# HEART DISEASE PREDICTION USING

# MACHINE LEARNING

### A PROJECT REPORT

#### Submitted by

MOHESHWARAN S 20BCB4022

PRIYADHARSHAN K 20BCB4030

RAAJ GANESH S S 20BCB4031

#### in partial fulfillment for the award of the degree

#### of

## **BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND BUSINESS SYSTEMS**

## **M KUMARASAMY COLLEGE OF ENGINEERING, KARUR**

## 

## **ANNA UNIVERSITY: CHENNAI 600 025**

##### **DEC-2022**

# HEART DISEASE PREDICTION USING

# MACHNIE LEARNING

### A PROJECT REPORT

***Submitted by***

**MOHESHWARAN S**

**PRIYADHARSAN K**

**RAAJ GANESH S S**

***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF TECHNOLOGY**

***in***

### COMPUTER SCIENCE AND BUSINESS SYSTEMS

**M KUMARASAMY COLLEGE OF ENGINEERING, KARUR**

## 

## **ANNA UNIVERSITY: CHENNAI 600 025**

### DEC-2022

.

###### **APPENDIX 2**

# ANNA UNIVERSITY: CHENNAI 600 025

## **BONAFIDE CERTIFICATE**

Certified that this project report **“HEART DISEASE PREDICTION USING MACHINE LEARNING”** is the bonafide work of **“MOHESHWARAN S (20BCB4022), PRIYADHARSHAN K (20BCB4030), RAAJ GANESH S S(20BCB4031)”** who carried out the project work under my supervision

Signature Signature

**Mrs. V. Banupriya/M.E.,(Ph.D).,**

**SUPERVISOR**

**Assistant Professor**

Department of Computer Science and Business Systems,

M. Kumarasamy College of Engineering, Thalavapalayam, Karur -639113

**Dr. R. RAJA GURU M.Tech, Ph.D.,**

**HEAD OF THE DEPARTMENT**

**Associate Professor**

Department of Computer Science and Business Systems,

M. Kumarasamy College of Engineering Thalavapalayam, Karur -639113.

**DEPARTMENT OF COMPUTER SCIENCE AND BUSINESS SYSTEMS VISION OF THE INSTITUTION**

* To emerge as a leader among the top institutions in the field of technical education

**MISSION OF THE INSTITUTION**

* Produce smart technocrats with empirical knowledge who can surmount the global challenges
* Create a diverse, fully-engaged, learner-centric campus environment to provide quality education to the students
* Maintain mutually beneficial partnerships with our alumni, industry, and Professional associations

**VISION OF THE DEPARTMENT**

* To produce competent industry relevant education, skillful research, technical and innovative computer science professionals acquaintance with managerial skills, human and social values.

**MISSION OF THE DEPARTMENT**

* To impart technical knowledge through innovative teaching, research, and consultancy.
* To develop and to promote student ability thereby to compete globally through excellence in education.
* To facilitate the development of academic-industry Collaboration.
* To produce competent engineers with professional ethics, technical competence and a spirit of innovation and managerial skills.

**PROGRAM EDUCATIONAL OBJECTIVES (PEOs)**

**PEO 1:** To acquire technical knowledge and proficiency required for the carrier advancement and higher education in the contemporary areas of computer science, business systems and various issues in the society.

**PEO 2**: To apply their competency in design and development of innovative solutions to adapt various emerging technological challenges for real world problems.

**PEO 3:** To demonstrate leadership qualities with high ethical standards and collaborated with other industries for the socio-economical growth of the country

**PROGRAM OUTCOMES (POs)**

Engineering students will be able to:

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriateconsideration for the public health and safety, and the cultural, societal, and environmentalconsiderations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, andsynthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilitiesrelevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineeringsolutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities andnorms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a memberand leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engagein independent and life-long learning in the broadest context of technological change.

**PROGRAM SPECIFIC OUTCOMES (PSOs)**

 **PSO1:** Ability to apply the analytical and business skills to provide sustainable solutions as an engineer/researcher for the real-time applications using Machine Learning, Internet of Things and Data analytics

 **PSO2:** Ability to practice ethical and human values with soft-skills qualities in computer science and business disciplines to emerge as an entrepreneur for the growth and development of the society.

# ABSTRACT WITH PO AND PSO MAPPING

|  |  |  |
| --- | --- | --- |
| **ABSTRACT** | **POs MAPPED** | **PSOs MAPPED** |
| Cardiovascular diseases are one of the life- | **PO1(3)**  **PO 2(3)**  **PO 3(2)**  **PO 4(2)**  **PO 5(2)**  **PO6(1)**  **PO 7(3)**  **PO 8(2)**  **PO 9(3)**  **PO 10(3)**  **PO 11(2)**  **PO 12(2)** |  |
| threatening diseases. There is a need for reliable | **PSO 1(3)** |
| and accurate system to diagnose the disease and | **PSO 2(2)** |
| to give proper treatment. Through the machine |  |
| learning algorithm and techniques, it helps to |  |
| diagnosis of heart related disease. Prediction |  |
| and diagnosis of heart disease become |  |
| challenging for doctors and hospitals. To reduce |  |
| the large scale of death, this system can be very |  |
| useful. As heart disease prediction is complex |  |
| task, there is a need to automate the prediction |  |
| process to avoid risk associated with it. A |  |
| complete genomic data analysis can easily be |  |
| done using machine learning models. |  |

Note: 1- Low, 2-Medium, 3- High

**SUPERVISOR HEAD OF THE DEPARTMENT**

# ABSTRACT

Heart diseases are one of the life-threatening diseases. There is a need for reliable and accurate system to diagnose the disease and to give proper treatment. Through the machine learning algorithm and techniques, it helps to diagnosis of heart related disease. Prediction and diagnosis of heart disease become challenging for doctors and hospitals. To reduce the large scale of death, this system can be very useful. As heart disease prediction is complex task, there is a need to automate the prediction process to avoid risk associated with it. A complete genomic data analysis can easily be done using Machine Learning models

