International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Predicting Ailment of Thyroid Using Classification and Recital Indicators

A. Kiruthika, P. Deepika, Dr. S. Sasikala, S. Saranya
Department of Computer Science, Hindusthan College of Arts And Science, Coimbatore, Tamil Nadu, India

ABSTRACT

Thyroid disease is one of the deadly diseases for human. Data mining is the popular area which helps to provide various methodologies to predict and identification of various diseases in health care domain. The medical data having vast amount of data and classifying these data is one the challenging task. Moreover data mining technique has been applied in various sectors the classification results of the medical dataset which helps the way of treatments to the patients.

Keywords : Thyroid Using Classification, Recital Indicators, Data Mining, Diseases,Health Care Domain, Hyperthyroidism and hypothyroidism, Levothyroxine, UCI Repository

I. INTRODUCTION

Thyroid hormones are created by the tyroid glad introduce two active thyroid hormones including levothyroxine and Triiodothyronine to control the human body's metabolism. These are important to help each cell in each tissue and organ to work correctly and give protein in the regulation of body temperature, and in overall energy production and regulation. Thyroid function affects every essential organ in the body. The significance of thyroid disorders should not do justice to thyroid storm (an episode of severe hyperthyroidism) and myxedema coma (the end stage of untreated hypothyroidism) may cause death in a substantial number of cases [3].

Thyroid is a butterfly-shaped gland, which is located at the bottom of the throat responsiblefor producing two active thyroid hormones, levothyroxine (T4) and triiodothyronine (T3) that affectsome functions of the body such as: stabilizing body temperature, blood pressure, regulating theheart rate etc. Reverse T3 (RT3) is manufactured from thyroxine (T4), and its role is to block theaction of T3. An abnormal function of the thyroid implies the occurrence of hyperthyroidism andhypothyroidism, two of the common thyroid affections.Hypothyroidism (underactive thyroid or lowthyroid) means that the thyroid gland doesn't produce enough of certain important hormones.Without an adequate treatment, hypothyroidism can cause various health problems such as: obesity,joint pain, infertility and heart disease. Hyperthyroidism (overactive thyroid) refers to a condition inwhich the thyroid gland produces too much of the hormone thyroxin.

DATA MINING

Data Mining is the process of semi-automatically analyzing large databases to find patterns. Classification is a data mining (machine learning) technique used to predict group membership for data instances.In this paper, J48, decision stump Algorithm is used to predicate thyroid disease.A data set is downloaded from UCI repository site is used for the experimental purpose. The entire work is carried out with WEKA open source software under Windows 7environment.

CLASSIFICATION

Classification is a process that is used to find a model that describes and differentiate data classes or concepts, for the purpose of using the model to predict the class of objects whose class label is unknown.

DECISION TREE

Berry and Linoff defined decision tree as "a structurethat can be used to divide up a large collection of records into successive smaller sets of records by applying a sequence of simple decision rules. With each successive division, the members of the resulting sets become more and more similar to one another."

Decision tree is similar to the flowchart in which every non-leaf nodes denotes a test on a particular attribute and every branch denotes an outcome of that test and every leaf node have a class label. The node at the top most labels in the tree iscalled root node. Using Decision Tree, decision makers can choose best alternative and traversal from root to leaf indicates unique classseparation based on maximum information gain.

There are three main steps for classification by using decision trees: The first step is the learning process. The model is constructed on the training data. Hence, this model is presented by classification rules. In the second step, a test is selected in order to calculate the model accuracy. The model is accepted according to the value of this test. If this value is considerably accepted, the model could be used for the classification of a new datum. At last, the third step includes the usage of the model for a classification or prediction of a new data (Figure 2).

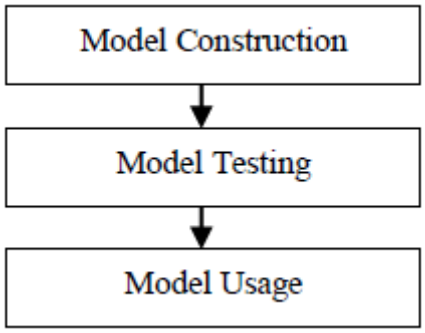


Figure 2. Steps for Classification and Prediction Process

Data Set Description

The hypothyroid dataset used in this work is collected from the website. The hypothyroid dataset consists of 3772 instances from which 3481 instances belongs to category negative,194 instances belongs to category compensated hypothyroid , 95 instances belongs to primary hypothyroid category while 2 instances belongs to category secondary hypothyroid. There are totally 30 attributes. The hypothyroid data set is given below.

