Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning

Reddy Prasad,Pidaparthi Anjali, S.Adil, N.Deepa

ABSTRACT---We are in a period of "Information Age" where the traditional industry can pressure the rapid shift to the industrial revolution for industrialization, based on economy of information technology Terabytes of data are produced and stored day-to day life because of fast growth in 'Information Technology'. Terabytes of data are produced and stored day-to day life because of fast growth in 'Information Technology'. The data which is collected is converted into knowledge by data analysis by using various combinations of algorithms. For example: the huge amount of the data regarding the patients is generated by the hospitals such as x-ray results , lungs results ,heart paining results, chest pain results , personal health records(PHRs) ., etc. There is no effective use of the data which is generated from the hospitals. Some certain tools are used to extract the information from the database for the detection of heart diseases and other functions is not accepted. The main theme of the paper is the prediction of heart diseases using machine learning techniques by summarizing the few current researches. In this paper the logistic regression algorithms is used and the health care data which classifies the patients whether they are having heart diseases or not according to the information in the record. Also I will try to use this data a model which predicts the patient whether they are having heart disease or not.

Keywords :-Heart Diseases; Machine Learning; Data Analysis; Logistic Regression Algorithms .

I. INTRODUCTION: DOMAIN :

To initiate with the work we can use different types of techniques and algorithms. In this paper, machine learningtechniques are used toincreasetheaccuracy rate. In machine learning technique we can use the followingalgorithm

1. Logistic regression
2. Naivebayes
3. Comparing and confusionmatrix

Sklearn Logistic regression

The logistic regression is also known as sigmoid function which helps in the easy representation in graphs. 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 haveto estimate the best and approximate coefficient and representit.

Naive bayes

Naive bayesalgorithm[1] which is mainly used,also helps for classification.The large datasets can be build by using the algorithm native bayes. It is a classification technique and it assumes as independence among each attribute. It will helps in the removing of correlated data.

Comparing and confusion matrices

The comparison a confusing matrices, it is the summary for the prediction of the result which we classified. Based on the classification of attributes the correct and incorrect predictions are marked with count values.A confusion matrix,which is represented in table format will explains about the performance of characterization model on the trained dataset.The most of the performance measures are calculated using this confusion matrix.

The most important organ of the human body is heart. The function of the heart is to pump the blood and circulates entire body [2]. It is protected by rib cage and it is surrounded by two layered tissue membrane called Pericardium. It is a four chambered organ which separates oxygenated and deoxygenated blood. Heart is having the five types of blood vessels:arteries,veins,capillaries, arterioles, venules and .The size of the human heart is about the size of the fist and weight approximately 300grams,the weight in females being about 25% lesser than males.Arteries and veins are present in heart which helps to collect the blood from all the parts and purifies it and the circulates to all the body parts. The nutrients and oxygen present in the body parts are provided by the blood and also will helps in the removal of metabolic wastes. Now a days the life span of human being is reduced due to the heart diseases. The factors which may lead to heart disease are obesity, high cholesterol,smoking, increase in blood pressure, diabetes and other factors.

As per the WHO(World Health Organization) records, each and every year millions of people died with different types of cardio attacks such as heart stroke, chest pain,etc. We here proposed the collection of relevant data from the hospitals wherethe data is so enormous. Now we have to

Revised Manuscript Received on February 14, 2019.

Reddy prasad, ug scholar, department of computer science and engineering, saveetha school of engineering, simats, chennai, tamil nadu, india. (email id:reddyprasadprince@gmial.com)

Pidaparthi anjali, ug scholar, department of computer science and engineering, saveetha school of engineering, simats, chennai, tamil nadu, india

S.adil, ug scholar, department of computer science and engineering, saveetha school of engineering, simats, chennai, tamil nadu, india

N.deepa, assistant professor, department of computer science and engineering, saveetha school of engineering, simats, chennai,tamil nadu, india (email id: ndeepa.sse@saveetha.com)

separate the data regarding the patients related to heart diseases. We train the data as per proposed algorithm of machine learning[3] by using logistic regression. For the purpose of detecting the heart disease , we can enter patient medical details into the trained data.

