Review of Heart Disease Prediction using Data Mining Classifications

Kamal Kant [1] Dr. Kanwal Garg[2]
[1]Scholar, M.Tech(CSE) [2]Assistant Professor
[1, 2] Deptt. Of Computer Science & Applications
[1, 2] Kurukshetra University, Kurukshtra (Haryana)

Abstract--- In the research data mining techniques will helpful to handle this technique of predictive model. Research will show the most effective parameter of the heart disease prediction and will get the scenario for least predictive value and most predictive value in whole data mining technique. Initially need to identify the exact state of the user entering parameter which can be frequent item although random value of dataset in data modeling .This research has a prototype Heart Disease Prediction using data mining techniques, namely Naïve Bayes which in artificial Neural Network concept. Results show that each technique has its unique strength in identify the objectives of the defined mining goal and expert system .Researchers are using some medical exponent attributes such as age, sex, blood pressure and blood sugar , glucose and some related factors can predict the likelihood of patients getting a heart disease with its exact probability. Basically this technique is expended on the defected and non-defected parameter which works as result class.

Keywords: - Back propagation, Data mining, Heart disease, Multilayer perceptron neural network, Neural Network, Naïve Byes

I. INTRODUCTION

Heart is the major part of human body. If operation of heart is not proper, it will affect the other body parts of human such as brain, kidney, bones etc. Heart disease is a disease that effects on the operation of heart and some internal bones. There are number of factors which increases risk of Heart disease which the author using in our research dataset. Some of them are listed below:

- Genetic heart disease
- Smoking factor
- Cholesterol week entity
- High blood pressure
- Obesity
- Lack of physical exercise during the age factor
- KFT

The World Health Organization (WHO) has estimated that 14 million deaths occur worldwide, every year due to the Heart diseases cause. In 2010, 17.9 million people died due to Heart Disease[1]. Over 90% of deaths in world are because of Heart disease. WHO prediction by 2030, almost 23.6 million people will die due to Heart disease as per forecasting. Predication should be done to reduce risk of Heart disease [1]. Diagnosis is usually based on signs, symptoms and physical examination of a patient. Almost all the doctors are predicting heart disease by learning and experience. The diagnosis of disease is uphill and negative task in medical field [2]. Predicting Heart disease from various factors or symptoms is a multi-layered issue which may lead to false presumptions and unpredictable effects. Healthcare industry today generates large amounts of complex data about patients, hospitals resources, disease diagnosis [3], electronic patient records, medical devices etc. The large amount of medical data is a key resource to be processed and filtering for knowledge extraction that enables support for cost-savings and decision making in evaluation factor [4].

There are some important factor to predict the future forecasting in heart disease prediction and for this study to whole dataset is needed, which is used to find the probable damping factor for the byse classifier .

II. LITERATURE REVIEW

An Heart Disease Prediction System is developed by using data mining techniques Naive Bayes, Neural Network, and Decision Trees was proposed by Sellappan Palani[5]. Each method has its own strength to get appropriate results. To build this system hidden patterns and relationship between them is used [6, 7]. It is web-based, user friendly & expandable. The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [8, 9]. The dataset contains records with 13 attributes in each record. The supervised networks i.e. Neural Network with back propagation algorithm is used for training and testing of data. Three different supervised machine learning algorithms. They are Naïve Bayes, K-NN, and Decision List algorithm. These algorithms have been used for analyzing the heart disease dataset [10]. Tanagra data mining tool is used for classifying these data. These classified data is evaluated using 10 fold cross validation and the results are compared. Decision tree is one of the popular and important classifier which is easy and simple to implement. It doesn't have domain knowledge or parameter setting [11]. It handle huge amount of dimensional data. It is more suitable for exploratory knowledge discovery. The results attained from Decision Tree are easier to interpret and read [12]. Naïve Bayes is a statistical classifier which assigns no dependency between attributes. To determine the class the posterior probability should be maximized. The advantages are one can work with the naïve bayes model without using any Bayesian methods. Here Naïve Bayes Classifiers performs well.

Decision Support in Heart Disease Prediction System Using Neural Network [5], 2007 Niti Guru et al proposed the prediction of various disease like Sugar, Heart disease, Blood Pressure with the use of neural networks. The Neural Network is tested and trained with 13 input variables such as smoke, Age, obesity, Blood Pressure, Angiography's report and the like.

III. RESEARCH OBSERVATION

The main objective of this research is to develop a prototype Health Care Prediction System using, Naïve Bayes .The System can discover and extract hidden knowledge associated with diseases (heart attack, cancer and diabetes)

All rights reserved by www.ijsrd.com

from a historical heart disease database[13,14]. However, they cannot answer complex queries like Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and Given patient records, predict the probability of patients getting a heart disease.

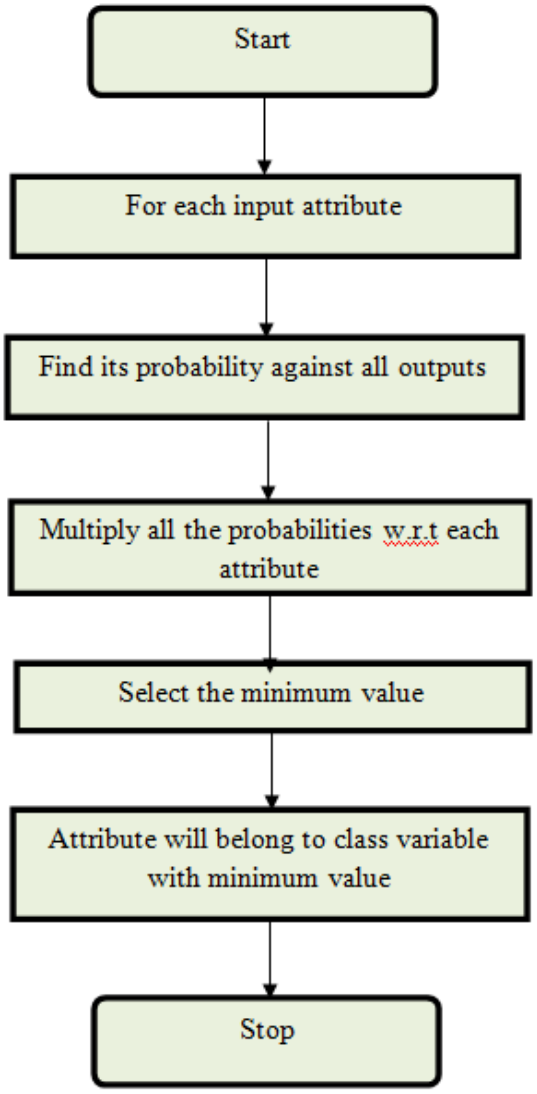

Fig. 1:

1. In the statistical probability the researchers gaining the predictive value using mining concept of naïve byes .
2. Aiming an expert system which identify for the diseases predicted values on the basis of iterations, then have to evaluate the resulting parameter on the basis of disease occurrence parameter at the end user.
3. The existing work identify the statistical probability in disease tracking system dataset attribute value And identify the major comparison in between to whole dataset in data mining that further software can be improved the system at the testing case.
4. The real time expert system requirement had more efficient result ,at the existing work basically identify only statistical value until don't have no such idea how to find out finite state for the Disease parameter.
5. The major parameter should contain finite state of value in finite automata at the time of machine design. Evaluation for the finite state should have detection rate which should be parallel for finite automata machine.

IV. METHODOLOGY

The objectives and requirements from a business perspective, converting this knowledge into a data mining problem definition, and designing a preliminary plan to achieve the objectives[9]. Data understanding phase uses the raw the data and proceeds to understand the data, identify its quality, gain preliminary insights, and detect interesting subsets to form hypotheses for hidden information[7]. Data preparation phase constructs the final dataset that will be fed into the modeling tools. This includes table, record, and attribute selection as well as data cleaning and transformation.

