International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Prediction of Heart Disease Using Logistic Regression

Abhijna Bhat[1], Pragathi[2], Pranamya M S[3], Smitha[4]

[1, 2, 3]UG student, Dept. of Computer Science and Engineering, Yenepoya Institute of Technology, Karnataka, India
[4] Assistant Professor, Dept. of Computer Science and Engineering, Yenepoya Institute of Technology, Karnataka, India

---

Abstract - *Over the last decade heart disease is one of the main reasons for death in the world. The heart is the next major organ comparing to the brain which has more priority in the human body. It pumps the blood and supplies to the whole body. Nowadays, heart diseases are increasing day by day due to lifestyle, hereditary. In this paper, a new method is proposed that aims at finding significant features by applying machine learning techniques resulting in improving accuracy. In this, we use a logistic regression classification algorithm to classify the risk level. We consider a dataset with features like age, gender, blood pressure, etc. Using this database, we have to create a model which predicts the heart disease using logistic regression*

Key Words: **Machine learning, predicting heart disease, prediction model, classification algorithm, cardiovascular disease**

1. INTRODUCTION

Heart disease is one of the most significant causes of death in the world. Each one has different values for Blood pressure, cholesterol, and pulse rate. According to medically proven results, the normal values of Blood pressure are 120/90, cholesterol is and pulse rate is 72. Many research works have been done to predict the heart disease using data mining classification techniques such as KNN (K Nearest Neighbors), Decision tree Algorithm, Neural Network, etc., Heart rate data were collected through the internet and collected in server and it takes one year to get enough data to make predictions Prediction of heart disease is a challenge in the area of clinical data analysis. Machine learning (ML) is effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. It is very difficult to identify heart disease because of several risk factors such as diabetes, high blood pressure, high cholesterol and many other factors. The nature of the cardiovascular disease is complex and hence, the disease must be handled carefully. Not doing so may affect the heart and in worst case premature death.

Here we use a classification algorithm called Logistic regression. It is, used when the value of the target variable is categorical. It is most commonly used when the data in question has binary output, so when it belongs to one class or another, or is either a 0 or 1. By the idea of using a regression model to solve the classification problem, we rationally raise a question of whether we can draw a hypothesis function to fit the binary dataset. The answer is that you will have to use a type of function, different from linear functions, called a logistic function, or a sigmoid function. It also provides high accuracy. In this algorithm first, the data should be imported and then trained. By using equation the logistic regression algorithm is represented in the graphs showing the difference between the attributes. From the training data, we have to estimate the best and approximate coefficient and represent it.

2. LITERATURE SURVEY

The amount of cardiovascular disease can exceed the control line. Heart disease is complicated and every year lots of people are dying with this disease. To solve this problem our system will help to detect the heart disease more accurately. [2] Prediction of a person's heart disease one year ahead is performed by studying the model heart rate data in the survey. Data is taken by using tools which is a drawback because if the tools are not available then the data cannot be obtained. Predictive of future heart disease in one year can increase a person's awareness of heart disease itself. But has less accuracy. [3] With the rampant increase in heart stroke rates at juvenile ages. It is inappropriate for a person to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of heart disease. [11] It is observed that the best classification technique which can be used in our domain is none other than Back Propagation since it's the only technique which is used for nonlinear relationships. But there is a drawback of being stuck in local minima. [12] Applying different classification algorithms, each with its own advantage on three separate databases of disease (Heart, Breast cancer, Diabetes) available in UCI repository for disease prediction. The feature selection for each dataset was accomplished by backward modeling using the p-value test. The results of the study strengthen the idea of the application of machine learning in early detection of diseases.

3. PROPOSED SYSTEM

The proposed system will fetch the data from the dataset which is classified based on patients having heart disease or not according to features in it. This proposed system can use the data to create a model which tries to predict if a patient has heart disease. In this system logistic regression is used as

a classifier. From the data we are having, it should be classified into different structured data based on the features of the patient's heart. From the availability of the data, we have to create a model which predicts the patient disease using logistic regression algorithm. We have to import the datasets, Read the datasets, the data should contain different variables like age, gender, chest pain, blood pressure. The data should be explored so that the information is verified. We need to create a temporary variable and also build a model for logistic regression. From the training data, we have to estimate the best and approximate coefficient and represent it.

