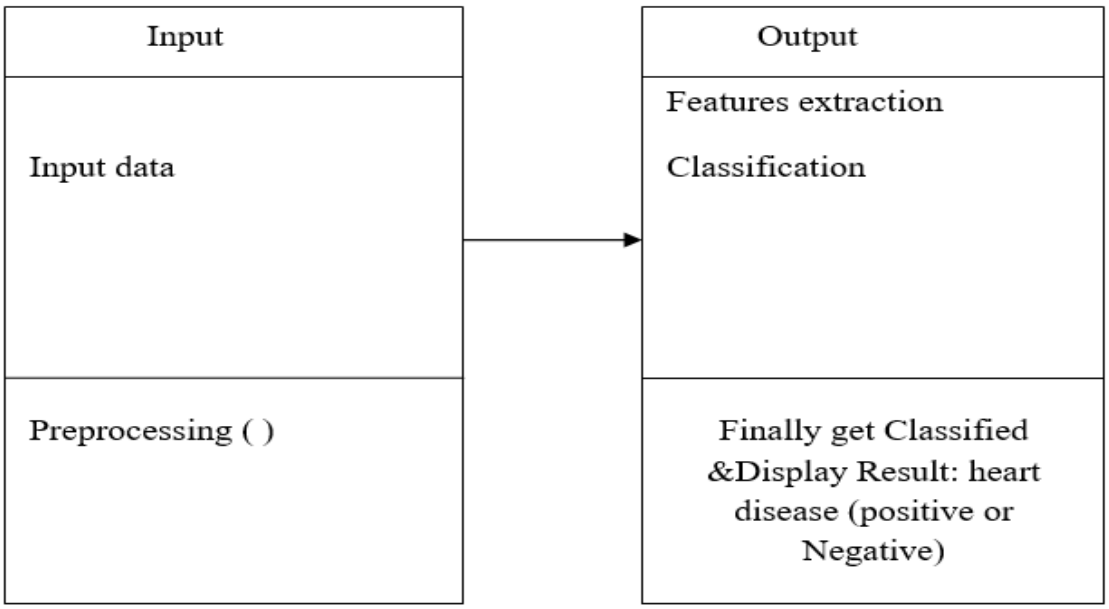

Fig 4.7 Class Diagram

INPUT DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?

- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

OBJECTIVES

1.Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2.It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

OUTPUT DESIGN

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows that include step-by-step activities and actions, as well as support chevalier, iteration, and concurrency. Activity diagrams can be used in the Unified Modeling Language to depict the business and operational step-by-step processes of system components. An activity diagram depicts the

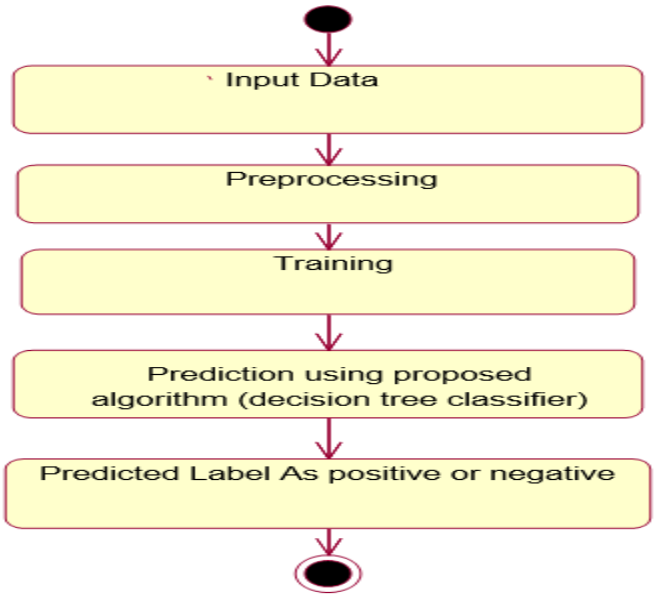

Fig 4.8 Activity Diagram

CHAPTER 5

IMPLEMENTATION

5.1 MODULES:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyse and Prediction
- Accuracy on test set
- Saving the Trained Model

5.2 MODULES DESCSRIPTION:

Data Collection:

This is the first step in collecting data for the building of a machine learning model. This is a vital phase that will have a cascading effect on the model's performance; the more and better data we collect, the better our model will perform.

