International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.02, February 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Evaluating the Performance of Classification Algorithms Based on Metrics over Different Datasets

D.Ramya

Department of Computer Science & Engineering,
Sri Venkateswara College of Engineering & Technology, (India)

ABSTRACT

Classification is a data mining (machine learning) technique used to predict group membership for data instances. Evaluation of information, gathered almost everywhere in our day to day life can help devising some efficient and personalized strategies. Classification is one of the fundamental tasks in data mining and has also been studied extensively in statistics, machine learning, neural networks and expert systems over decades. Classification is a well known data mining technique that tells the class of an unknown object. For this purpose, classification predicts categorical (discrete, unordered) labels. Many classification algorithms have been proposed by researchers in statistics, machine learning and pattern recognition. In this study, the performance evaluation of BayesNet, LWL, BFTree, J48, LADTree and NBTree classification algorithms are experimented. I am Scheming the Performance of selective classification algorithms over different chosen data sets based on evaluation metrics Precision, F-Measure and ROC Area.

***Keywords :* Classification, BayesNet, LWL, BFTree, J48, LADTree, NBTree**

I. INTRODUCTION

Data mining is the process of exploration and analysis, by automatic and semi automatic means of large quantities of data in order to discover meaningful patterns and rules. The six main data mining activities are classification, estimation, prediction, affinity grouping, clustering, estimation and visualization [1]. From the past few years, the fields of machine learning and data mining have been studied to a great extent and applied in various fields of studies. It is now realized among the research communities that the contribution of machine learning has become immense for the development of science and technology. Classification, which is one of the supervised machine learning methodologies, is related to one of the fundamental tasks in data mining and has also been studied extensively in statistics, neural networks and expert systems over decades [2].

Classification involves two phases-construction of a model for classification/prediction and testing & usage of it for determining the class labels/ prediction. In this paper, performance evaluation of BayesNet, LWL, BFTree, J48, LADTree and NBTree classification algorithms are experimented based on the five different standard UCI data sets.

II. LEARNING ALGORITHMS FOR CLASSIFICATION

BayesNet: A Bayes Net is a model. It reflects the states of some part of a world that is being modeled and it describes how those states are related by probabilities. The model might be of your house, or your car, your

body, your community, an ecosystem, a stock-market, etc. Absolutely anything can be modeled by a Bayes Net. All the possible states of the model represent all the possible worlds that can exist, that is, all the possible ways that the parts or states can be configured. The car engine can be running normally or giving trouble. It's tires can be inflated or flat. Your body can be sick or healthy, and so on.

LWL: Locally Weighted Learning is a class of function approximation techniques, where a prediction is done by using an approximated local model around the current point of interest.

BFTree: An alternating decision tree (ADTree) is a machine learning method for classification. It generalizes decision trees and has connections to boosting. An ADTree consists of an alternation of decision nodes, which specify a predicate condition, and prediction nodes, which contain a single number. An instance is classified by an ADTree by following all paths for which all decision nodes are true, and summing any prediction nodes that are traversed and Class for building a best-first decision tree classifier is known as BFTree [3].

J48: In case the instances belong to the same class the tree represents a leaf so the leaf is returned by labeling with the same class. The potential information is calculated for every attribute, given by a test on the attribute. Then the gain in information is calculated that would result from a test on the attribute. Then the best attribute is found on the basis of the present selection criterion and that attribute selected for branching [4].

LADTree: LAD Tree builds a classifier for binary target variable based on learning a logical expression that can discriminate between positive and negative samples in a data set. LAD Tree Classifier generates a multi-class alternating decision tree using the Logit Boost strategy. The LAD Tree algorithm applies logistic boosting algorithm in order to induce an alternating decision tree. In this algorithm, a single attribute test is chosen as a splitter node for the tree at each iteration. For each training instance, working response and weights are calculated and stored on a per-class basis. Then, it fits the working response to the mean value of the instances, in a particular subset, by minimizing the least-squares value between them. In this algorithm, trees for the different classes are grown in parallel. Once all the trees have been constructed, then it merges the trees into a final model [5].

NBTree: The NBTree algorithm is a hybrid of the Naïve Bayes and the Decision Tree algorithm. This tree is constructed recursively. But, the leaf nodes are Naive Bayes categorizers. The NBTree algorithm strives to approximate whether the generalization accuracy of Naive Bayes at each leaf is higher than a single Naive Bayes classifier at that node. A split is considered to be significant if relative reduction in the error is greater than 5% and there are a minimum of 30 instances in the node. For discrete valued attributes, the Naive Bayes method performs quite well. With the increase in data size, the performance also improves. But in case of continuous valued attributes, Naive Bayes method does not take into account the attribute interactions. Whereas, the decision trees do not give good performance when the data size is very large. These shortcomings are overcome by the NBTree algorithm [6].

