Multi Disease Prediction using Data Mining Techniques

2 authors:

K.Gomathi Kamaraj
Sri Krishna College of Arts and Science

9 PUBLICATIONS **71** CITATIONS

SEE PROFILE

D.Shanmuga Priyaa
Karpagam Academy of Higher Education

30 PUBLICATIONS **115** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Coronary heart disease prediction View project

Research paper View project

Multi Disease Prediction using Data Mining Techniques

K.Gomathi*, Dr. D. Shanmuga Priyaa**

Abstract

Data mining techniques are used for a variety of applications. In healthcare industry, data mining plays an important role in predicting diseases. For detecting a disease number of tests should be required from the patient. But using data mining technique the number of tests can be reduced. This reduced test plays an important role in time and performance. This paper analyzes data mining techniques which can be used for predicting different types of diseases. This paper reviewed the research papers which mainly concentrate on predicting heart disease, Diabetes and Breast cancer etc.

Keywords: Data Mining, Classification, Naive Bayes, J48, Decision tree.

Introduction

Data mining is the process of selecting, discovering and modeling huge amounts of data. This process has become an increasingly insidious activity in all areas of medical science research. Data mining has resulted in the discovery of useful hidden patterns from huge databases. Data mining problems are often solved using different approaches from both computer sciences, such as multi-dimensional databases, machine learning, soft computing and data visualization; and includes classification and regression techniques. Some of the research works are done in this side, but all of them are focusing on a few methods of analysis, diagnosis or prediction of this disease by using different tools and techniques, this work is focused on the early prediction of various diseases by using WEKA tool.

Heart Disease

Heart disease is the leading cause of death in the U.S. At any point in your life, either you or one of your loved ones will be forced to make decisions about some aspect of heart disease. Knowing about the structure and functioning of the heart, in particular how angina and heart attacks work, will enable to make informed decisions about your health. Heart disease can strike suddenly and need to make decisions quickly.

Several factors contribute to this damage. They include:

1. Smoking, including secondhand smoke
2. High amounts of certain fats and cholesterol in the blood
3. High blood pressure
4. High amounts of sugar in the blood due to insulin resistance or diabetes
5. Blood vessel inflammation

Diabetics

Diabetes mellitus, or simply diabetes, is a chronic disease that occurs when the pancreas is no longer able to make insulin, or when the body cannot make good use of the insulin it produces. Insulin is a hormone made by the pancreas, which acts like a key to let glucose from the food we eat pass from the blood stream into the cells in the body to produce energy. All carbohydrate foods are broken down into glucose in the blood. Insulin helps glucose get into the cells.

Not being able to produce insulin or use it effectively leads to raised glucose levels in the blood (known as

* Research Scholar, Department of Computer Science, Karpagam University, Coimbatore, Tamil Nadu, India.
E-mail: gomshri@gmail.com
** Professor, Department of Information Technology, Karpagam University, Coimbatore, Tamil Nadu, IndiaIndia

hyperglycemia). Over the long-term high glucose levels are associated with damage to the body and failure of various organs and tissues.

Fast facts on diabetes

1. Diabetes is a long-term condition that causes high blood sugar levels.

2. In 2013 it was estimated that over 382 million people throughout the world had diabetes (Williams's textbook of endocrinology).

3. Type 1 Diabetes - the body does not produce insulin. Approximately 10% of all diabetes cases are type 1.

4. Type 2 Diabetes - the body does not produce enough insulin for proper function. Approximately 90% of all cases of diabetes worldwide are of this type.

5. Gestational Diabetes - this type affects females during pregnancy.

6. The most common diabetes symptoms include frequent urination, intense thirst and hunger, weight gain, unusual weight loss, fatigue, cuts and bruises that do not heal, male sexual dysfunction, numbness and tingling in hands and feet.

7. If you have Type 1 and follow a healthy eating plan, do adequate exercise, and take insulin, you can lead a normal life.

8. Type 2 patients need to eat healthily, be physically active, and test their blood glucose. They may also need to take oral medication, and/or insulin to control blood glucose levels.

9. As the risk of cardiovascular disease is much higher for a diabetic, it is crucial that blood pressure and cholesterol levels are monitored regularly.

10. As smoking might have a serious effect on cardiovascular health, diabetics should stop smoking.

11. Hypoglycemia - low blood glucose - can have a bad effect on the patient. Hyperglycemia - when blood glucose is too high - can also have a bad effect on the patient.