Data can be collected via a variety of methods, including online scraping, manual interventions, and so on.

Heart Disease dataset taken from Kaggle (https://www.kaggle.com/ronitf/heart-disease-uci)

Dataset:

The dataset consists of 303 individual data. There are 14 columns in the dataset, which are described below.

1. Age: Displays the age of the individual.

2. Sex: Displays the gender of the individual using the following format:
   1=Male
   0=Female

3. Chest-Pain Type(cp): Displays the type of the Chest-Pain experienced by the individual using the following format:
   1=Typical Angina
   2=Atypical Angina
   3=Non Anginal Pain
   4=Asymptotic

4. Resting Blood Pressure(trestbps): displays the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholesterol (chol): displays the serum cholesterol in mg/dl (unit).
6. Fasting Blood Sugar(fbs): Compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar >120mg/dl then: 1(true), else:0(false)
7. Resting ECG(restecg): Displays resting electrocardiographic results
   0=Normal
   1=Having ST-T wave Abnormality
   2=Left Ventricular Hyperthrophy
8. Max heart rate achieved : displays the max heart rate achieved by an individual.
9. Exercise induced Angina:
   1=Yes
   0=No
10. ST depression induced by exercise relative to rest: displays the value which is an integer or float.
11. Peak Exercise ST Segment:
    1=unsloping
    2=Flat
    3=Downsloping
12. Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.
13. Thal: Displays the Thalassemia:

3=Normal

6=Fixed defect

7=Reversible defect

14. Diagnosis of heart disease: Displays whether the individual is suffering from heart disease or not:

0=Absence

1=Present

Data Preparation: Gather information and prepare it for training. Clean anything that may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)

Randomize data to remove the impacts of the sequence in which we gathered and/or otherwise prepared our data.

Perform other exploratory analysis, such as visualizing data to assist uncover meaningful associations between variables or class imbalances (bias alert!).

Sets for training and evaluation have been created.

Model Selection:

We used Decision Tree Classifier machine learning algorithm, We got a accuracy of 96.7% on test set so we implemented this algorithm.

Decision Tree Classification Algorithm:

Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however it is most commonly employed to solve classification issues. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier.

The Decision Node and the Leaf Node are the two nodes of a Decision tree. Leaf nodes are the output of those decisions and do not contain any more branches, whereas Decision nodes are used to make any decision and have several branches.

The decisions or tests are made based on the characteristics of the given dataset.

It's a graphical depiction for obtaining all feasible solutions to a problem/decision depending on certain parameters.

It's termed a decision tree because, like a tree, it starts with the root node and grows into a tree-like structure with additional branches.

We utilize the CART algorithm, which stands for Classification and Regression Tree algorithm, to form a tree.

A decision tree simply asks a question and divides the tree into subtrees based on the answer (Yes/No).

Below diagram explains the general structure of a decision tree:

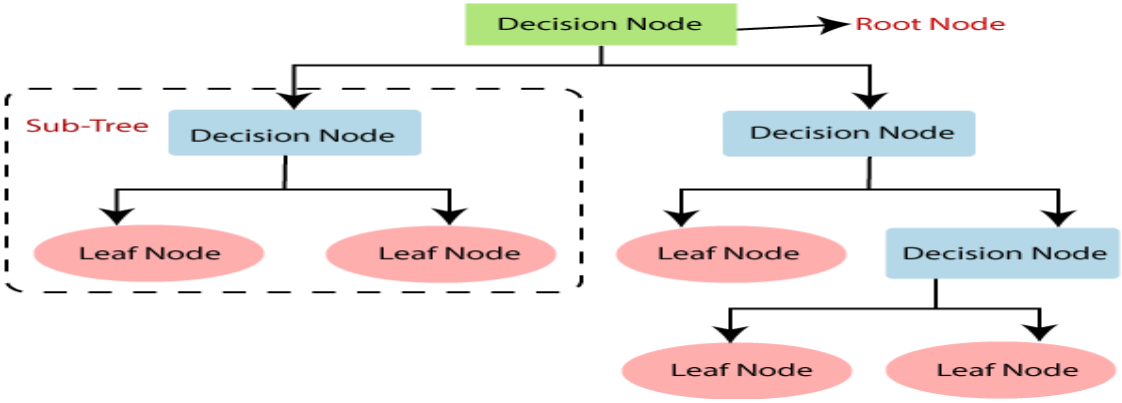

Fig 5.1 Decision Tree

Why use Decision Trees?

Machine learning uses a variety of methods, therefore picking the right approach for the given dataset and problem is the most important thing to remember while building a machine learning model. The following are two reasons to use the Decision Tree:

o Decision Trees are easy to understand because they usually mimic human thinking abilities when making a decision.

o Because the decision tree has a tree-like form, the rationale behind it is simple to comprehend.

Terminologies for Decision Trees

Root Node: The root node is the starting point for the decision tree. It represents the full dataset, which is then split into two or more homogeneous groups.

Leaf Node: Leaf nodes are the tree's final output nodes, and they can't be separated any further after that.

Splitting: Splitting is the process of separating the decision node/root node into sub-nodes based on the conditions specified.

Branch/Sub Tree: A tree that has been split into branches or subtrees.

Pruning: Pruning is the procedure of pruning a tree to remove undesired branches.

Parent/Child node: The root node of the tree is known as the parent node, while the remaining nodes are known as the child nodes.

Analyze and Prediction:

In the actual dataset, we chose only 7 features:

1. Age: Displays the age of the individual.
2. Sex: Displays the gender of the individual using the following format:
   1=Male
   0=Female
3. Chest-Pain Type(cp): Displays the type of the Chest-Pain experienced by the individual using the following format:
   1=Typical Angina
   2=Atypical Angina
   3=Non Anginal Pain
   4=Asymptotic
4. Resting Blood Pressure(trestbps): displays the resting blood pressure value of an individual in mmHg (unit)
5. Serum Cholesterol (chol): displays the serum cholesterol in mg/dl (unit).
6. Fasting Blood Sugar(fbs): Compares the fasting blood sugar value of an individual with 120mg/dl.
   If fasting blood sugar >120mg/dl then: 1(true), else:0(false)
7. Resting ECG(restecg): Displays resting electrocardiographic results
   0=Normal
   1=Having ST-T wave Abnormality
   2=Left Ventricular Hyperthrophy
8. Max heart rate achieved : displays the max heart rate achieved by an individual.

9. Exercise induced Angina:
   1=Yes
   0=No
10. ST depression induced by exercise relative to rest: displays the value which is an integer or float.
11. Peak Exercise ST Segment:
    1=unsloping
    2=Flat
    3=Downsloping

Accuracy on test set:

On the test set, we were able to get a 96.7 percent accuracy.

Saving the Trained Model:

When you're ready to put your trained and tested model into a production-ready environment, the first step is to save it as a.h5 or.pkl file with a library like pickle.

Check to see if your environment has Pickle installed.

