WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Prediction of Heart Disease Using Machine Learning Algorithms

[1]Sowmiya T, [2]Suriyapriya M, [3]Thenguzhali J, [4]Mrs. Jancy Rani K

[1,2,3] Student, Computer Science and Engineering, Agni College of Technology, Chennai-600 130.

[4] Assistant Professor, Computer Science and Engineering, Agni College of Technology, Chennai-600 130.

ABSTRACT:

Machine Learning is used in a wide variety of applications. Machine Learning could play an important role in predicting heart disease. Heart diseases are increasing in our daily lives and the cases of heart diseases are increasing and it is very important to predict any such kind of diseases beforehand. In most cases, the diagnosis of Cardio Vascular disease has become one of the most complicated tasks in the medical fields. Within this time, approximately on person dies due to upset. This research paper mainly focus on which paper is most likely to have a heart disease based on various medical conditions. Heart Disease Prediction is a difficult task, there is a requirement to automate the prediction process to classify the patient's risk level by using methods such as Naïve Bayes, Decision Trees, Logistic Regression, Support Vector Machines, Random Forests and to prevent the related risks and to warn patients before they are affected by heart disease. This document conducts comparative studies by analyzing the effectiveness of various aspects of machine algorithms. The test confirmed that the Logistic Regression algorithm has the highest accuracy of 90.06% compared to other machine learning algorithms.

I. INTRODUCTION

Heart is one of the important organ in the human body which pumps to various parts of the body. So that it is very important to keep our heart healthy in such a way. It is also very important to predict the heart diseases before it comes and it will save our money and more importantly it will also save our lives. The symptoms which arises in the patient are required to diagnose the disease. The reasons for the heart disease may be as follows it may be of Cholesterol, High Blood Pressure, Genetic heart disease, Gastroenterology problems and Smoking can lead to a heart disease. This disease can be treated by various medical treatments and it can also be treated by changing our day to day activities like quitting Smoking and alcohol. This will improve our life better. However predicting the heart disease is difficult Machine Learning can help in the early diagnose of a heart disease. In recent times, heart complaint is one of the reason for the death of many people. It is because of Smoking, drinking and tobacco that will lead to hypertension. So these kind of heart diseases will ruin many of the lives. So, the proposed work makes a trail to identify the heart diseases at very early stage to avoid such kind of consequences. The most important thing of this paper is to provide a tool to predict the heart complaint at the very earliest stage this will helpful for many people who are having heart disease at a very early stage to avoid severe consequences. Machine Literacy plays a major role to descry the retired separate patterns and thereby after reviewing the data presented after the application of machine learning technology useful for early prediction of cardiovascular diseases, this article presents an evaluation of the performance of read machines, logistic regression, and random forest for early prediction, similar to mean Bayesian decision trees.

2. LITERATURE SURVEY

1. T.Nagamani et al have proposed a system with the Map Reduce algorithm. The accuracy obtained in line with this paper for the 45 instances of testing set, was greater than the accuracy obtained using artificial neural network. Here, the accuracy was improved.
2. Sonam Nikhar et al has presented this research paper, Predicting Heart Disease Using Central Bayesian Machine Learning and Decision Tree Classifiers, which makes the assumption that decision trees pass Bayesian classification.
3. Avinash Golande and et al studies were carried out to examine the KNN Decision Trees available for heart disease classification and K-word algorithms available for classification and obtained the most accurate decision trees
4. Fadh Saleh Alotaibi designed a Machine Learning model and compared five different algorithms. In this study, trees with higher accuracy than Matlab and weka tools using the speed miner tool.
5. Anjan N Repaka et al presented a model of predictive performance of two classification models analyzed and compared with previous work to see if I improved the precision of our plan in finding percentage risk compared to all models.
6. Aakash Chauhan et al presented " Heart disease Prediction using Rule Learning ". This eliminates the manual task that additionally helps in extracting the information directly from the electronic records.We have applied FP Growth association mining on patient's dataset. This

helps in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of Coronary Sickness.

7. Ashir Javeed, Shijie Zhou et al (2017) designed "An Intelligent Learning System algorithm and Optimized Random Forest Model for Improved Heart Disease Detection ". It uses random search algorithm for diagnosing the Cardiovascular disease.
8. Lokanath Sarangi, Mihar Narayan Mohanty, Srikanta Pattnaik (2015) : This is a system for Cardiac Disease Prediction", designed a model by using genetic algorithm technique. The weights were fed as an input to the given network. The accuracy achieved was 90% by using the hybrid technique of and neural networks.

