Selection Of The Best Classifier From Different Datasets Using WEKA

Ranjita kumari Dash
Assistant Professor, Institute Of Technical Education and Research,
SOA University

Abstract

In today's world large amount of data is available in science, industry, business and many other areas. These data can provide valuable information which can be used by management for making important decisions. By using data mining we can find valuable information. Data mining is the popular topic among researchers. There is a lot of work that cannot be explored till now. But, this paper focuses on the fundamental concept of the Data mining that is Classification Techniques. In this paper, Naive Bays , Functions, Lazy, Meta, Nested dichotomies, Rules and Trees classifiers are used for the classification of data set. The performance of these classifiers analyzed with the help of correctly classified instances, incorrectly classified instances and time taken to build the model and the result can be shown statistical as well as graphically. WEKA data mining tool is used for this purpose. WEKA stands for Waikato Environment for Knowledge Analysis. Three datasets are used on which different classifiers are applied to check which classifier is giving the best result, where different measurements are taken.71 different classifiers are applied on this dataset. The dataset is in ARFF format.10 fold cross validation is used to provide better accuracy. Finally the classification technique which provides the best result will be suggested. The result shows that no single algorithm always performed the best for each dataset.

KEY TERM'S

Bays Net, J48, Mean Absolute Error, Naive Bays, Root Mean-Squared Error

1. Introduction

Data mining is the process of extracting patterns from data [10, 11]. It is seen as an increasingly important tool by modern business to transform data as the technology advances and the need for efficient data analysis is required. Data mining involves the use of data analysis tools to discover previously unknown, valid patterns and relationships in large data set. It is currently used in a wide range of areas like marketing, surveillance, fraud detection, and scientific discovery etc.

In this paper we process a cancer dataset and use different classification methods to learn from the test data set.

Classification is a basic task in the data analysis that requires the construction of a classifier, that is, a function that assigns a class label to instances described by a set of attributes. It is one of the important applications of data mining. This technique predicts categorical class labels. In this paper, we are giving the comparison of various classification techniques using WEKA. Our aim is to investigate the performance of different classification methods using WEKA. Classification of data is very typical task in data mining. There are large number of classifiers that are used to classify the data such as Bayes, function, lazy learners, Meta, rule based and Decision tree etc. The goal of classification is to correctly predict the value.

For Breast cancer, there is a substantial amount of research with machine learning algorithm [1]. Machine learning covers such a broad range of processes that it is difficult to define precisely [6]. Young women being diagnosed in their teens, twenties and thirties. Even if the percentage is very low compared to that of older women aged 40 years and older [7, 8, 9]. 1% of all diagnosed breast cancers are in men. We report the case of a 34-year-old woman affected by

breast cancer that had metastasized to the bone. Today, about one in eight women over their lifetime have been affected by breast cancer in the United States. In recent years, the incidence rate keeps increasing. However the appropriate methods to predict the breast cancer survival have not been established. In this study, we use those models to evaluate the prediction rate of breast cancer patients from the perspectives of accuracy.

2. WEKA

WEKA stands for Waikato Environment for Knowledge Analysis. WEKA is created by researchers at the University of Waikato in New Zealand. WEKA was first implemented in its modern form in 1997. The GNU General Public License (GPL) is used here. The figure of WEKA is shown in the figure .The software is written in the Java™ language and contains a GUI for interacting with data files .For working of WEKA, we do not need the deep knowledge of data mining for which WEKA a very popular data is mining tool. WEKA also provides the graphical user interface of the user and provides many facilities. In this paper, we are giving the comparison of various classification techniques using WEKA. WEKA is a state of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. The data file normally used by WEKA is in ARFF file format. ARFF stands for Attribute Relation File Format, which consists of special tags to indicate differentiating in the data file. WEKA implements algorithms for data pre-processing, classification, regression and clustering and association rules. It also includes visualization tools. It has a set of panels, each of which can be used to perform a certain task. The new machine learning schemes can also be developed with this package. WEKA is open source software issued under General Public License. The algorithms are applied directly to a dataset. The main features of WEKA includes

• 49 data pre-processing tools
• 76 classification/regression algorithms
• 8 clustering algorithms
• 15 attribute/subset evaluators + 10 search Algorithms for feature selection.
• 3 algorithms for finding association rules
• 3 graphical user interfaces

3. METHODS

This section describes the classification methods used in this paper. We discuss each method and explain how the method has been used in our experiment. For this Breast cancer dataset we have taken eight methods Bayes, Functions, Lazy, Meta, Misc, Nested dichotomies, Rules and Trees classifiers for the classification of data set.

