JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Prediction of Heart Disease Using Machine Learning Classifiers

[1]Vangalapudi Jahnavi, [2]Angaraju Chaitanya Sridhar, [3]A. Sai Chandra Rahul, [4]Jajimoggala Krishna Kumar, [5]Dr. Bode Prasad

Department of Information Technology
Vignan's Institute of Information Technology (VIIT), Visakhapatnam, India

Abstract:

The Heart is an important key organ of the human body which supplies blood into all the other constituents of the human body. A lot of people had been suffering and dying from various heart diseases making it the No 1 in the Cause of Death in the world. The people who are dying from it is being swiftly increasing every year. On average, it takes 17.9 millions of lives per year and still increasing enormously. For the diagnosis of heart disease it costs a huge amount of money for a person and many can't afford it, there has to be a practice that can be helpful for the people and with the help of this practice they can detect the chances of having a heart disease. Therefore, we came up with an idea of developing a web application that can be used by people which helps them to detect their chances of having a heart disease, this system will take various attributes. In this, we took various Machine Learning Classifiers which are considered as best in producing accurate results and implemented them.

Keywords: heart, disease, data, machine learning, classifiers.

Introduction:

As the chances of having a heart disease is quickly growing we are making it to be available each and every person by building it as a webpage. Many Health Organizations giving their best efforts for identifying and treating heart diseases, yet the number is still increasing. Therefore, we came up with a web application which can be used by everyone which helps them for identifying their chances of detection of heart disease, this system will take various attributes (Symptoms) for it, which are the results of few tests performed. In this, we are using 3 major machine learning classifiers that are Decision Trees, Random Forest and Extreme Gradient Boosting classifiers for making the prediction of heart disease. We are using the proposed model which is developed as the combination of above three classifiers which produces the best results.

Collecting the dataset from Kaggle and exploring the dataset as per the requirements. For the training and testing of these 3 models the dataset with 14 attributes which contains 1318 data records that is 1318 rows x 14 columns which later divided into 70%-30% where 70% of data records is used in training phase and another 30% is used in testing the trained model to get their accuracy. All three models had given 3 various accuracies with 99% for extreme Gradient Boosting, 98% for Random Forest and 97% for the Decision Trees.

Literature Survey:

Senthil Kumar Mohan et al,[1] proposed Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in which strategy that objective is to finding critical includes by applying Machine Learning bringing about improving the exactness in the expectation of cardiovascular malady. We produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with a linear model (HRFLM). Sonam Nikhar et al [2] has built up the paper titled as Prediction of Heart Disease Using Machine Learning Algorithms by This exploration plans to give a point by point portrayal of Naive Bayes and decision tree classifier that are applied in our examination especially in the prediction of Heart Disease. Abhay Kishore et al,[3] developed Heart Attack Prediction Using Deep Learning in which This paper proposes a heart attack prediction system using Deep learning procedures. Avinash Golande et al,[4] proposed Heart Disease Prediction Using Effective Machine Learning Techniques in which Specialists utilize a few data mining strategies that are available to support the authorities or doctors distinguish the heart disease. V.V. Ramalingam et Al,[5] proposed Heart disease prediction using machine learning techniques in which Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Lakshmana Rao et al,[6] Machine Learning Techniques for Heart Disease Prediction in which the contributing elements for heart disease are more (circulatory strain, diabetes, current smoker, high cholesterol, etc..). So, it is difficult to distinguish heart disease. Different systems in data mining and neural systems have been utilized to discover the seriousness of heart disease among people. Mr. Santhana Krishnan.J and Dr. Geetha.S, [7] Prediction of heart disease using machine learning algorithm This Paper predicts heart disease for Male Patient using Classification Techniques.

Existing system:

Mostly the models or algorithms which are used recently gives good accuracy, but those are not the best. Still it needs to be improved further. It uses the machine learning algorithms like decision trees, support vector machine etc. for the prediction of heart disease with average accuracy.

Proposed System:

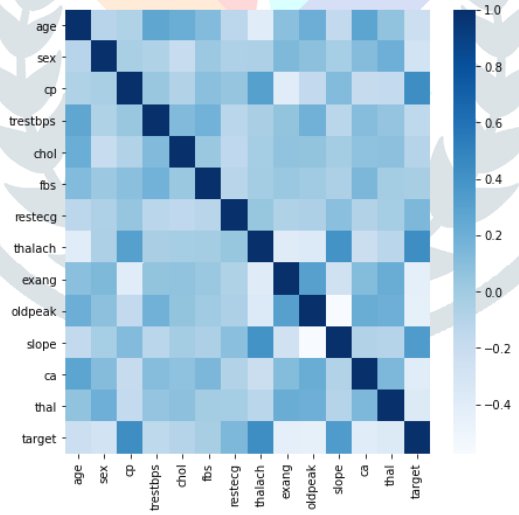
Although many heart disease prediction systems uses different machine learning classifiers with different accuracies produced, we came up with an idea of using machine learning classifier in our system which consists of a most recently introduced technology Extreme

