

Heart Disease Prediction System Using Decision Tree Classifier

¹Deepa V. Shete, ²Prof.Deshmane K. A.

Department of Computer Engineering, SVERI's College of Engineering, Pandharpur.

Abstract- As huge amount of information is produced in medical associations (healing facilities, therapeutic focuses) yet this information is not properly utilized. The health care system is "data rich" however "knowledge poor". There is an absence of successful analysis methods to find connections and patterns in health care data. Data mining methods can help as remedy in this circumstance. For this reason, different data mining techniques can be utilized. The paper intends to give details about various techniques of knowledge abstraction by using data mining methods that are being used in today's research for prediction of heart disease. In this paper, data mining methods namely, Naive Bayes, Neural network, Decision tree algorithm are analyzed on medical data sets using algorithms.

Keywords- Decision trees, Neural networks, Naive Bayes Classifier.

1. INTRODUCTION

The headway of information technology, framework coordination and additionally programming advancement, systems have formed an imaginative era of multifaceted computer framework. Information technology specialists have been offered few challenges by these frameworks. A case of such framework is the healthcare services framework. Recently, there has been an augmented attention to make utilization of the headway of data mining advances in healthcare frameworks. Thus, the target of the present effort is to find out the aspects of use of healthcare data for aid of people by method of machine learning furthermore data mining procedures. The main aim is to suggest an automated system for diagnosing heart diseases by taking into account earlier information and data.

A major challenge confronting healthcare associations i.e. hospitals, medicinal focuses are the procurement of quality services at reasonable expenses. Quality services suggest diagnosing patients accurately and overseeing medicines that are more effective. Poor clinical decisions can prompt to poor outcomes which are therefore unsatisfactory. Healthcare organizations can reduce costs by accomplishment of computer-based data and/or decision support systems. Healthcare Services data is very huge as it incorporates patient records, resource management information and updated information. Human services associations must have capacity to break down information. Treatment records of many patients can be stored Away in computerized way; furthermore data mining methods may help in finding out a few vital and basic inquiries related with healthcare organizations. Clinical choices are

frequently made focused around doctors' instinct and experience instead of on the knowledge rich information covered up in the database. This practice prompts undesirable biases, blunders and unnecessary medicinal expenses, which influence the quality of services given to the patients. Wu, et al proposed that combination of clinical choice backing with computer based patient records could decrease medical errors, enhance safety of patients, lessening undesirable practice variety, and enhance patient outcome. This suggestion is guaranteeing as the data demonstrating and analysis tools for example data mining, have the possibility to create a knowledge rich environment which can help to essentially enhance the nature of clinical decisions.

2. LITERATURE SURVEY

K Raj Mohan, Ilango Paramasivam, Subhashini, SathyaNarayan " Prediction and Diagnosis of CardioVascular Disease". [1] Traditional way of predicting heart disease is through physician's examination or a number of medical tests such as ECG Stress test, Heart MRI, CT etc., Computer based information along with advanced data mining techniques are used for appropriate results. The main aim of this study is to detect the various causes of cardiovascular diseases by means of machine-learning techniques with the help of clinical diagnosis. For detecting these image analysis data is used. The aim of this research work is to develop a framework for detecting causes by means of data mining and machine-learning techniques.

Ms. Priti V. Wadal, Dr. S. R. Gupta, "Predictive Data Mining For Medical Diagnosis:[2] The Healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. A wide variety of areas including marketing, customer relationship management, engineering, medicine, crime analysis, expert prediction, Web mining, and mobile computing, besides others utilize Data mining.

Aqueel Ahmed, Shaikh Abdul Hannan [3]. Heart disease is a major cause of morbidity and mortality in modern society. Medical diagnosis is extremely important but complicated

task that should be performed accurately and efficiently. Although significant progress has been made in the diagnosis and treatment of heart disease, further investigation is still needed. The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. There is a huge data available within the healthcare systems. However, there is a task of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous application in business and scientific domain. Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets. Disease diagnosis is one of the applications where data mining tools are proving successful results. This research paper proposed to find out the heart diseases through data mining, Support Vector Machine (SVM), Genetic Algorithm, rough set theory, association rules and Neural Networks.

K.Srinivas B.Kavihta Rani Dr. A.Govrdhan[4]. The healthcare environment is generally perceived as being 'information rich yet 'knowledge poor'. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system.

Deepali Chandna[5]. This thesis proposes new methods intended for investigating feature selection techniques as well as develop new machine learning algorithms designed for providing automatic Computer aided analysis and decision support system for heart disease diagnosis. The aim is to build up an integrated structure with a righteous workflow (constructing missing features values, feature selections, and classification algorithms). In prerequisites of features selection techniques, the research decided on features selection technique as a process to increase high superiority attributes to improve the mining process.

3. METHODOLOGY

Methodology with respect to each objective is briefly given below: In this prediction of heart disease, following classification models of analyzed:

1. *Decision trees.*

The decision tree approach is one of the most powerful techniques in classification in data mining. It builds the models in the form of tree structure. Mainly, dataset breaks in small sets and concurrently, an associated decision tree is formed. Decision trees can handle both numerical data and categorical data. For medical purpose, decision trees determine order in different attributes and a decision is then taken based on the attribute. There are various decision tree algorithms that are used. Most preferred algorithm is ID3 i.e. Iterative Dichotomized 3 by J. R.Quinlan. ID3 uses information gain and entropy to classify data in tree.

2. *Neural networks*

An artificial neural network is information processing method encouraged by biological nervous system. Dr. Robert Hech-Nielsen. He defines a neural network as: "a computing system made up of a number of simple, highly interconnected processing elements which process information by their dynamic state response to external inputs". (In "Neural Network Primer: Part I" by Maureen Caudill, AI Expert Feb 1989).

