HEART DISEASE PREDICTION USING MACHINE LEARNING

Richa Singh[1], Aniket Panpatte[2]

[1]ASM's Institute of Management & Computer Studies, Thane 400604, India
[2]ASM's Institute of Management & Computer Studies, Thane 400604, India

ABSTRACT :

In this modern living lifestyle heart disease has become most common among people due to multiple changes in day-to-day life. Heart disease can

be fatal, can risk one's life and can create many other severe situations leading to other diseases.

This model detects diseases using machine learning to give accurate results. This system has been beneficial to users, patients, doctors, physician

and ultimately health industry. Patients don't need to go for costly test they can predict disease using some simple tests with this model. This intelligent system retrieves data from hidden database system and compares the input values to give results accordingly.

Machine Learning algorithms are used here to test and give the probability of heart disease.

Introduction :

Machine Learning as an automated method used by systems to learn from data, identify useful patterns and minimize human interference in the decision-making process. In this paper, a comparative analysis of different classifiers has been performed for the classification of the Heart Disease datasets in order to correctly classify and or predict HD cases with minimal attributes.

The set contains many attributes including the class attribute, for patients collected from people around us, but in this paper, only a subset of 14 attributes is used, and each attribute has a given set value. Algorithms are used for performance of the selected classifications algorithms to classify the best and to predict, the heart disease. This proposed system can reserve problem by predicting the presence of heart condition in the patient.

Technology:

Clinical evaluation is necessary to keep you safe from death or other severe conditions. This system not only helps financially but can also give a patient an overview of precautions or suggests the patient a further treatment. Some of the first test doctor may order include Blood test, Chest X-Ray, CT scan which can cost up to 500-2000 Rs.

There can be also a chance that the report of the test can be incorrect, and if the doctor starts your treatment considering this report that can lead to severe condition or even death.

WHO states that there are 2.6 million deaths worldwide due to incorrect medical report. They say that this type of phenomenon can lead the patient to unnecessary medication or lack of medication.

Current Status of Development:

- •By going through various research papers attributes and dataset has been selected.
- •Various algorithm has been theoretically studied and been selected.
- •The main body of the project i.e the coding work is in progress.

Methodology

This is an experimental type research design with a quantitative research method. Basically, this experiment is carried out on the database of heart disease which is available publicly. Doctors and researcher trust the intelligent system which is made by Machine Learning techniques to develop screening tool for whether a patient has a risk of CHD or other heart diseases.

In this project we have used some algorithm related to machine learning which can find the illness from the entered data. The details added by the user is processed to check the different heart illness that could be associated and the result/report is shared/displayed to the user.

This system works by giving us a prediction of high risk or low risk depending on risk factors such as diabetes, high blood pressure, high cholesterol and many more. The user can show this report to the doctor and can start his further treatment. This software also shows some methods that can help the user to prevent the heart disease.

The model is trained using UCL heart disease dataset

Algorithm

NAIVE BAYES:
Naive Bayes classifier is one of the simple and most effective classification algorithms which helps in building the fast machine learning model that can make quick prediction. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

DECISION TREE:
Decision tree learning or induction of decision trees is one of the predictive modelling approaches used in statistics, data mining and machine learning. It uses a decision tree to go from observations about an item to conclusions about the item's target value.

SUPPORT VECTOR MACHINE:
In machine learning, support-vector machines are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis.

K-NN:
K-NN stands for k nearest neighbor. In statistics, the k-nearest neighbors' algorithm (k-NN) is a non-parametric classification method. It is used for classification and regression.

K mean clustering:
Means is a clustering algorithm in machine learning that can group an unlabeled datasets very quickly and efficiently in just a few iterations. It works by labeling all instances on the cluster with the closest centrefold.

Challenges

Heart Disease is that the domain of health and life science field which distinct from the IT field, also as ML, is an emerging field of AI that's a replacement technology to open new horizons. ML projects are always technology-driven that involves the particular product isn't easy. These are some high-level challenges are going to be faced during the developing the project.

 The followings are the challenges while developing this project as listed below:

•Within the survey of this project, the interview of HD doctors will difficult to conduct because of their busy time schedule. Thereby it'll make hard to spot the correct attribute of the HD.

•The main problem are built a classifier model which will predict if it's getting HD or not it means the way to train the classifier model in order that the system will get exact if the patient is possessing HD or not that's the main challenge of this project.

