IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9692

Heart Disease Prediction System, Machine Learning Techniques

Prajwal G Ugrani[1], A M Shivaram[2]

[1]Student, Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India.

[2]Assoc Professor,Dept. of MCA, Bangalore Institute of Technology, Bengaluru, India.

Abstract: Heart disease has become a major cause of death in this era. Heart disease claims the lives of about one person every minute. Because of the rapid rise of information technology, data is produced and must be reserved on a regular basis. Heart is considered to be the most important part of our body after the brain. Data analysis and machine learning are used to convert the obtained data into knowledge using a variety of techniques. Medical practitioners who are specialized in this field of heart diseases have some of their own limits, they cannot accurately predict the probability of developing heart diseases. The main goal of the research is to improve accuracy of the heart problems forecasting using the Logistic Regression model of the machine learning taking into consideration the health dataset which further classifies the patients if they have heart problems or not.

I. INTRODUCTION

Heart disease is one of the biggest killers worldwide. It will encompass all data using this information technology. Despite the fact that these diseases are the leading cause of mortality, they have been labelled the most achievable and preventable condition. A heart stroke is caused by a blockage in the arteries. It happens if the heart doesn't adequately circulate the blood around body. The leading cause of the heart problem is high blood pressure. Person's obesity, poor nutrition, a rise in cholesterol, and the lack of physical activities are just few of other causes of heart disease. As a result, prevention is critical. It is critical to be aware of cardiac disorders. In the field of accurate data analysis, predicting cardiac disease is an absolute must. Machine learning (ML) is extremely useful in assisting in the extraction of conclusions and guesses from the massive amounts of the data generated by healthcare industries. Because of many risk elements like diabetes, high BP (blood pressure), high cholesterol content, and other conditions, it might be difficult to diagnose heart disease. We utilised a categorization algorithm called Logistic regression in this case. It's commonly utilised when the target variable's value is categorical. It's most typically employed when the data has a binary output, such as whether it's in the range of 0 or 1. This also ensures excellent precision. We will try to estimate and depict the best and approximate coefficient using the training data.

II. LITERATURE SURVEY

The degree of cardiovascular disease that is present can be more than the control line. Heart disease is a difficult condition that kills a number of individuals every year. To address this issue, our technology will aid in the more accurate detection of cardiac disease. The model heart data in the survey is used to find if a person has heart problems or not in one year in advance. There are numerous studies in the literature that use machine learning and data mining to diagnose heart problems.

Ujma Ansari used the decision tree model to predict heart disease in 2012 and has achieved the high accuracy of 99 percent and inspiring us to utilize the superior version of the decision tree, the Random Forest. Regrettably, a study employs the dataset of 3000 case but does not provide the source for data. The website has about 303 dataset values, so we are not sure where the author got the other 3000. When compared to today's study, some articles published two to three years ago have a lower accuracy for predicting cardiac problems.

III. METHODOLOGY

This study looks at a number of elements that affect the human cardiovascular system. The procedure starts with data retrieval, followed by correlation analysis, data partitioning, and prediction using the logistic regression algorithm.

Python is a widely useful, adaptable, and well-known program language. It is one fantastic first language due to its compact, simple to use, and it is also a great language to have in the developer's stack because it could be used for anything from developing to programming developments and logic applications.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9692

A. DataRetrieval

Data retrieval is the initial step. The dataset will be used in this process. It's loaded into the Jupyter notebook programme. The information gathered is both category and numerical.

B. DataPreprocessing

The data preprocessing is a key phase in the machine learning because quality of datas and the useful detailsthatcanbegleanedfromthisdirectlyinfluencesourmodelcapacitytolearn.Asaresult,it'scritical that we preprocess our data before feeding it to ourmodel.

C. Splittingdataset

For training a Machine Learning model, one should hint at the target column in the dataset, and then he/she can break the dataset into 2 small datasets. Training set to train algorithm and the Test set to test the algorithm. We have divided our dataset into two parts: The first one uses the trained dataset which is a size of about 80%, and the test part has a size of 20%. That is how we have divided the dataset.

D. ModelTraining

This is the stage in which we apply and evaluate the algorithms we've chosen (in this case, Logistic Regression) to determine their correctness. The logistic regression procedure is used, and the results are analyzed. The logistic regression method of prediction will yield multiple data points that can be used to draw conclusions and make predictions. The logistic regression algorithm was proven to be an effective and efficient method in predicting the key causes of cardiovascular disease as the problem became more prevalent in this study.

