

# Heart Disease Prediction using Machine Learning Algorithm

Ravi Kumar Singh, Dr. A Rengarajan

Department of Master of Computer Applications, Jain Deemed to be University, Bengaluru, Karnataka, India

## ABSTRACT

Nowadays, Heart disease has become dangerous to a human being, it effects very badly to human body. If anyone is suffering from heart disease, then it leads to blood clotting. Heart disease prediction is very difficult task to predict in the field of medical science. Affiliation has predicted that 12 million people fail horrendously every year as a result of heart disease. In this paper, we propose a k-Nearest Neighbors Algorithm (KNN) way to deal with improve the exactness of heart determination. We show that k-Nearest Neighbors Algorithm (KNN) have better accuracy than random forest algorithm for viewing heart disease. The k-Nearest Neighbors Algorithm give more precise and exact outcome . We have taken 13 attributes in the dataset and a target attribute, by applying machine learning we achieved 84% accuracy in the heart disease detection.

**KEYWORDS:** Machine Learning, k-Nearest Neighbors classifier, Decision Tree classifier, Random Forest Classifier, Jupyter

**How to cite this paper:** Ravi Kumar Singh | Dr. A Rengarajan "Heart Disease Prediction using Machine Learning Algorithm" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-5 | Issue-2, February 2021, pp.183-187, URL: [www.ijtsrd.com/papers/ijtsrd38358.pdf](http://www.ijtsrd.com/papers/ijtsrd38358.pdf)



IJTSRD38358

Copyright © 2021 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



## INTRODUCTION

Heart disease forecast is one of the most notable point in the machine learning field for expectation. It clusters the blood to all aspects of the body. If the blood not siphons to every part of the body, at that point the brain and different organ will stop work and the person may die. It is hard to recognize heart disease on account of few factors, for example, diabetes, hypertension, high cholesterol, heart beat rate and various other factors. As per World Health Organization heart related disease are liable for taking 17.7 million lives every year, 31% of all over worldwide. In India, heart disease has become the main source of mortality. Heart disease has killed 1.7 million Indian in 2016, as indicated by the 2016 worldwide weight of infection report.

In clinical science coronary illness is one of the huge challenges, because a lot of parameter and technicality is involved for predicting this disease. Machine learning could be a superior decision for accomplishing high precision for heart disease as well as another disease and its diverse information types under different condition for predicting the heart disease calculation, for example, Naive Bayes, Decision Tree, KNN, Neural Network are utilized to predict risk of heart algorithm and its speciality such as Naive Bayes is utilized for predicating heart disease, while Decision Tree is utilized to give ordered report to the heart disease, though the Neural Network give chances to limit the mistake for predication of heart disease. All these procedures are utilized in old patient record for getting expectation about new patient. The expectation for heart disease encourages doctor to predict heart disease in early stage so that he can save millions of lives.

This overview paper is committed for a review in the field of machine learning technique in heart disease. Later aspects of the overview paper will discuss about different machine learning calculation for heart disease and their comparison on different parameters. It also shows future outline of machine learning calculation in heart disease. This paper gives a profound analysis in the field of predicting heart disease.

## RELATED WORKS

Heart is one of the main organ of human body, it plays vital function of blood siphoning in human body which is as fundamental as the oxygen of human body so there is consistently need of insurance of it, this is one of the main explanation behind the analysts to work on it. So there are number of specialists dealing with it. There is consistently need of examination of heart related things either analysis or expectation or you can say that assurance of heart disease. There are different fields like artificial intelligence, machine learning, data mining that contributed on this work. Here, we will discuss some of them.

Some of the analysts have taken a shot of information about the expectation of heart disease. Kaur et al. have worked on this and characterize how the interesting pattern and information are gotten from a huge dataset. They perform exactness correlation on different machine learning and information mining 453 methodologies for discovering which one is best among at that point and get the outcome on the kindness of SVM.

Zhao et al. (2017) built up a framework for heart disease classification by utilizing two datasets, one from Shanghai Shuguang Hospital and another in UCI coronary disease dataset. The model uses support Vector Machine calculation alongside PCA, CCA and DMPCCA which are utilized for include extraction and combination. The general investigation come about that DMPCCA gave the best outcome.

Ganesan et al. (2019) utilize IOT innovation for expectation and conclusion of heart disease by taking UCI dataset and applied J48 classifier, Logistic Regression, Multiplayer Perception, and SVM utilizing Java on Amazon cloud. In this examination J48 gives 91.48%, SVM gave 84.07%, LR gave

83.07%, and MPL gave 78.14% exactness and inferred that J48 beats every other calculation.