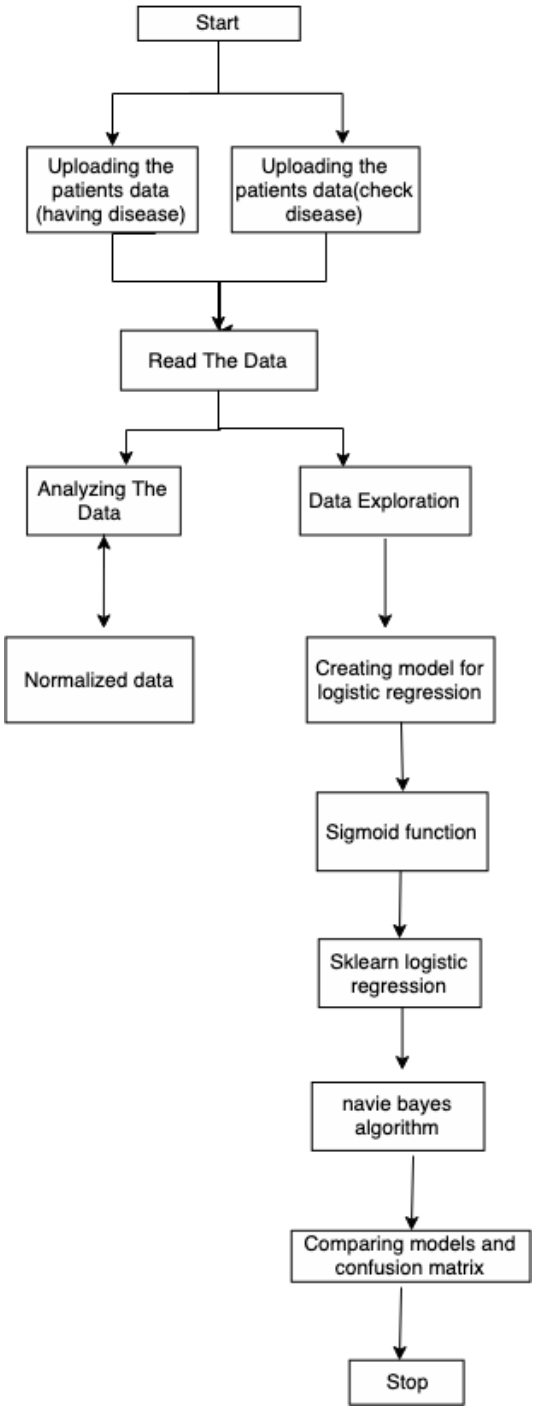
II. EXISTING SYSTEM :

The before all existing system[4] works on sets of both Deep learning and data mining[5]. The existing system modules generates comprehensive report by implementing the strong prediction algorithm The main aims of the existing system to compare and check the before patient whose having disease outputs and new patient disease and determine future possibilities of the heart disease to a particular patient By Implementing the above mentioned model we will get the goal of developing a system with increased rate of accuracy of estimating the new patient getting heart attack percentage. The model which is proposed for Heart Attack Prediction System is invented for using Deep learning[6] algorithms and approach. But by using all the existing systems the accuracy is very less.

III. PROPOSED SYSTEM:

This proposed system have a data which classified if patients have heart disease or not according to features in it. This proposed system can try to use this data to create a model which tries predict(reading data and data Exploration)[7] if a patient has this disease or not. In this proposed system, use logistic regression (classification) algorithm. By using sklearn library to calculate score. Implements Navie Bayes algorithm to getting accuracy result.Finally analysing the results by the help of Comparing Models and Confusion Matrix.

From the data we are having, it should be classified into different structured data based on the features of the patient heart. Fromtheavailabilityofthedata,wehaveto create a model which predicts the patient disease using logistic regression algorithm. First, we have to import the datasets. Read the datasets, the data should contain different variables like age,gender, sex, cp(chest pain),slope, target. The data should be explored so that the information is verified.Create a temporary variable and also build a model for logistic regression. Here, we use sigmoid function which helps in the graphical representation of the classifieddata.Byusinglogisticregression, naïve bayes the accuracy rateincreases.