Breast Cancer

Breast cancer is a malignant tumor that starts in the cells of the breast. A malignant tumor is a group of cancer cells that can grow up into (invade) surrounding tissues or spread (metastasize) to distant areas of the body. The disease occurs almost entirely in women, but men can get it, too.

Breast cancer facts

1. Breast cancer is the most common cancer among American women.

2. One in every eight women in the United States develops breast cancer.

3. There are many types of breast cancer that differ in their capability of spreading (metastasize) to other body tissues.

4. The causes of breast cancer are not yet fully known, although a number of risk factors have been identified.

5. There are many different types of breast cancer.

6. Breast cancer symptoms and signs include

   - a lump in the breast or armpit,
   - bloody nipple discharge,
   - inverted nipple,
   - orange-peel texture or dimpling of the breast's skin,
   - breast pain or sore nipple,
   - swollen lymph nodes in the neck or armpit,
   - Change in the size or shape of the breast or nipple.
   - Breast cancer is diagnosed during a physical exam, by self-examination of the breasts, mammography, ultrasound testing, and biopsy.
   - Treatment of breast cancer depends on the type of cancer and its stage (0-IV) and may involve surgery, radiation, or chemotherapy.

Data Mining Techniques

Naive Bayes

The Naive Bayesian classifier is based on Baye's theorem with independence assumptions between predictors. A Naive Bayesian model is simple to build, with no difficult iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Bayes theorem provides a way of manipulative the posterior probability,

1) P(c|x)is the posterior probability of class(target) given predictor(attribute).

2) P(c)is the prior probability of class.

3) P (x|c) is the likelihood which is the probability of predictor given class.

4) P(x)is the prior probability of predictor.

Decision Tree (J48)

Decision tree models are commonly used in data mining to examine data and induce the tree and its rules that will be used to make predictions[10].The prediction could be to predict categorical values (classification trees) when instances are to be placed in categories or classes. Decision tree is a classifier in the structure of a tree structure where each node is either a leaf node, indicating the value of the target attribute or class of the examples, or a decision node, specifying some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test. A decision tree can be used to classify an case by starting at the root of the tree and moving through it until a leaf node is reached, which provides the classification of the instance.

Experimental Results

In this study, the accuracy of three data mining techniques is compared. The goal is to have high accuracy, besides high precision and recall metrics. Although these metrics are used more often in the field of information retrieval, here we have considered them as they are related to the other existing metrics such as specificity and sensitivity. These metrics can be derived from the confusion matrix and can be easily converted to true-positive (TP) and false-positive (FP) metrics.

Technique	Diseases	Accuracy
Naïve Bayes	Heart Disease	79%
	Diabetes	77.6%
	Breast Cancer	82.5%
J48	Heart Disease	77.033%
	Diabetes	100%
	Breast Cancer	75.52%

Conclusion

In this study two different data mining classification techniques was used for the prediction of various diseases and their performance was compared in order to evaluate the best classifier. An important challenge in data mining and machine learning areas is to build precise and computationally efficient classifiers for Medical applications.

References

http://www.mayoclinic.org/diseases-conditions/heart-disease/basics/definition/con-20034056.

http://www.idf.org/about-diabetes.

http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-what-is-breast-cancer.

Kumara, M., Vohra, R., Arora, A. (2014). Prediction of diabetes using Bayesian network. *International Journal of Computer Science and Information Technologies*, 5(4), 5174-5178.

Thirumal, P. C., & Nagarajan, N. (2015). Utilization of data mining techniques for diagnosis of diabetes mellitus - A case study. *ARPN Journal of Engineering and Applied Sciences*, January, 10(1), 8-13.

Gomathi, K. (2012). An empirical study on breast cancer using data mining techniques. *International Journal of Research in Computer Application & Management*, July, 2(7), 97-102.

Witten, H. I., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, (2nd Ed), Morgan Kaufmann Publishers.

Witten, I. H. & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques, (2ndEd.) San Fransisco: Morgan Kaufmann.

WEKA: Data Mining Software in Java. Retrieved from http://www.cs.waikato.ac.nz/ml/weka/.

Delen, D., Walker, G., & Kadam, A. (2005). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, June, 34(2), 113-127.