



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Heart Disease Prediction using ML and AWS

Shanmugapriya, Pavan S, Manoj T, Anish N, Nishant Adityan

Assistant Professor, Dept. of CSE., Dr Ambedkar Institute of Technology, Bangalore, India

UG Student, Dept. of CSE., Dr Ambedkar Institute of Technology, Bangalore, India

UG Student, Dept. of CSE., Dr Ambedkar Institute of Technology, Bangalore, India

UG Student, Dept. of CSE., Dr Ambedkar Institute of Technology, Bangalore, India

UG Student, Dept. of CSE., Dr Ambedkar Institute of Technology, Bangalore, India

ABSTRACT: Cardiovascular diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. In this project, we have developed and researched about models for heart disease prediction through the various heart attributes of patient and detect impending heart disease using Machine learning algorithm logistic regression on the dataset available publicly in Kaggle Website, further evaluating the results using confusion matrix and cross validation. The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine.

KEYWORDS: UCI Machine Learning dataset, Linear Regression

I. INTRODUCTION

The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in an attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. If it is not diagnosed in the early stages itself the condition of the person will get worse and eventually a person might die due to the reluctance. The proposed solution is to develop a machine learning model to predict the possibility of heart disease by implementing a Logistic Regression algorithm and to make it available to everyone by hosting the system in amazon AWS. To determine significant risk factors based on medical dataset which may lead to heart disease. Our proposed system will take inputs from the users, there will be thirteen attributes which will be filled by the user and based on the values which are being filled by the users the system will predict whether the person is having a heart disease condition or not.

II. RELATED WORK

One of the research work which was published by Baban Et al. [1] has used dataset which is available from UCI Machine Learning Respiratory. Dataset consists of 303 sample data in which 14 input features as well as 1 output feature. Using this dataset they have proposed a system by using 3 Machine Learning algorithms which are Random forest, Support Vector Machine, Artificial Neural Network. So the result of their project after implementing 3 algorithms was found to be 84% for Support Vector Classifier, 83.5% for Neural Network, 80% for Random Forest Classifier. So they have concluded that the techniques that are studied and compared for finding the efficient and accurate systems.

In another work which was published by Harshit Jindal Et al. [2] have taken the data source which was similar to the previous paper that is from the UCI Machine Learning Respiratory. So the algorithms which they have used are KNN, Logistic Regression and Random Forest Classifier. So in their result they have told that KNN and Logistic Regression were having maximum accuracy of 88.5%. So they have concluded that by using Logistic Regression and KNN they got more accuracy than Random Forest Classifier.

In another work which was published by Pooja Anbuselvan [3] have taken the data source which was similar to the previous paper that is from the UCI Machine Learning Respiratory. So in this paper she has referred to 7 of the Machine Learning Algorithms which are Logistic Regression, Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, XG-Boost. So the aim of this project was to analyze the performance of various classification algorithms and in doing so find the most accurate algorithm for predicting whether the patient would develop a heart disease or not. After working with all the 7 algorithms she has found that Random Forest has the highest accuracy with 86.89% and the least accurate was found to be KNearest Neighbor.

In another work which was published by Rishabh Magar Et al [4] in which they have used Machine Learning Respiratory to test data on different data mining techniques. So they have created a web page machine application which is trained by using dataset. The user inputs specific medical data details to get accurate prediction for heart disease for the user. So they have used Support Vector Machine, Decision Trees, Naives Bayes, Logistic Regression as their Machine Learning Algorithm. So the result was found to be a Logistic Regression algorithm having the best accuracy of 82.89%. Decision tree and Naives bayes having 80.43% and 80.43% respectively and SVM having 81.57%.

III. PROPOSED ALGORITHM

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no).

Mathematically, a logistic regression model predicts $P(Y=1)$ as a function of X . It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc

In Logistic regression, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1)