Gradient Boosting along with Decision Trees and Random Forest. The dataset also play a vital role in our system. The dataset should be pure and do not produce any kind of overfitting of data which results in inaccurate results. For the removal of overfitting data from the dataset we use data preprocessing methods for cleaning of the dataset. After the dataset is clean we then split the data for both the training and testing of the machine learning classifiers. After the models are trained, we deploy them as a website which uses all the classifier to produce final prediction of chances for having a heart disease. With this system we can identify any disease affecting one's heart so that he/she can consult the doctors and cure it before it is too late.

Dimensionality Reduction:

It is essential to have a dimensionality reduction phase where the data preprocessing and feature extraction and selection is being performed. Many datasets consists of overfitting of data which should be removed as soon as possible. Data preprocessing is about the cleaning of unwanted data from the dataset before the training phase only this involves in actions to be performed for removal of null values or any values which are being out of range. Feature extraction is about drawing the new feature from the existing features. Feature selection is for the selection of the important features which contribute the maximum for prediction of the output as not all the features contribute equally.

The correlation of the data can be analyzed graphically as:



Methodology:

First the data will be going through the data preprocessing method as mentioned above. After that we will be feeding it to our models. In our model we used Decision trees, Random forest and Extreme Gradient Boosting (xgboost) are the three different machine learning algorithms. These algorithms are considered as ensemble techniques. We took these 3 classifiers as they produce the best results for our system and all three algorithms are combined as a model for producing the best results.

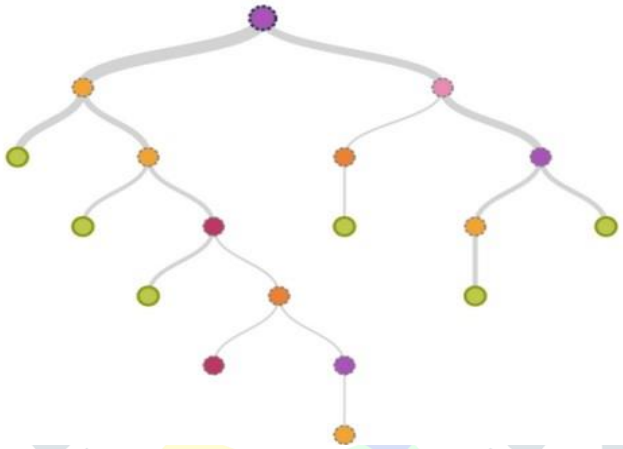
A. Decision tree:

As the name itself says tree, this algorithm is a tree based algorithm where every node is considered as a decision node where it has 2 paths for it. The paths are based on the condition present inside the decision node if the condition is true then it chooses one path otherwise it chooses a path assigned for false one. The leaf node will be a result. Decision tree also calculates the information gain and entropy of all the attributes present in the dataset provided. This algorithm when trained with the data achieved an accuracy of 97% after tested.

