HEART DISEASE SURVIVAL RATE USING ML

Yedduri Sowmya[1], Vardhineni Jayanth[2], M. Krishna Prasanna[3]

[1] Student, Information Technology, B V Raju Institute of Technology, Telangana, India
[2] student, Information Technology, B V Raju Institute of Technology, Telangana, India
[3] Assistant Professor, Information Technology, B V Raju Institute of Technology , Telangana, India

ABSTRACT

Heart disease is one of the most critical human diseases in the world and affects humanlife very badly. In heart disease, the heart is unable to push the required amount of blood to other parts of the body. Accurate and on time diagnosis of heart disease is important for heart failure prevention and treatment. The diagnosis of heart disease through traditional medical history has been considered as not reliable in many aspects. To classify the healthy people and people with heart disease, noninvasive- based methods such as machine learning are reliable and efficient.

Keyword: *Binary Classification, Heart Disease, Logistic Regression, Machine Learning*

1. INTRODUCTION

WHAT IS HEART DISEASE?

The term “heart disease” refers to several types of heart problems. The most common type is coronary artery disease, which can cause a heart attack. Other types of heart disease may involve the valves in the heart, or the heart may not pump well and cause heart failure. Some people are born with heart disease. Anyone, including children, can develop heart disease. It happens when a substance called plaque builds up in your arteries. Smoking, unhealthy eating and lack of exercise increase your risk of heart disease. High cholesterol, high blood pressure or diabetes can also increase your risk of heart disease.

What is Machine Learning?

The field of study known as machine learning enables computers to learn without being explicitly programmed. Using sample data or prior knowledge, machine learning involves programming computers to optimize a performance criterion. We have a model that is defined up to a certain point, and learning is the application of a computer program to optimize the model's parameters using training data or prior knowledge.

Types of Machine Learning

Supervised Machine Learning

To train the supervised learning models, we use the labelled dataset. The model is tested by supplying a sample of test data after training and processing to see if it predicts the proper output.

Unsupervised Machine Learning

Unsupervised models can be trained using a dataset that has no classifications or categorizations and no labels, and the algorithm must act on that data without any supervision.

Reinforcement Learning

In reinforcement learning, an agent creates interactions with its surroundings and absorbs information from feedback.. The agent gets feedback in the form of rewards; for instance, he gets a good reward for every good action and a terrible reward for every bad action.

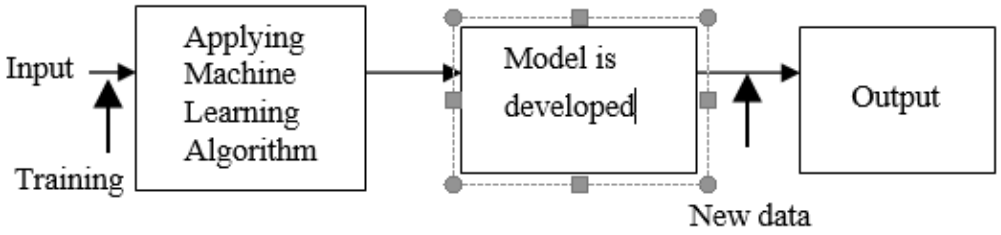

Fig-1*:* Machine Learning Block Diagram

2. METHODOLOGY

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such aswhether the cells are cancerous or not, a mouse is obese or not based on its weight, etc. Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets. Within machine learning, logistic regression belongs to the family of supervised machine learning models. It is also considered a discriminative model, which means that it attempts to distinguish between classes (or categories). Unlike a generative algorithm, such as naïve bayes, it cannot, as the name implies, generate information, such as an image, of the class that it is trying to predict (e.g. a picture of a cat).

- After evaluating results from existing methodologies, we have used python and pandas operations to perform heart disease classification for the data obtained.
- It provides easy to visual representation of dataset.
- Logistic regression is used to improve the accuracy of the result.