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER** | **TITLE** | **PAGE** |
| **No.** |  | **No.** |
|  | **Abstract** | **vii** |
|  | **List of Figures** | **ix** |
|  | **Acronyms/List of Abbreviations** | **x** |
| **1** | **Introduction** | **1** |
|  | 1.1 Overview | 2 |
|  | 1.2 Domain Introduction | 3 |
|  | 1.3 Objective | 4 |
| **2** | **Literature Survey** | **5** |
| **3** | **Feasibility Study** | **11** |
| **4** | **Project Methodology** | **13** |
|  | 4.1 Block diagram of CDP | 13 |
|  | 4.2 Module Description | 14 |
| **5** | **Results and Discussion** | **15** |
| **6** | **Conclusion** | **19** |
| **7**  **8** | **References**    **Appendix** | **20**    **22** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **FIGURE No.** | **TITLE** | **PAGE No.** |
| 1.1 | Heart Diseases Prediction | 1 |
| 3.1 | Feasibility study for Heart | 7 |
|  | Disease Prediction |  |
| 4.1 | Block Diagram of CDP | 9 |
| 5 | Output Screenshot | 13 |

# LIST OF ACRONYMS / ABBREVIATIONS

|  |  |
| --- | --- |
| LS | Logistic Regression Algorithm |
| DT | Decision Tree Algorithm |
| RM | Random Forest Algorithm |
| SVM | Support Vector Machine Algorithm |
| KNN | K-Nearest Neighbors Algorithm |
| GB | Gradient Boosting Algorithm |

**CHAPTER 1 INTRODUCTION**

The Heart Diseases Prediction is used to effectively predict if the patient is suffering from any kind of heart diseases. It is used to determine if a patient is diagnosed with heart disease or not, which is a binary outcome. By using Random forest, outcome is predicted.



# Figure: 1.1 HEART DISEASE PREDICTION

* 1. **OVERVIEW**

The Heart is the primary circulatory organ that pumps the blood all over the body and it plays the most crucial role in the circulatory system. If the heart does not function properly then it will lead to serious health issues sometimes it may even lead to death. So, prior prediction of heart disease is essential to prevent the heart disease from its initial stage.

The Heart Disease Prediction is used to predict whether the person is affected by heart disease or not. It is predicted by using the medical symptoms and records of the person. Through the past data, the disease will be predicted. The system uses 14 medical parameters such as age, gender, blood pressure, cholesterol, and sugar level for prediction. The diagnosis of heart disease is a challenging task, which can offer automated prediction about the heart condition of patient so that further treatment can be made effective.

# DOMAIN INTRODUCTION

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence). Machine learning algorithms build a model based on sample data, known as [training data](https://en.wikipedia.org/wiki/Training_data), in order to make predictions or decisions without being explicitly programmed to do. Machine learning algorithms are used in a wide variety of applications, such as in medicine, [email filtering](https://en.wikipedia.org/wiki/Email_filtering), [speech recognition](https://en.wikipedia.org/wiki/Speech_recognition), and [computer vision](https://en.wikipedia.org/wiki/Computer_vision), where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to [computational statistics](https://en.wikipedia.org/wiki/Computational_statistics), which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of [mathematical optimization](https://en.wikipedia.org/wiki/Mathematical_optimization) delivers methods, theory and application domains to the field of machine learning. [Data](https://en.wikipedia.org/wiki/Data_mining) [mining](https://en.wikipedia.org/wiki/Data_mining) is a related field of study, focusing on [exploratory data](https://en.wikipedia.org/wiki/Exploratory_data_analysis) [analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis) through [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning). Some implementations of machine learning use data and [neural networks](https://en.wikipedia.org/wiki/Neural_networks) in a way that mimics the working of a [biological brain](https://en.wikipedia.org/wiki/Brain). In its application across business problems, machine learning is also referred to as [predictive analytics](https://en.wikipedia.org/wiki/Predictive_analytics).

Machine learning programs can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step

# OBJECTIVE

In recent years, Heart diseases are found to be a primary source of death in the world, they are also the ones that can be controlled and managed effectively. The whole accuracy in management of a disease lies on the proper time of detection.

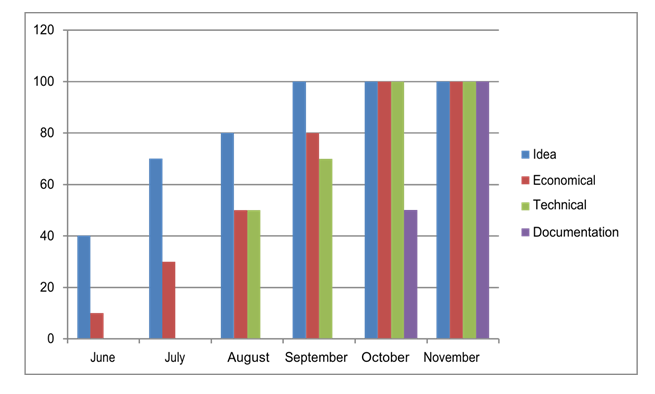
So, the main objective of our project is to predict whether the person is affected by heart disease or not. The goal of heart disease prediction is to determine if a patient should be diagnosed with heart disease or not, which is a binary outcome.