Next, let’s import the module and dump the model into.pkl file

CHAPTER 6

CODING

App.py File

```
from flask import Flask, render_template, request, redirect, url_for
import pickle
import numpy as np

app = Flask(__name__)
model = pickle.load(open('best_model.pkl', 'rb'))
scaler = pickle.load(open('scaler.pkl', 'rb'))

USERNAME = 'user'
PASSWORD = 'password'

@app.route('/')
def home():
    return render_template('home.html')

@app.route('/predict_page')
def predict_page():
    return render_template('index.html')

@app.route('/recommand_page')
def recommand_page():
    return render_template('recommand.html')
```

```
@app.route('/thank_you_page')
def thank_you_page():
    return render_template('thankyou.html')


@app.route('/predict', methods=['POST'])
def predict():
    if request.method == 'POST':
        try:
            features = [float(x) for x in request.form.values()]
            input_array = np.array(features).reshape(1, -1)
            input_scaled = scaler.transform(input_array)
            prediction = model.predict(input_scaled)


            if prediction[0] == 1:
                return render_template('doctors.html', message='✅ You are not healthy.  heart disease detected!')
            else:
                return render_template('result.html', message='✅ You are healthy. No heart disease detected!')
        except Exception as e:
            return render_template('result.html', message=f'⚠️ Error: {e}')


@app.route('/doctors')
def doctors():
    return render_template('doctors.html')
```

```
@app.route('/login', methods=['GET', 'POST'])
def login():
    if request.method == 'POST':
        username = request.form['username']
        password = request.form['password']

        if username == USERNAME and password == PASSWORD:
            return redirect(url_for('predict_page'))
        else:
            return "Invalid credentials. Please try again."
    return render_template('login.html')

if __name__ == '__main__':
    app.run(debug=True)
```

CHAPTER 7

SOFTWARE TESTING

SYSTEM TESTING

The goal of testing is to find mistakes. Testing is the practice of attempting to find all possible flaws or weaknesses in a work product. It allows you to test the functionality of individual components, sub-assemblies, assemblies, and/or a whole product. It is the process of testing software to ensure that it meets its requirements and meets user expectations, and that it does not fail in an unacceptable way. There are many different types of tests. Each test type is designed to fulfil a distinct testing need.

TYPES OF TESTS

Unit testing

Unit testing entails creating test cases to ensure that the program's internal logic is working properly and that programme inputs result in valid outputs. Validation should be performed on all decision branches and internal code flow. It is the testing of the application's individual software units. Before integration, it is done after each individual unit is completed. This is intrusive structural testing that relies on prior knowledge of the structure. Unit tests are used to test a specific business process, application, or system configuration at the component level. Unit tests guarantee that each individual path of a business process follows the published specifications and has clearly defined inputs and outputs.

Integration testing

Integration tests are used to see if two or more software components can work together as a single application. Testing is mainly concerned with the basic output of screens or fields and is event driven. Integration tests show that, while the components were individually satisfying, the combination of components is proper and consistent, as demonstrated by successful unit testing. Integration testing is a type of testing that focuses on uncovering issues that occur from combining components.

Functional test

Functional tests demonstrate that the functions being tested are available in accordance with the business and technical requirements, system documentation, and user manuals.

The following items are the focus of functional testing:

Valid Input: the approved kinds of valid input must be identified.

Invalid Input: classifications of invalid input must be detected and discarded.

Functions: The functions that have been identified must be used.

Output: The application outputs that have been identified must be tested.

Interfacing systems and procedures must be invoked.

Functional tests are organized and prepared around requirements, important functions, or unique test cases. Furthermore, comprehensive coverage of Business process flows, data fields, established procedures, and subsequent processes must be considered for testing. Additional tests are identified and the effective value of present tests is determined before functional testing is completed.

System Test: System testing guarantees that the complete integrated software system complies with the specifications. It checks a setup to ensure that the results are known and predictable. The configuration-oriented system integration test is an example of system testing. Process descriptions and flows are used in system testing, with an emphasis on pre-driven process connections and integration points.

White Box Testing

White Box Testing is a type of software testing in which the software tester is familiar with the software's inner workings, structure, and language, or at the very least its purpose. It serves a purpose. It is used to test regions that a black box level cannot reach.

Black Box Testing

Testing software without knowing the inner workings, structure, or language of the module being tested is known as black box testing. Black box tests, like most other types of tests, require a definite source document, such as a specification or requirements document. It's a type of testing in which the programme being tested is treated as if it were a black box. It is impossible to "look" into it. The test accepts inputs and responds to outputs without taking into account how the software functions.

7.1 Unit Testing:

Unit testing is normally done as part of a combined code and unit test phase of the software development lifecycle; however, it is not uncommon for coding and unit testing to be done separately.

Test strategy and approach

Field testing will be done by hand, and functional tests will be meticulously documented.

Test objectives

• All field entries must function correctly.

• The indicated link must be used to activate the pages.

• There must be no delays in the entering screen, messages, or responses.

