

ANALYSIS OF HEART DISEASE USING MACHINE LEARNING TECHNIQUES

Samson Cherlapally, K.Srilatha Reddy

PG Scholar, Department of Computer Science and Engineering,
Assistant Professor, Department of Computer Science and Engineering,
Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, Telangana 500097

Article History: Received: 11 November 2020; Accepted: 27 December 2020; Published online: 05 April 2021

Abstract : There is a very large number in the health sector and special methods are also used systematically. Data exchange is one of the most commonly used methods. Heart disease is one of the leading causes of death in the world. This system predicts the possibility of heart disease. The results of this system provide a 100% risk of heart disease. The data used are categorized according to medical criteria. The system evaluates these parameters using data extraction methods used in Python using two basic machine learning algorithms, the Solution Tree Algorithm, and the algorithm that demonstrates the best accuracy in heart disease.

Keywords: Supervised Classification, KNN, CART, C4.5, CHAID, J48, ID3.

1. INTRODUCTION

Data entry is access to information and knowledge of big data. Activity data is the latest step in reporting from a database. There are many databases, depots, and depots around the world. Databases are primarily used to access confidential information from large databases. The database is also called Knowledge Acquisition Database (KDD). Data entry includes four main methods, namely division, inclusion, retrieval, and corporate rules. Communication techniques allow you to quickly search for big data.

In many areas, special data analysis is needed to extract important data from too much data. Areas such as medicine, business, and education have a lot of information, so this source of information can be used to find the most important information and methods. Communication techniques can be implemented by machine learning algorithms. Each technology can be expanded with a special machine learning model. This system uses a database for heart disease. The main purpose of this system is to predict the patient's heart rate. This is done using the classification data extraction method.

The partition method is used to divide all the data into two groups: yes and no. The partition method is applied to a set of data using a machine partition algorithm, namely, part of the decision tree and model dividing Naive Bayes. This model is used to improve the accuracy of classification techniques. This model also includes classification and prediction methods. This example is implemented in the Python programming language.

SCOPE

The purpose of this project is to prevent heart disease. In this project, we use different machine learning algorithms and build through food and marine programs. Computer graphics are not a box

PURPOSE OF PROJECT

The main idea of this article is to predict heart disease using data entry techniques. The main methods used in the forecast are KNN, CART, C4.5, CHAID, J48, ID3 algorithms, and Bayesi techniques. The system takes 13 medical properties as information and plants and demonstrates a more accurate way to extract data from this information.

2. LITERATURE SURVEY

Heart disease prediction using data mining techniques

Many factors affect the human heart in everyday life. Many problems arise quickly and new heart disease is diagnosed quickly. In today's stressful world, the heart, which is the main organ in the human body circulating in the blood, is vital to life. The health of the human heart depends on the experience of human life and it all depends on human behavior and professionalism. There may also be some genetic factors that drive a form of heart disease from one generation to the next. According to the World Health Organization, every year more than 12 million people die each year from heart disease, also known as heart disease. The term "heart disease" includes many different diseases that have a specific effect on the heart and blood vessels. Twenty-three years old is still suffering from heart disease. Many factors, such as poor eating habits, insomnia, anxiety, depression and obesity, poor nutrition, family history, high blood pressure, high cholesterol, inactivity, and medical conditions, can increase the risk of heart disease. for young people. family history, smoking, and high blood

*Corresponding author: Samson Cherlapally,
PG Scholar, Department of Computer Science and Engineering,
Teegala Krishna Reddy Engineering College, Meerpet, Hyderabad, Telangana 500097

