

## **Prediction of cardiac patients using IOT software and Machine Learning Techniques**

**Authors: Shavi Rishi<sup>1</sup> and Sikander Singh Cheema<sup>2</sup>**

<sup>1</sup>Research Scholar, Dept. of Computer Science & Engineering, Punjabi University, Patiala.

<sup>2</sup>Assistant Professor, Dept. of Computer Science & Engineering, Punjabi University, Patiala.

### **Abstract**

We all know the heart attack will kill you in three attempts but even a few days in the first attempt can be risky even also. If we test our safety on a daily basis every day then we Its capacity to diagnose too many common diseases everything used to be precious. Many of us lose theirs existence of heart attack. This is due to their diet, age and less Physical exercise, plus a lot more. Cardiac attack isn't simple to pin point. The Internet of Things (IoT) is the popular things in future, it can affect communication, health monitoring system, business. Using networked devices, our physical and mental health knowledge will bring about a positive change in the healthcare landscape.

Recommend that in this effort heartbeat tracking and warning program for heart attacks use the Internet of Things. It helps surveillance to enter the population who don't have easy access to an appropriate health tracking program. The data collected can then be processed using various machine learning techniques, and then exchanged with medical professionals who can make suitable decisions by wireless communication. These provide the variety of communication capabilities that are offered by the transmitting data existing row rates and body temperatures. Heart disease can be controlled using machine learning. There is any

nonlinear aspect of Cleveland Heart Disease data collection. Improvement of heart attack prediction technique by applying the Random Forest Machine Learning Model correctly this will benefit patients with a lot less cost of getting a prompt diagnosis.

### **Objective**

To build a Heart Disease prediction system to overcome the shortcomings of the prior Heart Disease detection techniques.

### **Keywords**

Heart attack, Machine learning techniques, PYTHON.

### **Related Work**

Dilip Roy Chowdhury has suggested treatment of h eart disease by neural network sets. His approach p roduces a novel paradigm by integrating later prob abilities from Multi-Layered Perceptron. They applied the procedure on the Cleveland heart disease data collection with SAS base tools and reached accuracy of 88.07 per cent. Limitation:- They used 94 attributes(which is a great number) to train the NN which gives a lesser accuracy of 75%.

Milan Kumari has suggested treatment of Mining Classification Methods in Cardiovascular Disease

Prediction . His approach produces Decision Tree, Artificial Neural Network . 10-fold cross validation approach was used in our studies to calculate objective estimation of these predictive models. For RIPPER error rates according to his findings.

Limitation:- Accuracy was closer to 80 % which could be improved.

### **Proposed Work**

Firstly, I collect the data set of heart prediction disease. Then Select the dependent attributes from the total number of attributes. Apply Machine Learning algorithms according to the nature of dataset. Then calculate Cross-validate the algorithms accuracy. Approache Use for prediction of heart attack:-

### **Machine learning algorithms used (using sklearn library of python).**

#### **1) Linear Regression:-**

Linear regression is one of the most common and well-understood algorithms in statistics and machine-learning. The majority of data science enthusiasts and fanatics of machine learning continue their journey with linear regression algorithms. In this post, we'll look at how the linear regression algorithm operates and how it can be used successfully to create stronger models in your machine learning projects.

Linear Regression is one of the algorithms in machine learning where the outcomes are estimated by the use of known parameters associated with the output. This is used instead of attempting to assign values into groups, to estimate values within a continuous spectrum. The established parameters are used

to create a steady, continuous slope that is used to estimate the unknown or the outcome.

#### **2) Logistic Regression :-**

This is a method to evaluate a data set with a dependent variable and

one or more independent variables to estimate the outcome of a binary variable, meaning that it can only produce two outcomes. The dependent variable, in nature, is empirical. Dependent variable is also referred to as target variable and the determinants are called outcome variable. Logistic regression is a particular case of linear regression in which we only predict the result in a categorical variable. Using the log function it predicts the probability of the event.

**3) Decision Tree:-** A decision tree is a supervised support method that uses a decision graph similar to a tree or model and its future consequences, including the effects of fortuitous events, capital and service costs. It is one way to see an algorithm which describes a situation only. Tree-based learning algorithms are among the effective supervised methods to instruction, which are frequently used. Tree-based approaches allow for statistical modeling High accuracy, durability

and ease of understanding. They are like linear models Pretty good for modeling non-linear connections. They are adaptable to any problem issue (classification or regression) at hand.