Features to be tested

• Check that the entries are in the correct format

• No duplicate entries should be allowed

• All links should lead to the correct page.

7.2 Integration Testing

The progressive integration testing of two or more integrated software components on a single platform to induce failures caused by interface faults is known as software integration testing. The integration test's goal is to ensure that components or software applications, such as those in a software system or – one step higher – software applications at the company level, work together flawlessly.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

7.3 Acceptance Testing

Acceptance by users Testing is an important aspect of any project, and it necessitates active engagement from the end user. It also guarantees that the system complies with the functional specifications.

Test Results: All of the above-mentioned test cases were successful. No flaws were discovered.

CHAPTER 8

RESULTS AND SCREENSHOTS

Home page:

Fig 8.1 Home Page

Login Page:

Fig 8.2 Login Page

Enter username and password:

Fig 8.3 Enter Username and Passward

Prediction page:

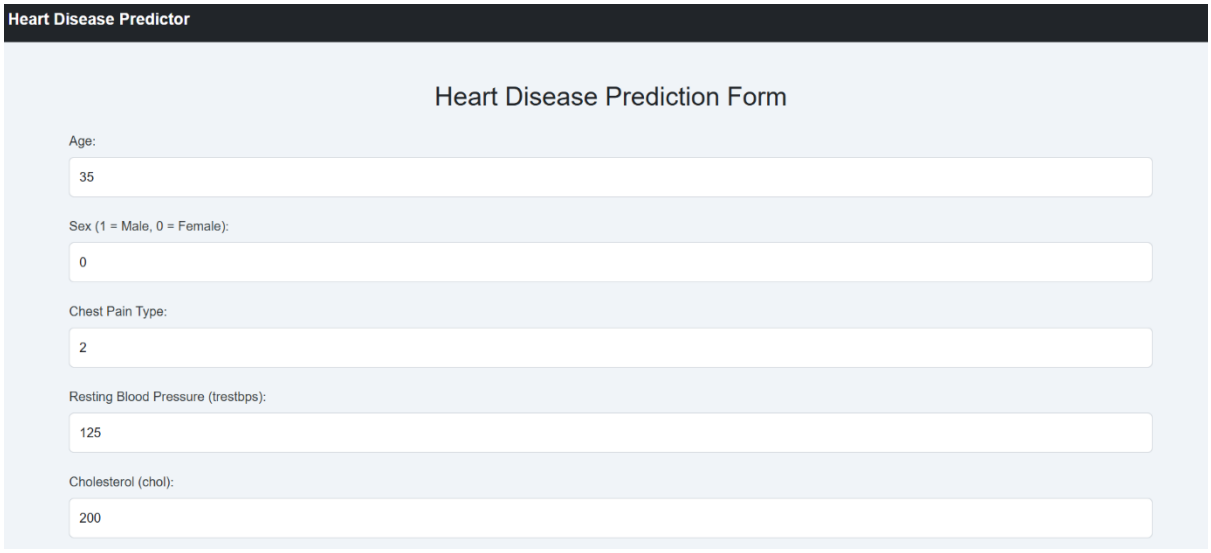

Fig 8.4 Prediction page

Click on predict button and results will display as:

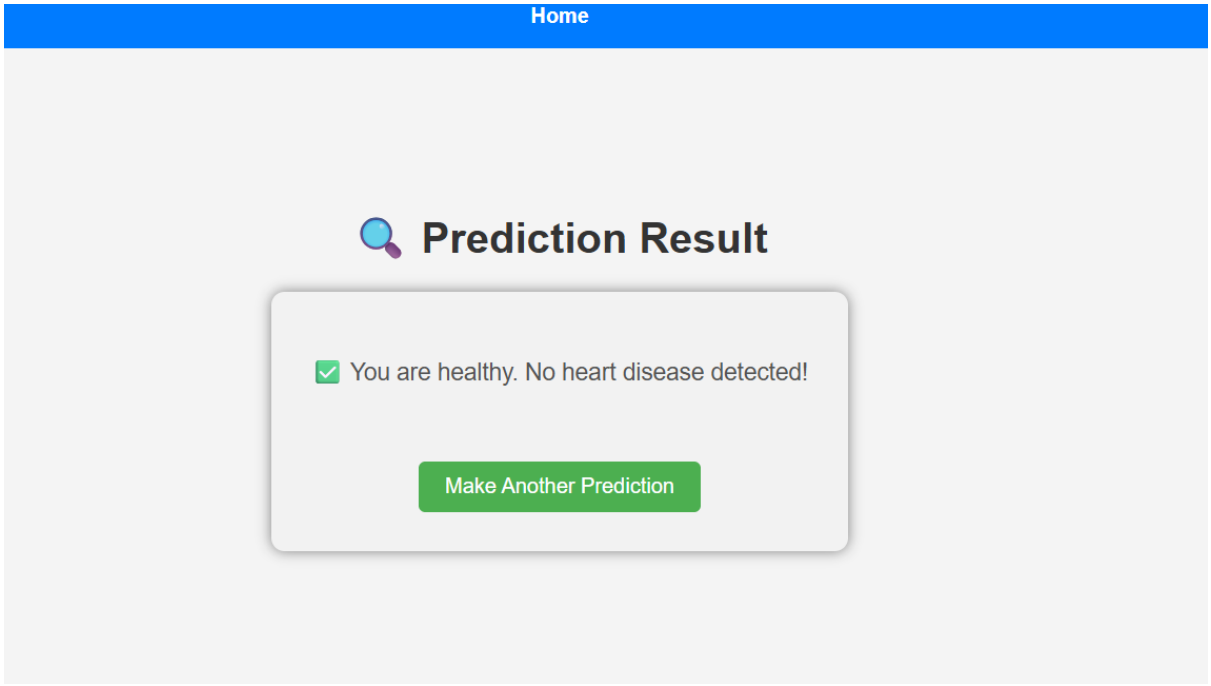

Fig 8.5 Negative results (Not suffering from heart disease)

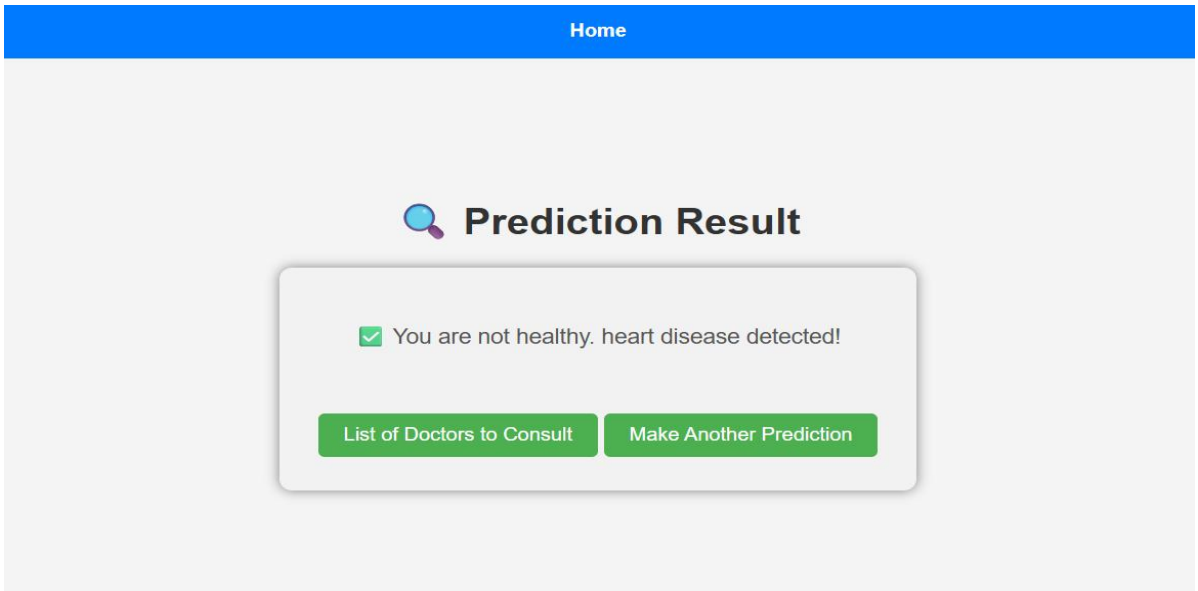

Fig 8.6 Positive results (suffering from heart disease)

If Prediction of Heart disease is true, the List of Cardiologists will be shown and displays.