pressure. Diagnosis of heart disease in medicine is one of the most difficult tasks. All of these factors are regularly monitored by physicians when analyzing and diagnosing patients. Symptoms of heart disease are usually associated with discomfort. Some tips are usually not heard by ordinary people. However, the most common symptoms are chest pain, shortness of breath, and nausea. Chest pain, which occurs in most cases of the heart, is called angina pectoris or angina and occurs when part of the heart is deprived of oxygen. Angina can be caused by stressful events or physical exertion and is usually no longer than 10 minutes. Myocardial infarction can also be the result of a variety of heart attacks. The symptoms of a heart attack are similar to angina pectoris if mild and usually more severe. The symptoms of a heart attack can sometimes be the same as a heart attack. You may experience heartache and abdominal pain, as well as heaviness in the chest. Other symptoms of a heart attack include pain that spreads throughout the body, such as from the chest to the arms, neck, thighs, abdomen, or groin, dizziness, nausea, vomiting, and vomiting. Heart failure also occurs as a result of heart disease, which occurs when breathing stops and the heart becomes very weak in blood circulation. Other heart diseases, especially in the elderly and diabetics, have no symptoms. The term "congenital heart disease" covers a wide range of conditions, but the most common symptoms are fatigue, fatigue, rapid heartbeat, and shortness of breath. However, these symptoms can appear until the age of 13. In these cases, diagnosis becomes a difficult task that requires a lot of experience and care. The risk of heart attack or the risk of heart disease, if detected earlier, can help patients to take care. Recently, the healthcare industry has developed a lot of data about patients and diagnostic reports that are primarily used to predict heart attacks in the world. If there is information about heart disease, machine learning methods are used for analysis. The definition is the process of extracting data from a set of records that make appropriate decisions for future analysis or forecasting. Confidential information is stored and cannot be retrieved without the use of data. Distribution is one of the methods of data extraction that can be used to provide future results or predictions based on historical data. The extraction of medical data has led to the integration of distribution channels with independent organizations in data sharing, leading to the study of hidden maps in medical data used to predict the future of patients. By providing medical data, you can learn more about the patient's history and provide clinical support through analysis. This system is very important in the clinical analysis of patients. Simply put, medical data mining uses a classification algorithm, which is an important part of determining the likelihood of a heart attack. By studying distribution algorithms, you can predict a person's heart condition. In this essay, learning a controlled machine is used for prediction. The proposal uses a comparative analysis of three data sharing algorithms, namely Random Forest, Tree Decision, and Naive Bayes. The analysis is conducted at different levels of security services and is carried out in violation of the percentage of price relevant to the percentage of the violation. The StatLog package from the UCI machine learning block works in this scientific report to diagnose heart disease. When teaching several data on heart disease, predictions are made using a classification model based on a classification algorithm.

Using Data Mining Techniques to Predict Diabetes and Heart Diseases

Data communication is a new field of study, whose main task is to enable the science of several data. It is a process of selecting, retrieving, and modeling large amounts of data to allow anonymous sampling. Modern life and business lead to poor health and improvement as a result of the rise of non-communicable diseases such as heart disease and diabetes. Non-communicable diseases are a direct result of inactivity, inactivity, and inactivity of humans. Heart disease and diabetes are some of the most dangerous killers in society. The purpose of this study is to create a program that can be used by physicians and other physicians to predict or repeat non-communicable diseases. Because most patient data is not used for that purpose, it is important to use this data. Data collection and analysis is done in two stages: the first is data collection and the second is a literature review. Second, important data are collected to help physicians understand the importance of resources in predicting SPS. Sophisticated programs should contain appropriate secondary information so that doctors can use programs that completely repeat or repeat the presence of non-communicable diseases. "The purpose of the research is to develop programs that doctors can trust to prevent non-communicable diseases. Researchers have developed new programs that can be used to collect data about patients and provide predictable data.

3. OVERVIEW OF THE SYSTEM

A. Existing System:

The figures are subject to difficulty, data loss, and incompatibility. Therefore, it is very important to process the data first. We put filters on the data and remove it from the dirty and aggressive data contained in the data. Filtering of controlled and idle assets (working assets) and samples (sample manipulation) were all used in WEKA 2016 (version 3.9.0). In this study, a multi-layer filtering process is applied to the data that balances the unbalanced data.

Disadvantages:

1. WEKA solutions for learning toy machines and upgrading. This is a package that cannot take full advantage of the WEKA machine learning project
2. It only supports small data.
3. Insufficient memory error when large data is available.