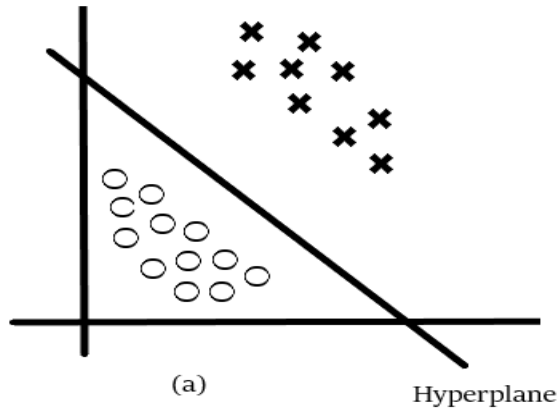
Decision Tree algorithms include as the CART (Classification and Regression Trees). Decision Tree algorithms are referred to as CART (Regression and Classification Trees).

#### **4) Support Vector Machine:-**

Supporting vector machines or SVM is a supervised learning

algorithm that can be used as supporting vector classification (SVC) and supporting vector regression (SVR) for the classification and regression problems. It is used for smaller datasets, because processing takes too long. We'll be focussing

ng on SVC in this package. Find a condition which follows: There is a stalker who sends you emails and now you want to develop a function(hyperplane) that will specifically separate the two instances, so that it will be marked as a spam if you get an email from the stalker the following is an example of two situations in which you draw the hyperplane, which one you will choose and why? Take a minute to assess the situation .....



ensemble approach that is stronger than a single decision tree, and by multiplying the result it reduces the overfitting.

