International Journal of Research (IJR)

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 9 Issue 02
February 2022

A REVIEW ON PREDICTION OF HEART CONDITION USING MACHINE LEARNING

Mr. Harshit Saxena[1]
M. Tech Research Scholar
Department of Computer Science and Engg.
Lakshmi Narain College of Technology, Bhopal
Email:-hsaxena20@gmail.com

Prof. Nidhi Singh[2]
Assistant Professor
Department of Computer Science and Engg.
Lakshmi Narain College of Technology, Bhopal
Email:-nidhis@lnct.ac.in

ABSTRACT - Each disorder is split by memory loss, human body language, and other abilities that affect a person's ability to perform everyday activities. Lack of awareness isn't going to cause these diseases. It's commonest in people 50 years aged and older and younger. the foremost risk factor is that the sudden shock caused by the amazing things around them. Sleep death is that the second commonest explanation for death for those over the age of 55 that suffer from a stroke. Nowadays, because of lifestyle and heredity, health illness is increasing day by day. Especially, heart disease has become quite common lately, meaning people's lives are in peril . Everyone has different values for sign , cholesterol and pulse. But according to clinically proven results, normal values of important sign are 120/90, cholesterol 100-129 mg / dL, pulse 72, fasting blood sugar level 100 mg / dL, pulse 60-100 ppm, ECG is normal, the width of the foremost vessels is 25 mm (1 inch) within the mains and eight μm within the capillaries. this text presents a survey of varied classification techniques used to predict each person's risk level supported age, sex, sign , cholesterol, and pulse . The "disease prognosis" system, supported predictive modeling, predicts a user's disease supported the symptoms the user presents as a system input. The system analyzes the symptoms presented by the user as input and provides the probability of the disease as an output. Disease prognosis is performed by implementing 5 techniques just like the Nave Base, KNN, Decision Tree, linear regression and Random Forest Mechanisms. These techniques is employed for the calculation of the permutations of the disease. Therefore, the standard predictive accuracy probability is 83% obtained.

KEYWORDS—Machine Learning, Heart Disease, Naive Bayes, Logistic Regression, Linear Regression, and Random Forest.

INTRODUCTION -Many factors that affect the human heart in lifestyle. Many problems occur at a rapid pace, and new cardiovascular diseases are quickly identified. In today's stressful world, there's a crucial element during a human body that transmits blood through the body for blood flow and its health must be preserved for a healthy life. Disorder has become the foremost important explanation for death worldwide. according to the earth Health Organization, cardiovascular diseases cause 17.7 million lives annually, accounting for 31% of worldwide deaths. In India too, cardiovascular diseases are the leading explanation for death [1].Risk factors for heart disease include: high cholesterol, high sign , diabetes, smoking, excessive alcohol intake, being overweight or obese, history of coronary artery disease

Symptoms of a Heart Attack:

- Wheezing Chest, chest pains and numbness
- Pain may spread to the left or right arm or neck, jaw, back, or abdomen
- Fatigue
- Cold sweat and instability
- Rapid or irregular heartbeat
- Burn heart burning or abnormal pain

Classification of heart disease:

- Coronary artery disease
- Myocardial infarction
- Congestive heart failure
- Ro strokes and so on

SYMPTOMS

- Chest Pain
- Heart Attack
- Congestive Cardiopathy
- Congestive Coronary Failure
- Stroke

MACHINE LEARNING-Machine learning may be a field of study that permits computers to find out without being explicitly planned. ML is one among the foremost exciting technologies. Because it is obvious from the name, it gives the machine a really similar look to humans: the power to find out machine learning is getting used intensively today, perhaps in additional places than one might expect. Instrumentation available in mechanical gradients:

- Logistic Regression
- Naive Bayes
- Linear Regression
- Random Forest

ALGORITHMS-The logistic regression is additionally called the sigmoid function, which makes it easier to represent in graphs. It also provides greater accuracy. This method requires importing data first then training. Using the equation, the logistic regression algorithm is represented within the graphs that distinguish between attributes. Logistic regression may be a classification method want to assign observations to a group of distinct classes. Unlike rectilinear regression that outputs continuous numerical values, the logistic regression returns its output using the logistic sigmoid function, which may then be converted to 2 or more distinct classes.