B. Proposed System:

This system uses a database for heart disease. The main purpose of this system is to predict the patient's heart rate. This is done using the classification data extraction method. The partition method is used to divide all the data into two groups: yes and no. The partition method is applied to a set of data using a machine partition algorithm, namely, part of the decision tree and model dividing Nayve Bayes. This model is used to improve the accuracy of classification techniques. This model also includes classification and prediction methods. Python programming language is used for this model.

Advantages:

- The main idea of this article is to predict heart disease using data entry techniques. The most important methods used in forecasting are decision trees such as KNN, CART, C4.5, CHAID, J48, ID3Algorithm, and Naive Bayes algorithms.
- This system takes 13 medical properties as information and plant and shows the most accurate method for data extraction with this data entry.

C. Modules Description

Supervised Classification (Training Dataset)

Tasks are divided into two parts, namely training and test data at 70:30. Learning algorithms are based on training data and a set of experimental data is calculated based on training.

Supervised Classification (Test Dataset)

The experimental group contains 30% of all data. Guided learning methods are used for experimental data and the results compared to the actual results.

4. OUTPUT SCREEN SHOTS

Dataset Size

(303, 14)

Dataset Head

Out[5]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Description

Out[6]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca
count	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000	303.000000
mean	54.366337	0.683168	0.966997	131.623762	246.264026	0.148515	0.528053	149.646865	0.326733	1.039604	1.399340	0.729373
std	9.082101	0.466011	1.032052	17.538143	51.830751	0.356198	0.525860	22.905161	0.469794	1.161075	0.616226	1.022606
min	29.000000	0.000000	0.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.000000	0.000000	0.000000
25%	47.500000	0.000000	0.000000	120.000000	211.000000	0.000000	0.000000	133.500000	0.000000	0.000000	1.000000	0.000000
50%	55.000000	1.000000	1.000000	130.000000	240.000000	0.000000	1.000000	153.000000	0.000000	0.800000	1.000000	0.000000
75%	61.000000	1.000000	2.000000	140.000000	274.500000	0.000000	1.000000	166.000000	1.000000	1.600000	2.000000	1.000000
max	77.000000	1.000000	3.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.200000	2.000000	4.000000

Dataset Tail

Out[7]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
185	44	1	0	112	290	0	0	153	0	0.0	2	1	2	0
94	45	0	1	112	160	0	1	138	0	0.0	1	0	2	1
124	39	0	2	94	199	0	1	179	0	0.0	2	0	2	1
93	54	0	1	132	288	1	0	159	1	0.0	2	1	2	1
206	59	1	0	110	239	0	0	142	1	1.2	1	1	3	0

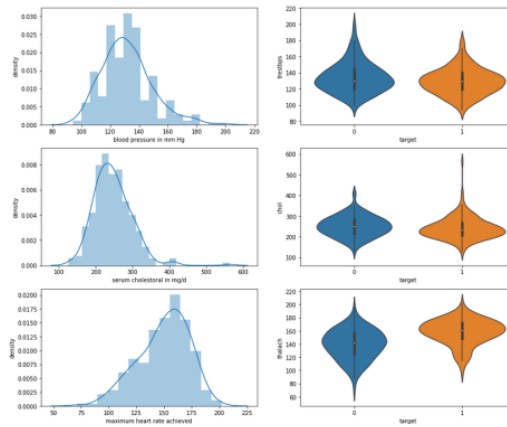
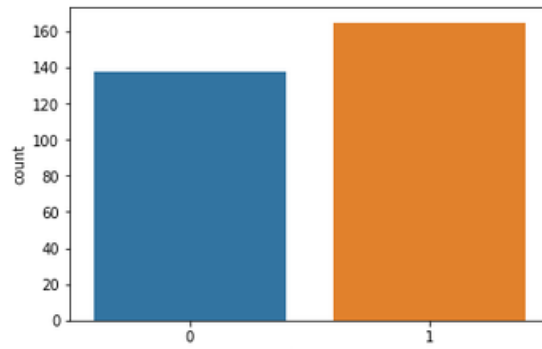
```
Out[8]: age      0
        sex      0
        cp      0
        trestbps 0
        chol     0
        fbs     0
        restecg  0
        thalach  0
        exang    0
        oldpeak  0
        slope    0
        ca      0
        thal     0
        target   0
        dtype: int64
```

