

Implementation of Heart Diseases Prediction System using Combination of XGBoost, Logistic Regression and Random Forest

Prof. Pushpa Chutel¹, Govind Thawkar², Abhishek Gupta³
Aman Thakur⁴, Abhishek Shukla⁵, Anish Karhade⁶, Prashil Meshram⁷

Department of Computer Science and Engineering¹⁻⁷

G. H. Rasoni Institute of Engineering and Technology, Nagpur, Maharashtra, Bharat

Abstract: *In this modern times, Heart Disease prediction is one of the most critical tasks in the world. In recent times, a lot of people have died due to heart disease. Machine learning plays a very important role in training and testing the huge amount of data in the medical field. Heart disease prediction is a crucial task to create and evaluate the prediction process to avoid heart disease and alert the patient before he/she suffers from disease. This research predicts the chances of Heart Disease and says whether the patient has heart disease or not by implementing different machine learning techniques such as Decision Tree, Logistic Regression. Finally, this study shows a result of heart disease and Results are obtained and comparative experiments have shown that the proposed approach can be utilized to give the prediction to the patient. This research predicts the chances of Heart Disease and says whether the patient has heart disease or not by implementing different machine learning techniques such as Decision Tree, Logistic Regression. Finally, this study shows a result of heart disease and Results are obtained and comparative experiments have shown that the proposed approach can be utilized to give the prediction to the patient.*

Keywords: Machine Learning, Heart Disease, KN, Logistic Regression, Random Forest Heart Risk, Classification Algorithm.

I. INTRODUCTION

The work proposed in this model focuses mainly on various methods that are employed in heart disease prediction. In the human body the heart is the main role and it regulates the blood to the whole body. Basically if the heart can't regulate proper blood it causes a huge problem to the body. Any misleading things can affect the heart disease and also the chance of getting a heart stroke. In today's modern era, heart disease is one of the primary reasons for common deaths in this generation due to their luxury and unhealthy lifestyle like huge alcohol, fast food fat food and smoking and stress. World Health Organization said that in every year lakhs of people are suffering from this heart disease and they are losing their lives. A good and healthy measures can save from the heart disease earlier. The main effective is need to improve to create prediction system and help the poor to save from the lives. Heart diseases are found as the prime source of death in the world due to modern era luxury and unhealthy food. The main goal of this model is to provide a tool for doctors to detect heart disease at an early stage. This model will help to prevent and detect the patients earlier from the heart disease

II. LITERATURE SURVEY

Paper 1. In this paper we were introduced to a ML model comparing five types of different algorithms. A Rapid Miner tool was used which resulted in higher accuracy compared to Matlab software and Weka tools for data mining. In this research the results of Decision Tree, Logistic Regression, Random forest, Naive Bayes and SVM classification algorithms were used. Decision tree algorithm comes with the highest accuracy[1].

Paper 2. The aim of this study was to identify the most important risk factors for cardiovascular disease (CVD) in patients with metabolic-associated fatty liver disease (MAFLD) using machine learning (ML) techniques. 191 MAFLD

patients had their blood biochemically analysed, and subclinical atherosclerosis was evaluated. ML techniques, such as multiple logistic regression classifier, univariate feature ranking, and principal component analysis (PCA), were used to construct a model to determine people who had the highest risk of heart disease related complications [2].

Paper 3. The suggested system, which use a quantum neural network to learn and recognise patterns of CVD, was experimentally validated and compared with the FRS using data from 689 people who experienced symptoms of CVD and a validation dataset from the Framingham research. [3].

Paper 4. This paper introduced a model designed by applying Artificial Neural Networks. This paper used the electronic health record data from real-world datasets related to patients' heart disease to perform the analysis and predict the heart disease. We implemented a one-hot encryption model that diagnoses events and heart risk failure events victimization, the essential principles of an expanded memory in the neural network model. By analyzing the results, we predicted to reveal the importance of respecting the results of nature in the records. [4]

III. PROPOSED METHODOLOGY

This paper shows the analysis of various machine learning algorithms, the algorithms that are used in this paper are XGBoost, Logistic Regression and Random Forest Classifiers which can be helpful for practitioners or medical analysts for accurately diagnose Heart Disease. The methodology is a process which includes steps that transform given data into recognized data patterns for the knowledge of the users. The proposed methodology (Figure 1.) includes steps, where first step is referred as the collection of the data than in second stage it extracts significant values than the 3rd is the preprocessing stage where we explore the data. Data preprocessing deals with the missing values, cleaning of data and normalization depending on algorithms used. After pre-processing of data, classifier is used to classify the pre-processed data the classifier used in the proposed model are XGBoost, Logistic Regression, Random Forest Classifier. Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics. Here in this model, an effective heart disease prediction system gas been developed using different classifiers.

