

# HEART DISEASE PREDICTION

A

PROJECT REPORT

SUBMITTED

BY

Mr. Vaibhav Pawar	2127052
Mr. Avdhoot Kumbhar	2127037
Mr. Yashraj Devrat	2127011
Mr. Shubham Kesar	2127029
Mr. Tushar Mamadge	2127040

IN PARTIAL FULFILLMENT FOR THE REQUIREMENT OF PROJECT BASED LEARNING-II

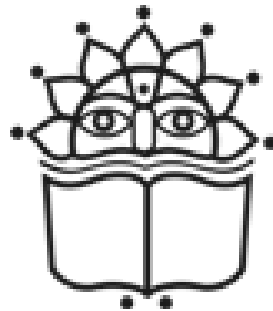
OF

**Second Year of Artificial Intelligence and Data Science**

Under the guidance of

**Prof R.V.Panchal**

(Assistant Professor)



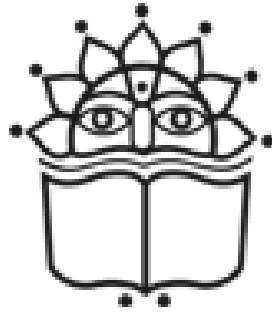
**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA  
SCIENCE**

VIDYA PRATISHTHAN'S KAMALNAYAN BAJAJ INSTITUTE OF

ENGINEERING AND TECHNOLOGY

Bhigwan Road, Vidyanagari

Baramati-413133



Vidya Pratishthan's  
Kamalnayan Bajaj Institute of Engineering and Technology, Baramati  
**Department of Artificial Intelligence and Data Science**

## **Certificate**

THIS IS TO CERTIFY THAT FOLLOWING STUDENTS

<b>Mr. Vaibhav Pawar</b>	<b>2127052</b>
<b>Mr. Avdhoot Kumbhar</b>	<b>2127037</b>
<b>Mr. Yashraj Devrat</b>	<b>2127011</b>
<b>Mr. Shubham Keskar</b>	<b>2127029</b>
<b>Mr. Tushar Mamadge</b>	<b>2127040</b>

HAVE SUCCESSFULLY COMPLETED THEIR PROJECT WORK ON

### **Heart Disease Prediction**

DURING THE ACADEMIC YEAR **2021-2022** IN THE PARTIAL FULFILLMENT TOWARDS  
THE COMPLETION OF **PROJECT BASED LEARNING-II** IN **ARTIFICIAL INTELLI-  
GENCE AND DATA SCIENCE**

Project Guide  
(Name of Guide/Supervisor)

Head, Deptt. of AI & DS  
(Digambar Padulkar)

Principal  
(Dr. R. S. Bichkar)

---

Internal Examiner

External Examiner

# Acknowledgments

It gives us great pleasure in presenting the project report on **Heart Disease Prediction**. We would like to take this opportunity to thank our internal supervisor **Mr.R.V.Panchal** for giving us all the help and supervision that we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to **Mr.Digambar Padulkar**, Head of Artificial Intelligence and Data Science Engineering Department, VPKBIET for his indispensable support, suggestions.

**Mr. Vaibhav Pawar**  
**Mr. Avdhoot Kumbhar**  
**Mr. Yashraj Devrat**  
**Mr. Shubham Keskar**  
**Mr. Tushar Mamadge**

# Abstract

In the medical field, the diagnosis of heart disease is the most difficult task. The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart disease prediction. In case of heart disease, the correct diagnosis in early stage is important as time is the very important factor. Heart disease is the principal source of deaths widespread, and the prediction of Heart Disease is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The main idea behind this work is to study diverse prediction models for the heart disease and selecting important heart disease feature using Random Forests algorithm. Random Forests is the Supervised Machine Learning algorithm which has the high accuracy compared to other Supervised Machine Learning algorithms such as logistic regression etc. By using Random Forests algorithm we are going to predict if a person has heart disease or not.