3.1. NAIVE BAYES CLASSIFIER

Bayes methods are also used as one of the classification solutions in data mining. In our work we use six main Bayesian methods namely AODE, AODEsr, Naive Bayes, Bayesian net, Naive Bayes simple and Naive Bayes updateable, that are implemented in WEKA software for classification .Naive Bayes is an extension of Bayes theorem in that it assumes independence of attributes[3]. This assumption is not strictly correct when considering classification based on text extraction from a document as there are relationships between the words that accumulate into concepts. Problems of this kind, called problems of supervised classification, are ubiquotous.Naive Bayes sometimes also called as idiot's Bayes, simple Bayes and independence Bayes.This is important for several reasons.

It is easy to construct without any need for complicated iterative parameter estimation schemes. This means it may be readily applied to huge datasets. It is robust, easy to interpret, and often does surprisingly well though it may not be the best classifier in any particular application.

3.2. FUNCTION CLASSIFIER

Function classifier uses the concept of neural network and regression. Here two examples from neural network and regression will be taken for discussing the scenario[2]. A multilayer perceptron is a free forward artificial neural network model that maps sets of input data onto a set of appropriate output. It is a modification of the standard linear perceptron in that it uses three or more layers of neurons with nonlinear activation functions and it is more powerful than the perceptron in that it can distinguish data that is not linearly separable or separable by a hyperplane[4].A multilayer perceptron has distinctive characteristics. The model of each neuron in the network includes a non linear activation function. The network contains one or more layers of hidden neurons that are not part of the input or output of the network. These hidden

neurons enable the network to learn complex tasks by extracting progressively more meaningful features from the input patterns. The network exhibits a high degree of connectivity determined by the network. A change in the connectivity of the network requires a change in the population of synaptic connections on their weights[5].

3.3. RULES CLASSIFIER

Association rules are used to find interesting correction relationship among all the attributes. They may predict more than one conclusion. The number of records an association rule can predict correctly is called coverage. Support is defined as coverage divided by total number of records[5]. Accuracy is the number of records that is predicted correctly expressed as a percentage of all instances that are applied to the methods of this algorithm are Conjunctive Rule, Decision table,DTNB,JRip,NNge,Oner,Rider and Zero. Rules are easier to understand than large trees. One root is created for each path from the root to the leaf. Each attribute value pair along a path forms a conjunction. The leaf holds the class prediction. Rules are mutually exclusive. These are learned one at a time .Each time a rule is learned ,the tuples are covered by the rules are removed.

3.4. LAZY CLASSIFIER

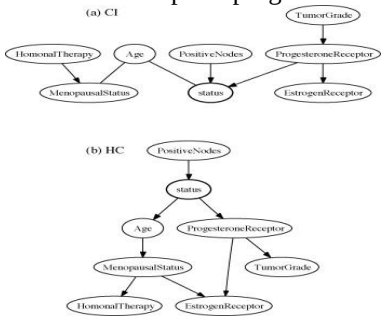
When making a classification or prediction, lazy learners can be computationally expensive. They require efficient storage techniques and well suited to implementation on parallel hardware. They offer little explanation or insight into the structure of the data. Lazy learners however, naturally support incremental learning. They are able to model complex decision spaces having hyper polygonal shapes that may not be as easily describable by other learning algorithms. The methods of this algorithm are IBI, IBK,K- Star, LBK and LWL.

3.5. META CLASSIFIER

Meta classifier includes a wide range of classifier. When the attributes have a large number of values because the time and space complexities depend not only on the number of attributes, but also on the number of values for each attribute.

3.6. DECISSION TREES

Decision tree induction has been studied in details in both areas of pattern recognition and machine learning [13, 14]. This synthesizes the experience gained by people working in the area of machine learning and describes a computer program called ID3.



4. DISCUSSION AND RESULT

By investigating the performance on the selected classification methods or algorithms namely Bayes ,Function, Lazy ,Meta ,Rules ,Misc ,nested dichotomies and Trees we use the same experiment procedure as suggested by WEKA. The 75% data is used for training and the remaining is for testing purposes.

In WEKA, all data are considered as instances and features in the data are known as attributes. The simulation results are partitioned into several sub items for easier analysis and evaluation. On the first part, correctly and incorrectly classified instances will be partitioned in numeric and percentage value and subsequently time taken to build model will be in second .The results of the simulation are shown in Tables. These are the graphical representation of the simulation result. On the basis of comparison done over accuracy and error rates the classification techniques with highest accuracy are obtained for this dataset in given different machine learning tools. We can clearly see that the highest accuracy is 75.52% and the lowest is 51.74%.In fact, the highest accuracy belongs to the Meta classifier. The total time required to build the model is also a crucial parameter in comparing the classification algorithm. In this experiment, we can say that a single conjunctive rule learner requires the shortest time which is around 0.15 seconds compared to the others.