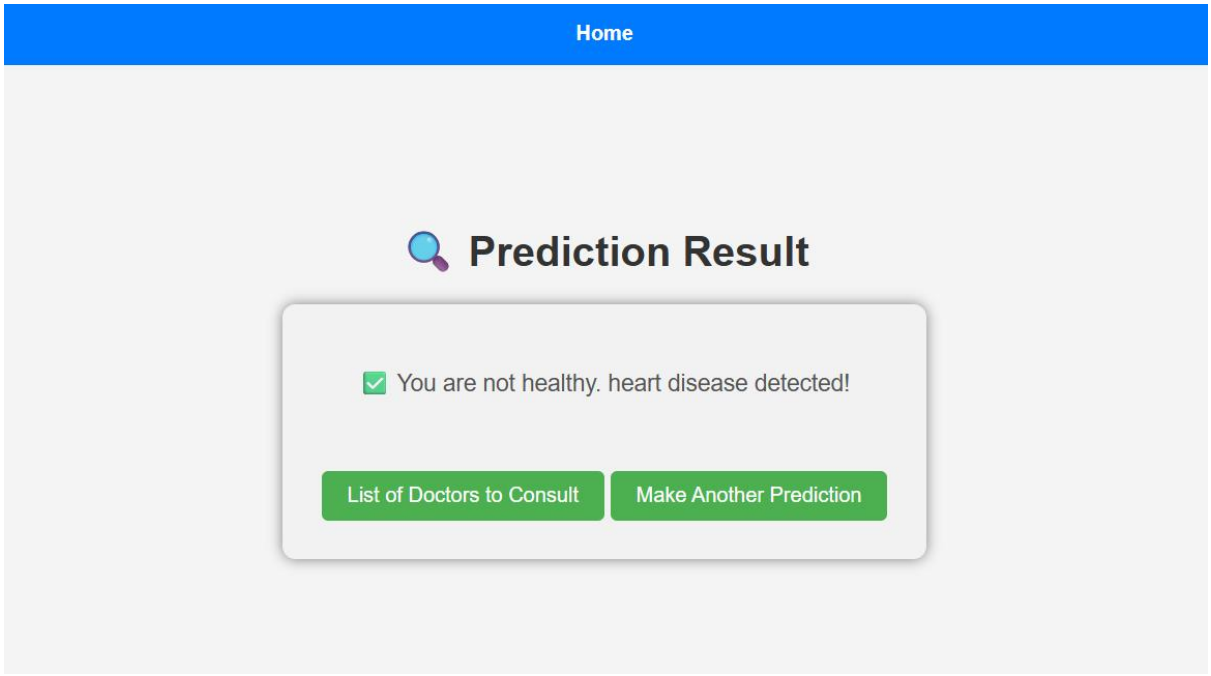

Fig 8.7 List of Doctors to consult button

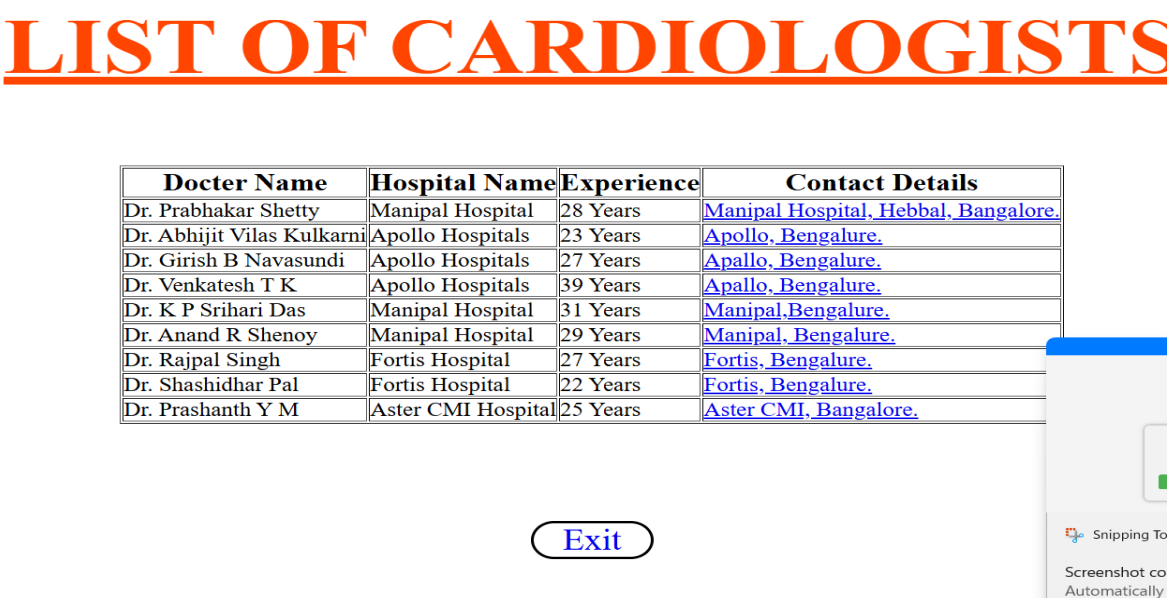

Fig 8.8 List of Cardiologists

CONCLUSION

The Decision Tree Classifier is an effective algorithm for regression and classification approaches that uses an ensemble learning mechanism. The procedure generates N decision trees and returns the class that is the average of all the decision trees' outputs. As a result, early-stage prediction accuracy is efficiently accomplished. The processing of healthcare data, specifically data connected to the heart, will aid in the early detection of heart disease or aberrant heart conditions, resulting in the prevention of long-term fatalities. In today's world, predicting heart disease is a huge difficulty. If a patient or user is unable to contact a doctor, he or she can use this application to anticipate disease by just entering the report information. And can decide whether to continue or not a doctor or not.

FUTURE ENHANCEMENTS

This application can be improved in the future by adding new features, such as sending a message to all of the user's family members if the user is diagnosed with heart disease. The information should also be forwarded to the local hospital. Another option that should be provided is online doctor consultation with the closest doctor.

It's worth noting that ML applications based on numerous efficient algorithms are used not only in disease prediction and diagnosis, but also in radiology, bioinformatics, and medical imaging diagnosis, among other fields.