**CHAPTER 2**

# LITERATURE SURVEY

In medical field, predicting heart disease is treated as one of the intricate tasks. There is a necessity to develop a decision support system to forecast the cardio vascular disease in a patient. Machine learning plays a vital part in disease prediction. In this paper, various machine learning methods were used to predict the heart disease and their performances were compared. The results obtained show the superiority of the Random forest algorithm**.[1]** UCI Heart Disease dataset have used to test ML techniques along with conventional methods (i.e. random forest, support vector machine, K-nearest neighbor), as well as deep learning models (i.e. long short-term-memory and gated-recurrent unit neural networks). To improve the accuracy of weak algorithms we explore voting based model by combining multiple classifiers. The strength of the proposed ensemble approach such as voting based model is compelling in improving the prognosis accuracy of anemic classifiers and established adequate achievement in analyze risk of heart disease.**[2]**The proposed work predicts the probabilities of heart condition and classifies patient's risk level by implementing different data processing techniques like Naive Bayes, Decision Tree, Logistic Regression and Random Forest. Thus, this paper presents a comparative study by analysing the performance of various machine learning algorithms. The trial result verifies that Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented**.[3]**The main aim is to analyze the machine learning algorithms based on the percentage of various performance metrics (such as, Accuracy, Precision and Recall). The machine learning methodology is proposed. The most suitable algorithm for each metrics is predicted. It is analyzed using the specific variables in the dataset by using the python programming as well as different supervised machine learning algorithms which include, Decision Tree, Logistic Regression, KNN and Random Forest**.[4]** In the proposed method, anomalous changes in multiple HRV parameters are monitored by means of a convulsion prediction model, and convulsion alarms are issued when abnormal changes in HRV are detected. The convulsion prediction model is constructed based on multivariate statistical process control (MSPC), a well-known anomaly detection algorithm in machine learning. In this study, HRV data were collected from four cynomolgus monkeys administered with multiple doses of pentylenetetrazol (PTZ) and picrotoxin (PTX), which are GABA receptor antagonists, as convulsant agents. **[5]** Designed an automated, population-specific ASCVD risk calculator using machine-learning (ML) methods and electronic medical record (EMR) data, and compared its predictive power with that of the PCE calculator. Prediction models built using ML techniques improved ASCVD prediction and reduced the number of screenings required to predict ASCVD when compared with the PCE calculator, alone. Combining LT and CS features in the ML models significantly improved ASCVD prediction compared with using CS features, alone**.[6]**Compared and reported the various Classification, Data Mining, Machine Learning, Deep Learning models that are used for prediction of the Cardio-Vascular diseases. The survey is organized as threefold: Classification and Data Mining Techniques for CVD, Machine Learning Models for CVD and Deep Learning Models for CVD prediction. The performance metrics used for reporting the accuracy, the dataset used for prediction and classification, and the tools used for each category of these techniques are also compiled and reported in this survey**.[7]**Using various statistical techniques and principal component analysis, we identify the most important factors for stroke prediction. We conclude that age, heart disease, average glucose level, and hypertension are the most important factors for detecting stroke in patients. Furthermore, a perceptron neural network using these four attributes provides the highest accuracy rate and lowest miss rate compared to using all available input features and other benchmarking algorithms. As the dataset is highly imbalanced concerning the occurrence of stroke, we report our results on a balanced dataset created via sub-sampling techniques**.[8]** The different features and many well-known classification methods are used to implement the prediction model to predict the heart disease and diabetes. The proposed method utilizes ensemble approach for achieving a higher degree of accuracy rates for by using classification algorithms and feature selection methods. The proposed method implements voting classifier that has sigmoid SVC, AdaBoost, and Decision tree algorithms.Also implements the traditional classifiers and presents the comparison of different models in terms of accuracy. The web application is also developed for users to avail its services very easily and make it convenient for their use, particularly in the prediction of heart and diabetes collectively**.[9]** Huge amount of patient related data is maintained on monthly basis. The stored data can be useful for source of predicting the occurrence of future disease. Some of the data mining and machine learning techniques are used to predict the heart disease, such as Artificial Neural Network (ANN), Decision tree, Fuzzy Logic, K-Nearest Neighbour(KNN), Naïve Bayes and Support Vector Machine (SVM). This paper provides an insight of the existing algorithm and it gives an overall summary of the existing work**.[10]** In this study, the ML-based risk stratification tool was able to accurately assess and stratify the risk of 3-year all-cause mortality in patients with HF caused by CHD. A combination of ML and SHAP could provide an explicit explanation of individualized risk prediction, allowing physicians to intuitively understand the influence of key features in the model, thus helping clinicians better understand the decision-making process for disease severity assessment **[11].** A ML model developed using data from the UNOS database showed satisfactory predictive accuracy for 1-year mortality after HT in adults with congenital heart disease. Explainability analysis helps interpret the results in a clinical manner [**12].** The use of ML algorithms when working with small databases (around 4000 participants) is relatively simple. These algorithms can enhance the diagnostic and prognostic capacity of more traditional regression techniques. R-Studio is a powerful tool for conducting complex ML analytics with high reliability in creating a record of all changes. Mixed research teams, comprising healthcare professionals and computer scientists or mathematicians, are optimal for the conceptualization and development of ML projects **[13].** It works with categorical variables along with that it will break particular categorical columns into dummy columns with 1s and 0s. Apart from this, this part also mentioned that the data output of this application is used for many medical parameters such as age, gender, blood pressure, cholesterol, and obesity for prediction and requirement of the software used in the development application **[14].** The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of KNN, Logistic Regression and Random Forest for predicting heart disease using UCI machine learning repository dataset. The result of this study indicates that the Logistic regression algorithm is the most efficient algorithm with accuracy score of 89% for prediction of heart disease **[15].** A Decision tree approach is used to predict the disease. The data in the dataset is pre-processed to make it suitable for classification. This paper proposes a scalable system for heart disease monitoring system. This system focuses on applying real-time classification model on heart disease attributes for continuous monitoring of the patient’s health **[16].** In this project, the techniques are Random Forest and Logistic Regression: we have analyzed that the Random Forest has better accuracy as compared to Logistic Regression. Our purpose is to improve the performance of the Random Forest by removing unnecessary and irrelevant attributes from the dataset and only picking those that are most informative for the classification task **[17].** In this, HTML, CSS and Django framework of Python were used to build interactive web application. And this web application with machine learning algorithms forms a robust model to predict a heart disease. This would help the end users get a preliminary prediction about the condition of their heart **[18].** The algorithms used in building the given model are Logistic regression, Random Forest Classifier and KNN. The accuracy of our model is 87.5%. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not. By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much **[19].** The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in  high risk patients and in turn reduce the complications, which can be a great milestone in the field  of medicine. This project resolved the feature selection i.e. backward elimination and RFECV  behind the models and successfully predict the heart disease, with 85% accuracy. The model used  was Logistic Regression. Further for its enhancement, we can train on models and predict the  types of cardiovascular diseases providing recommendations to the users, and also use more  enhanced models **[20].**This system can help medical practitioner in efficient decision making based on the given parameter. We have train and test the system using 10 fold method and find the accuracy of 86.3 % in testing phase and 87.3 % in training phase and because this model demonstrates the better results and helps the area specialists and even individual related with the field to get ready for a superior determine and give the patient to have early determination results as it performs sensibly well even without retraining**.[21]** A remarkable performance was achieved by the SMO classifier using the chi-squared attribute evaluation method. Eventually, we noticed that there was a significant improvement in the prediction performance with appropriate attribute selection and tuning the hyper parameters of the classifiers. Although the performance of the classifiers looks satisfactory, a smaller dataset of 303 instances, 10 machine learning classifiers, and 3 feature selection methods were used in this research. There is a huge scope to explore various machine learning algorithms and feature selection techniques. In the future, we intend to combine multiple datasets to obtain a higher number of observations and conduct more experiments by selecting appropriate attributes to improve the classifier’s predictive performance**.[22]** The main goal of this research study is to observe the impact of feature selection techniques on the performance of ML models. This analysis was performed for CVD and Framingham heart disease datasets which are available online. In this research, first, we performed a data preprocessing step in which data transformation, cleansing and balancing were involved. Secondly, we used a filter-based feature selection technique namely the ANOVA-F test to identify the most important features from the datasets for an effective heart disease prediction. Using the ANOVA-F test most relevant features with outcomes from both datasets were identified using the individual feature scores. We observed that features like age, hypertension, glucose, previous heart disease, and blood pressure were found to reflect the most important risk factors for heart disease except the traditional factors using both datasets. Furthermore, the classification experiments were performed with full as well as the reduced feature sets to analyze the effect of selected features on the prediction accuracy of various ML prediction models. Using the full feature set the highest accuracy achieved was 0.73 for CVD and 0.66 for the Framingham heart disease dataset. After using the reduced feature set the accuracy increased to 0.75 and 0.71 for both datasets**.[23]** The purpose of the experimental studies was to find suitable Classifiers using Python-based Deep Learning Neural Networks and a few selected Machine Learning Algorithms using verified normal and abnormal heart sounds of five classes as described in the text. Classifier Accuracy, Precision, Recall, and F1-Scores are evaluated with the same heart sound database in this entire experiment. It has been observed that the Random Forest Classifier can be chosen as the best Machine Learning Classifier though it has some complexity when the size of training data becomes very large. CNN-based modified deep learning exception network is most suitable as an AI classifier of Heart sound though it takes much time in the learning phase compared to machine learning. However, once learned, it gives a speedy result. In the subsequent research, the objective will be to use CNN based deep learning model and random forest classifier to develop a low-cost IoT-enabled Vaccular Heart disease screening system for rural use**.[24]** . Based on the dataset used, the implementation of 20-fold Cross-validation and 10-fold Cross-validation with DT and SVM reported the same accuracy of the implementation of 5-fold Cross-validation with SGD, which is 87.69%. The early detection of health problems may reduce the significant cause of death related to Ambient Assisted Living. Moreover, the existence of an automatic method of detecting heart problems guarantees the development of a low-cost method for the detection of heart disease. Next, the use of this type of sensors with machine learning methods makes more critical the research in this area. The accuracy values give high strength to this study and increase the speed and capacity of the sensor electrocardiogram. With the use of automatic detection methods, it speeds up the detection and ability to generalize its use by different fringes of society. It also supports health professionals in the early detection of symptoms. With the combination of these types of methods, we can change and help in the development of systems. These systems may be diverse, including a pacemaker with even more capacity, and devices to assist people with identified problems**.[25]** Due to the nature of the raw data collected with a sampling frequency of 2000 kHz, the data were first reviewed and preprocessed under the supervision of a specialist. The analysis was made by extracting various heart sound information using statistical, discrete wavelet, and information theory features. Then we reduced the dimension of the dataset obtained by PCA, LDA, and GA by selecting appropriate features and classifying the sound using classification algorithms such as GBC, RFC, and SVC. The analysis and the codes related to the feature extractions and GA were written in MATLAB. We also utilized Python packages (Pandas, Numpy, Matplotlib, Scikitlearn, Scipy, etc.) for all the classification algorithms and data analysis. We also provided extensive comparative analysis to come up with the best technique in our experiments. The comparisons were made by performing all the processes and procedures on sets of preprocessed heart data**.[26]** The accuracy of the proposed models may vary and it depends on the quality of dataset used, tool used by various researchers, the number of attributes and records in the dataset along with the preprocessing techniques used in the model. It depends on whether it is a hybrid model or not and whether the model make use of feature selection or not. From comparison table, we can conclude that the researcher who produced the highest accuracy was Malav that uses Hybrid approach with combining K-means clustering algorithm and ANN by making use of WEKA tool and dataset was taken from Cleveland UCI repository. The dataset must be preprocessed for getting good results. Also, a suitable algorithm must be used when developing a prediction model**.[27]** This work will be useful in identifying the possible patients who may suffer from heart disease in the next 10 years. This may help in taking preventive measures and hence try to avoid the possibility of heart disease for the patient. So the doctors can closely analyze when a patient is predicted as positive for heart disease, then the medical data for the patient. An example would be - suppose the patient has diabetes that may be the cause for heart disease in future and then the patient can be given treatment to have diabetes in control, which in turn may prevent the heart disease.**[28]** With the increasing number of deaths thanks to heart diseases, it's become mandatory to develop a system to predict heart diseases effectively and accurately. The motivation for the study was to seek out the foremost efficient ML algorithm for detection of heart diseases. This study gives a brief description about the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart condition using the dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of more than 90% for prediction of heart disease**.[29]** Heart disease prediction which uses Machine learning algorithm provides users a prediction result if the user has heart disease. Recent advancements in technology made machine learning algorithms to evolve. In this proposed method Random Forest Algorithm was used because of its efficiency and accuracy. This algorithm is also used to find the heart disease prediction percentage by knowing the correlation details between diabetes and heart diseases.**[30]**