 •ML may be a new technology that gives an answer to the contemporary problem. it's very young technology environment that each big company like Google, Facebook, and Microsoft are working ML project in their own environment. So, this purposed project are going to be faced with the matter during the deployment.

Conclusion :

Population is growing in an exponential way. Death rate due to cardiovascular diseases is also increasing. The only solution to control this is to predict the heart disease and medicate it before it gone worse. By applying different machine learning algorithms to see what difference comes when it is applied to the data, we got accurate results. The first approach is normal datasets which is acquired for direct used of classification, whereas second approach is checking the data with feature selection which are taken care of and there is no outliers detection.

The results which are achieved are quite promising and then in the third approach the datasets was normalized taking care of the outliers and feature selection; the results achieved are much better than the previous techniques in other researches by researchers, and when compared with other research accuracy, our results are quite promising. In this paper, we proposed three methods in which comparative analysis was done and promising results were achieved.

Many researchers have previously suggested that we should use ML where the datasets is not that large, which is proved in this paper. The methods which are used for comparison are confusion matrix, precision, specificity, sensitivity, and F1 score. For the 13 features which were in the datasets, K Neighbors classifier performed better in the ML approach when data pre-processing is applied.

If a large datasets is present, the results can increase very much in ML model. Researchers are using Machine learning techniques on this data to diagnosis heart disease. In future this model needs to be improved in all terms and some more latest technologies in prediction can help people predict heart disease cost effectively.

REFERENCES :

From internet and Books some of the links are mentioned below:

[1] https://www.codecademy.com/pro/get-started

[2] https://www.coursera.org/ml course

[3] www.unite.ai/what-is-k-nearest-neighbors/

[4] https://www.upgrad.com/blog/naive-bayes-explained

[5] towardsmachinelearning.org/decision-tree-algorithm/

[6] https://www.javatpoint.com/machine-learning-algorithms

[7] www.tutorialspoint.com/r/r_logistic_regression.htm

[8] https://www.acls.net

[9] www.ibm.com/cloud/learn/random-forest

Critical Reviews Mentioned in this Paper:

[1] Vijayarani Mohan:Disease Prediction in Data Mining Technique -A survey-2013

[2] Hlaudi Daniel Masethe: Prediction of Heart Disease Using Classification Algorithm -2014

[3] Andrea D'Souza: Heart Disease Prediction Using Data Mining Techniques- 2015

[4] Jaymin Patel : Heart Disease Prediction using ML and Data Mining Techniques   -2015

[5] Cong Long: Expert System with Application- 2015

[6] T. Santhanam: Heart Disease Prediction using Hybrid Genetic Fuzzy Model- 2015

[7] Purushottam: Efficient Heart Diesase Prediction System- 2016

[8] Alireza Alinezhad: Data Mining Apriori Algorithm for HDP-2016

[9] Kalyani kadam: Prediction of HD Using K-Mean and ANN as Hybrid approach to improve accuracy-2017

[10] Marimuthu Muthuvel & M.Abhinaya&V.Pavitrha: A Review on Heart disease prediction using Machine Learning and Data Analysis Approach- 2018

[10] Marimuthu Muthuvel & M.Abhinaya&V.Pavitrha: A Review on Heart disease prediction using Machine Learning and Data Analysis Approach- 2018

[11]V V Ramalingam: Heart Disease Prediction using ML Techniques-A survey 2018

[12] N. Satish Chandra reddy: Classification and feature selection approaches by ML-HDP☐2019

[13] Vikas Boddu: Neural Network based intelligent system for HDP-2019

[14] Senthilkumar Mohan: Effective HDP using Hybrid ML Techniques-2019

[15] Alireza Alinezhad & Mirupouya Mirmazaffari: Data Mining Apriori Algorithm for Heart   Diseasae Prediction- 2019

[16] R Venkatesh & C Balasubramanian & M Kaliappan: Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique-2019

[17] DR. S. Anitha: Heart Disease Prediction Using Data Mining Techniques- 2019

[18] Rami Mustafa A Mohammad: Prediction of Coronary Heart Disease using Machine Learning: An Experimental Analysis-2020

[19] Komal Kumar Napa: Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifier- 2020

[20] Heart Dataset (heart.csv) https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset