



EFFECTIVE PREDICTION OF HYPOTHYROID USING VARIOUS DATA MINING TECHNIQUES

Mrs.K.Sindhya

Assistant Professor, Department of Information Technology, Nirmala College for Women, Coimbatore.

ABSTRACT

Thyroid is a common disease around the world. So there is an emerge to identify the disease for treatment. In older days, prediction process were very slow and the accuracy level was too poor. Data mining is one of the effective area where researchers doing implementations rapidly. Data mining in healthcare is providing a great result in various disease prediction. The accuracy of prediction using data mining techniques are high and the cost of prediction is low. The other main advantage is time taken for prediction is very less. In this paper, I implemented classification algorithms on thyroid data set and produced result.

KEYWORDS: Data mining, Classification, Thyroid, J48, Naïve Bayes, Random Forest.

INTRODUCTION

Data mining takes a major role in Research which provides effective results in various fields and especially in health care also. It is more important to diagnosis diseases in early stage. Data mining techniques are efficient to predict the disease with high accuracy. These results are more useful to the patients who can start immediate treatment. Classification is used to assign the collected data items to the target class. These classification techniques are widely used because of its effectiveness in the prediction of diseases in healthcare domain.

Thyroid disease are one of the deadly disease which affect the human life worldwide. Thyroid disease

is a medical condition that affects the function of thyroid gland which is located in the front neck of the human body. The main effect of this thyroid is to produce hormones which spreads through the bloodstream to all the organs of human body to affect the growth of the organs[1]. The shape of thyroid gland is like butterfly[2].

CLASSIFICATION ON THYROID DISEASE

The thyroid Diseases are majorly classified into two types.

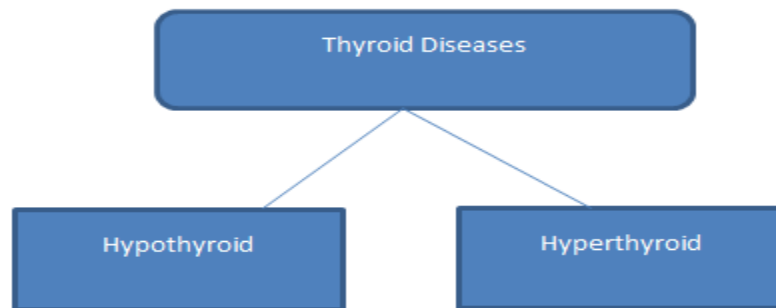


Fig.No:1 Classification of Thyroid Disease



1.Hypothyroid

Hypothyroid is a kind of thyroid which means the insufficient production of hormones of thyroid gland. This insufficient growth can cause lot of problems in human body like too much of weight gain, dry skin and dry hair, heavy menstrual periods etc. These symptoms are different from one person to another.

2.Hyperthyroid

The term hyper refers that too much or heavy. So hyperthyroid indicates that the massive growth of hormones of thyroid glands. This hyperthyroid is often called as overactive thyroid. This may cause too much of weight loss even though the sufficient intake of food, increased heart rate, shorter menstrual periods etc.[3]. Table 1 shows the symptoms of Hypothyroid and Hyperthyroid.

Hypothyroid	Hyperthyroid
Memory loss	Irritability
Dry hair	Hair loss
Muscle aches	Muscle weakness
Slow Heart beat	Rapid Heart beat
Heavy menstrual periods	Short Menstrual periods
Weight gain	Weight loss
Constipation	Sleeping difficulties

Tab.No:1 Symptoms of Hypothyroid and Hyperthyroid

RESEARCH METHODOLOGY

(i) Data set Description

The dataset was taken from UCI machine learning repository. The name of the dataset is hypothyroid. This dataset contains 3772 instances. In this dataset

3481 instances are in negative category and 194 instances are in compensated hypothyroid category, 95 instances are in primary hypothyroid category and 2 instances are in secondary hypothyroid category.