### Cross Validation Techniques used:

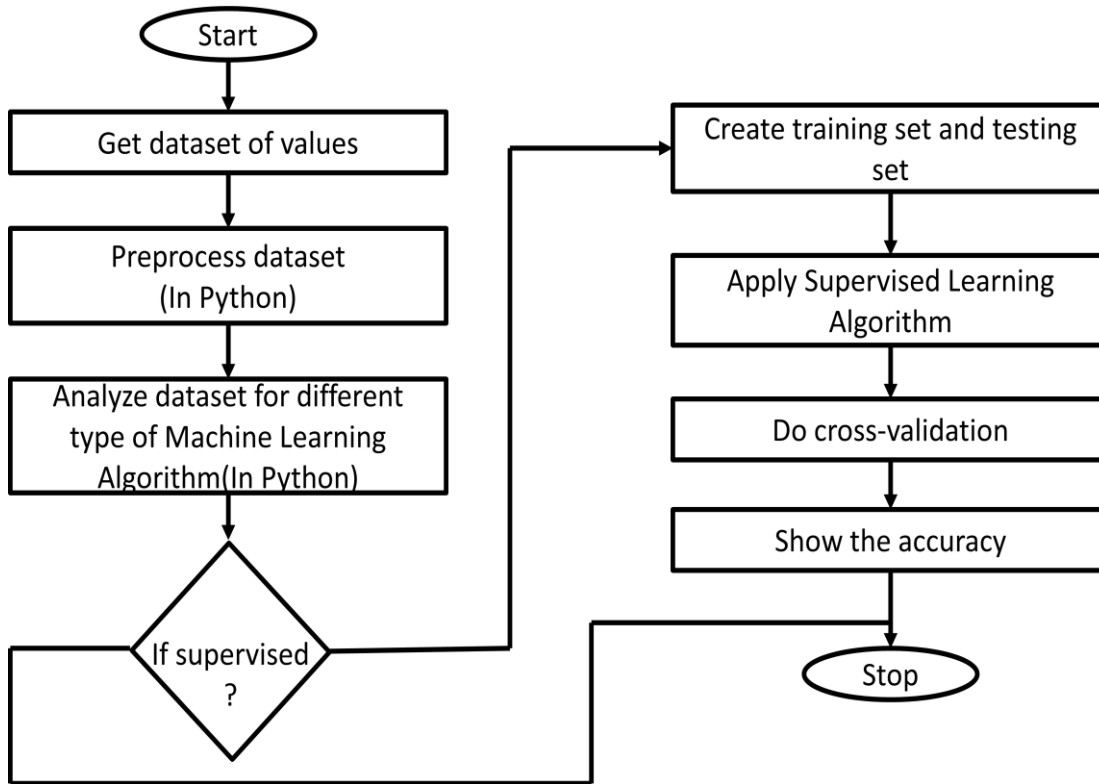
- 1) 3-fold
- 2) 5-fold
- 3) 10-fold

### IMPLEMENTATION

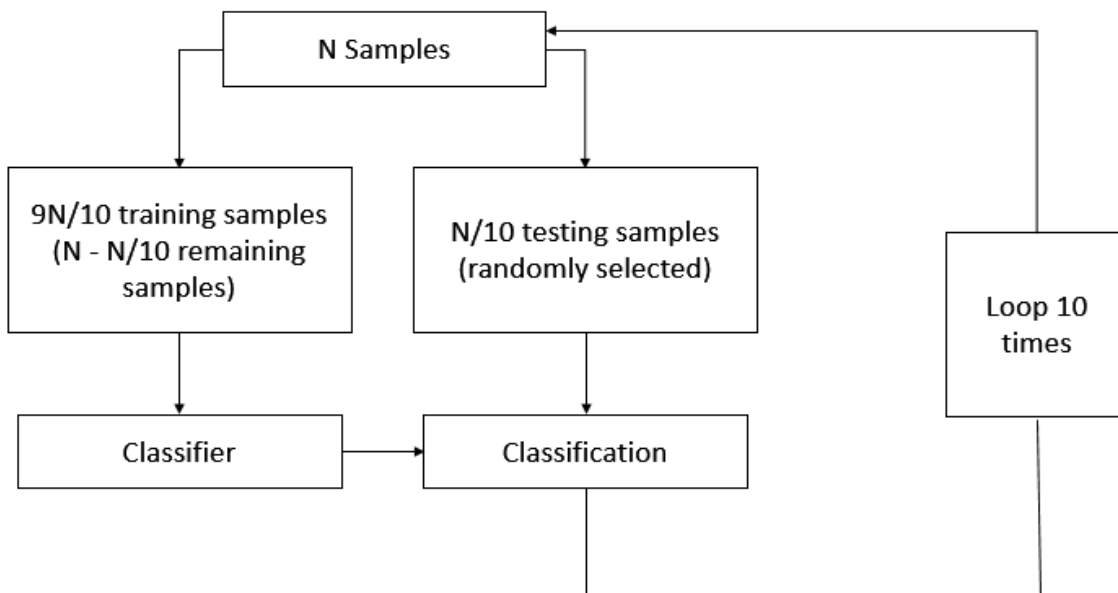
- 1) Start
- 2) Collect the dataset.
- 3) Select the dependent attribute.
- 4) Preprocess the data.
- 5) Analyze dataset for different type of machine learning algorithm.
- 6) If supervised :
  - 6.1) Create training set and testing set using 10-fold cross-validation.
  - 6.2) Apply supervised learning algorithm.
  - 6.3) Do testing using the test set.
  - 6.4) Show the accuracy.
7. Stop

### 5) Random Forest:-

Random forest is a supervised learning algorithm, used both for classification and regression. However, it is used mainly for grouping issues. Since we know a forest consists of trees and more trees means a more healthy forest. Similarly, random forest algorithm builds decision trees on data samples and then gets the prediction from each of them and eventually selects the best solution by voting. It is an ensemble



## VALIDATION TECHNIQUE



## Conclusion

Our dataset set had Non-Linear dependency. By correctly adjusting the parameters of Random Forest we were able to achieve better accuracy. We had lesser amount of dataset so 10-fold cross-validation gave us better result. Our solution strategy is not so Robust in nature, every time it needs a few adjustment in parameter. The accuracy achieved is satisfactory but can be future improved. We can use data warehouses in hospital so that the amount of data increases and a greater accuracy could be achieved.

## References

- [1] Schneider A, Hommel G, Blettner M. Linear regression analysis: Part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2010;107:776-82.
- [2] Chan YH. Biostatistics 201: Linear regression analysis. Age (years). *Singapore Med J* 2004;45:55-61.
- [3] Meenal Saini, Niyati Baliyan and Vineeta Bassi, "Prediction of Heart Disease Severity with Hybrid Data Mining".
- [4] Aditya Methaila, Prince Kansal, Himanshu Arya, Pankaj Kumar, " Early Prediction of Heart Diseases Using Data Mining Techniques ".
- [5] Asha Rajkumar, Mrs. G.Sophia Reena, "Diagnosis Of Heart Disease, Using Data-mining Algorithm".
- [6] Durga Kinge , S. K. Gaikwad, "Survey on data mining techniques for disease prediction".
- [7] Cemil Colaka, Esra Karamanb, M. Gokhan Turtayb., " Application of knowledge discovery process on the prediction of stroke".
- [8] Yanwei X, Wang J, Zhao Z, Gao Y (2007) Combination data mining models with new medical data to predict outcome of coronary heart disease. In: Proceedings international conference on convergence information technology, pp 868–872
- [9] Le Duff F, Munteanb C, Cuggiaa M, Mabob P (2004) Predicting survival causes after out of hospital cardiac arrest using data mining method. *Stud Health Technol Inform* 107(Pt 2):1256–1259
- [10]. Saad Talib Hasson and Zahraa Yaseen, "Simulating Road Modeling Approaches in VANET Environment Using Net Logo", *Research Journal of Applied Sciences*, Vol. 11, Issue 10, Page No. 1130-1136.
- [11]. Mohammad M. Masud Hind Al Falasi and Nad bver Mohamed. "Trusting the Same- Using Similarity to Establish Trust Among Vehicles". In: *International Conference on Collaboration Technologies and Systems*, 2015, pp. 64–69.
- [12]. Wenjia Li and Houbing Song. "ART: An Attack- Resistant Trust Management Scheme for Securing Vehicular Ad Hoc Networks". In: *IEEE Trans. Intell. Transport. Syst.* 17.4 (2016), pp. 960–969