In the binary logistic regression model, the variable has two degrees (categorized). Outputs that have two values are modeled by way of polynomial logistic regression and therefore the order of the logical regression if a couple of classes are ordered (for instance proportional odds ordinal logistic model). The logistic regression mannequin fashions the likelihood of output supported enter and would not function statistical classification (it is not a classifier), even though it are frequently wont to create a classifier, for occasion with the aid of deciding on a cutoff price and classifying the inputs as a class rather of reducing Yadu, one beneath the other; this is frequently a general thanks to create a binary classifier. Unlike linear technique of least squares, the coefficients are commonly no longer computed by way of a closed shape expression; See pattern fit. Logistic regression used to be first developed and popularized as a generic statistical model, establishing with Joseph Bergson, Bergson (1944), the place he named "logit".

Naive Bayes-Naive Bayes Algorithm (NB) may also be a classification algorithm it truly is used when the dimensionality of the input is extraordinarily high. A Naive Bayes classifier considers the existence of a particular characteristic throughout a type as unrelated to the presence of the different feature. This is regularly supported the Bayes theorem. Bayes' theorem is as follows: P (Y / X) = P (X / Y) P (X) [3].

Linear Regression-The rectilinear regression is developed within the statistical field and studied as a mannequin for perception the connection between input and output quantity variables, however is borrowed from desktop learning.It is a statistical algorithm and machine learning algorithm. Rectilinear regression is that the first sort of multivariate analysis, which has been rigorously studied and used extensively in practical applications. [4] Because the models that linearly depend upon their unknown parameters are better suited to models that don't linearly correlate with their parameters, it's easier to work out the statistical properties of the resulting estimators.

Random Forest-The identical random forest algorithm or random forest classifier can be used for each the classification algorithm and consequently the regression task. Random forest classification algorithm deals with missing values, once we have more trees within the forest, the random forest classification algorithm doesn't overestimate the model, the random forest classification algorithm for categorical values you'll sample it. An extension of this method was developed by Leo Freeman and Adele Cutler, who registered "Random Forest" as a trademark (until 2019, owned by Minitab Inc.). The extension combines Freeman's "packing" idea and randomselection of features, first introduced by Ho, then by Amit and Zeman to make a group of decision trees with variants independently controlled.

EXISITNG SYSTEM-In the present work, the Hybrid Random Forest with Linear Model (HRFLM) is employed. Numerous studies are conducted on limiting feature selection for algorithmic use. In contrast, the HRFLM method uses all features with none constraints on feature selection. Here we conduct experiments that are wont to identify the features of a mechanical hybrid system with a hybrid system. Experimental results show that our proposed hybrid system features a strong ability to predict disorder compared to existing methods.

PROPOSED SYSTEM-In the proposed system, data refinement is performed to get rid of null values from the dataset. The subsequent data visualization process is performed to point out the dataset values and therefore the bar chart. Finally, the classification process decision is administered using the tree and SVM algorithm, where the test phase depends on the training phase. The accuracy of the algorithm is predicted using the confusion matrix.

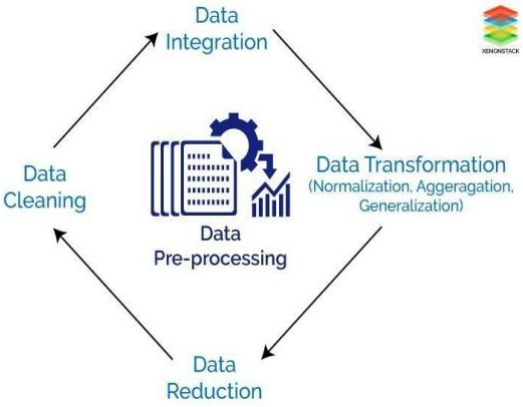
ARCHITECTURE DIAGRAM-Firstly, the information is being collected from the patients. Now the info is being cleansed where the info pre- processing process occurs. Data pre-processing means the info are going to be checked if there are any null values or wrong values entered and can be made to zero for a fresh entry. Next process the info

is being visualized, in data visualization there are three main properties, they're what? How? Why? Here we visualize the info to make a tabular column for descriptive and predictive analysis. Now there are two processes which can occur after data visualization, Firstly, the prediction process occurs using the multi and binary logistic regression classes. Subsequent one is, it predicts using both classification and regression, but here we are using only classification algorithm because the anticipated results will have more accuracy as compared to using both the regression and classification. supported the logistic regression classes and classification algorithm, the two-algorithm used here are referred to as Logistic Regression and Random Forest Algorithm. Now both the algorithm is being sent to testing and training process. Data Training is a crucial step within the machine learning process. Data Training is employed to rework the information during a helpful and proficient organization. The data can have numerous unessential and missing parts. To deal with this part, information cleaning is finished. It includes the strategy for treatment of the missing information and in this manner the uproarious information and so on. The testing process approaches when training, every tree learns from a random pattern of the data points. The samples are drawn with replacement, referred to as bootstrapping, which suggests that some samples are going to be used a couple of instances for the duration of a single tree. At check time, predictions are made with the aid of averaging the predictions of each choice tree.