Neural network is organized into number of layers consisting of huge number of element that are highly interconnected i.e. neurons that have an activation function. Different patterns are generated with input layer that communicates with one or more hidden layers and finally output layer is generated. Mostly, ANN consists of 'Learning rule' that modify the weight of connections. Learning in neutral net can be of both types i.e. supervised learning and unsupervised learning. An artificial neural network consists of three layers i.e. input layer, hidden layer and output layer. The principal layer is the input layer and last layer is output layer. Between input and output layer there may be extra layer i.e. hidden layer. a neural network can easily be trained to perform different functions by adjustments in values of weight among elements

3. *Naive Bayes Classifier :*

Naive Bayes classifier is a simple probabilistic classifier that depends on Bayes' theorem with strong i.e. naive independence assumptions. It is also be called as "independent feature model". In general terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Naive Bayes classifiers are trained to work in supervised learning. Naive Bayes classifier mainly pre assumes the effect of a variable value on predefined class that is not dependent on value of other variable. This is called as property of class conditional independence. It is particularly suited when the dimensionality of the inputs is high. Naive Bayesian is mainly used to form models with predictive capabilities.

Bayes' Theorem:

Probability (B given A) = (Probability (A and B)/ Probability (A))

Assume X as a data tuple. Let H be any hypothesis. P (H|X) be posterior probability of the H that is conditioned on X. In the same way, P (X|H) is the posterior probability of X condition on H.

$P(H|X) = (P(X|H)P(H)/ P(X))$

P(H) is prior probability of H.

4. IMPLEMENTATION

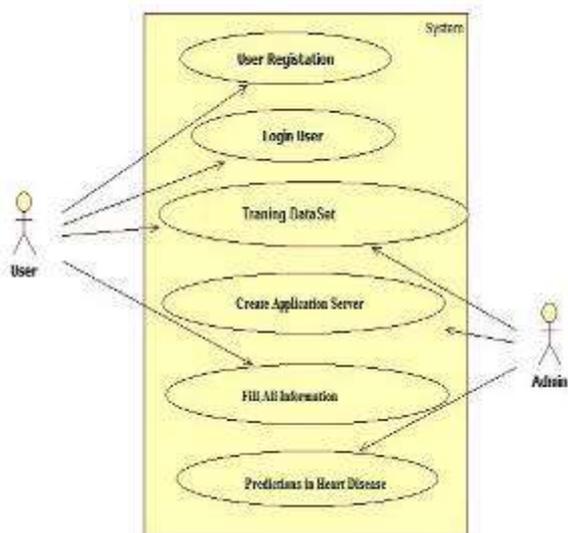


Fig 1: Use Case Diagram

Heart is vital part or an organ of the body. Life is subject to proficient working of heart. In the event that operation of heart is not proper, it will influence the other body parts of human, for example, mind, kidney, etc. Heart is simply a pump, which pumps the blood through the body. In the event that if blood in body is insufficient then many organs like cerebrum suffer and if heart quits working by, death happens inside minutes. Life is totally subject to effective working of the heart. The term Heart sickness alludes to illness of heart & vessel framework inside it.

There are number of elements which build the danger of Heart infection:

- Family history of coronary illness
- Smoking
- Poor eating methodology
- High pulse
- Cholesterol
- High blood cholesterol
- Obesity
- Physical inertia

4. CONCLUSION

In this paper we studied how data mining techniques brings with set of techniques to find out hidden patterns for making decision in healthcare organizations. We focused on classification methods of data mining used in data discovery. Different classification techniques of data mining have merits and demerits for data classification and knowledge extraction. Furthermore, neural networks, decision trees or naïve Bayes can be studied in more detail to implement an algorithm that is helpful in healthcare organizations.

REFERENSES

[1] K Raj Mohan, Ilango Paramasivam ,Subhashini, SathyaNarayan “ Prediction and Diagnosis of Cardio Vascular Disease – A Critical Survey”, 2014 World Congress on Computing and Communication Technologies.

[2] Ms. Priti V. Wadal, Dr. S. R. Gupta, “Predictive Data Mining For Medical Diagnosis: An Overview Of Heart Disease Prediction “International Conference on Industrial Automation and Computing (ICIAC- 12-13th April 2014)

[3] Aqueel Ahmed, Shaikh Abdul Hannan, “Data Mining Techniques to Find Out Heart Diseases: An Overview”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), September 2012.

[4] K.Srinivas B.Kavihta Rani Dr. A.Govrdhan, “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”, (IJCS) International Journal on Computer Science and Engineering, 2010.

[5] Deepali Chandna, “Diagnosis of Heart Disease Using Data Mining Algorithm”, 1678-1680, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014

[6] K.Sudhakar, Dr. M. Manimekalai, “ Study of Heart Disease Prediction using Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014.

[7] Qeethara Kadhim Al-Shayea, “Artificial Neural Networks in Medical Diagnosis”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

[8]Ms.Rupali R.Patil,” Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014.

[9] Dilip Roy Chowdhury, MridulaChatterjee & R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2012.

[10] JareeThongkam, GuandongXu, Yanchun Zhang and Fuchun Huang 'Breast Cancer Survivability via AdaBoost Algorithms'HDKM,2008, wollongon,Australia.

[11] Hsinchun Chen, Sherrilynne S. Fuller, Carol Friedman, and William Hersh, "Knowledge Management, Data Mining, and Text Mining In Medical Informatics", Chapter 1, eds. Medical Informatics: Knowledge Management And Data Mining In Biomedicine, New York, Springer, pp. 3-34, 2005.

[12] Charly, K.: “Data Mining for the Enterprise”, 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7, 295 304, 1998.