# Contents

<b>Acknowledgments</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1 Synopsis</b>	<b>1</b>
1.1 Project Title . . . . .	1
1.2 Technical Keyword . . . . .	1
1.3 Problem Statement . . . . .	1
1.4 Abstract . . . . .	1
1.5 Goals and Objective . . . . .	2
1.6 Plan of Project Execution . . . . .	2
<b>2 Technical Keywords</b>	<b>4</b>
2.1 Area of Project . . . . .	4
2.2 Technical Keywords . . . . .	4
<b>3 Introduction</b>	<b>5</b>
3.1 Motivation of Project . . . . .	5
3.2 Literature Survey . . . . .	5
<b>4 Problem Definition and Scope</b>	<b>7</b>
4.1 Problem Statement . . . . .	7
4.2 Goals and Objective . . . . .	7
4.2.1 Statement of Scope . . . . .	7
4.2.2 Methodology of problem Solving . . . . .	8
4.2.3 Outcome . . . . .	8
4.2.4 Application . . . . .	8
4.2.5 Constraints . . . . .	8
<b>5 Project Plan</b>	<b>9</b>
5.0.1 Problem Estimates . . . . .	9
5.0.2 Project Resources . . . . .	9
5.0.3 Project Task Set . . . . .	9
5.0.4 TEAM ORGANIZATION . . . . .	10

<b>6</b>	<b>Software Requirement Specification</b>	<b>11</b>
6.0.1	User Profiles . . . . .	11
6.0.2	Use Case View . . . . .	11
6.0.3	Gantt Chart . . . . .	12
6.0.4	Class Diagram . . . . .	13
<b>7</b>	<b>Conclusion</b>	<b>14</b>
<b>8</b>	<b>References</b>	<b>15</b>

# Synopsis

## 1.1 Project Title

Heart Disease Prediction

## 1.2 Technical Keyword

Classification, Machine Learning, Heart Disease Prediction, Training and Testing.

## 1.3 Problem Statement

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

## 1.4 Abstract

In the medical field, the diagnosis of heart disease is the most difficult task. The diagnosis of heart disease is difficult as a decision relied on grouping of large clinical and pathological



data. Due to this complication, the interest increased in a significant amount between the researchers and clinical professionals about the efficient and accurate heart disease prediction. In case of heart disease, the correct diagnosis in early stage is important as time is the very important factor. Heart disease is the principal source of deaths widespread, and the prediction of Heart Disease is significant at an untimely phase. Machine learning in recent years has been the evolving, reliable and supporting tools in medical domain and has provided the greatest support for predicting disease with correct case of training and testing. The main idea behind this work is to study diverse prediction models for the heart disease and selecting important heart disease feature using Random Forests algorithm. Random Forests is the Supervised Machine Learning algorithm which has the high accuracy compared to other Supervised Machine Learning algorithms such as logistic regression etc. By using Random Forests algorithm we are going to predict if a person has heart disease or not.

## 1.5 Goals and Objective

The Goals and Objective of this project is to create a model that can predict the patient's heart disease status. Another Objective is to explore the data we have been given and find key insights into Heart Disease that could be helpful for the Medical community going forward.

## 1.6 Plan of Project Execution

TASK	START DATE	Duration
1) Identifying Problem Statement	2/10/2022	10
2) Gathering Resources about Project	2/21/2022	7
3) Planning	3/14/2022	7
4) Designing	3/22/2022	6
5) PPT of Project	3/29/2022	3
6) Coding	3/31/2022	4
7) Project Report	4/12/2022	8

## Heart Disease Prediction



# Technical Keywords

## 2.1 Area of Project

Project Consists Of Machine Learning Field.Different Machine Learning Techniques are used to analyze the Heart Disease Datasets.

## 2.2 Technical Keywords

1]Classification- Different Classification Technique are Applied to Classify the Dataset.