With the help of figures we are showing the working of various algorithms used in WEKA. We are showing also advantages and disadvantages of each algorithm. Every algorithm has their own importance and we use them on the behaviour of the data. Deep knowledge of algorithms is not required for working in WEKA. This is the main reason WEKA is more

suitable tool for data mining applications. This paper shows only the clustering operations in the WEKA, we will try to make a complete reference paper of WEKA.

Table for best algorithms:-

Name of algorithm	Correctly classified instance	Incorrectly classified instances	Time taken to build the model
Bayesnet	72.028	27.972	0.03
Simple logistic	75.1748	24.8252	1.44
K-Star	73.4266	26.5734	0
Filtered classifier	75.5245	24.4755	0
Ordinal classifier	75.5245	24.4755	0.01
Misc	69.9301	30.0699	0
Decision Table	73.4266	26.5734	0.5
J48	**75.5245**	**24.4755**	**0.01**

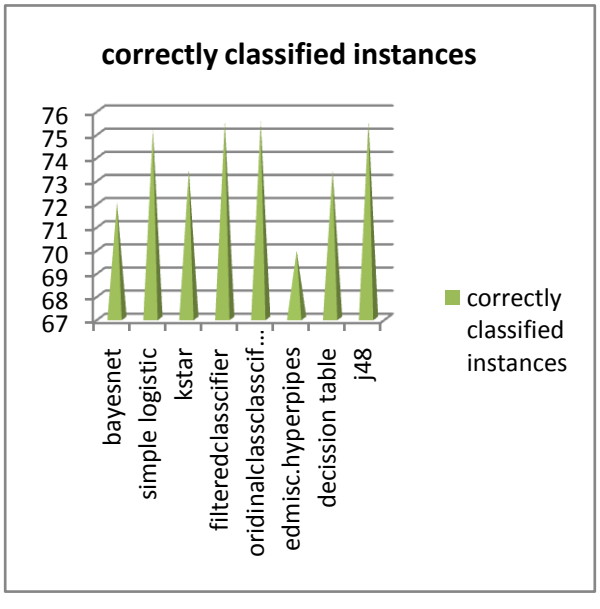


Figure no-1

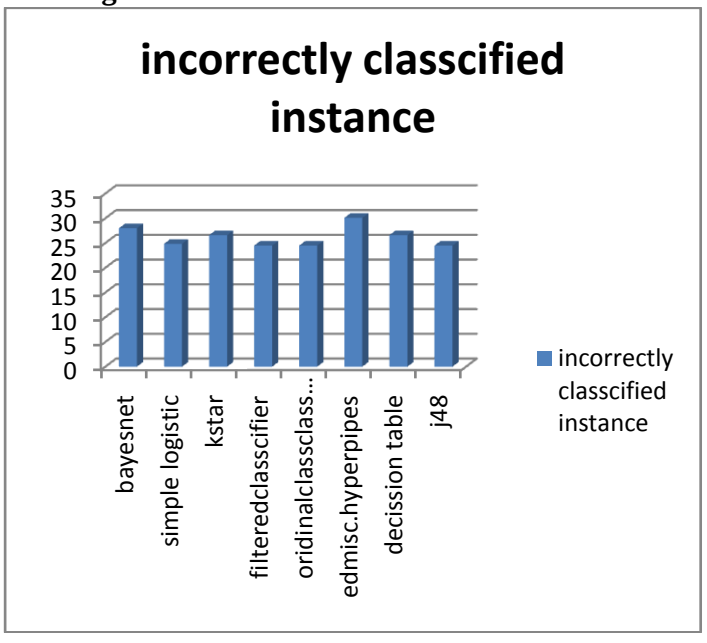


Figure no-2

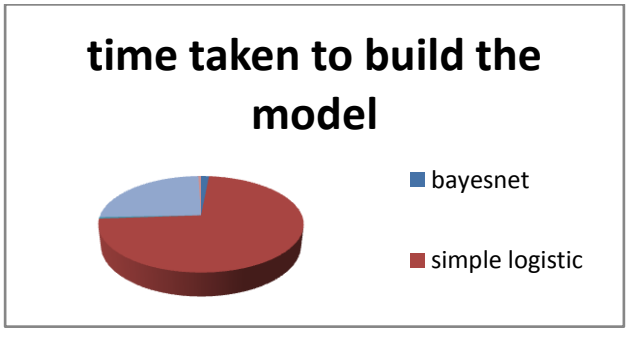


Figure no-3

4.2 Comparison between LUNG dataset, HEART dataset, DIABETES DATASET

Algorithm	Correctly Classified Instances in %	Incorrectly Classified Instances in %	TP Rate	FP Rate	Time taken to build model in seconds (s)
Multilayer Perceptron	100	0	0.75	0.436	0.2
Multiclass Classifier	77.2135	22.7865	0.772	0.321	0.02
SPegasos	77.7344	22.2656	0.777	0.327	0.19