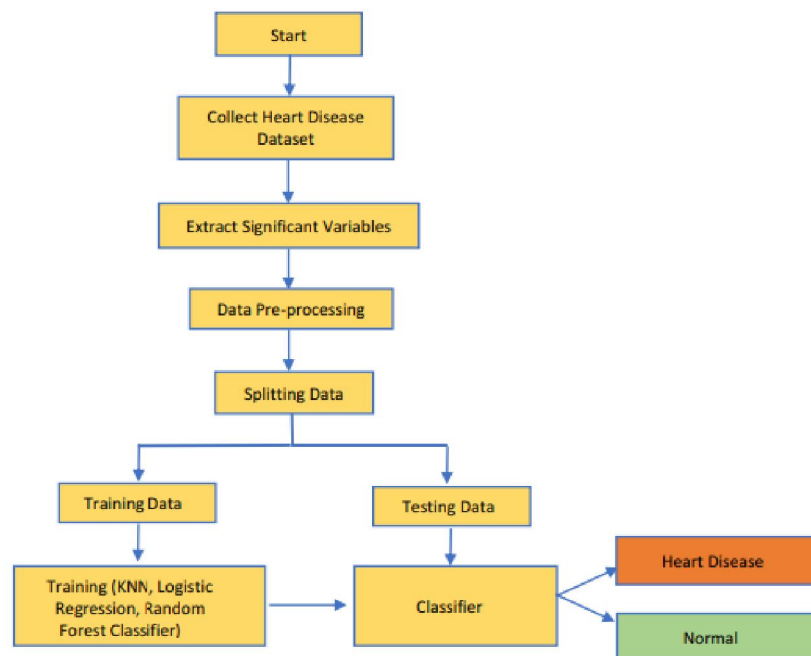
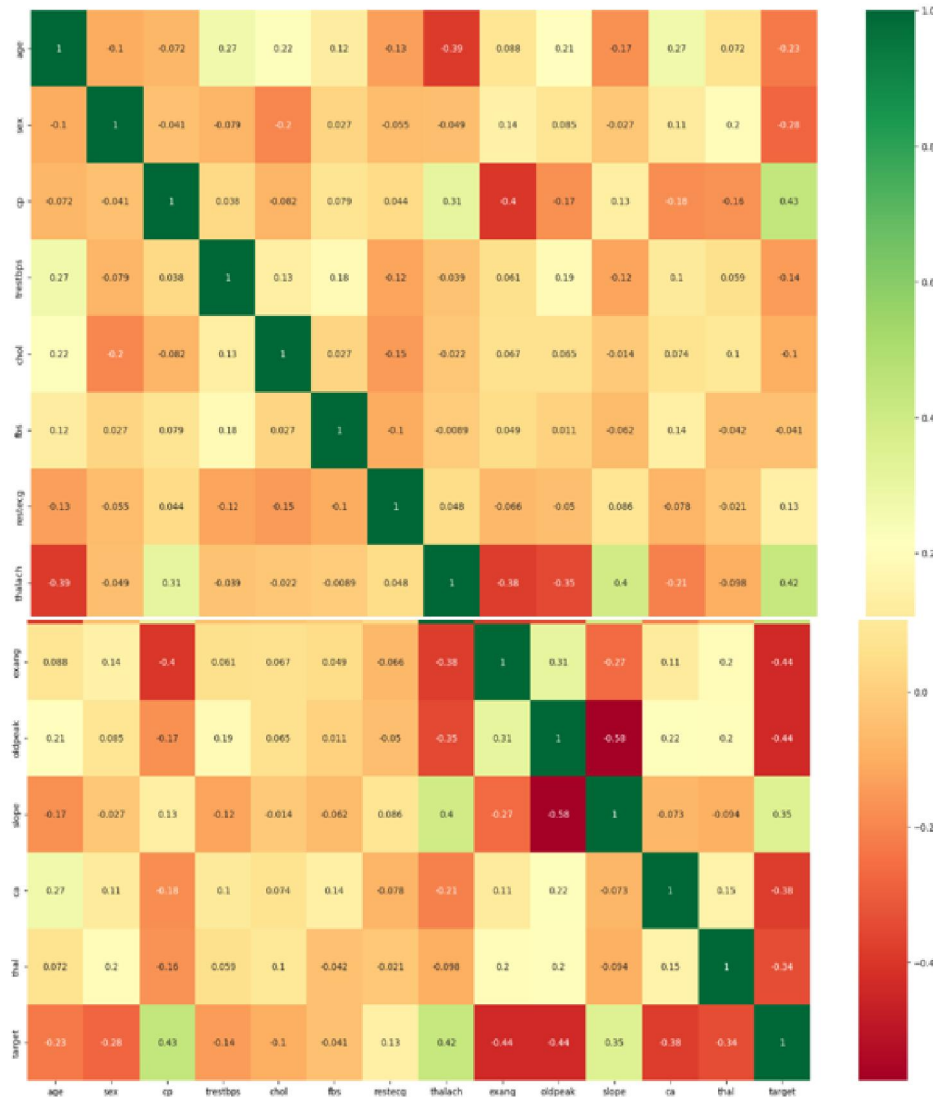