IV. RESULT

A. The dataset's first fiverows:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Table view of the first 5 records in the dataset:

index	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

B. The dataset's last fiverows:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9692

Table view of the last 5 records in the dataset:

index	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1020	59	1	1	140	221	0	1	164	1	0.0	2	0	2	1
1021	60	1	0	125	258	0	0	141	1	2.8	1	1	3	0
1022	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
1023	50	0	0	110	254	0	0	159	0	0.0	2	0	2	1
1024	54	1	0	120	188	0	1	113	0	1.4	1	1	3	0

C. Distribution of targetvariable:

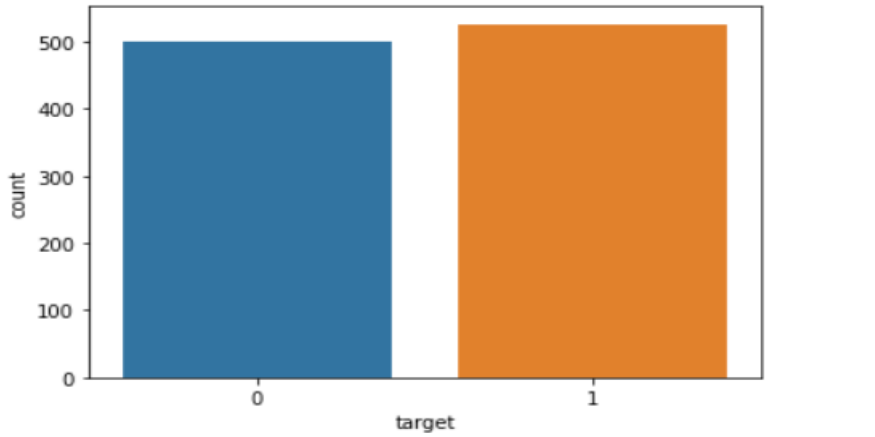
```
heart_data['target'].value_counts()
```

```
1    526
0    499
Name: target, dtype: int64
```

1 --> Defective Heart

0 --> Healthy Heart

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fd7e28fca90>
```



D. Accuracy on Trainingdata:

```
print('Accuracy on Training data : ', training_data_accuracy)
```

```
Accuracy on Training data :  0.8524390243902439
```

E. Accuracy on Testingdata:

```
print('Accuracy on Test data : ', test_data_accuracy)
```

```
Accuracy on Test data :  0.8048780487804879
```

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9692

F. Building a Predictivesystem:

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

# change the input data to a 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
```

V. CONCLUSION

The cardiovascular disease detection model in this project was created with the help of the ML classification model Logistic Regression approach. This project predicted people who would develop cardiovascular problem by drawing out patient medical records that may lead to deadly heart disease from a dataset that carries patient medical history like blood sugar levels, blood pressure, and the other factors. The number of heart ailments could well outnumber the existing scenario and reach an all-time high. Heart disease is complicated, and thousands of people die each year as a result of it. Manually calculating the chances of a person developing the heart problem based on risk factors already shown is extremely difficult. Examining various data cleaning and data mining strategies for preparing and building a dataset suitable for data mining, people may apply machine learning in logistic regression algorithm to guess whether or not a patient has heartdisease.

REFERENCES

[1] Avinash Golande, Pavan Kumar T. Machine learning techniques are being used to forecast heartdisease.
[2] Himanshu Sharma, M A Rizvi. A survey on the use of machine learning algorithms to predict cardiac disease.
[3] Jaymin Patel, Prof. Tejpal Upadhyay and Dr. Samir Patel. Machine learning and data mining techniques are used to forecast heartdisease.
[4] Monika Gandhi, 2015. The international conference of advanced computational analysis and knowledge management, employing data mining tools to predict cardiacdisease
[5] Sarath Babu,2017. International conference on electronics, communication, and aerospace technologies, employing data mining to diagnose heartproblems.
[6] A H Chen, 2011. HDPS stands for heart disease prediction system, and computing in cardiology in 2011 was a bigdeal.
[7] Purushottam, 2015. A decision tree-based approach for predicting cardiac disease.International conference on computing, communication &automation.