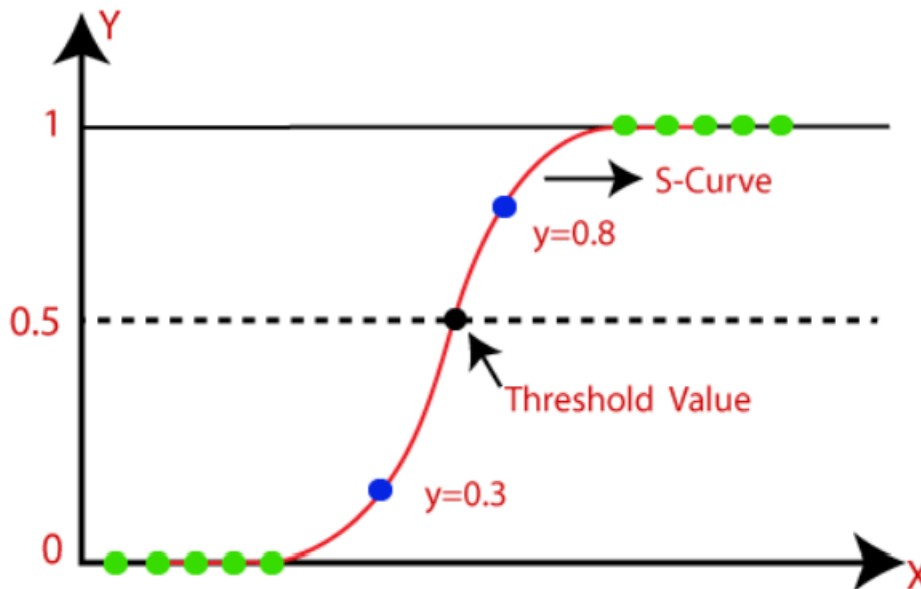


Figure 1: Graph for Logistic Regression

The general mathematical equation for logistic regression is,

$$y = \frac{1}{1 + e^{-(a + b_1x_1 + b_2x_2 + b_3x_3 + \dots)}}$$

Where, y is the dependent variable.

a and b are the coefficients. x is the independent variable.

where k is constant and n is path loss factor which is generally between (2-4) [8].

IV. SIMULATION RESULTS

The result for our proposed system will predict whether the person is having an heart disease or not. The Logistic Regression in the backend of our model will run the data which is being provided by the user and looking at the attributes the Logistic Regression will determine whether the person is having an heart disease or not.