III. DATASETS

In the classification algorithms, data sets are transformed into training sets and test sets in order to build a model and use it for the classification purpose respectively. The training set involves the various attributes

having one as classifying attribute. On the other hand the test set includes the same attributes with the unseen tuples of data that the model is going to classify the instances [7].

IV. EXPERIMENT

To conduct experiment, I used six classifiers namely BayesNet, *LWL, BFTree, J48, LADTree and NBTree* on the five different UCI data sets (Breast-Cancer, Weather, Labor, Diabetes and Iris2D). The following table shows the description of considered datasets.

Table 1: Data Description

DATASETS	INSTANCES	ATTRIBUTES	TYPE
Breast-Cancer	286	10	Nominal
Weather	14	5	
Labor	57	17	Numeric
Diabetes	768	9	
Iris2d	150	43	

For the performance issue of classifiers, I focused on the evaluation parameters: precision, F-measure and ROC Area.

V. RESULT

The Fig. 1 shows the Accuracy comparison of six selective classification algorithms over different chosen data sets.

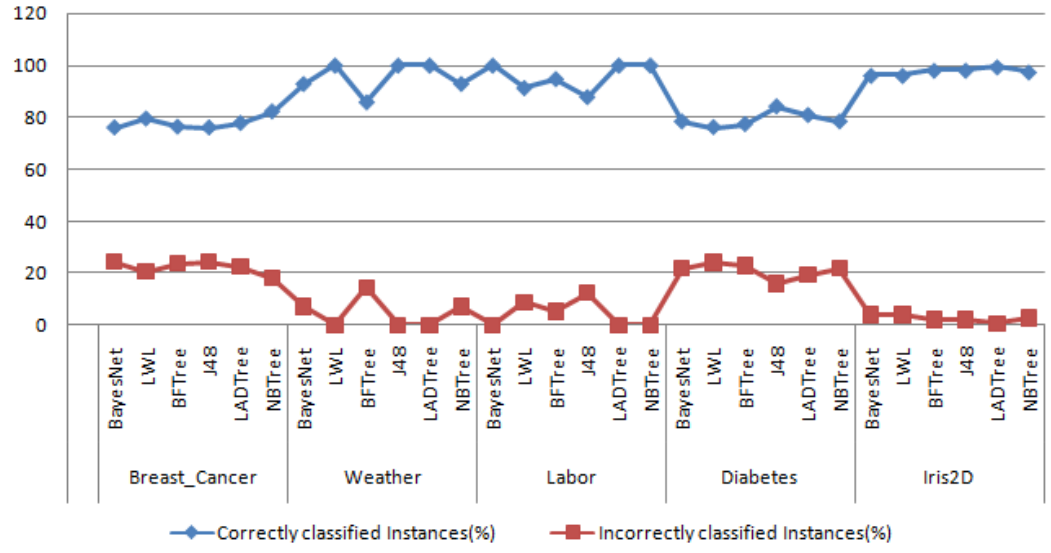

Fig. 1. : Accuracy comparison of different classifiers

On the other hand, the figures named as Fig. 2, Fig. 3 and Fig. 4 shows the Performance of selective classification algorithms over different chosen data sets based on evaluation metrics Precision, F-Measure and ROC Area respectively.

International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.02, February 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