DATA SN	DESCRIPTION	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

Tab.No:2 Dataset Description**(ii) Methodology used**

Classification is the most popular technique of data mining which produces the best results of prediction. For this work, J48, naïve Bayes and Random Forest algorithms were taken for implementation.

• J48 Classification Model:

This model is a tree based model which used Divide-and-Conquer method for prediction. By using this model, we can get a tree structure which contains one root node and many leaf nodes.

• Naïve Bayes Model:

This Naïve Bayes is a collective model of various classification algorithms based on Bayes theorem. This model is one of the powerful model in classification which is used for high accuracy in prediction.

• Random Forest Model:

Random forest is one of the effective classification model which contains collection of decision trees. This Random Forest was trained using Bagging method which produces number of decision trees to produce a forest like model.

RESULTS & DISCUSSION

The implementation was done with the help of WEKA tool. The dataset was taken from UCI repository.

1. Preprocessing

Preprocessing is the initial step of every data mining process. By using preprocessing we can identify the missing values, irrelevant data in the dataset and redundancy of the dataset. By removing the noisy data we can get preprocessed dataset for implantation. After that the Dataset was loaded into WEKA tool.

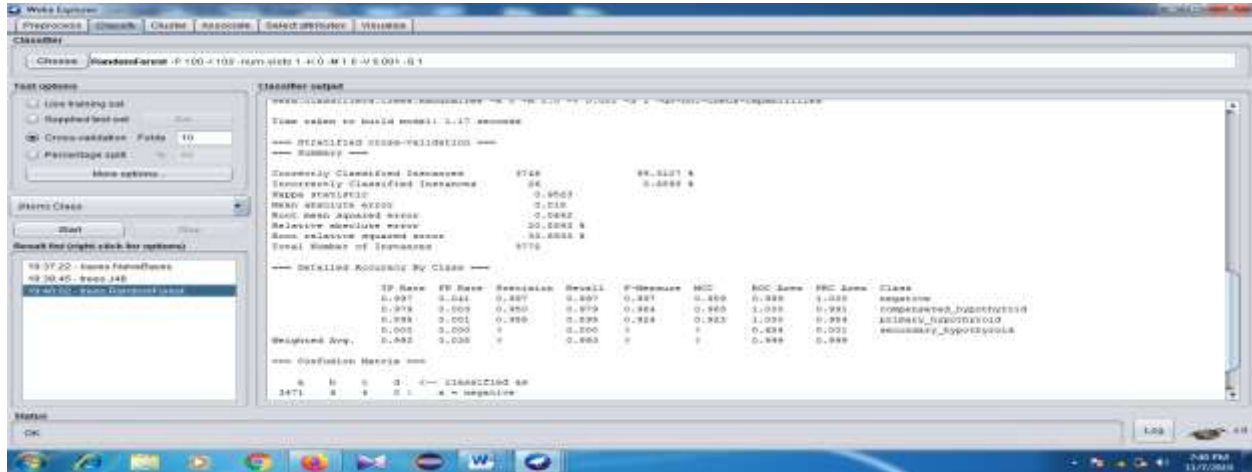


Fig.No: 6 Implementation using Random Forest Model

Comparison of Results:

The efficiency of used model is evaluated by the accuracy the reached and the time taken to reach.

1.Accuracy Comparison

Based on the Prediction accuracy, the Random Forest Model produced high accuracy than the other two models used.

Model Used	Accuracy Reached
Naïve Bayes	95%
J48	99%
Random Forest	99.3%

**Tab.No:3 Accuracy Comparison
Accuracy**

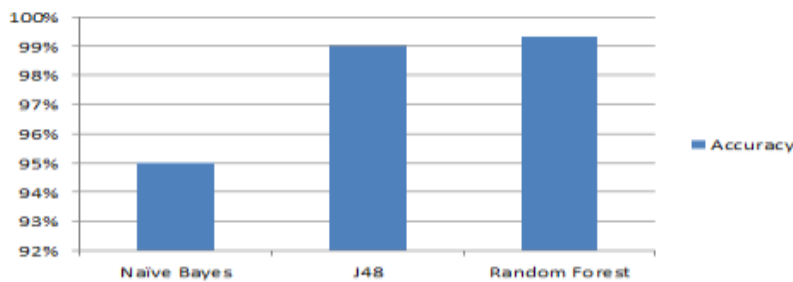


Fig.No:7 Accuracy Comparison Report

2.Time Comparison:

Based on the time consumption to construct a model, Naïve Bayes model has taken very short period of time for prediction than the other two models.