### PROJECT SCOPE AND OBJECTIVES

The primary goal of this examination is to develop a heart forecast framework. The system can find information related with heart disease from the historical heart data set to implement the classifier that classifies the disease according to the contribution of the client and reduce the cost of the medical test. The scope of the project is to execute machine learning calculation to bigger dataset helps to improve the accuracy of results. Utilizing of machine learning procedure gives more exact outcomes than more experienced doctor. By this clinical choice with computer-based patient record could decrease medical error and improve patient result.

### Literature survey

SI. no	Authors	Year	Description
1	Palaniappan and Awang	2008	The authors proposed to develop a model Intelligent Heart Disease Prediction System (IHDP) utilizing information mining procedures to be specific Naive Bayes, Decision Tree, and Neural Network.
2	Bhatla and Jyoti	2012	The authors proposed that neural network was best survey in information mining methods to anticipate heart disease.
3	Chaurasia and Pal	2013	The creators proposed three mainstream information mining calculation CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and Decision Table (DT) separated from a choice tree to foresee heart disease.
4	Boshra Brahmi et al.	2015	The authors proposed to utilize diverse characterization procedures in coronary illness determination like J48 Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes (NB) and SMO to classify dataset.
5	K. Vembandasamy et al.	2015	The authors proposed Naive Bayes algorithm in data mining technique which serves diagnosis of heart disease patient.
6	S. Seema et al.	2016	The authors propose an efficient mechanism to predict heart disease by mining the data from health record.
7	K. Gomathi et al.	2016	The authors proposed to analysis information mining methods to foresee various kinds of sicknesses like heart disease, diabetes and bosom disease and so on.
8	Ayon Dey et al.	2016	The authors proposed of this examination is to dissect directed AI calculation to anticipate heart disease.

### Requirement Analysis

#### Tools

##### Anaconda

Anaconda is an open-source appropriation for python and R programming language. It is utilized for information science, machine learning, profound learning, and so on. With the availability of more than 300 libraries for information science, it turns out to be genuinely ideal for any developer to work on anaconda for information science. Anaconda helps in improved bundle the board and sending. Anaconda accompanies the wide assortment variety of tools to effectively gather information from different source using different machine learning and machine learning calculations. It is developed and maintained by Anaconda.inc., which was developed by Peter Wang and Travis Oilphant in 2012.

#### Hardware Requirements

- Operating System: Windows 10
- Processor: Intel(R)Pentium(R) CPU N3710 @1.60GHz 1.60GHz
- System Type: 64-bit operating system, x64-based processor
- Installed Ram: 4.00 GB

### Software Requirements

#### Jupyter Notebook

The Jupyter Notebook is an open-source web application that permits you to make and offer chronicles that contain live code, condition, perceptions and story text. Utilization include: information cleaning and transformation, mathematical simulation, measurable displaying, information representation, machine learning, and significantly more.

#### Python

Python is a universally useful deciphered, intelligent, object-arranged and elevated level programming language. It was developed by Guido van Rossum during 1985-1990. Like Perl, python source code is additionally accessible under the GNU General Public License (GNL). Its **Error! Bookmark not defined.** and **object-oriented** approach aim to help **programmers** write clear, logical code for small and large-scale projects.

#### Python Libraries

- Numpy
- Pandas
- Matplotlib
- Sklearn

**Material and Methods****Dataset Used for Research**

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: shows the kind of chest-torment experienced by the individual utilizing the accompanying organization t :  
1 = typical angina  
2 = atypical angina  
3 = non — anginal pain  
4 = asymptotic
4. Resting Blood Pressure: shows the resting pulse estimation of a person in mmHg (unit)
5. Serum Cholestrol: shows the serum cholesterol in mg/dl (unit)
6. Fasting Blood Sugar: looks at the fasting glucose estimation of a person with 120mg/dl. In the event that fasting glucose > 120mg/dl at that point: 1 (valid)
7. Resting ECG: displays resting electrocardiographic results  
0 = normal  
1 = having ST-T wave abnormality  
2 = left ventricular hypertrophy
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 = upsloping  
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.

**Classification Techniques**

Procedures In AI and measurements, grouping is a directed learning approach in which the PC program gains from the information and afterward utilizes this figuring out how to

order groundbreaking perceptions. At the end of the day, the preparation dataset is utilized to acquire better limit conditions which can be utilized to decide each target class; when such limit conditions are resolved, next undertaking is to foresee the objective class

Machine learning is a field of study and is concerned with algorithms that learn from examples. There are many different types of classification tasks that you may encounter in machine learning and specialized approaches to modelling that may be used for each.

**K-Nearest Neighbor Algorithm (KNN)**

K nearest neighbors is one of the easiest machine learning calculation is dependent on supervised learning procedure. K-NN calculation accepts the closeness between the new case and available cases and put the new case into the classification that is generally like the accessible classification. K-NN calculation can be utilized for regression just as for classification issue. K-NN is a non-parametric calculation, which implies it doesn't make any presumption on hidden information.

**Pros:**

- Basic Algorithm and consequently simple to decipher the forecast. Quick calculation time.
- Used for both classification and regression.

**Cons:**

- Does not work well for large dataset.
- Prediction is very costly.
- Poor at classifying data points in a boundary where they can be classified one way or another.

**Random Forest Classifier**

Random Forest is one of the most prestigious and most remarkable machine learning calculations. It is one sort of machine learning calculation that is called Bagging or Bootstrap Aggregation. So, as the access an incentive from an information test, for example, mean, the bootstrap is very powerful statistical approach. Here, lots of information are taken, the mean is determined, after that all the mean value are averaged to give a superior expectation of the mean value. In bagging, a similar strategy is utilized, but instead of estimating the mean of each information test, decision tree is commonly utilized.

**Advantage of Random Forest:**

- Random Forest Algorithm is exact outfit learning calculation.
- Random Forest runs efficiently for large scale data sets.
- It can handle hundreds of input variables.

**Disadvantage of Random Forest:**

- Features need to have some predictive power else they won't work.
- Forecasts of the trees should be uncorrelated.
- Appears as black box.

**Decision Tree Classifier**

Decision Tree Classifier is a basic and generally utilized grouping procedure. It applies a waterway forward plan to take care of the grouping issue.. Decision tree classifier represents a progression of deliberately made inquiries concerning the characteristics of the test record. Decision

Trees (DTs) are a non-parametric directed learning method used for classification and regression. It is a Supervised Machine Learning where the information is constantly part as indicated by a specific boundary.

**Decision Tree consists of:**

- **Nodes:** Test for the estimation of a specific quality.
- **Edges/ Branch:** Compare to the result of a test and associate with the following hub or leaf.
- **Leaf nodes:** Terminal hubs that anticipate the result (speak to class marks or class appropriation).

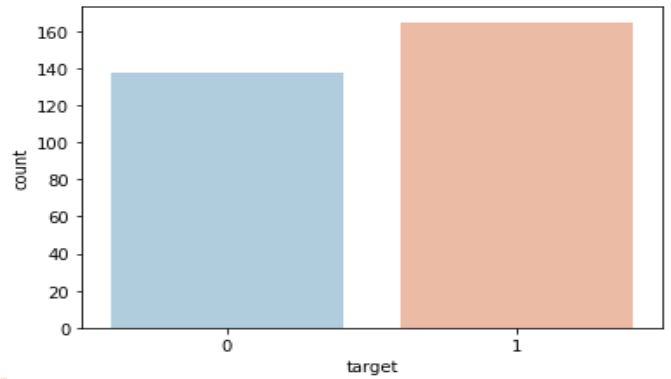
**Experiment**

**The Proposed Method**

Heart disease is the main source of death among all the diseases, even cancer. The quality of people facing heart disease is on a raise every year. The prompts for its initial finding and treatment. Because of absence of source in the medical field, the prediction of heart disease might be a issue. Use of suitable technology can be useful to the medical society and patient. The issue can be settled by embracing machine learning techniques. In my project, I would be taking a shot at basic machine learning classification model. And using this model I could prepare my model utilizing the information which comprise of different attribute like age, sex, cp, blood pressure, skin thickness and so on and based

on this attribute I would anticipate the outcome for a patient whether he is experiencing heart disease or not. This paper has Random Forest classifier, KNN (K-Nearest neighbour classifier) & Decision Tree classifier – three techniques for the effective prediction of heart disease. It analyses the efficiency & accuracy of the three techniques to choose them the best.

The figure below shows the number of the heart disease cases.