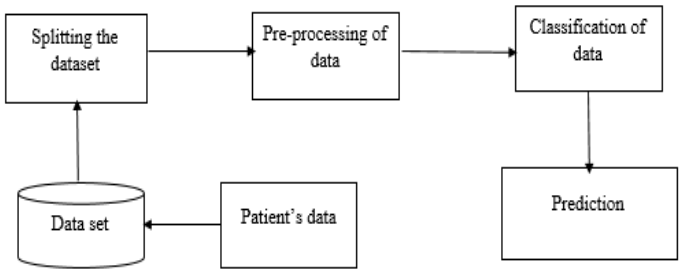

Fig -1: system architecture

The system is implemented using Python. In this system first, the dataset is collected. Here we are using a dataset which contains 13 attributes. For any given data set related to heart disease, different graphs will be plotted based on the attribute.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
44	1	1	120	263	0	1	173	0	0	2	0	3	1
52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
58	0	3	150	283	1	0	162	0	1	2	0	2	1
50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
58	0	2	120	340	0	1	172	0	0	2	0	2	1
66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
42	1	0	140	226	0	1	178	0	0	2	0	2	1

Fig -2: heart disease dataset

For instance, if a user needs a graph for the target attribute. The below-shown figure will be displayed. The system works as shown in the architecture (Fig 1). To begin with, heart disease patient's data should be collected. In this, the dataset is consisting of 13 attributes. Pre-processing of data is carried out to delete the null values and rows. The dataset is classified based on patients having heart disease or not according to features in it. This system can use the data to create a model which tries to predict if a patient has heart disease. In this system, we use logistic regression as a classifier. From the data we are having, it should be classified into different structured data based on the features of the patient's heart. From the availability of the data, we have to create a model which predicts the patient disease using logistic regression algorithm. We have to import the datasets and Read the datasets. The dataset contains different variables like age, gender, chest pain, blood pressure etc. a temporary variable and a model is created for logistic regression, SVM, kNN after creating the model should be trained. By using training data best and approximate coefficient is estimated and represented. Then all the modules are evaluated and the accuracy will be found. After that conclude by comparing the accuracy of the model with high performance and the graph is displayed.

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment

The objective of this project is to prevent the complexity that may occur in the future due to heart disease. The designed heart disease prediction system helps to check out whether a person is affected by heart disease or not. To reduce the risks of heart disease by predicting the disease at an early stage so that it can be cured. The main objective is to improve the performance accuracy of heart disease prediction.

4. RESULT AND ANALYSIS

The system will get the data from the patient's dataset. Logistic regression will be used for classification finally the accuracy will be obtained.

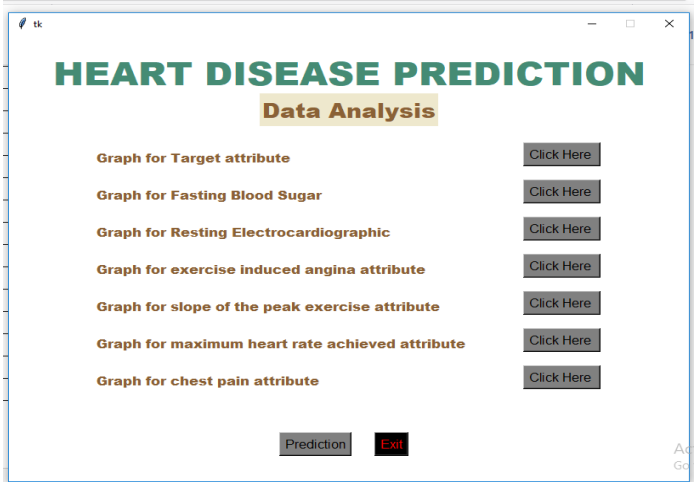

Fig -3: home page

The user can click the button on the home page for displaying the graph. For any given data set related to heart disease, different graphs will be plotted based on the attribute.