BIBLIOGRAPHY

1. Kaur H Beant and Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", IJRITCC, vol. 2, no. 10, pp. 3003-08, 2014.
2. M.M. Kirmani and S.I. Ansarullah, "Prediction of heart disease using decision tree a data mining technique", IJCSN Int. J. Comput. Sci. Netw., vol. 5, no. 6, pp. 885-892, 2016.
3. Salam Ismaeel, Ali Miri et al., "Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis", IEEE Canada International Humanitarian Technology
4. Tahira Mahboob, Rida Irfan and Bazelah Ghaffar et al" Evaluating ensemble prediction of coronary heart disease using receiver operating characteristics".
5. Ammar Asjad Raja, Irfan-ul-Haq, Madiha Guftar Tamim Ahmed Khan "Intelligence syncope Disease Prediction Framework using DM-techniques" FTC 2016 –Future Technologies Conference 2016.
6. M.A. Jabbar, B.L. Deekshatulu, and Priti Chandra, "Intelligent heart disease prediction system using random forest and evolutionary approach", Journal of Network and Innovative Computing, Vol. 4, pp.174-184, 2016.
7. N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Machine Learning, (1997).
8. Ayon Dey, Jyoti Singh, N. Singh "Analysis of supervised machine learning algorithms for heart disease prediction" International journal of computer applications, published in 2006.
9. Q. K. Al-Shayea, "Artificial neural networks in medical diagnosis," International Journal of Computer Science Issues, published in 2011.