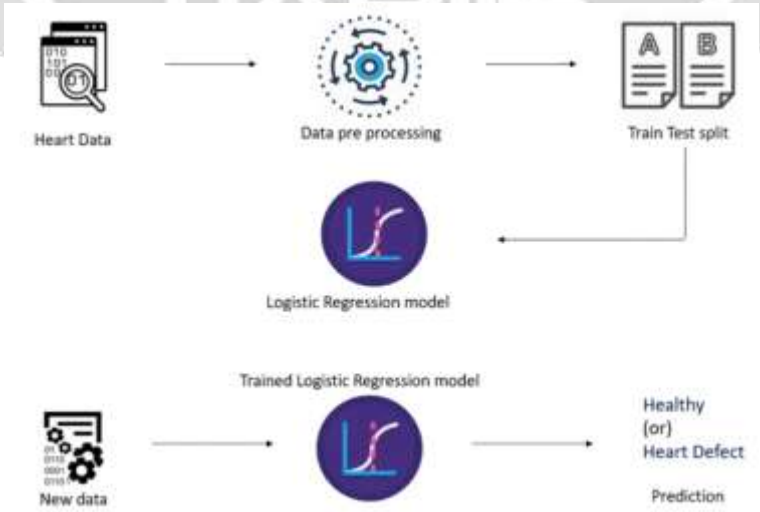

Fig-2*:* Proposed Architecture

The proposed architecture's modules include:

1) Upload the Heart Disease Dataset: This module is used to upload the dataset for the application

Name of the dataset	:	Heart Disease Dataset
Source	:	Kaggle
Types of Attributes	:	continuous and binary
No of Attributes	:	13 including class variable
Instances	:	1025

List of Attributes:

Attribute ID	Description
1	Age
2	Sex
3	Chest Pain
4	Trestbps
5	Cholesterol
6	Fasting Blood Sugar
7	Resting ElectroCardiographic Results
8	Maximum Heart Rate Achieved
9	Exercise Induced Angina
10	Old peak
11	Slope
12	Number of major vessels
13	Thal

. 2) Collection of data set: Initially, we collect a dataset for our heart disease prediction system. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model.

3) Selection of attributes: Attribute includes the selection of appropriate attributes for the prediction system. This is used to increase the efficiency of the system. Various attributes of the patient like age, gender, cholesterol,etc are selected for the prediction.

4) Preprocessing of data: Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. Preprocessing of data is required for improving the accuracy of the model.

5) Classification: One of the Simplest and best ML classification algorithm is Logistic Regression. The LR is the supervised ML binary classification algorithm widely used in most application. It works on categorical dependent variable the result can be discrete or binary categorical variable 0 or 1. To predict the cardiac disease logistic regression ML model is used, firstly the LR model are trained with five splitting condition and tested with test data for prediction to get the best accuracy and to find the models behavior. The algorithm results category of 1 and 0 for presence and absences of cardiac disease.

3. RESULTS

After entering the inputs when we click submit button then the results will be visible to the user, whether he/she is going to be affected with heart disease or not. The proposed system's results are shown in the figures below.

```
input_data = (51,1,0,140,298,0,1,122,1,4.2,1,3,3)

# change the input data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array as we are predicting for only on instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if (prediction[0]==0):
   print('The Person does not have a Heart Disease')
else:
  print('The Person has Heart Disease')
```

```
[0]
The Person does not have a Heart Disease
```

Fig 3: Screenshot 1

```
input_data = (71,0,0,112,149,0,1,125,0,1.6,1,0,2)

# change the input data to numpy array
input_data_as_numpy_array = np.asarray(input_data)

# reshape the numpy array as we are predicting for only on instance
input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

prediction = model.predict(input_data_reshaped)
print(prediction)

if (prediction[0]==0):
   print('The Person does not have a Heart Disease')
else:
  print('The Person has Heart Disease')
```

```
[1]
The Person has Heart Disease
```

Fig 4: Screenshot

4. CONCLUSIONS

This We proposed a method for heart disease prediction using machine techniques, these results showed a great accuracy standard for producing a better estimating result. Heart disease prediction is challenging and very important in medical field. However, the mortality rate can be drastically can be controlled if disease is detected at early stage and preventive measures are adopted as soon as possible.

5. REFERENCES

[1] Dataset regarding heart disease in kaggle
https://www.kaggle.com/ronitf/heart-disease-uci

[2] M. Saw, , T. Saxena, S. Kaithwas, R. Yadav and N. LalEstimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning

[3] World Health Organization and J. Dostupno, cardiovascular diseases: key facts, vol. 13, no. 2016, p. 6, 2016.