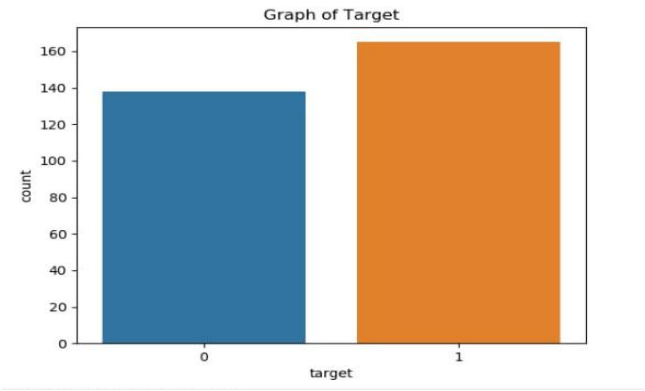

Fig -4: Target attribute

The accuracy score is compared and logistic regression has the highest accuracy. The user can also click the button click here to see the graph for the comparison of the accuracy score between the algorithms. Fig 4 shows the graph for comparing the accuracy score

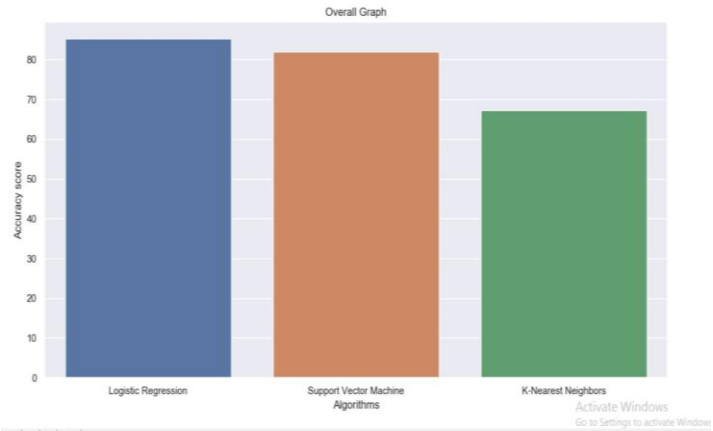

Fig -5: Graph for comparison of accuracy score

After creating a logistic regression model, SVM model and KNN model it should be trained using the dataset. After that, by evaluating all the modules the accuracy can be found then Concluding by comparing the accuracy of the model with high performance

5. CONCLUSION

The number of Heart diseases can exceed the control line and reach to the maximum point. Heart disease is complicated and every year lots of people are dying with this disease. By using different systems we can still have some of the major drawbacks. That is these work mainly focus only on the application of classification techniques and algorithms for heart disease prediction, by all these studies of various data cleaning and mining techniques that prepare and build a dataset appropriate for these algorithms. So that we can use this Machine Learning techniques with logistic regression for predicting if the patient has heart disease or not. The future scope of the paper is the prediction of heart diseases by using advanced techniques and algorithms in less time complexity.

REFERENCES

[1] M.A. Jabbar ,2016;Heart disease prediction system based on hidden naviebayesclassifier;2018 International conference on circuits , controls, communications and computing(14C)

[2] Rifki Wijaya,2013.Preliminary design of estimation heart disease by using machine learning ANN within one year; communication technology and electric vehicle technology , bandung-bali, Indonesia.

[3] AditiGavhane, 2018. Prediction of heart disease using machine learning, ISBN: 978-1-5386-0965-1.

[4] Sana Bharti, 2015 .Analytical study of heart disease prediction comparing with different algorithms; International conference on computing, communication and automation (ICCA2015).

[5] Monika Gandhi, 2015. Prediction in heart disease using techniques of data mining, International conference on futuristic trend in computational analysis and knowledge management(ABLAZE- 2015)

[6] Sarath Babu,2017.Heart disease diagnosis using data mining technique, international conference on electronics, communication and aerospace technology ICECA2017

[7] A H Chen, 2011. HDPS: heart disease prediction system; 2011 computing in cardiology

[8] Dinesh Kumar G, 2018. Prediction of cardiovascular disease using machine learning algorithms, preceding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India.

[9] Purushottam, 2015. Efficient heart disease prediction system using decision tree; International conference on computing, communication &automation.

[10] AnkitaDewan, 2015. Prediction of heart disease using a hybrid technique in data mining classification 2015

[11] Volume 2018 |A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms Amin Ul Haq,Jian Ping Li,Muhammad Hammad Memon, Shah Nazir and Ruinan Sun

[12] Application of Machine Learning in Disease Prediction Pahulpreet Singh Kohli__; 2018 4th International Conference on Computing Communication and Automation (ICCCA)