International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 10, Issue 2, No 1, 2022. ISSN: 2348-6600
Reference ID: IJCS-425 PAGE NO: 2849-2852

Data Mining Techniques in Heart Disease Detection

Vishnu Shankara M A[#1], Mr.N.Ganapathiram[*2]

[#]*Student, B.Sc Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamil Nadu, India -641021*
vishnushankara46@gmail.com

[*]*Assistant Professor, Department of Computer Science, Rathinam College of Arts and Science, Coimbatore, Tamil Nadu, India -641021*
ganapathiram.cs@rathinam.in

Abstract **-** Mining information and knowledge from large databases has been recognized by many researchers as a key research topic in database systems and machine learning, and by many industrial companies as an important area with an opportunity of major revenues. Researchers in many different fields have shown great interest in data mining. Several emerging applications in information providing services, such as data warehousing and on-line services over the Internet. It also call for various data mining techniques to better understand user behavior, to improve the service provided, and to increase the business opportunities. This paper presents a majority voting ensemble method that is able to predict the possible presence of heart disease in humans. The prediction is based on simple affordable medical tests conducted in any local clinic. Moreover, the aim of this project is to provide more confidence and accuracy to the Doctor's diagnosis since the model is trained using real-life data of healthy and ill patients. The model classifies the patient based on the majority vote of several machine learning models in order to provide more accurate solutions than having only one model. Finally, this approach produced an accuracy of 90the hard voting ensemble model.

Keywords—Data Mining, Large Databases, Data Warehousing, Machine Learning.

I. INTRODUCTION

Recently, our capabilities of both generating and collecting data have been increasing rapidly. The wide- spread use of bar codes for most commercial products, the computerization of many business and government transactions, and the advances in data collection tools have provided us with huge amounts of data. Millions of databases have been used in business management, government ad- ministration, scientific and engineering data management, and many other applications. It is noted that the number of such databases keeps growing rapidly because of the availability of powerful and affordable database systems.

Consequently, data mining has become a research area with increasing importance [30], [70], [76]. Data mining, which is also referred to as knowledge discovery in dufuhases, means a process of nontrivial extraction of implicit, previously unknown and potentially useful in- formation (such as knowledge rules, constraints, regularities) from data in databases [70]. There are also many other terms, appearing in some articles and documents, carrying a similar or slightly different meaning, such as knowledge mining from databases, knowledge extraction, data archaeology, data dredging, data analysis, etc. By knowledge discovery in databases, interesting knowledge, regularities, or high-level information can be extracted from the relevant sets of data in databases and be investigated from different angles, and large databases there!by serve as rich and reliable sources for knowledge generation and verification.

Objective of the project

In the present era, heart disease rates have dramatically increased to become the leading cause of death in the United States upon adults due to the widespread of unhealthy habits [1]. These include a declination in physical activity since the technology trend is moving towards replacing human physical activity and unhealthy eating habits which are directly linked to increasing the risk of having heart diseases. Starting off with the definition of a Heart Disease, according to [2] the National Heart, Lung, and Blood Institute states that heart disease is a disruption to the heart's normal electrical system and pumping functions. Where the disease makes it harder for the heart muscle to pump blood efficiently. Furthermore, according to the World Health Organization

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 10, Issue 2, No 1, 2022. ISSN: 2348-6600
Reference ID: IJCS-425 PAGE NO: 2849-2852

(WHO), 17.9 million people die each year from cardiovascular diseases which correspond to 31% of all deaths around the world [3].

II. SYSTEM DEVELOPMENT

Existing System:

Very few systems use the available clinical data for prediction purposes and even if they do, they are restricted by the large number of association rules that apply. Diagnosis of the condition solely depends upon the Doctors' intuition and patient's records.

Disadvantages:

1. Detection is not possible at an earlier stage.
2. In the existing system, practical use of various collected data is time consuming.

PROPOSED SYSTEM:

The proposed system acts as a decision support system and will prove to be an aid for the physicians with the diagnosis. The algorithm, Fuzzy c means uses clustering and makes use of clusters and data points to predict the relativity of an attribute. Each data point is associated with multiple clusters depending upon the membership degrees.