Table no -2(lung dataset,heardataset,diabetes dataset)

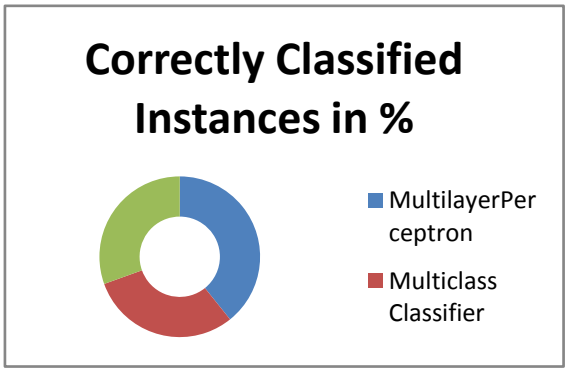


Figure no-4

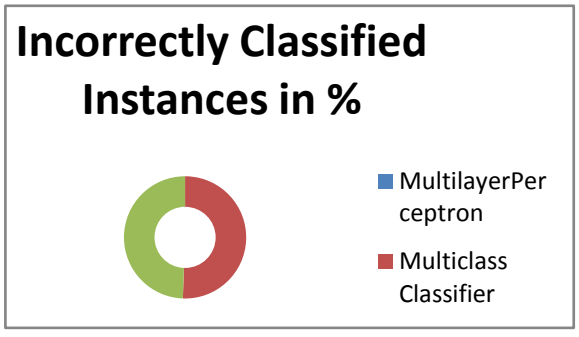


Figure no-5

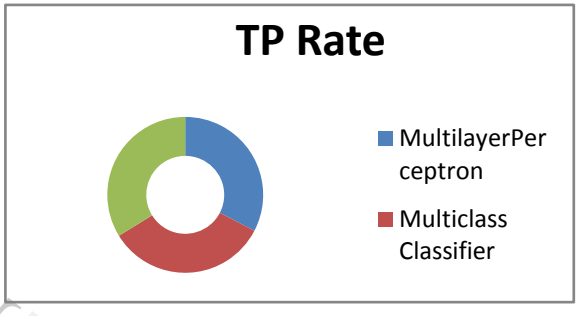


Figure no-6

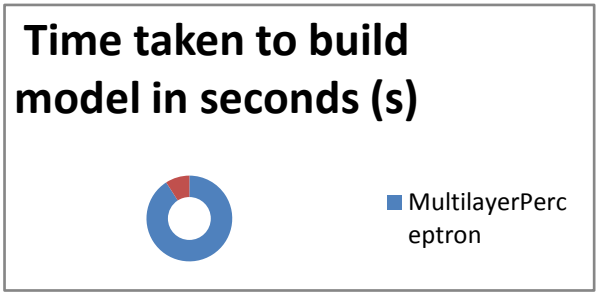


Figure no-7

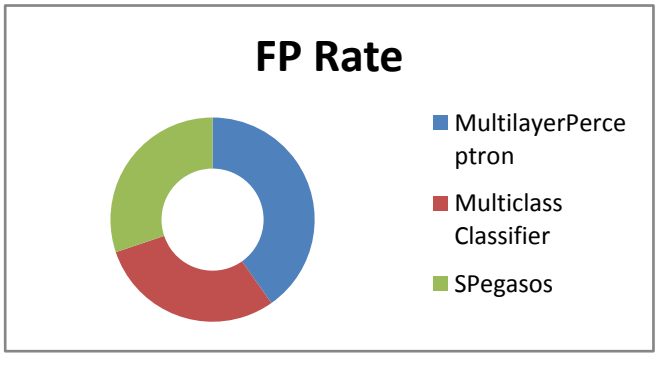


Figure no-8

5. References

[1] D.Lavanya, Dr.K.Usha Rani,..," Analysis of feature selection with classification: Breast cancer datasets",Indian Journal of Computer Science and Engineering (IJCSE),October 2011.

