Volume 4, Issue 6, June 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Applying Decision Tree Algorithm to Predict Jaundice

S. Gomathi*
I CT & Sri Krishna Arts and Science College,
India

B.Suchitra
I CT & Sri Krishna Arts and Science College,
India

Abstract— The World wide web plays a vital role in every day's life. As technology arises, the need for web is also necessary and large amount of data is available in Web. New Technologies and Techniques are used in order to predict and to provide solutions for problem reality, The fields such as medicine, telecommunication, banking, business intelligence plays a significant role and researches focus to bring out easy elucidations for new issue. This paper focus on the prediction and to create awareness to the users about jaundice. Data mining techniques are used to predict jaundice and various algorithm are stated and it is compared .The Prediction is not based on infants but it relates to all ages This paper helps people to know about the disease, how it is caused and the symptoms and also say about how techniques are used to predict the diseases in earlier stages.

Keywords— Data mining, classification, decision tree, bilirubin, RBC cells, liver, bile ,intestine

I. INTRODUCTION

Jaundice is a common disease for all ages. It is based on the bilirubin level. The byproduct of the normal breakdown of old red blood cells is called bilirubin. When the bilirubin level is high, the people will get affected from jaundice.RBC plays a vital role in human body. In human, normally bilirubin passes through the liver and is removes the waste through the intestines as bile which is yellow coloured fluid which helps in digestion of fats. The common disease can causes in all age groups, but it is categorized into two types. First it is focus for infants and latter to adults. In infants, the premature babies who have deliberate to process are susceptible to jaundice. Jaundice is common for breast feeding babies who aren't get enough milk and mothers who naturally produce substances which leads to high level of bilirubin. According to AAP(Academic to Pediatrics)the high bilirubin causes hearing loss and also brain damage in some cases. Soon after birth, the child should be examined for jaundice [1].If the infant is suffering from high level of bilirubin, they must be provided with phototherapy or exchange transfusions. Jaundice stages are grouped into three types they are: (1) Pre-Hepatic, (2) Hepatic (3) Post- Hepatic .These three stages are related to liver and they said to be before secretion to the liver, arises within the liver and arises after bilirubin is excreted from the liver [2].

Data Mining is a collection of techniques for efficient automated discovery of unknown, novel, valid, understandable and useful patterns in voluminous databases. The patterns are actionable so they are used in decision making[2].

II. CLASSIFICATION

Classification is the separation or ordering of objects into classes. Classification can be grouped into two types. They are apriori classification and posterior classification. The classes in the classification is created by looking at the data as well as without looking at that data.

III. DECISION TREE

One of the important method in classification is called decision tree. It is a tree like structure where each node represent a test on a attribute value and each branch represent an outcome of the test. The leaves in the tree is represented as classes. The decision tree model can be both descriptive and predictive. the decision tree in data mining consists of two types they are classification tree and regression tree. The classification tree is used , when the predicted outcome is the class in which the data belongs. Regression tree is used, when the predicted outcome is considered as a real number. The term classification and regression tree(cart) is introduced by Breiman et al [4].it is based on the umbrella term. The classification and regression trees have some similarities between them. And the differences is based on how to split the data. There are some techniques which can be construct more than one decision tree. They are a) bagging: it is an early ensemble method.

This method is used to focus to build multiple decision tree by iterating with the resampling training data with replacement [3]. B) random forest : it uses a number of decision tree, in order to improve the classification rate. C) boosted trees : it is used for regression and classification type problems [5,6]. D) rotation forest : tree is trained by applying pca on a random subset of input features [7]. There are many algorithms for decision tree. They are 1) id3 : it is known as iterative dichotomizes. It is invented by Ross quilan [25].the dataset are used, based on dataset they generate a decision tree. It is mainly used in machine learning and natural language processing. 2) c4.5 : it is an successor of id3.this algorithm is used for classification and it is also known as statistical classifier. It uses the concept of information entropy.

The implementation is done on j48.the j48 is an open source java implementation of c4.5 algorithm . it supports in weka data mining tool. 3) chaid : (chi- squared automatic interaction detection)chaid is one of the type of decision tree technique, it is based upon adjusted significance testing. This technique was developed in south Africa.chaid can be used for prediction as well as classification, and also for detection of interaction between variables. 4) mars : (multi variate adaptive regression splines) it is the form of regression analysis. It is a non-parametric regression technique. It is an extension of linear models ,that automatically models non-linearities and interactions between variables[8].

IV. LITERATURE REVIEW

Ahmad taher azar et al. [1] developed a prototype Intelligent Heart Disease Prediction System using data mining techniques, like Naïve Bayes, Neural Network Decision Trees. IHDPS can answer complex what if queries, where traditional decision support systems fails to answer. Using medical profiles such as sex, age, blood sugar and blood pressure it can predict the likelihood of patients getting a heart disease [1].

D.Lavanya and Dr.k.usha rani discussed that Data mining can contribute with important benefits to the blood bank sector. J48 algorithm and WEKA tool have been used for the complete research work. Classification rules performed well in the classification of blood donors, whose accuracy rate is approximately 89.9% [2].

Alsabi k et.al., [3] studied about the inductive learning algorithm (Decision Tree) which is implemented in the intelligent system. The main objective of the research is to construct the decision tree using java language until the appropriate classification is reached. According to Ahmad taher azar et al[1] suggest that the decision tree can be used in order to detect the cancer and they used decision support tool. They specifies in order to focus for large datasets and to categorize datasets, decision tree is widely used. D.Lavanya and Dr.k.usha rani [2] suggest that for medical diagnosis the decision tree are very useful. they used decision tree for diagnosis of tumor and also for ovarian cancer and heart diagnosis. According to Johannes Mair et al (16) the decision tree algorithm work well in order to find out the accurate myocardial infarction.and acts as a diagnostic aid.

V. PROPOSED WORK

Decision tree supports TOP-DOWN greedy algorithm. The Decision tree algorithm focus for sample of training data. If some of the attributes are continuous valued, they should be discretized.and attributes are splitted. It is done until leaves of the decision tree has no subchild. Two rules are used to evaluate they are (i) Rules based on information theory, (ii) Rules based on gini Index. The information theory algorithm is also called as entropy. The algorithm is based on claude shannon's idea[2]. If there is a uncertainty then information can be used. The information is defined as –pi log pi, where pi is the probability of event. Information of an event that is likely to have several possible outcomes is given by

$$I=\sum(-pi \log pi).$$

VI. CONCLUSION

In this paper, we focused for various algorithm used in decision tree and also states how decision tree plays a significant role in medical field. The common attributes are used for diagnosis of jaundice. Further the algorithm can be implemented to show the accuracy rate. In future the algorithm are applied, and implementation will be based on scalability and optimum solution is obtained when compared to other algorithms in data mining

REFERENCES

[1] Ahmad taher azar "Decision tree classifiers for automated medical diagnosis" ,springer,December 2013,volume 23,issue 7-8,pp 2387-2403 .

[1] D.lavanya and Dr.k.usha rani ," Performance evaluation of decision tree classifier on medical datasets", International journal of computer application (0975-8887) vol 26 no 4 ,July 2011.

[2] Alsabi K, Ranks. S ,Singh V (1998)clouds: " A decision tree classifier for large data sets" in: proceedings of the fourth international conference on knowledge discovery and data mining(kdd- 98),august ,27-31.aaai press,new York city,usa ,pp 2-8

[3] Breiman,Leo; Friedman,j.h; olshen,r.a;stone,c.j(1984) classification and regression tree.Monterey ,ca: Wadsworth and brooks Cole advanced books and software.ISBN 978-0-412-04841-8.

[4] Brieman l(1996)bagging predictor." Machine learning",24": pp 123-140

[5] Friedman,j.h,(1999)stochastic gradient boosting, Stanford university.

[6] Hastie,t,tibshirani,r.,Friedman,j.h(2001)." The elements of statistical learning : data mining",inference and prediction, newyork: springer verlag

[7] Rodriquez, J.J and kuncheva, l.I and Alonso,c.j(2006), rotation forest: " a new classifier ensemble method , IEEE transaction on platform analysis and machine intelligence " , 28(10): 1619-1630.

[8] Quilan,j.r.1986.induction of decision tree, machine learning,1.1(march 1986),81-106.

[9] Fried man,j.H(1991),"Multivariate adaptive regression splinter", the annals of statistics 19:1.