Advantages:

1. High performance and accuracy rate.
2. FCM is very flexible and is widely in various domains with high rates of success.

III. PROPOSED MODULES

Exploratory Data Analysis

It is a method of understanding the data. With the help of different type of graphs and plots, user can easily identify the values in the data.

Data Pre-Processing

It is the method of cleaning, filtering and arranging the data. In this method, we transform the data in the way we want it.

Model Training

For Heart disease detection, I planned to work with machine learning algorithms. With ML algorithms, this model can be trained.

Model Evaluation and Testing

When the model trained with the help of input data, we need to give some new data to check its accuracy level. When the accuracy level is high, our model is success.

Software Description

Front End: Machine Learning

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning.[5][6] Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain.[7][8] In its application across business problems, machine learning is also referred to as predictive analytics.

Back End: Python Program

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly procedural), object-oriented and functional programming. It is often described as a "batteries included" language due to its comprehensive standard library.

Python is meant to be an easily readable language. Its formatting is visually uncluttered and often uses English keywords where other languages use punctuation. Unlike many other languages, it does not use curly brackets to delimit blocks, and semicolons after statements are allowed but rarely used. It has fewer syntactic exceptions and special cases than C or Pascal.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 10, Issue 2, No 1, 2022. ISSN: 2348-6600
Reference ID: IJCS-425 PAGE NO: 2849-2852

System Implementation

Among various life-threatening diseases, heart disease has garnered a great deal of attention in medical research. There are several factors that increase the risk of heart disease, such as smoking habit, body cholesterol level, family history of heart disease, obesity, high blood pressure, and lack of physical exercise. The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some advanced data mining techniques are used. In this study, an effective heart disease prediction system (EHDPS) is developed using neural network for predicting the risk level of heart disease.

IV. EXPERIMENTAL RESULTS

Figure 1.1:

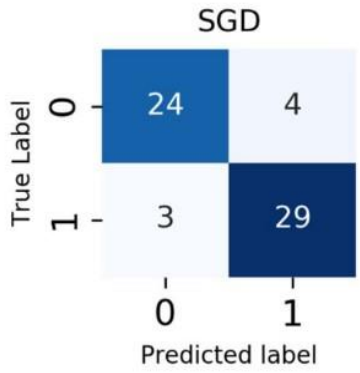


Figure 1.2:

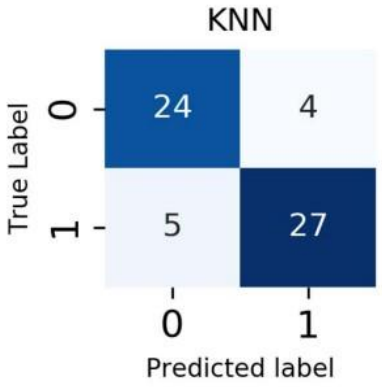


Figure 1.3:

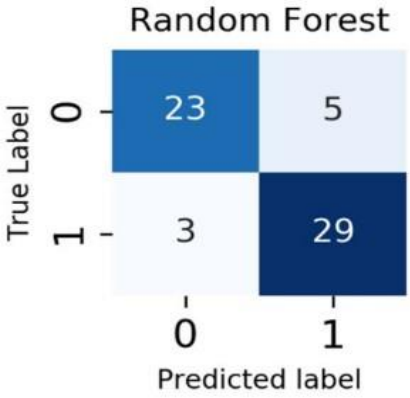


Figure 1.4:

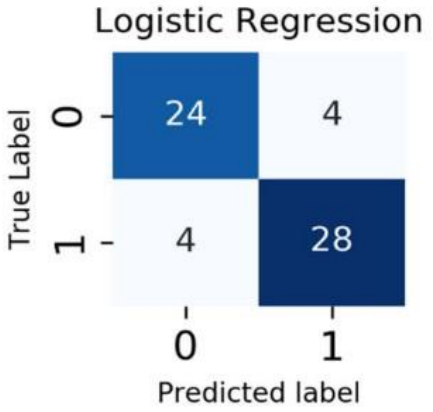


Table 1.1:

Model Name	Accuracy
SGD Classifier	88%
KNN Classifier	87%
Random Forest Classifier	87%
Logistic Regression Classifier	87%
Hard Voting Ensemble Method	90%

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 10, Issue 2, No 1, 2022. ISSN: 2348-6600
Reference ID: IJCS-425 PAGE NO: 2849-2852

Figure 1.5:

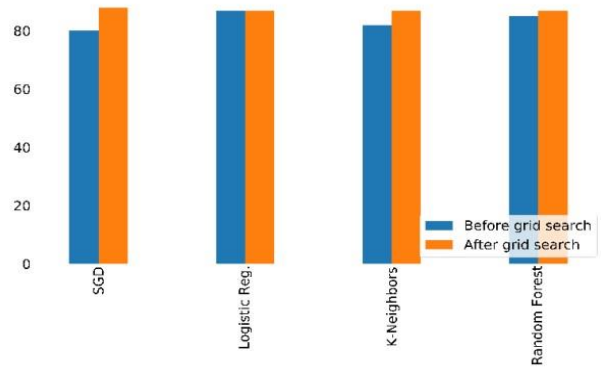
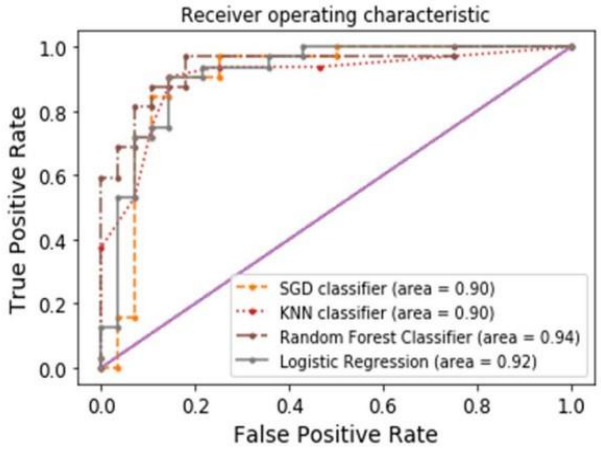


Figure 1.6:



V. CONCLUSION

It is important to note that it takes time to get valid information from data. Therefore, if you are after making your business grow rapidly, there is a need to make accurate and quick decisions that can take advantage of grabbing the available opportunities in a timely manner. Data mining is a rapidly growing industry in this technology-oriented world. Everyone nowadays requires their data to be used in an appropriate manner and with the right approach in order to obtain useful and accurate information.

FUTURE ENHANCEMENT

The proposed web-based application system approach is a friendly user interface, scalable and trustful that can be adopted and implemented in remote areas to try to act like human diagnostic expertise for treatment of heart slight illness. The system can be enlarged in the sense that an system should be trained using local dataset collected from the clinic or hospital acceptable number of records or attributes can be incorporated as well as the new significant rules can be generated using the applied techniques. For instance, the system has been using 14 attributes and 315 records from the hospitals (Clinique la Providence and Hôpital de Reference Nationale) databases. Considering the symptoms variation of a particular disease may vary according to the region, the system should be trained using local dataset collected from the clinic or hospital.

REFERENCES

[1] A. N. Nowbar, J. P. Howard, J. A. Finegold, P. Asaria, and D. P. Francis, "2014 Global geographic analysis of mortality from ischemic heart disease by country, age, and income: Statistics from World Health Organisation and United Nations," International journal of cardiology, vol. 174, pp. 293-298, 2014.

[2] S. Damodaran, “Liver Disease Prediction Using Bayesian Classification," 2014.

[3] C. S. Dangare and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," International Journal of Computer Applications, vol. 47, pp. 44-48, 2012.

[4] A. K. Sen, S. Patel, and D. Shukla, "A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level," International Journal Of Engineering And Computer Science ISSN, pp. 2319-7242, 2013.

[5] P. Harrington, Machine learning in action: Manning, 2012.