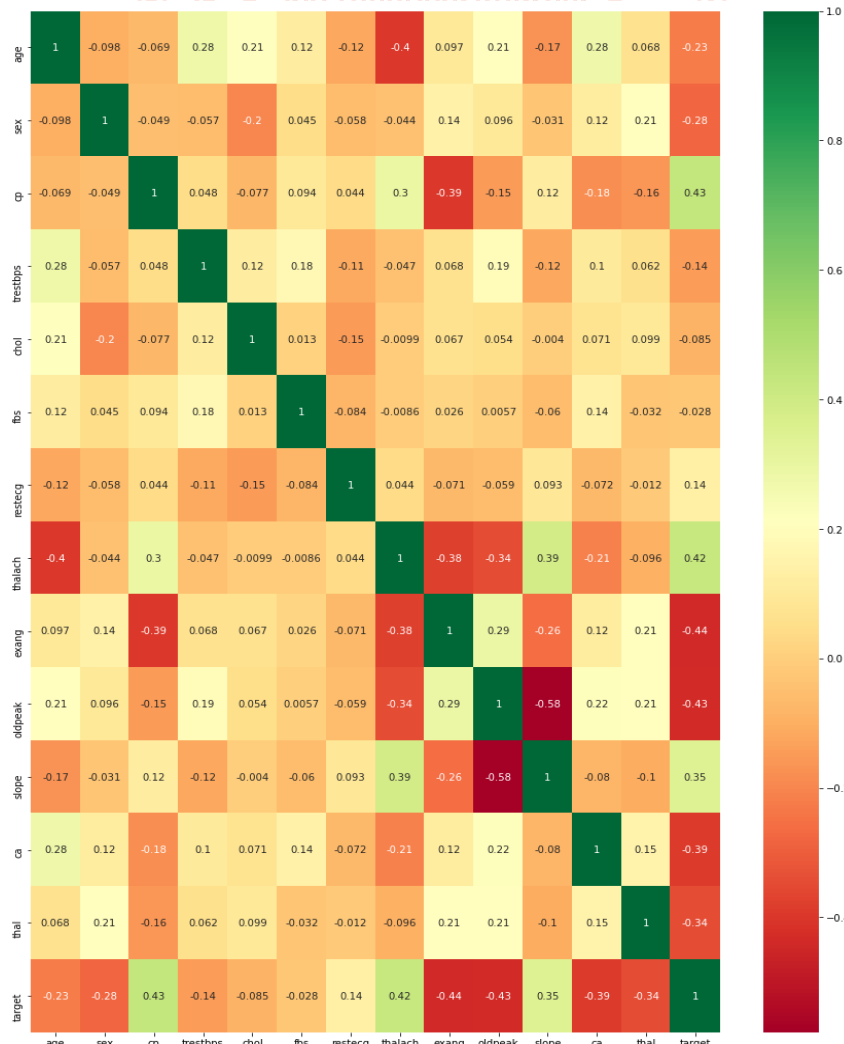


0 = absence 1 = present

**Result and Discussion**

**Correlation Matrix**

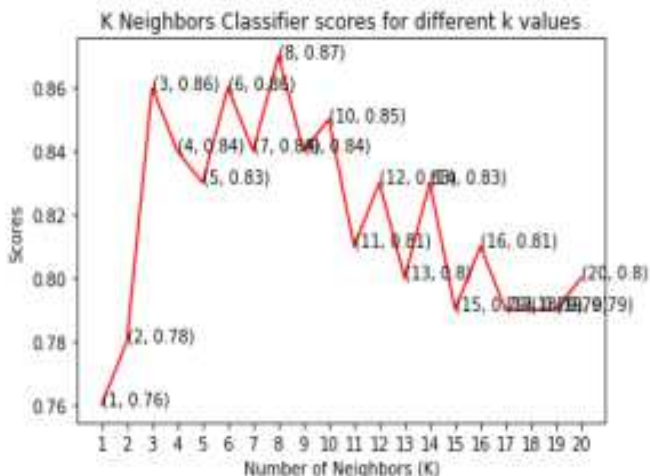
Let's see the correlation matrix of features. From this graph, we can observe that some features are highly correlated and some are not.