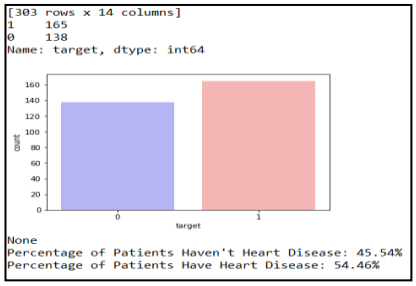


Implementation of Logistic Regression Algorithm :

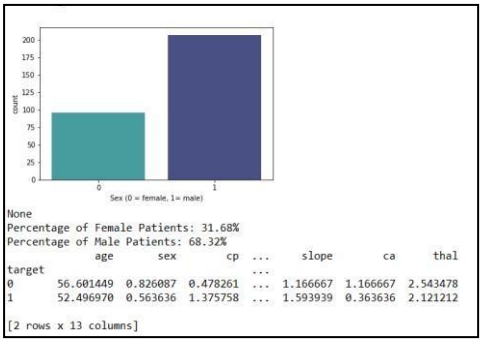
```
In [2]: runfile('C:/Users/Balaji/Desktop/r/k.py', wdir='C:/Users/Balaji/Desktop/r')
['heart.csv', 'heartDiseaseAndAges.png', 'k.py']
    age  sex  cp  trestbps  chol  fbs  ...  exang  oldpeak  slope  ca  thal  target
0    63    1   3       145   233    1  ...      0      2.3      0   0     1       1
1    37    1   2       130   250    0  ...      0      3.5      0   0     2       1
2    41    0   1       130   204    0  ...      0      1.4      2   0     2       1
3    56    1   1       120   236    0  ...      0      0.8      2   0     2       1
4    57    0   0       120   354    0  ...      1      0.6      2   0     2       1
5    57    1   0       140   192    0  ...      0      0.4      1   0     1       1
6    56    0   1       140   294    0  ...      0      1.3      1   0     2       1
7    44    1   1       120   263    0  ...      0      0.0      2   0     3       1
8    52    1   2       172   199    1  ...      0      0.5      2   0     3       1
9    57    1   2       150   168    0  ...      0      1.6      2   0     2       1
10   54    1   0       140   239    0  ...      0      1.2      2   0     2       1
11   48    0   2       130   275    0  ...      0      0.2      2   0     2       1
12   49    1   1       130   266    0  ...      0      0.6      2   0     2       1
13   64    1   3       110   211    0  ...      1      1.8      1   0     2       1
14   58    0   3       150   283    1  ...      0      1.0      2   0     2       1
15   50    0   2       120   219    0  ...      0      1.6      1   0     2       1
16   58    0   2       120   340    0  ...      0      0.0      2   0     2       1
17   66    0   3       150   226    0  ...      0      2.6      0   0     2       1
18   43    1   0       150   247    0  ...      0      1.5      2   0     2       1
19   69    0   3       140   239    0  ...      0      1.8      2   2     2       1
20   59    1   0       135   234    0  ...      0      0.5      1   0     3       1
21   44    1   2       130   233    0  ...      1      0.4      2   0     2       1
22   42    1   0       140   226    0  ...      0      0.0      2   0     2       1
23   61    1   2       150   243    1  ...      1      1.0      1   0     2       1
24   40    1   3       140   199    0  ...      1      1.4      2   0     3       1
```

Read the Data

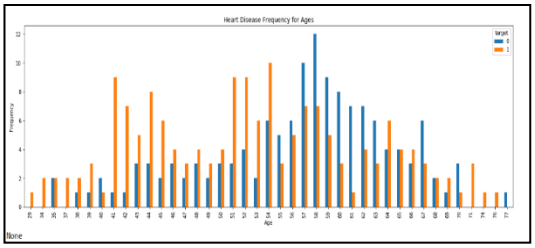
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication



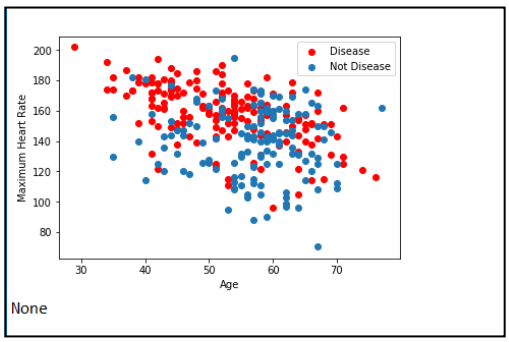
Fixing Targets about the Disease



Separate the data about sex, age



Heart disease Frequency for ages



Checking Disease

Result Accuracy% in algorithms:

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train.T, y_train.T)
print("Accuracy of Naive Bayes: {:.2f}%".format(nb.score(x_test.T,y_test.T)*100))
```

```
Accuracy of Naive Bayes: 86
```

Accuracy% result in Navie Bayes

```
from sklearn.tree import DecisionTreeClassifier
dtc = DecisionTreeClassifier()
dtc.fit(x_train.T, y_train.T)
print("Decision Tree Test Accuracy {:.2f}%".format(dtc.score(x_test.T, y_test.T)*100))
```

```
Decision Tree Test Accuracy 78.69%
```

Acuuracy% result in Decision Tree

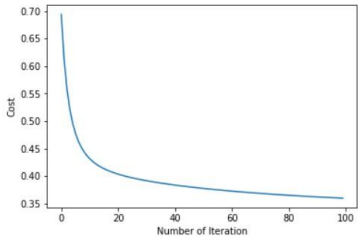
```
# KNN Model
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 2)  # n_neighbors means k
knn.fit(x_train.T, y_train.T)
prediction = knn.predict(x_test.T)

print("{} NN Score: {:.2f}%".format(2, knn.score(x_test.T, y_test.T)*100))
```

```
2 NN Score: 77.85%
```

Accuracy% result in Knn Algorithm

```
logistic_regression(x_train,y_train,x_test,y_test,1,100)
```

```
iteration: 100
cost: 0.3597736123664534
```



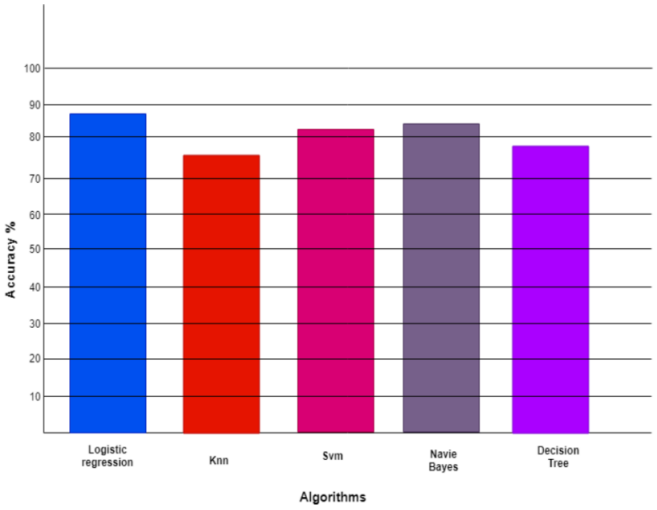

```
Manuel Test Accuracy: 86.89%
```

Accuracy% result in logistic regression algorithm

```
lr = LogisticRegression()
lr.fit(x_train.T,y_train.T)
print("Test Accuracy {:.2f}%".format(lr.score(x_test.T,
```

```
Test Accuracy 86.89%
```

Testing accuracy for logistic regression algorithm



Accuracy% result in graph between Algorithms

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. CONCLUSION

The amount of Heart diseases can exceed the control line and reach to maximum point. Heart disease are complicated andeachandeveryyearlotsof people are dying with this disease By using this all systems one of the major drawbacks of these works is mainly focus only to the application of classify techniques and algorithms for heart disease prediction, by all these studying various data cleaning and mining techniques that prepare and build a dataset appropriate for data mining. So that I can use this Machine Learning in that logistic regression algorithms by predicting if patient has heart disease or not. Any non-medical employee can use this software and predict the heart disease and reduce the time complexity of the doctors.

V. FUTURE WORK

Today's, world most of the data is computerized, the data is distributed and it is not utilizing properly. By Analyzing the available data we can also use for unknown patterns. The primary motive of this research is the prediction of heart diseases with high rate of accuracy. For predicting the heart disease we can use logistic regression algorithm, naviebayes, sklearn in machine learning. 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 naviebayes classifier;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 on year; communication technology andelectricvehicle 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 andautomation(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, procedingof 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. Nikhil Gawande ,2017. Heart diseases classification using convolutional neural network; 2012 2nd International conference on communication and electronics systems(ICCES)
11. AnkitaDewan, 2015. Prediction of heart disease using a hybrid technique in data mining classification 20152nd