I. MODULES

- Data Collection
- Data Pre Processing
- Data Visualization
- Random Forest Classification
- Training Process
- Testing Process
- Logistic Regression Classes
- Training Process
- Testing Process
- Machine Learning Model

DATA PRE-PROCESSING-Data learning is a crucial step within the process of machine learning. Data pre-processing is employed to convert data into useful and efficient form. The info may contain many inappropriate and lacking parts. To cope with this area, statistics is cleaned up. This consists of dealing with lacking data, noisy data, and so on.



DATA VISUALIZATION-Data visualization may be a technique that uses a series of static and interactive displays during a specific environment, enabling people to know and interpret large amounts of knowledge. Data is usually displayed in narrative form, displaying patterns, trends and interactions which will go unnoticed.

RANDOM FOREST ALGORITHM-Random Forest may be a supervised learning process. Random forest classification is often used for both regression and regression problems, and random forest is often utilized in regression problems. But we have used random forests in classification during this project, so we'll only consider the classification area.

DATA TRAINING MODULE-Data training is a crucial step within the machine learning process. Data training is employed to convert data into effective and efficient form. The info may contain many irrelevant and missing parts.

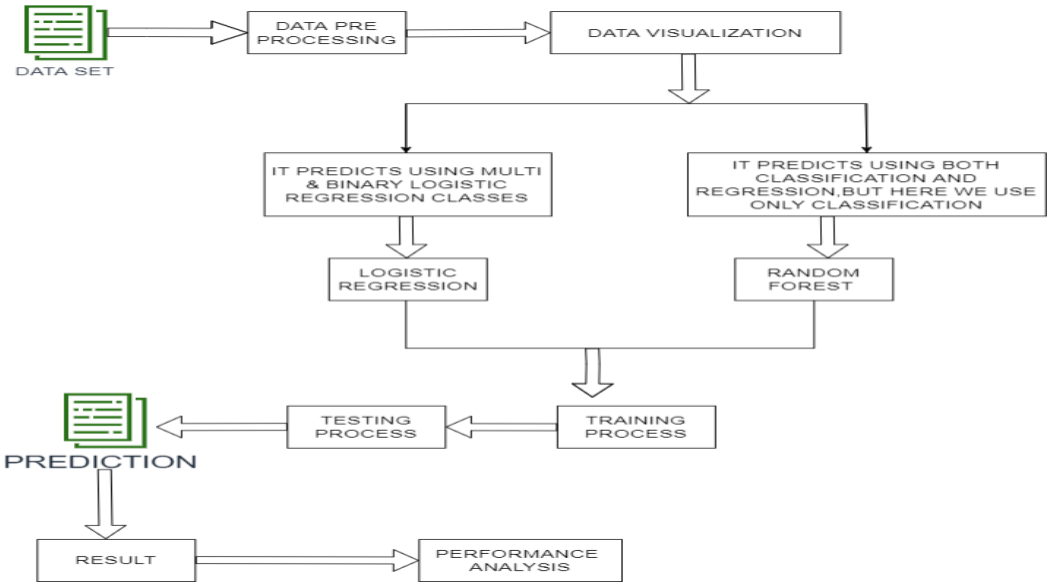
DATA TESTING MODULE-During training, each tree during a random forest learns from a sampling of knowledge points. Models are drawn by alternatives, called bootstrapping, which suggests that some models are used multiple times during a tree. During testing, predictions are made by averaging the results of every decision tree

International Journal of Research (IJR)
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 9 Issue 02
February 2022

LOGISTIC REGRESSION ALGORITHM-Logistic regression may be a classification method won't to assign observations to a group of distinct classes. Unlike rectilinear regression that outputs continuous numerical values, the logistic regression returns its output using the logistic sigmoid function, which may then be converted to 2 or more distinct classes.

DATA TRAINING MODULE-Data training is a crucial step within the machine learning process. Data training is employed to convert data into effective and efficient form. The information may contain many irrelevant and missing parts. To handle this area, data is cleaned up. It involves the method of handling missing data and quieter data.