This figure shows the correlation matrix

**K-Nearest Neighbors Classifier:**

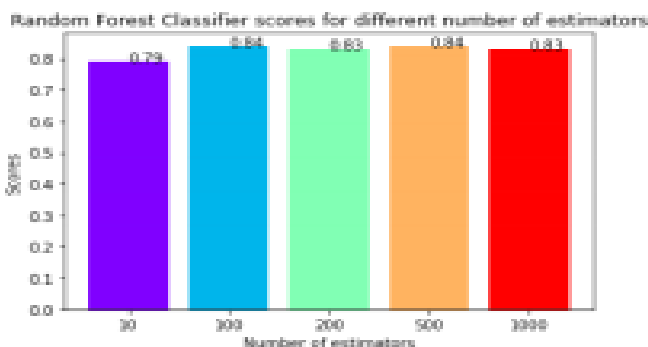
K Nearest Neighbors is a non-parametric strategy utilized for grouping. It is lazy learning figuring where all computation is surrendered until gathering. It is otherwise called case based learning calculation, where the capacity is approximated locally. This algorithm is used when the amount of data is large and there are non-linear decision boundaries between classes. KNN explains a categorical value using the majority votes of nearest neighbors. Not only for classification, KNN can be used for function approximation problem.



This figure shows the K Neighbors Classifier scores

**Random Forest Classifier:**

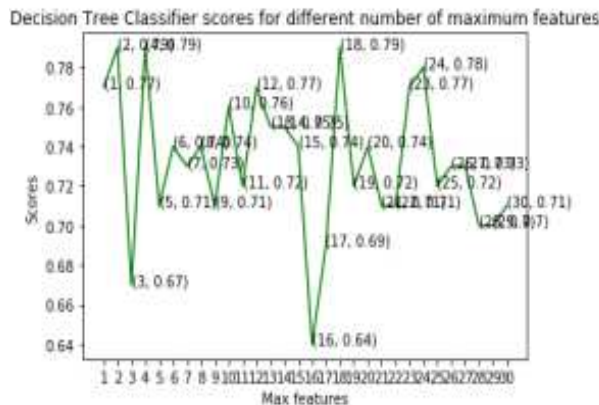
Random forest is a regulated learning calculation. It very well may be utilized for order and relapse. It is straightforward and simple to execute. A backwoods is contained trees. This classifier makes choice trees on haphazardly chose information tests, gets forecast from each tree and chooses the best arrangement by methods of casting a ballot. The random forest composed of multiple decision trees. It creates a forest of trees.



This figure shows the Random Forest Classifier scores.

**Decision Tree Classifier**

This classifier falls under the category of supervised learning. It very well may be utilized to take care of relapse and characterization issues. We can utilize this calculation for issues where we have ceaseless yet in addition unmitigated info and target highlights. It is the best machine learning calculation utilized for depicting the tree in a graphical way.



This figure shows the Decision Tree Classifier scores

**Conclusion**

Machine Learning plays an important role in various fields such as Healthcare, Stocks & Marketing, Banking, Weather Forecast and so on. With the help of KNN Algorithms it become easy to evaluate and fetch meaningful information from them. In KNN by using the various K- values of the K-NN classifier the accuracy of the model increases simultaneously, this study aims to accurately predicting whether a given patient is suffering from diabetes or not. Finally the accuracy of my model comes close to 84 % and for any new patient it could easily predict whether the patient is having diabetes or not.

**Bibliography**

- [1] S. E.-S. S. I. D. K. A. A. F Ali, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," 2020.
- [2] C. T. G. S. S Mohan, "Effective heart disease prediction using hybrid machine learning techniques," 2019.
- [3] M. R. M. I. M. I. S Nashif, "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system," 2018.
- [4] Y. H. K. H. L. W. L. W. M Chen, "Disease prediction by machine learning over big data from healthcare communities," 2017.
- [5] A. A. AA Soofi, "Classification techniques in machine learning: applications and issues," 2017.
- [6] S. S. K Deepika, "Predictive analytics to prevent and control chronic diseases," 2016.
- [7] D. P. K Gomathi, "Multi Disease Prediction using Data Mining Techniques," 2016.
- [8] J. S. N. S. A Dey, "Analysis of supervised machine learning algorithms for heart disease prediction with reduced number of attributes using principal component analysis," 2016.
- [9] M. S. B Bahrami, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques," 2015.
- [10] R. S. E. D. K Vembandasamy, "Heart diseases detection using Naive Bayes algorithm," 2015.
- [11] E. A. Y. K. AF Otoom, "Effective diagnosis and monitoring of heart disease," 2015.
- [12] S. P. V Chaurasia, "Early prediction of heart diseases using data mining techniques," 2013.
- [13] S. S. G Parthiban, "Applying machine learning methods in diagnosing heart disease for diabetic patients," 2012.
- [14] K. J. N Bhatla, "An analysis of heart disease prediction using different data mining techniques," 2012.
- [15] R. A. S Palaniappan, "Intelligent heart disease prediction system using data mining techniques," 2008.