**CHAPTER 3 FEASIBILITY STUDY**

****

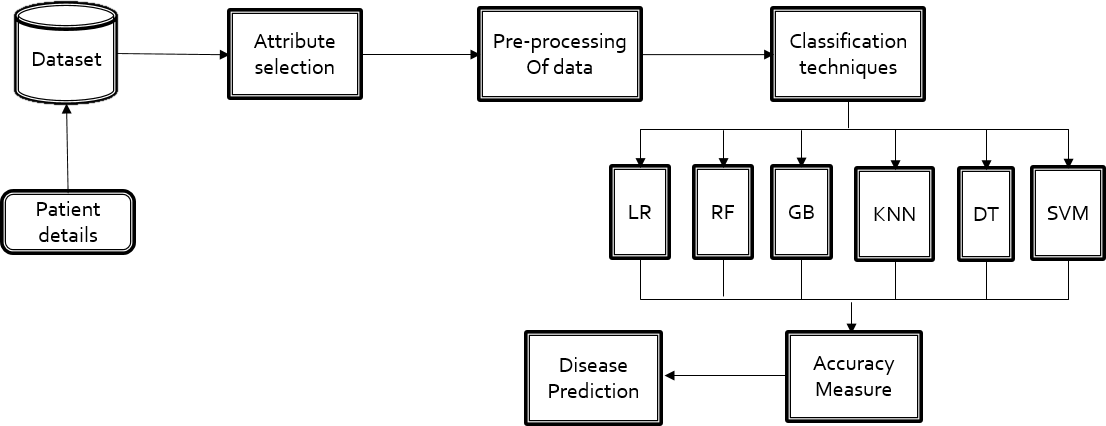
# Figure 3.1 Feasibility Study for Heart Disease Prediction

* 1. Idea: We have planned to design the Heart Disease Prediction by using Machine learning algorithms in Python.
  2. Economic Feasibility This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into research and development of the system is limited, the expenditure must be justified.
  3. Technical Feasibility: This is carried out to check the technical feasibility that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources the developed system must have a modest requirement and are required for implementing this system.
  4. Operational Feasibility: The system working is quite easy to use and learn due to its simple and attractive interface. User requires no special training for operating the system.
  5. Social Feasibility: The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must be threatened by the system. His level of confidence must be increased so that he is able to make some constructive criticism which is welcomed.
  6. Documentation: The documentation is completed after getting approval of

superior.

### CHAPTER 4

**PROJECT METHODOLOGY**



### Figure: 4.1 BLOCK DIAGRAM OF HEART DISEASE PREDICTION

This block diagram contains the details about Heart Disease Prediction. The dataset contains the patient details. Attribute selection is the first step in which the best attribute is selected. Then Pre-Processing of the dataset is done. Then the accuracy of the algorithms is checked using the classification technique. The algorithm with high accuracy rate is used in the Heart Disease Prediction. Finally, we can easily predict whether the person have heart disease or not.

# Module Description

### Attribute selection:

* Attribute or Feature selection includes the selection of appropriate attributes for the prediction system.
* This is used to increase the efficiency of the system.
* Various attributes of the patient like gender, age, cerebral palsy, cholesterol, restecg etc... are selected for the prediction.

### Analysis of Algorithm:

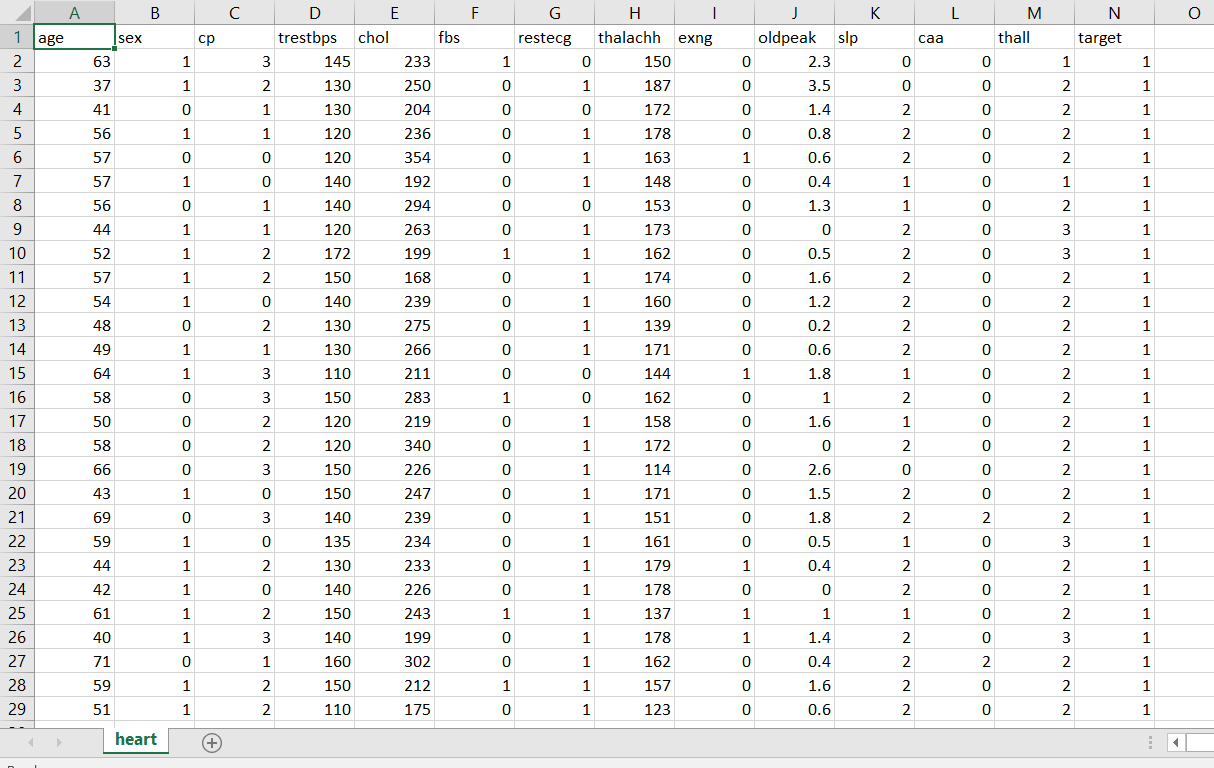
* Various machine learning algorithms like Support Vector Machine, Linear Regression, Decision Tree, Random Forest, Logistic Regression and Gradient Boosting are used for classification.
* By comparative analysis which is performed among algorithms and concluded that Logistic Regression gives more accuracy.

### Heart Disease Prediction:

* After the analysis of algorithm, the next module is Heart Disease Prediction.
* In this module, it is predicted whether the person any kind of have heart disease.

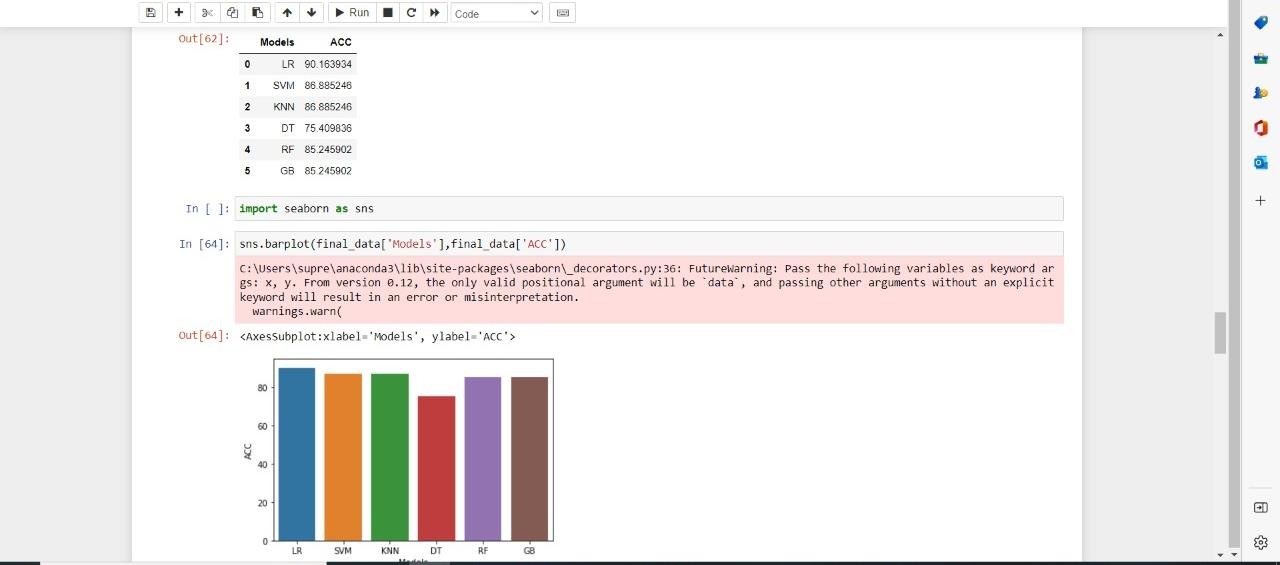
### CHAPTER 5

**RESULTS AND DISCUSSION**



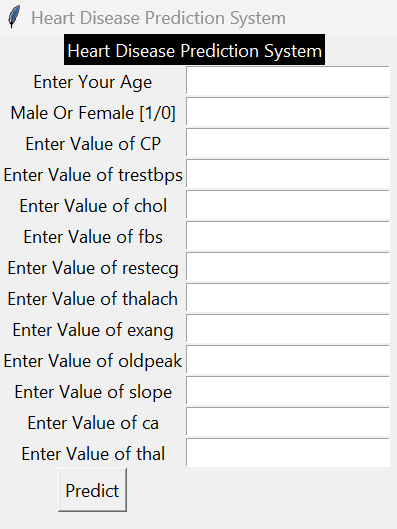
# Figure 5.1 Screenshot of Dataset

This figure deals with the dataset that contains the medical report of the person.

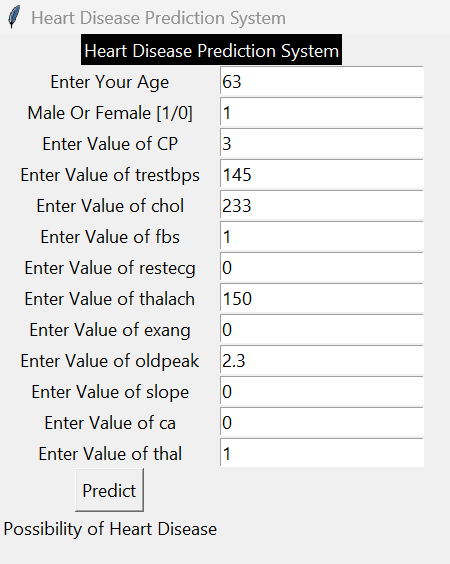


# Figure 5.2 Screenshot of Accuracy Prediction Module

This figure shows the accuracy prediction of the algorithms. The algorithm with high accuracy is chosen.



# Figure 5.3 Screenshot of Final Module



**Figure 5.4 Screenshot of Heart Disease Prediction**

### CHAPTER 6

**CONCLUSION AND SCOPE FOR FUTURE WORKS**

Heart disease is one of the major concerns for society today. It is difficult to manually determine the odds of getting heart disease based on risk factors. However, Machine Learning techniques are useful to predict the output from existing data. So the “Heart Disease Prediction” is used to predict the possibility of having heart disease.

### REFERENCES

**[1]** Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012072). IOP Publishing.

**[2]** Javid, I., Alsaedi, A. K. Z., & Ghazali, R. (2020). Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method. *International Journal of Advanced Computer Science and Applications*, *11*(3).

**[3]** Krishnan, S., & Geetha, S. (2019, April). Prediction of Heart Disease Using Machine Learning Algorithms. In *2019 1st international conference on innovations in information and communication technology (ICIICT)* (pp. 1-5). IEEE.

**[4]** Sanni, R. R., & Guruprasad, H. S. (2021). Analysis of performance metrics of heart failured patients using python and machine learning algorithms. *Global transitions proceedings*, *2*(2), 233-237.