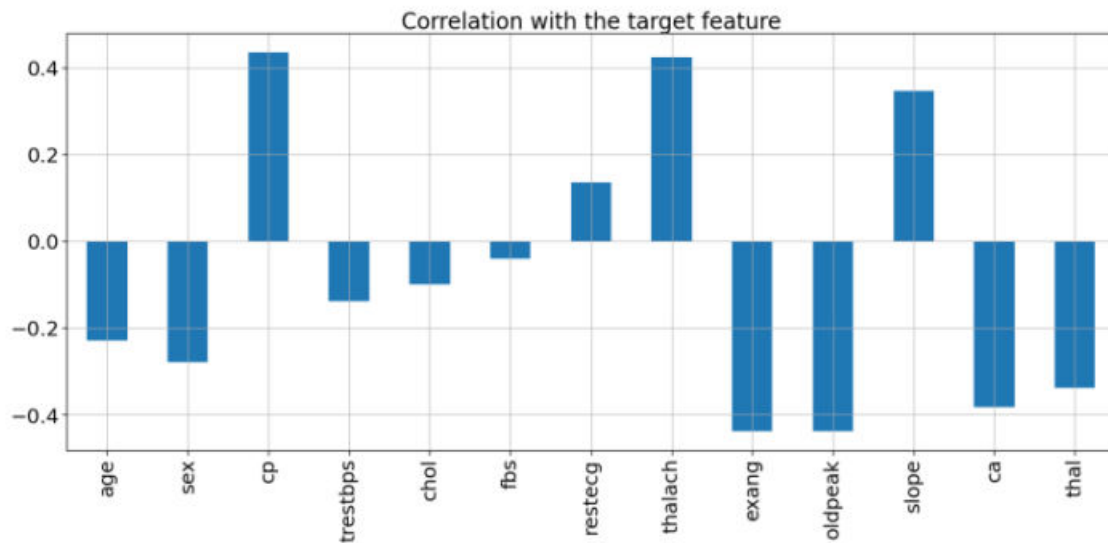


Figure 2 Correlation with the target feature

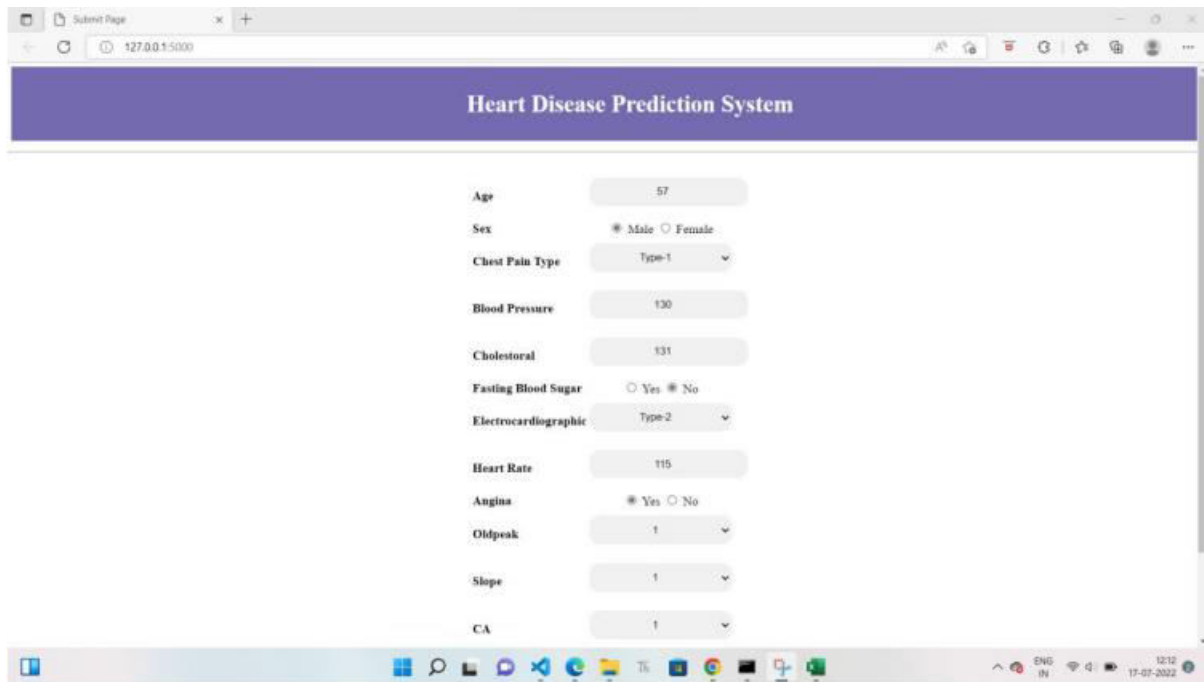


Figure 3: Web page of the proposed system

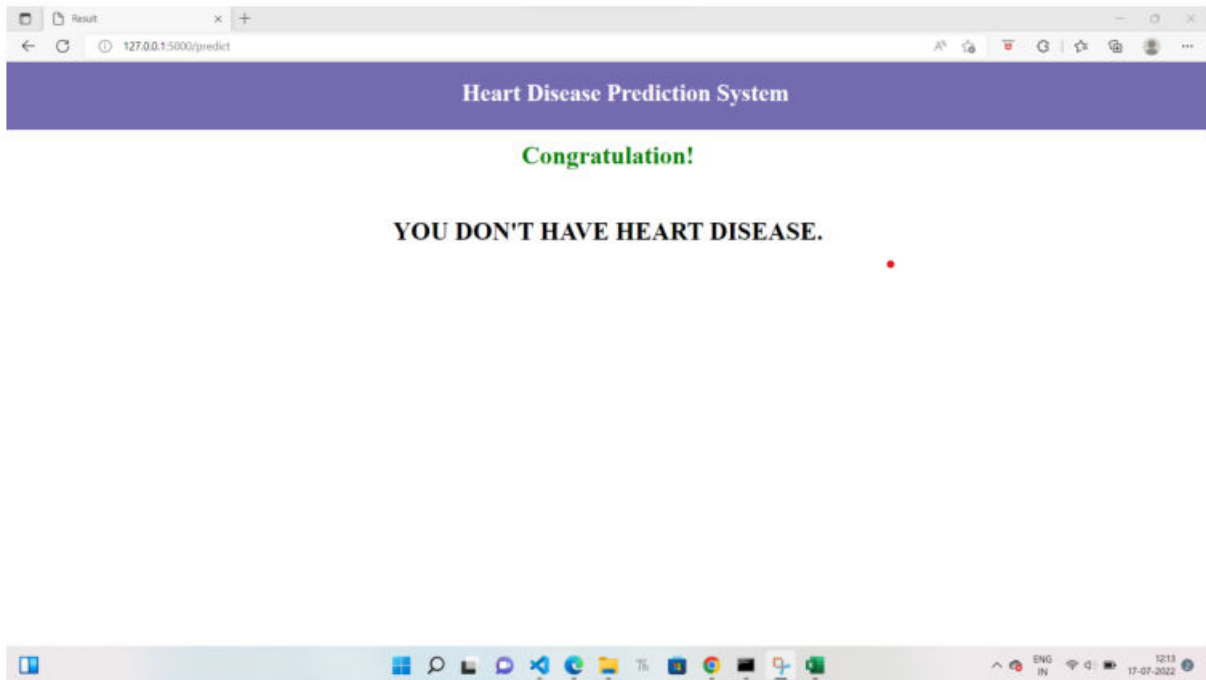


Figure 4: Result for having heart disease

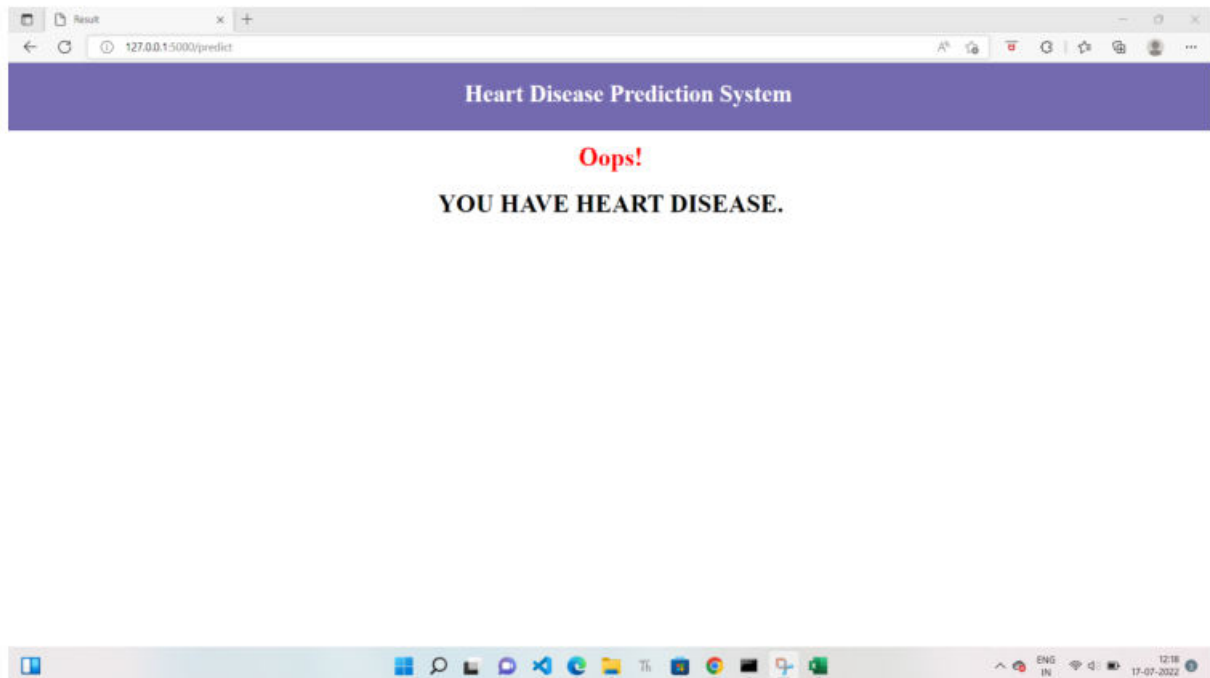


Figure 5: Result for not having heart disease

V. CONCLUSION AND FUTURE WORK

The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. So, in our project we have used two algorithms which are Logistic Regression and K-Nearest Neighbors which resulted in favor of Logistic Regression with the accuracy score of 85% and for KNN it was 78%. So, we decided to continue with Logistic

Regression for our project development. We deployed our system on AWS beanstalk so it could be available for everyone. For future enhancement, we have only used two of the machine learning algorithms out of many which are there. Exploring many algorithms would be an enhancement and also using different dataset which has different attributes other than the one which we have used.

REFERENCES

1. Heart Disease Prediction using machine learning techniques -Pooja Anbuselvan
2. Heart disease prediction using machine learning algorithms -Harshit Jindal, Sarthak Agrawal, Rishabh Khara, Rachna Jain and Preeti Nagrath.
3. Heart Disease Prediction Using Machine Learning-Baban.U. Rindhe¹ , Nikita Ahire² , Rupali Patil³ , Shweta Gagare⁴ , Manisha Darade⁵.
4. Heart Disease Prediction Using Machine Learning. Rishabh Magar, Rohan Memane, Suraj Raut.
5. A.N.Repaka, S.D. Ravikanti and R.G. Franklin, "Design and Implementing Heart Disease Prediction Using Naive Bayesian," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019.
6. S.Mohan, C.Thirumalai and G.Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access,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), CHENNAI, India, 2019.
8. A.Gavhane, G.Kokkula, I.Pandya and K.Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018.
9. P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018.
10. C.T. and A. Choudhary, "Heart Disease Diagnosis using a Machine Learning Algorithm," 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, India, 2019.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.165



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details