2]Machine Learning - Different Machine Learning Techniques are involved in Project.for example Classification of Heart Disease Dataset With Different Parameters.

3]Heart Disease Prediction - Using Different classifier doing a Heart Disease Prediction.

4]Training And Testing - Training and Testing Technique are used to find Accuracy.

# Introduction

## 3.1 Motivation of Project

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using classification algorithms namely logistic Regression and Random Forest are used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better understanding and help them identify a solution to identify the best method for predicting the heart diseases.

## 3.2 Literature Survey

According to Ordonez [1] the heart disease can be predicted with some basic attributes taken from the patient and in their work have introduced a system that includes the characteristics of an individual human being based on totally 13 basic attributes like sex, blood pressure, cholesterol and others to predict the likelihood of a patient getting affected by heart disease. They have added two more attributes i.e. fat and smoking behaviour and extended the research dataset. The data mining classification algorithms such as Decision Tree, Naive Bayes, and Neural Network are utilized to make predictions and the results are analysed on Heart disease database. Yilmaz, [2] have proposed a method that uses least squares support vector machine (LS-SVM) utilizing a binary decision tree for classification of cardiocogram to find out the patient condition. Duff, et al. [3] have done a research work involving five hundred and thirty-three patients who had suffered from cardiac arrest and they were integrated in the analysis of heart disease probabilities. They performed classical statistical analysis and data mining analysis using mostly Bayesian

networks. Frawley, et al. [4] have performed a work on prediction of survival of Coronary heart disease (CHD) which is a challenging research problem for medical society. They also used 10-fold cross-validation methods to determine the impartial estimate of the three prediction models for performance comparison purposes. Lee, et al. proposed a novel methodology to expand and study the multi-parametric feature along with linear and nonlinear features of Heart Rate Variability diagnosing cardiovascular disease. They have carried out various experiments on linear and non-linear features to estimate several classifiers, e.g., Bayesian classifiers, CMAR, C4.5 and SVM. Based on their experiments, SVM outperformed the other classifiers.

# Problem Definition and Scope

## 4.1 Problem Statement

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expensive. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

## 4.2 Goals and Objective

### 4.2.1 Statement of Scope

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions

### 4.2.2 Methodology of problem Solving

There are four phases that involve in the spiral model: 1) Planning phase - Phase where the requirement are collected and risk is assessed. This phase where the title of the project has been discussed with project supervisor. From that discussion, Heart Prediction System has been proposed. The requirement and risk was assessed after doing study on existing system and do literature review about another existing research.

2) Risk analysis Phase - Phase where the risk and alternative solution are identified. A prototype are created at the end this phase. If there is any risk during this phase, there will be suggestion about alternate solution.

3) Engineering phase - At this phase, a software are created and testing are done at the end this phase.

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

### 4.2.3 Outcome

1] It is very helpful to medical Community to Predict Accurate results.

2] By using classification models user can analyze the Dataset.

3] Then Software will display overall feature wise user opinion in the form of graphs.

### 4.2.4 Application

1] In Medical Community.

2] Analysis of Heart Disease.

### 4.2.5 Constraints

Data Mining techniques does not help to provide effective decision making.

# Project Plan

## 5.0.1 Problem Estimates

- 1.Planning- Requirement gathering like Software Requirement Specifications.
2. Modeling- Contains analysis and Design.
3. Construction- Design and Development of product with testing at the end.
- 4.Evaluation- Finally the evaluation of the output of project.

## 5.0.2 Project Resources

- 1.People: There are total Five Peoples in a Group.We have divide Project working among us.
- 2.Hardware: Storage:50-60GB (Minimum)  
RAM:2 GB (Minimum)  
Processor: intel i3 (Minimum)
3. Software:  
Platform : Operating System: Windows 7 or above, Ubuntu 12 or above  
IDE: Google Colab  
Programming Language- Python

## 5.0.3 Project Task Set

Major Tasks in the Project stages are:

- Task 1: Requirements
- Task 2: Defining roles and responsibilities
- Task 3: Developing the schedule and cost baselines
- Task 4: Designing modules and coding
- Task 5: Testing and analyzing project quality
- Task 6: Assessment of technical risks.