DATA TESTING MODULE-During training, each tree during a random forest learns from a sampling of knowledge points. Models are drawn by alternatives, called bootstrapping, which suggests that some models are used multiple times during a tree. During testing, predictions are made by averaging the results of every decision tree.



REVIEW CHART

S.No	Author	Year	Algorithm used	Dataset
1	V.V. Ramalingam, Ayantan Dandapath, M.Karthik Raja	2018	Support Vector Machines, K-Nearest Neighbour NaïveBayes, Decision Trees (DT), Random Forest (RF)	Cleveland Dataset
2	Nagaraj M. Lutimath, Chethan C, Basavaraj S Pol	2019	Naïve Bayes Classification and Support Vector Machines	UCI Machine Learning Data Set
3	R.G.P.V. Bhopal	2016	A Hill Climbing Algorithm	Cleveland

4	Praveen Kumar Reddy M, T Sunil Kumar Reddy, Balakrishnan, Syed Muzammil Basha Ravi Kumar Poluru	2019	Decision Tree Algorithm Support Vector Machine Algorithm	Kernel
5	Reddy Prasad, Pidaparthi Anjali, S.Adil, N.Deepa	2019	Logistic Regression Algorithms	Uci Machine Learning Data Set

CONCLUSION-In the above paper we made a comparative study of heart condition using various machine learning techniques. The above review, it are often concluded that there's an enormous scope for machine learning mechanisms of heart condition. Each of the above steps worked fine in some cases, but in another cases. The most purpose of this study is to predict heart condition using different machine learning methods.

REFERENCES

[1]. 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.[1]

[2]. Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, K. (2018, March). Prediction of heart disease using machine learning. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1275-1278). IEEE.[2]

[3]. Saw, M., Saxena, T., Kaithwas, S., Yadav, R., & Lal, N. Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning.[3]

[4]. Poluru, R. K. Heart Disease Prediction Using Machine Learning Algorithm.[4]

[5]. Ghumbre, S. U., & Ghatol, A. A. (2012). Heart Disease Diagnosis Using Machine Learning Algorithm. In Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012 (pp. 217-225). Springer, Berlin, Heidelberg.[5]

[6]. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.

[7]. Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari, V. (2018, March). Prediction of cardiovascular disease using machine learning algorithms. In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-7). IEEE.

[8]. J. S. Singh Navdeep, -No Title,| Intell. Hear. Dis. Predict. Syst. using CANFIS Genet. Algorithm, Int. J. Biol. Med. Sci., no. International Journal of Advance Research, Ideas and Innovations in Technology© 2018,

[9]. R. G. S. Rajkumar Asha, -No Title,| Diagnosis Hear. Dis. Using Datamining Algorithm, vol. Issue 10 V, no. Global Journal of Computer Science and Technology, p. Page 38, 2010.

[10]. S. Radhimeenakshi and G. M. Nasira, -Prediction of Heart Disease using Neural Network with Back Propagation,| Int. J. Data Min. Tech. Appl., vol. 4, no. 1, pp. 19– 22, 2016.

[11]. -Prediction of Heart Disease using Artificial Neural Network,| VFAST Trans. Softw. Eng., pp. 102–112, 2019.

[12]. Z. M. A. Mehta Rupa G. ,Rana Dipti P., -No Title,| A Nov. Fuzzy Based Classif. Data Min. using Fuzzy Discret., no. World Congress on Computer Science and Information, 2009.

[13]. A. Shinde, S. N. Kale, R. M. Samant, A. S. Naik, and S. A. Ghorpade, -Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network,| Int. J. Comput. Appl., vol. 166, no. 7, pp. 975– 8887, 2017.

[14]. N. Al-Milli, -Backpropogation neural network for prediction of heart disease,| J. Theor. Appl. Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[15]. G. S. Kumari Milan, -No Title,| Comp. Study Data Min. Classif. Methods Cardiovasc. Dis. Predict., vol. Vol. 2, no. IJCST Vol. 2, Iss ue2, June 2011, 2011.

[16]. Ordonez, -Improving Heart Disease Prediction using Constrained Association Rules,| Tech. Semin. Present. Univ. Tokyo, 2004.

[17]. M. C. and P. M. Franck Le Duff, CristianMunteanb, -Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method,| Stud. Health Technol. Inform., vol. Vol. 107, no. 2, p. No. 2,

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 9 Issue 02
February 2022

pp. 1256–1259, 2004.