Entropy and information gain can be calculated as

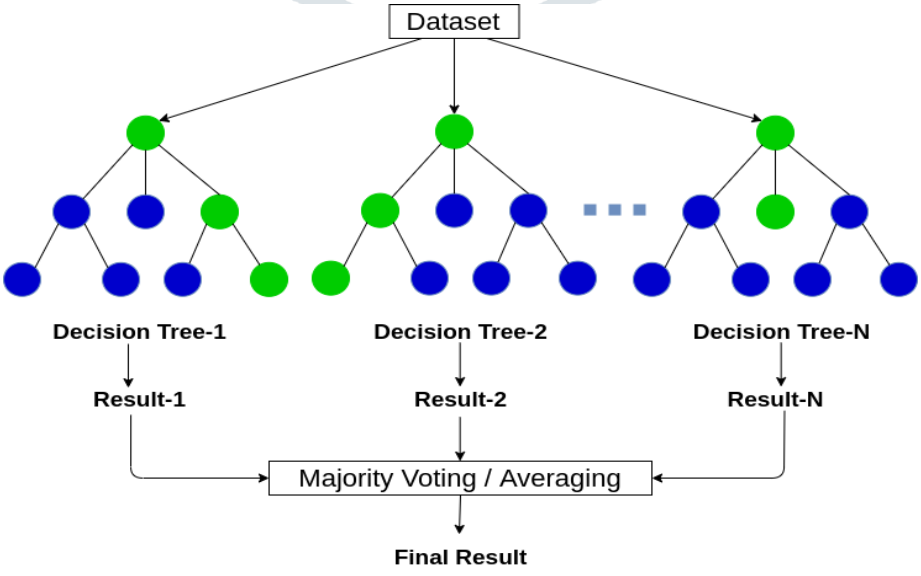
$$E = -\sum_{i=1}^{N} p_i log_2 p_i$$

$$Gain = E_{parent} - E_{children}$$



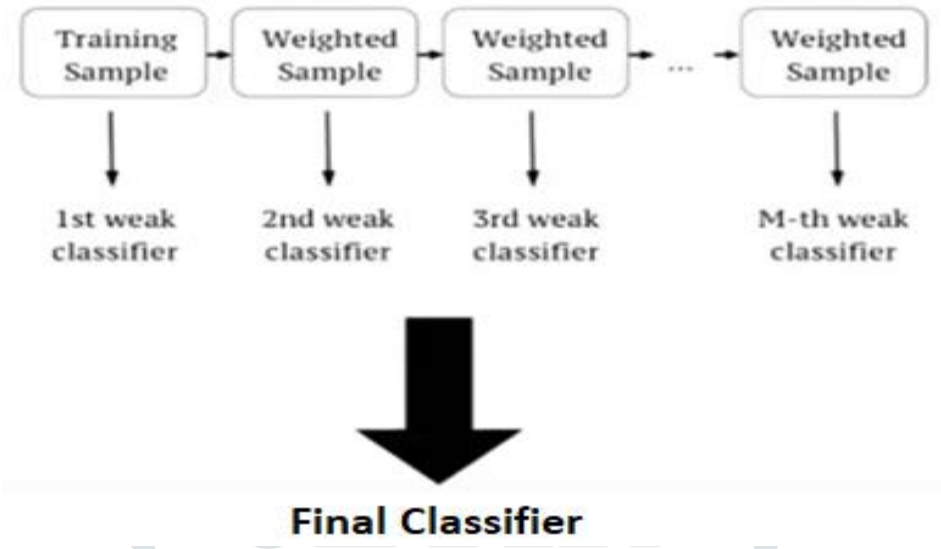
B. Random Forest:

It is a combination of trees which are arranged in a format of bagging. The bagging means splitting the data into number of subsets which equals to the number of decision trees present in it and train them all and produce the result using voting of the results we got from all the trees. We achieved an accuracy of 98% of training data after being tested.



C. eXtreme Gradient Boosting:

Xgboost is alike of random forest and also known as extreme Gradient Boosting but the formation of the trees makes the change. It follows the boosting concept for the formation. Boosting concept is about forming the trees in a serialized manner. This algorithm when trained with the data achieved an accuracy of 99% after tested.



Results:

In our model, we got an accuracy of 99% by our proposed model and xgboost classifier, 98% accuracy by Random Forest, 97% accuracy by Decision Trees which is greater than the previous models and the result will be displayed using the proposed model.

Conclusion:

We have built the machine learning model using the help of datasets. In this model we used decision tress, random forest and xgboost classifiers and mainly the models involve of predicting whether a person has heart disease or not. We got best accuracy with models and prediction of results is done using our proposed model, which is developed as the combination of three classifiers.

Future work:

In future this model can develop and add more additional features to the project. It can be developed under the combination of different prediction models and we can get the best output.

References:

[1] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[2] Nikhar, Sonam, and A. M. Karandikar. "Prediction of Heart Disease Using Machine Learning Algorithms." *International Journal of Advanced Engineering, Management and Science* 2, no. 6: 239484.

[3] Kishore A, Kumar A, Singh K, Punia M, Hambir Y. Heart attack prediction using deep learning. International Research Journal of Engineering and Technology (IRJET). 2018 Apr;5(04):2395-0072.

[4] Avinash Golande, Pavan Kumar T," Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.

[5] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja," Heart disease prediction using machine learning techniques: a survey", International Journal of Engineering & Technology, 7 (2.8) (2018) 684-687.

[6] A.Lakshmanarao, Y.Swathi, P.Sri Sai Sundareswar," Machine Learning Techniques For Heart Disease Prediction", International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.

[7] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019, pp. 1-5, doi: 10.1109/ICIICT1.2019.8741465.

[8] V. Ramalingam, V, Ayantan Dandapath, & M Karthik Raja. "Heart disease prediction using machine learning techniques : a survey." *International Journal of Engineering & Technology* [Online], 7.2.8 (2018): 684-687. Web. 15 May. 2022.

[9] A. H. Chen, S. Y. Huang, P. S. Hong, C. H. Cheng and E. J. Lin, "HDPS: Heart disease prediction system," *2011 Computing in Cardiology*, 2011, pp. 557-560.

[10] R. Valarmathi, T. Sheela, Heart disease prediction using hyper parameter optimization (HPO) tuning, Biomedical Signal Processing and Control, Volume 70,2021, 103033, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2021.103033.