Table 1

DATA DESCRIPTION SN	ATTRIBUTE NAME
1	age
2	sex
3	on thyroxine
4	query on thyroxine
5	on antithyroid medication
6	sick
7	pregnant
8	thyroid surgery
9	I131 treatment
10	query hypothyroid
11	query hyperthyroid
12	lithium
13	goitre
14	tumor
15	hypopituitary
16	psych
17	TSH measured
18	TSH
19	T3 measured
20	T3

21	TT4 measured
22	TT4
23	T4U measured
24	T4U
25	FTI measured
26	FTI
27	TBG measured
28	TBG
29	referral source
30	Class

II. METHODOLOGY

a. Preprocessing

Data preprocessing is a data mining technique. It is used to reduce the volume of data. There are many data reduction techniques are available such as data compression, numerosity reduction, dimensionality reduction and discretisation. In our work, we have used dimensionality reduction to select the subset of attributes from original data.

b. Classification

Classification is one of the data mining Technique. It is used to group the instances which belong to same class. It is a supervised learning, in which predefined training data is available. Most popular data mining classification techniques are decision trees and neural networks.

Decision tree

Decision tree is one of the classification technique in data mining. It is tree-like graph. [5] The internal node denotes a test on attribute, each branch represents an outcome of the test, and the leaf node represent classes. It isa graphical representation of possible solutions based on condition from these solutions optimum course of action is carried out. In our work, we have used two decision tree classifier such as decision stump and J48 to classify the hypothyroid data set.

1. J48 Algorithm

J48 is a tree based learning approach. It is developed by Ross Quinlan which is based on Iterative Dichotomiser 3 (ID3) algorithm.[5] J48 uses divide-and-conquer algorithm to split a root node into a subset of two partitions till leaf node (target node) occur in tree. Given a set T of total instances the following steps are used to construct the tree structure.

Step 1: If all the instances in T belong to the same group class or T is having fewer instances, than the tree is leaf labeled with the most frequent class in T.

Step 2: If step 1 does not occur then select a test based on a single attribute with at least two or greater possible outcomes. Then consider this test as a root node of the tree with one branch of each outcome of the test, partition T into corresponding T1 , T2 , T3 ....., according to the result for each respective cases, and the same may be applied in recursive way to each sub node.

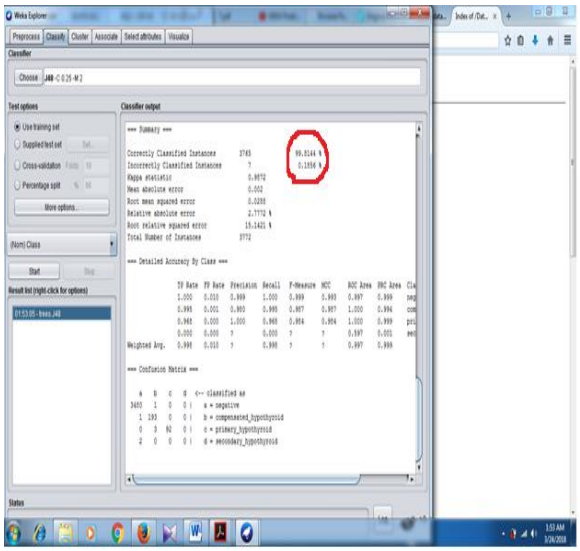
Step 3: Information gain and default gain ratio are ranked using two heuristic criteria by algorithm J48.