V. BAYES' RULE

More generally, the above is just an application of Bayes' Theorem.

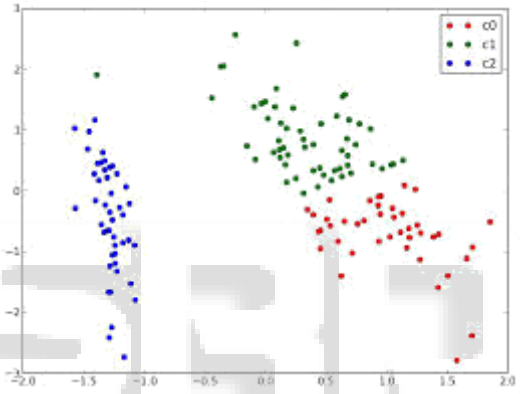

Fig. 2:

– Probability of event H given evidence E:

$$Pr(H \mid E) = \frac{Pr(E \mid H) * Pr(H)}{Pr(E)}$$

– A priori probability of H= Pr(H)
  - Probability of event before evidence has been seen
– A posteriori probability of H= Pr[H|E]
  - Probability of event after evidence has been seen
– Classification learning: what's the probability of the class given an instance?
  - Evidence E = instance
  - Event H = class value for instance
– Naive Bayes assumption: evidence can be split into independent parts (i.e. attributes of instance!

$$Pr(H \mid E) = \frac{Pr(E1 \mid H) * Pr(E2 \mid H) * .... * Pr(En \mid H) * Pr(H)}{Pr(E)}$$

Return the classification with highest probability

– Probability of the evidence Pr(E)
  - Constant across all possible classifications;
  - So, when comparing N classifications, it cancels out

All rights reserved by *www.ijsrd.com*

A. Numerical errors

From multiplication of lots of small numbers

- Use the standard fix: don't multiply the numbers, add the logs

B. Missing values

Missing values are a problem for any learner. Naive Bayes' treatment of missing values is particularly elegant.

- During training: instance is not included in frequency count for attribute value-class combination
- During classification: attribute will be omitted from calculation.

VI. CONCLUSION

The most effective model to predict patients with heart disease appears to be Naïve Bayes followed by Neural Network and Decision Trees. Five mining goals are defined based on business intelligence and data exploration. The goals are evaluated against the trained models. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions .The main objective of this research is to develop a prototype Heart Disease Prediction using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network.

REFERENCES

[1] Charly, K.: "Data Mining for the Enterprise", 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7, 295-304, 1998.
[2] Ho, T. J.: "Data Mining and Data Warehousing", Prentice Hall, 2005.
[3] Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: "CRISP-DM 1.0: Step by step data mining guide", SPSS, 1-78, 2000.
[4] Giudici, P.: "Applied Data Mining: Statistical Methods for Business and Industry", New York: John Wiley, 2003.
[5] Shantakumar B.Patil,Y.S.Kumaraswamy," Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656
[6] Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.
[7] Mehmed, K.: "Data mining: Concepts, Models, Methods and Algorithms", New Jersey: John Wiley, 2003.
[8] Fayyad, U: "Data Mining and Knowledge Discovery in Databases: Implications for scientific databases", Proc. of the 9th Int. Conf. on Scientific and Statistical Database Management, Olympia, Washington, USA, 2-11, 1997.
[9] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Technique, 978-1-4244-1968-5/08/$25.00 ©2008 IEEE
[10] Shantakumar B.Patil, Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research, ISSN 1450-216X, Vol.31 No.4 (2009), pp.642-656.
[11] Mohd, H., Mohamed, S. H. S.: "Acceptance Model of Electronic Medical Record", Journal of Advancing Information and Management Studies. 2(1), 75-92, 2005.
[12] Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", http://mlearn.ics.uci.edu/databases/heart-disease/, 2004.
[13] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
[14] Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

All rights reserved by *www.ijsrd.com*