Model used	Time Taken
Naïve Bayes	0.04 sec
J48	0.2 sec
Random Forest	1.17 sec

Tab.No:4 Time Comparison



CONCLUSION

The effectiveness of a model is basically evaluated by two aspects. One is prediction accuracy and the second is time taken for the prediction. Based on our results, Naïve Bayes had taken only 0.04 sec for prediction. But the accuracy level is low than J48 and Random Forest. If we had considered the prediction accuracy, Random Forest model produced 99.3% accuracy. But the time taken to build the model is high than the other two models. So we can conclude that J48 is the best model for hypothyroid prediction because the accuracy it produced is 99% which is also one of the highest accuracy and the time taken for this model is 0.2 sec which is comparatively low than the Random Forest model.

REFERENCES

1. Dr.B.Srinivasan, K.Pavya, "Diagnosis of Thyroid Disease using Data Mining Techniques: A Study", *International Research Journal of Engineering and Technology*, Vol-3, Issue-11, 2016.
2. S.Sangeetha, K.Palanivel, "Anticipating Thyroid Disorders Using Data Mining Techniques" *International Journal of Computer Science and Mobile Computing*, Vol-7, Issue-4, 2018.
3. <http://www.foxnews.com/health/2012/02/10/hypothyroidism-versus-hyperthyroidism.html>(accessed dec 2015).
4. A.Kiruthika, P.Deepika, S.Saranya, Dr.S.Sasikala, "Predicting Ailment of Thyroid Using Classification and Recital Indicators", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Vol-3, Issue-3, 2018.
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.Rasitha Banu, 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 Data mining Techniques Involves In Heart Disease Prediction-A Survey" in *International Journal of Scientific Research* ISSN NO. 2277-8179 September 2014.
10. P.Deepika, P.Vinothini, "Heart Disease Analysis and Prediction Using Various Classification Models- A Survey" *PARIPEX-Indian Journal of Research*, Volume-4, Issue-3, March 2015.
11. P.Deepika, A.Kiruthika, S.Saranya, Dr.S.Sasikala, "A Novel Classification and Prediction Algorithm for Heart Disease Identification" *International Journal of Engineering and Science Invention*, Volume-7, Issue-2, Feb 2018.
12. P.Deepika, S.Saranya, S.Sasikala, A.Kiruthika, Dr.S.Jansi, "Anticipating Heart Disease using C4.5 Classification Augmented with Feature Selection" *International Journal for Research & Development in Technology*, Volume-6, 2016.
13. P. Deepika, S.Saranya, Dr.S.Sasikala, Haritha R.A, Aravind G.Nair, "The Study of Fuzzy Cognitive Maps in Identification and Prediction of Disease" *IOSR Journal of Engineering*, Volume-6, Nov 2018.
14. S.Saranya, P.Deepika, Dr.S.Sasikala, "Comprehensive Review on Heart Disease Prediction using Optimization Techniques", *EPRA International Journal of Research and Development (IJRD)*, Volume-5, Issue-1, January 2020.
15. P.Deepika, S.Saranya, Dr.S.Sasikala, "An Effective Archetype Design of Heart Disease Anticipation using Optimization Techniques", *EPRA International Journal of Research and Development (IJRD)*, Volume-4, Issue-11, November 2019.
16. S.Sasikala, S.Jansi, S.Saranya, P.Deepika, A.Kiruthika, "Anticipating the Chronic Kidney Disorder(CKD) using Performance Optimization in AdaBoost and Multilayer Perceptron" *Imperial Journal of Interdisciplinary Research*, Volume-3, 2017.