3. DATASETS

According to WHO World Health Organization Cardio Vascular Diseases and Heart diseases are the leading to death around 17.9 million lives per year. The attributes are as follows:

The List of attributes used in this project to predict the heart disease accuracy:

1. Age
2. Sex
3. Chest pain type
4. Resting blood pressure
5. Serum Cholesterol in mg/dl
6. Fasting blood sugar>120 mg/dl
7. Resting electrocardiographic results
8. Maximum heart rate achieved
9. Exercise induced angina
10. Oldpeak-depression induced by exercise relative to rest
11. The Slope of the peak exercise ST segment
12. Number of major vessels
13. Thal
14. Target
15. dtype

IV. EXISTING AND PROPOSED SYSTEMS

EXISTING SYSTEM

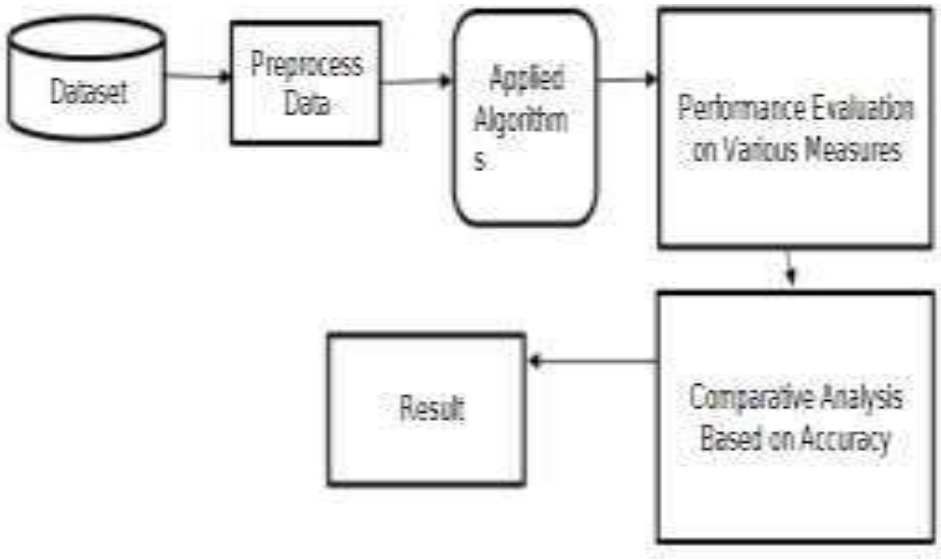
Heart disease is now killing people like a silent killer without a system, many tools and devices that can be designed to predict and diagnose heart disease. EHDPS I one of them that predicts the use of all heart disease.Designs using machine learning algorithm are used to train datasets and based on accuracy measurement and other actions.Succesfully predicting heart disease testing is done using the maximum algorithm.Efficiency, we use the distribution model proposed by the pipeline, the machine learning algorithm is recommended in the system, used to get the mot accurate results.

PROPOSED SYSTEM

In this proposed system, Machine Learning algorithms

It have been used to obtain the results with the maximum accuracy. been combined into a single generalized Dataset. The Datasets are further classified into train data and test data in the ratio of 7:3.

SYSTEM ARCHITECTURE



III. MODULES

1.Data Collection

2.Data Preprocessing

3.Train and Test data

4.Logistic Regression

5.Deployment

1. Data Collection

The Machine Learning System collects the data of person that collects the medical parameters such as age, sex, blood pressure, Cholesterol and obesity for prediction of a heart disease.

2. Data Preprocessing

Data Preprocessing is making all the data suitable for a Machine Learning model. It is the very first step of creating a Machine Learning model. When we create a Machine Learning Project, it is not always a case that we encounter the clean and formatted data.The Validation techniques in the Machine Learning are used to get the current or the true error of the dataset. If the dataset is large means we may not need for the Validation techniques. In real world, it is not the true representative of the population of a given dataset while working with the samples of data. The Sample of the data is used to provide evaluation of model to fit on the training dataset. The Validation set is used to evaluate a given model, but this is for continuous or frequent evaluation, Data collection, Data analysis, Data content and can add uo to a time-consuming to-do list. During the process of data identification it helps us to understand the data and its properties and this knowledge will help you to choose a better model suited for this project.

3.Train and test data

Train or test dataset is the way to test the accuracy of the machine learning model, we divide the data into train and test dataset 80% for training and 20% for testing, we use training dataset to train model and test dataset.