2. Decision Stump

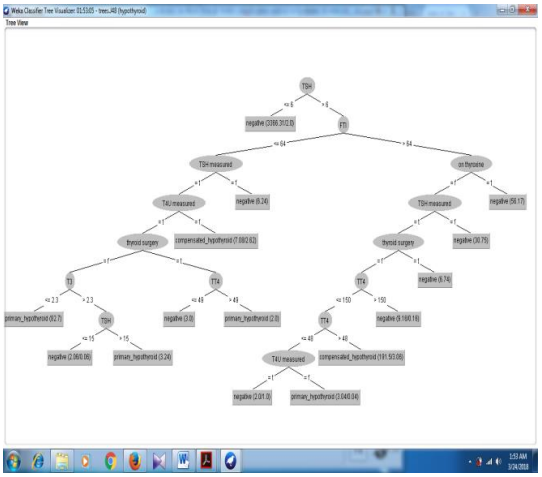
A **decision stump** is a machine learning model consisting of a one-level decision tree. That is, it is a decision tree with one internal node (the root) which is immediately connected to the terminal nodes (its leaves). A decision stump makes a prediction based on the value of just a single input feature. Sometimes they are also called 1-rule.

WEKA RESULTS

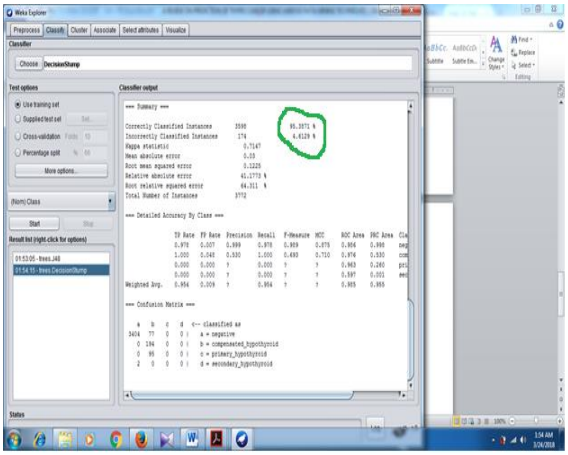
J48 Decision tree is used for classification using WEKA. The accuracy is 99.8%