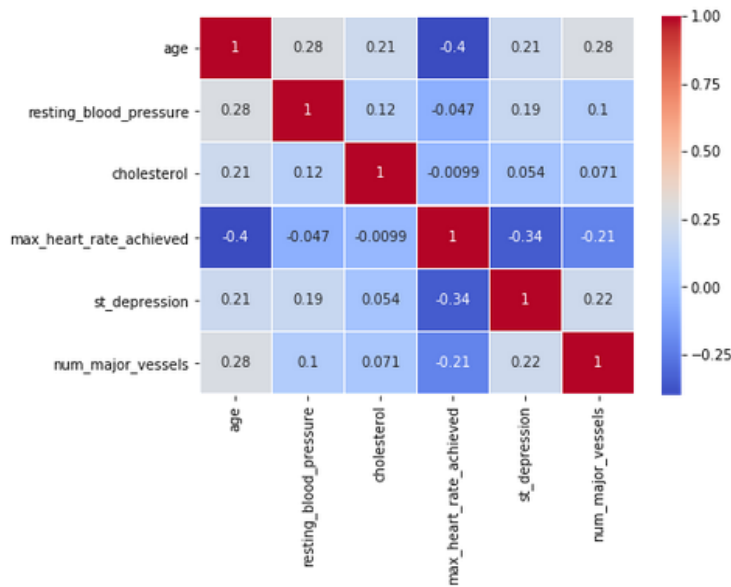
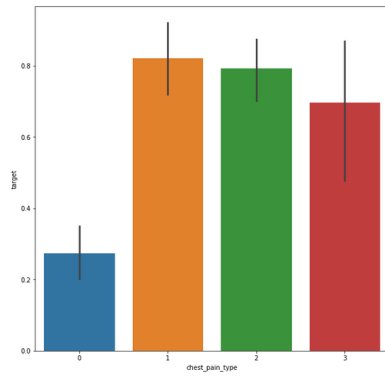
```
In [9]: data.isnull().sum().sum()
```

```
Out[9]: 0

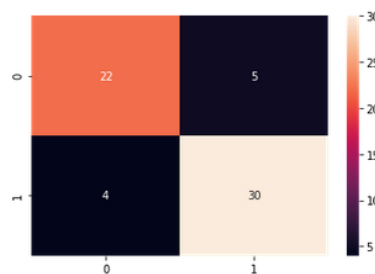
        target      1.000000
        exang      0.436757
        cp         0.433798
        oldpeak    0.430696
        thalach    0.421741
        ca         0.391724
        slope     0.345877
        thal       0.344029
        sex        0.280937
        age        0.225439
        trestbps   0.144931
        restecg    0.137230
        chol       0.085239
        fbs        0.028046
        Name: target, dtype: float64
```

```
1    165
0    138
Name: target, dtype: int64
```





Out [190]: 0.8524590163934426



5. CONCLUSION AND FUTURE SCOPE

CONCLUSION

In this paper, two data extraction algorithms are used in the data to diagnose patient heart disease and analyzed using a distribution model, namely Naïve Bayes distribution and tree distribution solutions. All of these algorithms are applied to the same data level as analyzing the accuracy of the algorithm. The model tree of the solution predicts 91% accuracy for people with heart disease, while its classification Bayve Bayes predicts heart disease with 87% accuracy. Let me finish

Part of the decision tree

Bayesian real money ID 91% 87% 85% 86% 87% 88% 89% 90% 91% 92% PERCENT

Accuracy of classification methods

The best and best of these projects he said is the recognition of medical data algorithms to complete its process.

FUTURE ENHANCEMENTS

Systems developed by machine learning algorithms can be used to predict or detect other diseases. His work can be extended or enhanced to automatically diagnose heart disease, as well as other machine learning algorithms.

6. REFERENCES

- Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.
- Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.
- Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.