**[5]** Nagata, S., Fujiwara, K., Kuga, K., & Ozaki, H. (2021). Prediction of GABA receptor antagonist-induced convulsion in cynomolgus monkeys by combining machine learning and heart rate variability analysis. *Journal of Pharmacological and Toxicological Methods*, *112*, 107127.

**[6]** Li, Q., Campan, A., Ren, A., & Eid, W. E. (2022). Automating and improving cardiovascular disease prediction using Machine learning and EMR data features from a regional healthcare system. *International Journal of Medical Informatics*, *163*, 104786.

**[7]** Swathy, M., & Saruladha, K. (2022). A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques. *ICT Express*, *8*(1), 109-116.

**[8]** Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics*, *2*, 100032.

**[9]** Dhande, B., Bamble, K., Chavan, S., & Maktum, T. (2022). Diabetes & Heart Disease Prediction Using Machine Learning. In *ITM Web of Conferences* (Vol. 44, p. 03057). EDP Sciences.

**[10]** Marimuthu, M., Abinaya, M., Hariesh, K. S., Madhankumar, K., & Pavithra, V. (2018). A review on heart disease prediction using machine learning and data analytics approach. *International Journal of Computer Applications*, *181*(18), 20-25.

**[11]** Wang, K., Tian, J., Zheng, C., Yang, H., Ren, J., Liu, Y., ... & Zhang, Y. (2021). Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Computers in Biology and Medicine*, *137*, 104813.

**[12]** Emfietzoglou, M., Siouras, A., Van den Eynde, J., Moustakidis, S., Doulamis, I., Giannakoulas, G., ... & Kampaktsis, P. (2022). A MACHINE LEARNING MODEL FOR THE PREDICTION OF 1-YEAR MORTALITY AFTER HEART TRANSPLANTATION IN ADULTS WITH CONGENITAL HEART DISEASE. *Journal of the American College of Cardiology*, *79*(9\_Supplement), 507-507.

**[13]** Beunza, J. J., Puertas, E., García-Ovejero, E., Villalba, G., Condes, E., Koleva, G., ... & Landecho, M. F. (2019). Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease). *Journal of biomedical informatics*, *97*, 103257.

**[14]** Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, *2*, 100016.

**[15]** Krishnan, S., & Geetha, S. (2019, April). Prediction of Heart Disease Using Machine Learning Algorithms. In *2019 1st international conference on innovations in information and communication technology (ICIICT)* (pp. 1-5). IEEE.

**[16]** Asabe, M., Shilwant, S., Dolare, N., Chorghade, S., & Pathak, K. R. (2020). Heart Attack Prediction and Analysis System Using Decision Tree Algorithm. *Advancement in Image Processing and Pattern Recognition*, *3*(2).

**[17]** Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In *2020 international conference on electrical and electronics engineering (ICE3)* (pp. 452-457). IEEE.

**[18]** Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, P. (2020). Heart disease prediction using machine learning. *International Journal of Research and Technology*, *9*(04), 659-662.

**[19]** Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012072). IOP Publishing.

**[20]** 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.

**[21]** Saxena, K., & Sharma, R. (2016). Efficient heart disease prediction system. *Procedia Computer Science*, *85*, 962-969.

**[22]** Reddy, K. V. V., Elamvazuthi, I., Aziz, A. A., Paramasivam, S., Chua, H. N., & Pranavanand, S. (2021). Heart disease risk prediction using machine learning classifiers with attribute evaluators. *Applied Sciences*, *11*(18), 8352.

**[23]** Pathan, M. S., Nag, A., Pathan, M. M., & Dev, S. (2022). Analyzing the impact of feature selection on the accuracy of heart disease prediction. *Healthcare Analytics*, *2*, 100060.

**[24]** Roy, T. S., Roy, J. K., & Mandal, N. (2022). Classifier identification using Deep Learning and Machine Learning Algorithms for the detection of Valvular Heart diseases. *Biomedical Engineering Advances*, 100035.

**[25]** Pires, I. M., Marques, G., Garcia, N. M., & Ponciano, V. (2020). Machine learning for the evaluation of the presence of heart disease. *Procedia Computer Science*, *177*, 432-437.

**[26]** Zeinali, Y., & Niaki, S. T. A. (2022). Heart sound classification using signal processing and machine learning algorithms. *Machine Learning with Applications*, *7*, 100206.

**[27]** Rubini, P. E., Subasini, C. A., Katharine, A. V., Kumaresan, V., Kumar, S. G., & Nithya, T. M. (2021). A cardiovascular disease prediction using machine learning algorithms. *Annals of the Romanian Society for Cell Biology*, 904-912.

**[28]** Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology*, *7*(2.8), 684-687.

**[29]** Jeyaganesan, J., Sathiya, A., Keerthana, S., & Aiyer, A. (2020). Diagnosis And Prediction Of Heart Disease Using Machine Learning Techniques. *Ilkogretim Online-Elementary Education Online*, *19*(2), 1817-1827.

**[30]** Omar, S., Mohamed, N., & Elbendary, N. (2021, August). A Cardiovascular Disease Prediction Using Machine Learning Algorithms. In *The International Undergraduate Research Conference* (Vol. 5, No. 5, pp. 177-179). The Military Technical College.