Tree View

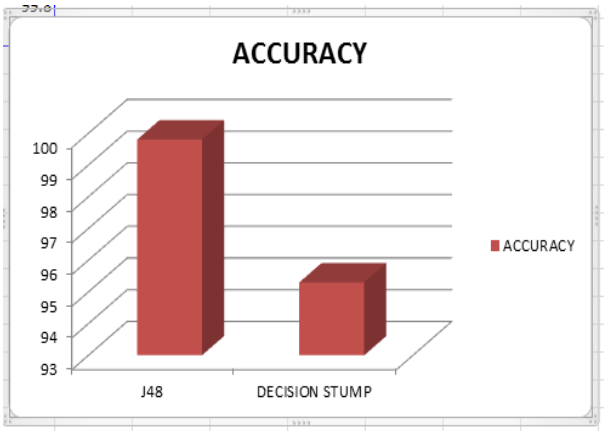


DecisionStump tree is used for classification using WEKA

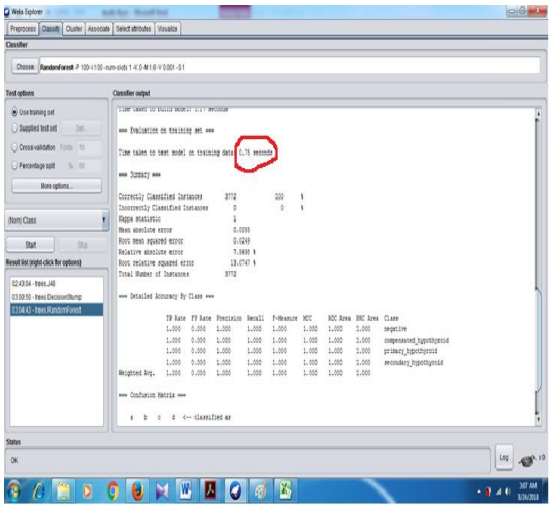


Comparision Table and CHART

CLASSIFIER	ACCURACY
J48	99.8%
DECISION STUMP	99.5

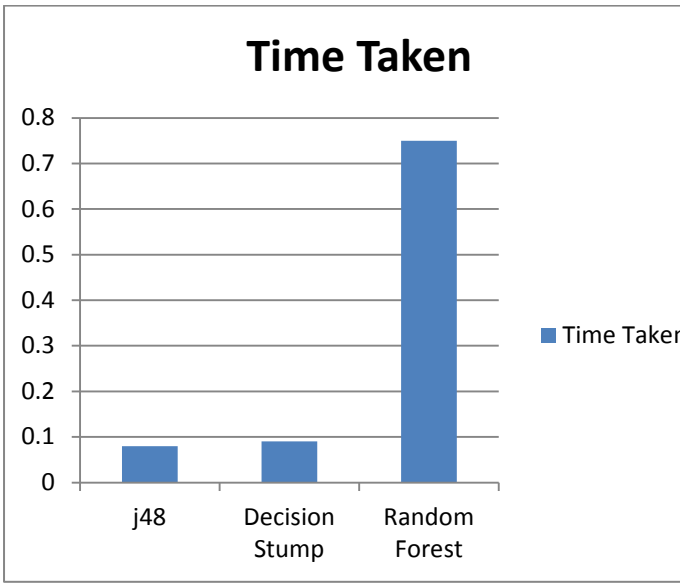


RANDOM FOREST tree is used for thyroid classification using WEKA. The accuracy achieved is 100%. But the time taken to provide the results are 0.75 seconds.



TIME DURATION TABLE

CLASSIFIER	TIME TAKEN
J48	0.08 Seconds
DECISION STUMP	0.09 Seconds
RANDOM FOREST	0.75 Seconds



III. CONCLUSION

Diagnosis of disease is a very challenging task in the field of health care. Many data mining techniques are used in decision making process. In this work, we have used J48, Random Forest Tree and decision stump data mining classification techniques which are used to classify the hypothyroid disease. The performance of classifiers are evaluated through the confusion matrix in terms of accuracy and Time Taken to produce the result. The J48 Algorithm gives 99.8% which is providing better Accuracy than decision stump tree accuracy and also J48 Algorithm took very minimum time than Decision stump and Random Forest. Here we conclude that J48 Decision Tree is more efficient in finding results.

IV. REFERENCES

[1]. P. K. Sharpe, H. E. Solberg, K. Rootwelt, et al; Artificial neural networks in diagnosis of thyroid function from vitro laboratory tests, Clin. Chem, 39 (1993) 2248-2253.

[2]. G. Serpen, H. Jiang, L. Allred, Performance analysis of probabilistic potential function neural network classifier, In Proceedings of artificial neural networks in engineering conference, St. Louis, MO, 7 (1997) 471-76.

[3]. M. Brameier, W. Banzhaf, A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining, IEEE Transactions on Evolutionary Computation, 5 (1) (2001) 17. http://dx.doi.org/10.1109/4235.910462

[4]. L. Ozyılmaz, T. Yıldırım, Diagnosis of thyroid disease using artificial neural network methods, In: Proceedings of ICONIP'02 9th international conference on neural information processing, (2002) 2033-36.http://dx.doi.org/10.1109/ICONIP.2002.1199031

[5]. L. Pasi, Similarity classifier applied to medical data sets, In International conference on soft computing, Helsinki, Finland & Gulf of Finland & Tallinn, Estonia, (2004).

[6]. "UCI Machine Learning Repository of machine learning database", University of California, school of Information and Computer Science, Irvine. C.A. Available from: http://www.ics.uci.edu/.

[7]. Dr.G.RasithaBanu, Baviya, "A study on Thyroid disease using Data Mining Technique". IJTRA Journal, Volume -3,Issue- 4,page no- (376-379),August 2015.

[8]. K.Thenmozhi, P.Deepika," Heart Disease Prediction Using Classification with Different Decision Tree Technique" International Journal of Engineering Research and General Science, Volume 2, Issue 6, October-November, 2014.

[9]. K.Thenmozhi, P.Deepika "DIFFERENT DATAMINING TECHNIQUES INVOLVES INHEART DISEASE PREDICTION-A SURVEY" in International Journal of ScientificResearch ISSN NO. 2277-8179 September 2014.