## 5.0.4 TEAM ORGANIZATION

## Team Structure

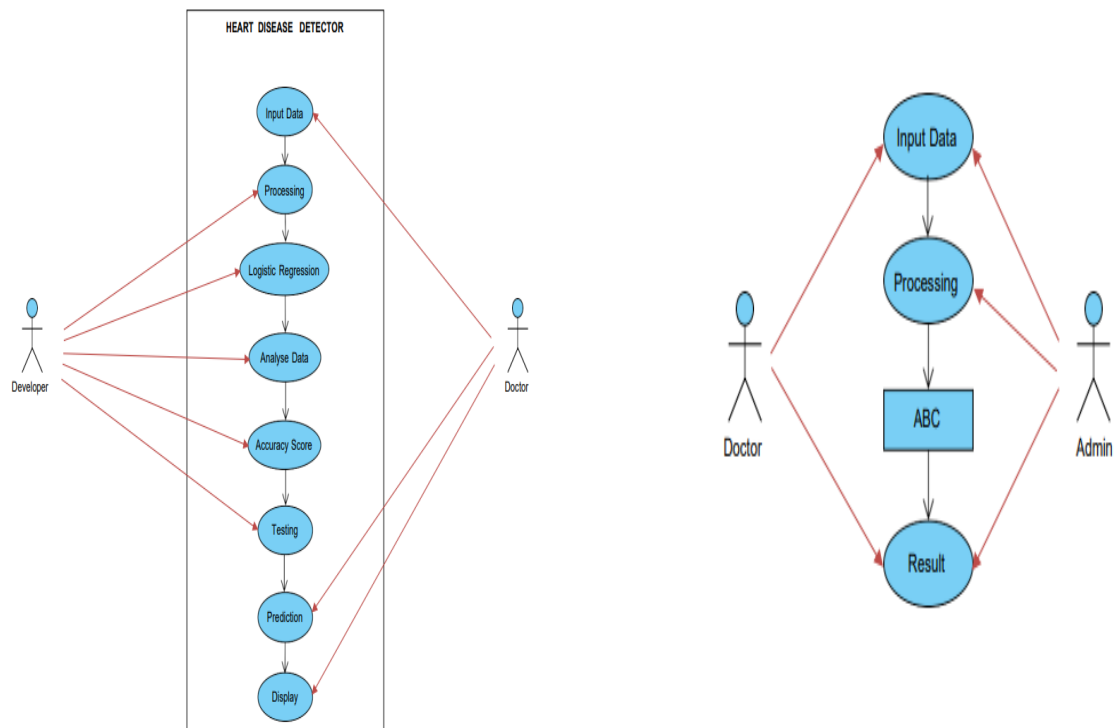
Sr.No	Name	Task
1	Mr.R.V.Panchal	Internal Supervisor
2	Mr.	External Supervisor
3	Vaibhav Pawar	Requirement Gathering and Project Design
4	Vaibhav Pawar	Project Implementation
5	Avdhoot Kumbhar	Project Implementation
6	Yashraj Devrat	Project Implementation
7	Shubham Keskar	Project Implementation
8	Tushar Mamadge	Analyzer

# Software Requirement Specification

## 6.0.1 User Profiles

1. User- User will provide required parameter.
2. Developer - Developer develops the tool to predict Heart Disease based on features. Satisfy the user requirement and keeps their record.