### APPENDIX

import pandas as pd data=pd.read\_csv("heart.csv") data.isnull().sum() data\_dup=data.duplicated().any() data\_dup data=data.drop\_duplicates() data\_dup=data.duplicated().any() data\_dup

cate\_val=[] cont\_val=[]

for column in data.columns:

if data[column].nunique()<=10: cate\_val.append(column)

else:

cont\_val.append(column) cate\_val

cont\_val cate\_val

data['cp'].unique() cate\_val.remove('sex') cate\_val.remove('target')

data=pd.get\_dummies(data,columns=cate\_val,drop\_first=True) data.head()

from sklearn.preprocessing import StandardScaler st=StandardScaler() data[cont\_val]=st.fit\_transform(data[cont\_val]) data.head()

X=data.drop('target',axis=1)

y=data['target']

from sklearn.model\_selection import train\_test\_split X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.2,random\_sta te=42)

from sklearn.linear\_model import LogisticRegression log=LogisticRegression()

log.fit(X\_train,y\_train) y\_predl=log.predict(X\_test)

from sklearn.metrics import accuracy\_score accuracy\_score(y\_test,y\_predl)

from sklearn import svm svm=svm.SVC() svm.fit(X\_train,y\_train) y\_pred2=svm.predict(X\_test) accuracy\_score(y\_test,y\_pred2)

from sklearn.neighbors import KNeighborsClassifier knn=KNeighborsClassifier()

knn.fit(X\_train,y\_train) y\_pred3=knn.predict(X\_test) accuracy\_score(y\_test,y\_pred3) score=[]

for k in range(1,40): knn=KNeighborsClassifier(n\_neighbors=k) knn.fit(X\_train,y\_train) y\_pred=knn.predict(X\_test) score.append(accuracy\_score(y\_test,y\_pred))

import matplotlib.pyplot as plt plt.plot(score)

plt.xlabel("K Value") plt.ylabel("Acc")

plt.show() knn=KNeighborsClassifier(n\_neighbors=2) knn.fit(X\_train,y\_train) y\_pred=knn.predict(X\_test) accuracy\_score(y\_test,y\_pred) data=pd.read\_csv('heart.csv') data=data.drop\_duplicates() X=data.drop('target',axis=1) y=data['target']

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.2,random\_sta te=42)

from sklearn.tree import DecisionTreeClassifier dt=DecisionTreeClassifier() dt.fit(X\_train,y\_train) y\_pred4=dt.predict(X\_test) accuracy\_score(y\_test,y\_pred4)

from sklearn.ensemble import RandomForestClassifier rf=RandomForestClassifier()

rf.fit(X\_train,y\_train) y\_pred5=rf.predict(X\_test) accuracy\_score(y\_test,y\_pred5)

from sklearn.ensemble import GradientBoostingClassifier gbc=GradientBoostingClassifier() gbc.fit(X\_train,y\_train)

y\_pred6=gbc.predict(X\_test) accuracy\_score(y\_test,y\_pred6)

final\_data=pd.DataFrame({'Models':['LR','SVM','KNN','DT','RF','GB'],'AC C':[accuracy\_score(y\_test,y\_predl)\*100,accuracy\_score(y\_test,y\_pred2)\*10 0,accuracy\_score(y\_test,y\_pred3)\*100,accuracy\_score(y\_test,y\_pred4)\*100

,accuracy\_score(y\_test,y\_pred5)\*100,accuracy\_score(y\_test,y\_pred6)\*100]

})

final\_data

import seaborn as sns sns.barplot(final\_data['Models'],final\_data['ACC']) X=data.drop('target',axis=1)

y=data['target']

from sklearn.linear\_model import RandomfoForestClassifier

rf= RandomfoForestClassifier()

rf.fit(X\_train,y\_train) import pandas as pd new\_data=pd.DataFrame(

{

'age':52,

'sex':1, 'trestbps':125, 'chol':212,

'fbs':0,

'restecg':1, 'thalach':168, 'exang':0,

'oldpeak':1.0,

'slope':2,

'ca':2,

'thal':3, 'target':2

},index=[0]) new\_data

p=rf.predict(new\_data) if p[0]==0:

print("No Disease")

else:

print("Disease") import joblib

joblib.dump(rf,'model\_joblib\_heart') model=joblib.load('model\_joblib\_heart') model.predict(new\_data)

from tkinter import \* import joblib

def show\_entry\_fields(): p1=int(e1.get())

p2=int(e2.get())

p3=int(e3.get())

p4=int(e4.get())

p5=int(e5.get())

p6=int(e6.get())

p7=int(e7.get())

p8=int(e8.get())

p9=int(e9.get()) p10=float(e10.get()) p11=int(e11.get()) p12=int(e12.get()) p13=int(e13.get())

model = joblib.load('model\_joblib\_heart') result=model.predict([[p1,p2,p3,p4,p5,p6,p7,p8,p8,p10,p11,p12,p13]])

if result == 0:

Label(master, text="No Heart Disease",fg="red").grid(row=31) else:

Label(master, text="Possibility of Heart Disease",fg="red").grid(row=31)

master = Tk()

master.title("Heart Disease Prediction System")

label = Label(master, text = "Cardium Disease Prediction System"

, bg = "black", fg = "white"). \ grid(row=0,columnspan=2)

Label(master, text="Age").grid(row=1)

Label(master, text="Male Or Female [1/0]").grid(row=2) Label(master, text="ChestPain[0-3]").grid(row=3) Label(master, text="Bloodpressure").grid(row=4) Label(master, text="Cholesterol").grid(row=5) Label(master, text="Bloodsugar[0/1]").grid(row=6) Label(master, text="Electrocardiograph[0/1]").grid(row=7) Label(master, text="Max heartrate").grid(row=8)

Label(master, text="Exercise induced angina[0/1]").grid(row=9) Label(master, text="Distance of SD").grid(row=10) Label(master, text="ST elevation[0-2]").grid(row=11) Label(master, text="Major vessels[0-3]").grid(row=12) Label(master, text="Pain after nitrate tablet").grid(row=13)

e1 = Entry(master) e2 = Entry(master) e3 = Entry(master) e4 = Entry(master) e5 = Entry(master)

e6 = Entry(master) e7 = Entry(master) e8 = Entry(master) e9 = Entry(master) e10 = Entry(master) e11 = Entry(master) e12 = Entry(master) e13 = Entry(master)

e1.grid(row=1, column=1) e2.grid(row=2, column=1) e3.grid(row=3, column=1) e4.grid(row=4, column=1) e5.grid(row=5, column=1) e6.grid(row=6, column=1) e7.grid(row=7, column=1) e8.grid(row=8, column=1) e9.grid(row=9, column=1) e10.grid(row=10, column=1) e11.grid(row=11, column=1) e12.grid(row=12, column=1) e13.grid(row=13, column=1)

Button(master, text='Predict', command=show\_entry\_fields).grid() mainloop()