Figure 1. Proposed Model

IV. IMPLEMENTATION AND OUTCOME

From these results we can see that although most of the researchers are using different algorithms such as SVC, Decision tree for the detection of patients diagnosed with Heart disease, XGBoost, Random Forest Classifier and Logistic regression yield a better result to out rule them. The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by XGBoost and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from previous researches. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took.



Our project also tells us that Logistic Regression and XGBoost outperforms Random Forest Classifier in the prediction of the patient diagnosed with a Heart Disease. This proves that XGBoost and Logistic Regression are better in diagnosis of a heart disease

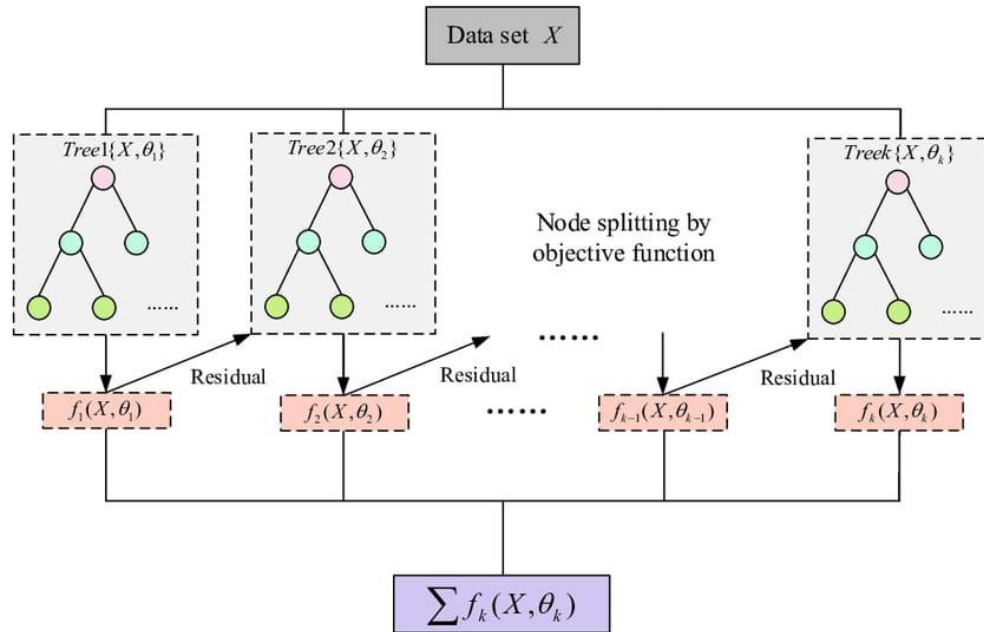


Figure. FlowGraph of XGBoost

V. CONCLUSION

This component will help in predicting the severity of the heart stroke/cardiovascular disease. After the successful model user will input data, the weights will be cross checked with the given inputs. The prediction of this heart disease system will consist of 13 attribute values that will be input to the system. The target value is zero or one The predicted will be generated in the form of a ‘yes’ or ‘no’ format considering all the risk factors whether they lie in the criteria as per the model is trained.

REFERENCES

[1] “Implementation of Machine Learning Model to Predict Heart Failure Disease”(Fahd Saleh) (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019
 [2] “Risk factors for cardiovascular disease in patients with A machine learning approach”.(Drozdź, K.; Nabrdalik.) IEEE Access 2022
 [3]“Early prediction of Paroxysmal measures of heart rate variability” (Narin, A.; Isler, Y.; Ozer, M.) 2016 Medical Technologies National Congress (TIPTEKNO)
 [4] Bo Jin, Chao Che (2018)“Predicting the Risk of Heart Disease With EHR”