## 6.0.2 Use Case View



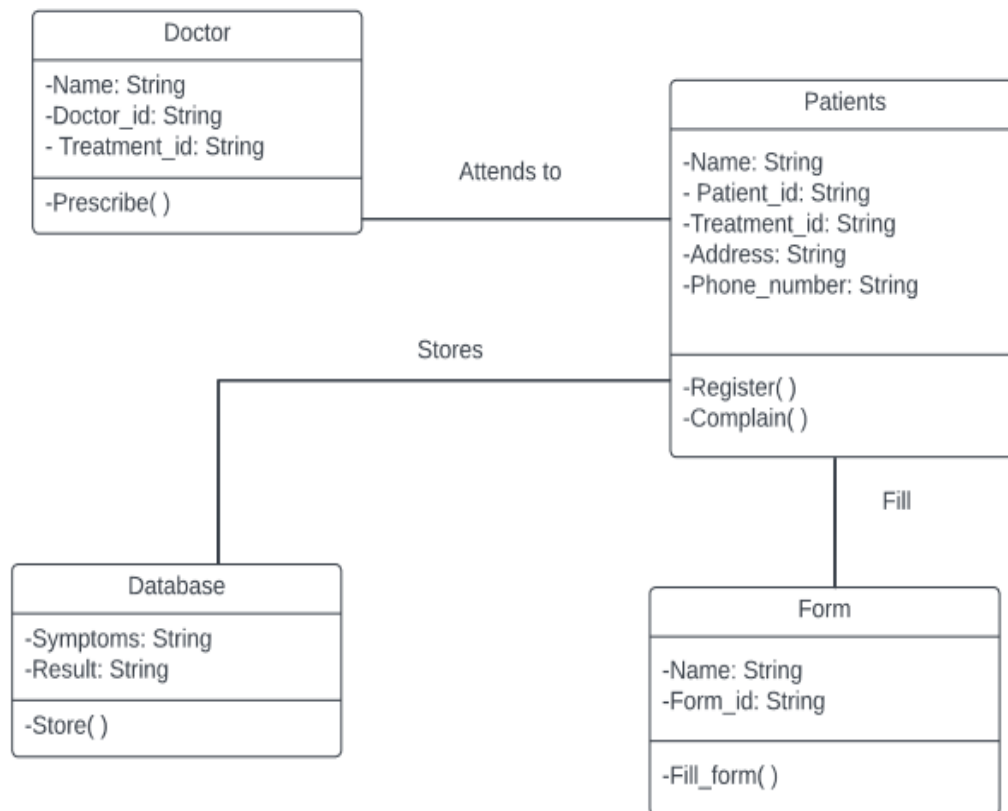
## 6.0.3 Gantt Chart

TASK	START DATE	Duration
1) Identifying Problem Statement	2/10/2022	10
2) Gathering Resources about Project	2/21/2022	7
3) Planning	3/14/2022	7
4) Designing	3/22/2022	6
5) PPT of Project	3/29/2022	3
6) Coding	3/31/2022	4
7) Project Report	4/12/2022	8

## Heart Disease Prediction



## 6.0.4 Class Diagram



## Conclusion

Heart disease prediction is challenging and very important in medical field. However, the mortality rate can be drastically controlled if the disease is detected at early stage and preventive measures are adopted as soon as possible. Hybrid Random Forest with Linear Model(HRFLM) Proved to be quite accurate in the prediction of heart disease.

## References

- [1] Brown N, Young T, Gray D, Skene A M Hampton J R (1997). Inpatient deaths from acute myocardial infarction, 1982-92: analysis of data in the Nottingham heart attack register. *BMJ*, 315(7101), 159-64.
- [2] Folsom A R, Prineas R J, Kaye S A Soler J T (1989). Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women. *International journal of epidemiology*, 18(2), 361-7.
- [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. 557-60). IEEE.
- [4] Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CANFIS and genetic algorithm." *International Journal of Biological, Biomedical and Medical Sciences* 3.3 (2008).
- [5] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. *IEEE antennas and propagation magazine*, 58(5), 84-92.