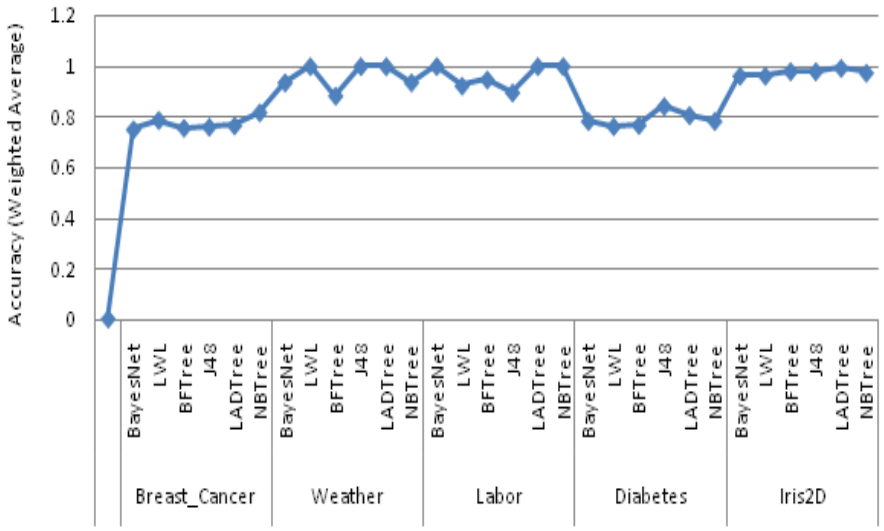

Fig. 2. : Average precision of learning classifiers over chosen datasets

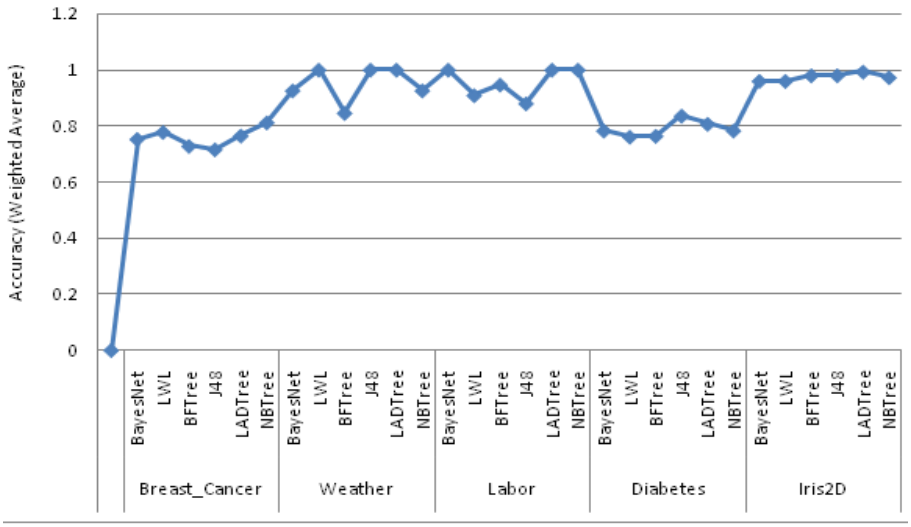

Fig. 3. : Average F-Measure of learning classifiers over chosen datasets

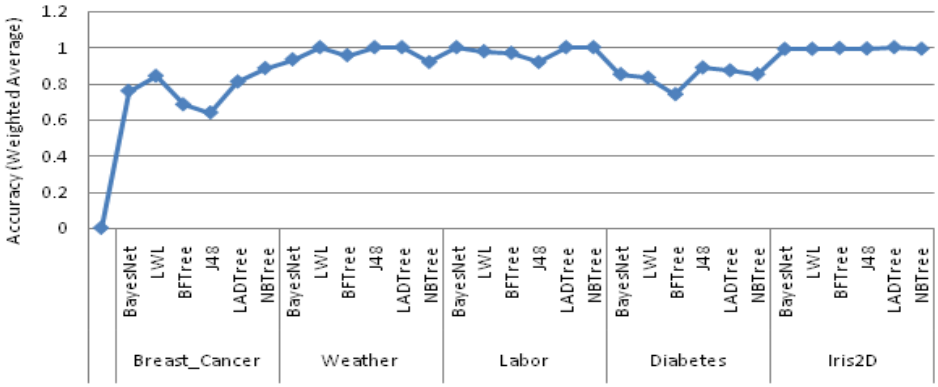

Fig. 4. Average ROC Area of learning classifiers over chosen datasets

VI. CONCLUSION

In this paper, various data classification techniques such as BayesNet, LWL, BFTree, J48, LADTree, NBTree have been discussed. I consider the precision, F-measure and ROC Area as evaluation metrics for conducting my experiment on the performance evaluation of different classifiers over five different UCI datasets (i.e., Breast_Cancer, Weather, Labor, Diabetes and Iris2D).

REFERENCES

[1] Dr. D. Durga Bhavani, A. Vasavi, P.T. Keshava : "*Machine Learning: A Critical Review of Classification Techniques*", International Journal of Advanced Research in Computer and Communication Engineering (2278-1021) Vol. 5, Special Issue 3, November 2016.

[2] D.Lavanya ,Dr. K.Usha Rani : "*Performance Evaluation of Decision Tree Classifiers on Medical Datasets*", International Journal of Computer Applications (0975 – 8887)Volume 26– No.4, July 2011.

[3] Abhaya Kumar Samal, Subhendu Kumar Pani," *Comparative Study of J48, AD Tree, REP Tree and BF Tree Data Mining Algorithms through Colon Tumour Dataset*", IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 03, 2016 | ISSN (online): 2321-0613.

[4] Gaganjot Kaur , Amit Chhabra "*Improved J48 Classification Algorithm for the Prediction of Diabetes*",International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014.

[5] Lakshmi Devasena C, "*proficiency comparison of ladtree And reptree classifiers for credit Risk forecast*", International Journal on Computational Sciences & Applications (IJCSA) Vol.5, No.1, February 2015.

[6] Rupali Malviya, Brajesh K. Umrao,"*Comparison of NBTree and VFI Machine Learning Algorithms for Network Intrusion Detection using Feature Selection*", International Journal of Computer Applications (0975 – 8887) Volume 108 – No. 2, December 2014.

[7] Solomon Getahun Fentie, Abebe Demessie Alemu, Bhabani Shankar D. M. : "*A Comparative Study on Performance Evalution of Eager versus Lazy Learning Methods*", *IJCSMC(*ISSN 2320–088X*), Vol. 3, Issue. 3, March 2014, pg.562 – 568.*