[2] E.Osuna, R.Freund, and F. Girosi, "Training support vector machines: Application to face detection". Proceedings of computer vision and pattern recognition, Puerto Rico pp. 130–136.1997.

[3] Buntine, Theory refinement on Bayesian networks. In B. D. D'Ambrosio, P. Smets, & P.P. Bonissone (Eds.), In Press of Proceedings
of the Seventh Annual Conference on Uncertainty Artificial Intelligent (pp. 52-60). San Francisco, CA

[4] S. V. Chakravarthy and J. Ghosh (1994), Scale Based Clustering
using Radial Basis Function Networks, In Press of Proceeding of
IEEE International Conference on Neural Networks, Orlando, Florida.
pp. 897-902. 5. M. D. Buhmann (2003), Radial Basis Functions: Theory and Implementations,

[5] Howell, A.J. and Buxton, H. (2002). RBF Network Methods for Face
Detection and Attentional Frames, Neural Processing Letters (15),
Pp.197-2114. Daniel Grossman and Pedro Domingo's (2004). Learning Bayesian
Network Classifiers by Maximizing Conditional Likelihood. In Press
of Proceedings of the 21st International Conference on Machine
Learning, Banff, Canada.

[6] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, Centres for Disease Control and Prevention, and National Cancer Institute; 2012.

[7] Lyon IAfRoC: World Cancer Report. International Agency for Research on Cancer Press 2003:188-193.

[8] Elattar, Inas. "Breast Cancer: Magnitude of the Problem", Egyptian Society of Surgical Oncology Conference, Taba, Sinai, in Egypt (30 March – 1 April 2005).

[9] Daniel F. Roses (2005). Clinical Assessment of Breast Cancer and Benign Breast Disease, In: Breast Cancer: Vol. 2, Ch. 14, M. N. Harris [editor], Churchill Livingstone, and Philadelphia.

[10] S. Aruna et al. (2011). Knowledge based analysis of various statistical tools in detecting breast cancer.

[11] J. Han and M. Kamber, (2000) "Data Mining: Concepts and Techniques," Morgan Kaufmann.

[12] William H. Wolberg, M.D., W. Nick Street, Ph.D., Dennis M. Heisey, Ph.D., Olvi L. Mangasarian, Ph.D. computerized breast cancer diagnosis and prognosis from fine needle aspirates, Western Surgical Association meeting in Palm Desert, California, November 14, 1994.

[13] Chen, Y., Abraham, A., Yang, B.(2006), Feature Selection and Classification using Flexible Neural Tree. Journal of Neurocomputing 70(1-3): 305–313.

[14] K. Golnabi, et al., "Analysis of firewall policy rules using data mining techniques," 2006, pp. 305-315.

[15] Duda, R.O., Hart, P.E.: "Pattern Classification and Scene Analysis", In: Wiley-Interscience Publication, New York (1973)

[16] Bishop, C.M.: "Neural Networks for Pattern Recognition". Oxford University Press, New York (1999).

[17] Vapnik, V.N., The Nature of Statistical Learning Theory, 1st ed., Springer-Verlag, New York, 1995.

[18] Ross Quinlan, (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.

[19] Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A. (1998). Discovering Data Mining: From Concept to Implementation, Upper Saddle River, N.J., Prentice Hall.

[20] E.Osuna, R.Freund, and F. Girosi, "Training support vector machines: Application to face detection". Proceedings of computer vision and pattern recognition, Puerto Rico pp. 130–136.1997.

[21] Vaibhav Narayan Chunekar, Hemant P. Ambulgekar (2009). Approach of Neural Network to Diagnose Breast Cancer on three different Data Set. 2009

[22] D. Lavanya, "Ensemble Decision Tree Classifier for Breast Cancer Data," International Journal of Information Technology Convergence and Services, vol. 2, no. 1, pp. 17-24, Feb. 2012.

[23] Yoav Freund, Robert E. Schapire, (1999) "Large Margin Classification Using the Perceptron Algorithm." In: Machine Learning, 37(3).

[24] J.D.M.Rennie, L.Shih, J.Teevan, and D.R.Karger, 2003. "Tackling the poor assumptions of naive bayes text classification." In ICML2003, pages 616–623.

[25] Kanako Komiya, Naoto Sato er. Al., "Negation Naïve Bayes for Categorization of Product Pages on the Web", Proceedings of recent advances in Natural Language Processing, pages 586-591, Hissar, Bulgaria, 12-14 September 2011

[26] Cheng J. Greiner, R, (2001). "Learning Bayesian Belief Networks Classifiers: Algorithms and Systems, In Stroulia, E. & Marwin, S.(ed.), AI 2001,, 141-151, LNAI 2056