Applications of Machine Learning algorithms:

Once the information has been collected and created for modelling, we use our five Machine Learning algorithms which we implement to predict the heart disease. We trained our model using different algorithms. We used the model Logistic regression once the information has been created for modelling we employ our five machine learning classification algorithms which we are visiting to implement to predict the heart condition. We trained our model using different algorithms it classifies the information into binary form means only in 0 and 1 which means 0 identify positive or negative of a heart disease. We trained our model by logistic regression and the accuracy of 85%

Used Python Packages

1.SK Learn

2.Numpy 3.Pandas

4.Matplotlib pyplot

IV. Logistic Regression

It estimates the probability of an event occurs/like voted or not voted, it is based on a given dataset of independent variables. Since the result is a probability 0 and 1. In Logistic Regression it is Binary.

```
[ ] import numpy as np
    import pandas as pd
    from sklearn.model_selection import train_test_split
    from sklearn.linear_model import LogisticRegression
    from sklearn.metrics import accuracy_score
```

Training accuracy	85%
Testing accuracy	80%

MODEL EVALUATION

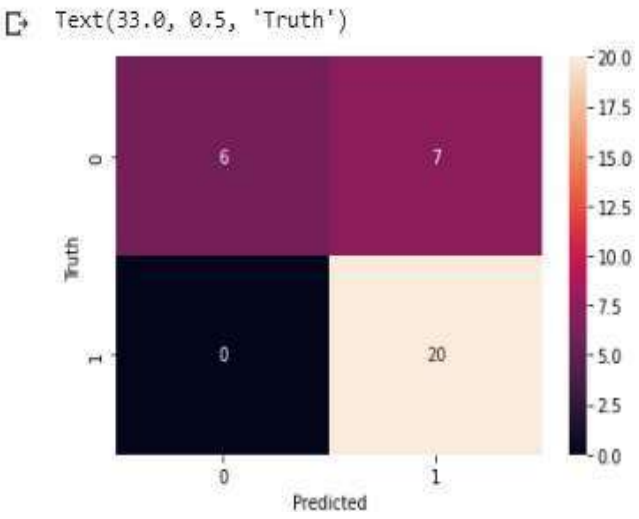
ACCURACY SCORE

```
[ ] # accuracy on training data
    X_train_prediction = model.predict(X_train)
    training_data_accuracy = accuracy_score(X_train_prediction, Y_train)

    print('Accuracy on Training data : ', training_data_accuracy)

    Accuracy on Training data :  0.8512396694214877

[ ] # accuracy on test data
    X_test_prediction = model.predict(X_test)
    test_data_accuracy = accuracy_score(X_test_prediction, Y_test)

[ ] print('Accuracy on Test data : ', test_data_accuracy)

    Accuracy on Test data :  0.8196721311475411
```

V. Deployment

The Deployment is done using Flask(WEB FRAMEWORK). It is classified as a micro-framework written in Python. It has no database abstraction layer and form validation. However flask supports extensions that can add the extra features of the application as if they were implemented in Flask itself. The extensions exist for object relational appears form validation, upload handling and various open authentication technologies and several common framewrok and related tools.

RESULTS:



OUTPUT:

Future Works

Heart Disease is dependent on a lifestyle attributes as well as past medical history. Here in this paper, We have considered seven lifestyle attributes and three medical conditions. In the future, for better performance of the model more medical attributes can be considered such has to be found by using large amount of data, So it is used to get a result in a single way of testing, Instead of testing the single data.

Conclusion:

With the increasing number of deaths because of heart diseases, it's become mandatory to develop a system to predict heart diseases effectively and accurately. We used Decision trees, Logistic Regression, Random Forest, and pure Bayesian algorithms to predict disease using a machine learning repository dataset. The results of this study show that the Random Forest algorithm is the best algorithm for predicting future heart diseases with an accuracy of 90.06%.Performance is often improved by developing a website that supports the Random Forest algorithm, such that it can help builf the algorithm. Better results and helps doctors to predict heart disease more effectively and efficiently.

References:

1. T.Nagamani, S Logeswari, B Gomathy heart disease prediction using mapreduce algorithm,International Journal IJITEE ISSN: 2278-3075 issue 8 January 3rd 2019.
2. Avinash Golande, Pavan Kumar T, heart disease prediction in International Technology and Engineering Vol 8,pp.944-950,2019.
3. Nagaraj M Lutimath, Chethan C, Basavaraj S Pol.,uses machine learning to predict cardiovascular diseases global cutting-edge technology and engineering journal,8,(2S10), pp 474-477,2019.
4. B.Rjeily, G. Badr, E.Hassani, A. H., and E.Andre uses clinical data on heart disease, machine learning the way forward mining paradigm in medicine,2019,pp.7199.
5. Fajr Ibrahem Alarsan, and Mamoon Younes analysis and classification of diseases in the heart using heartbeat and machine learning algorithms big data log Data,2019;6:81.
6. Dubey A ,Patel R,Choure K,heart disease prediction for data mining International Journal of Advanced Technology and Engineering Exploration.;(1):1-6,2014.
7. Peter,T.J, Somasundaram, K investigate heart disease using Classification of Data Mining Techniques (ICAESM),2012.
8. Mamatha Alex P and Shaji use Data Mining predicting and diagnosing Heart Disease patients techniques International Communication and Marketing,2019.