International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 225-229
e-ISSN:2278-621X

COMPARISON OF MACHINE LEARNING ALGORITHMS IN WEKA

Clive Almeida[1], Mevito Gonsalves[2] & Manimozhi R[3]

Abstract- Machine learning provides the computer to learn without being programmed explicitly. Its focus is on the development and designing of computer programs that can access the data and make use of it to learn themselves. By learning and understanding pattern recognition and conceptual learning theory in machine learning, artificial intelligence explores the study and construction of algorithms that can learn and make predictions. In this paper we discuss and compare Naïve Bayes, Multi-level Perceptron and J48 which are machine learning algorithms with respect to accuracy.
Keywords– Machine learning, artificial intelligence, mean, standard deviation, accuracy

1. INTRODUCTION

Today where otherwise human assistance was needed wen make use of different machine learning algorithms. Due to high level of accuracy machine learning methods have been applied for classification as an alternative to statistical methods. Data mining binds together statistics, database management, machine learning and areas which aim to get information from large amount of data. Exploration, model building and verification or validation are parts of data mining. Logical thinking steps that are required to make decision are taken into consideration for machine learning concept which is same as the working of human brain.

In this paper, we have compared and studied four machine learning algorithms like Naïve Bayes, Multi-level Perceptron and J48 with respect to accuracy using different data sets that are available.

2. RELATED WORK

In recent times, many people are working or have worked with machine learning algorithms to compare and understand the algorithms. The learning methods that designed and developed in the previous decade have good performance if the predictions are calibrated after training [1]. Based on different Data sets and different parameters on accuracy are taken into consideration to compare three algorithms Naive Bayes, Multilevel perceptron and J48. All the algorithms are given a set of input in WEKA to get the output which are later compared to each other.

Survey of Machine Learning Algorithms

2.1 Naïve Bayes algorithm –

Naïve Bayes classifier are simple probabilistic classifiers which are based on Bayes theorem with strong assumption between the feature which is independent. By evaluating a closed form expression maximum likelihood training can be achieved it takes linear time, other than iterative approximation used in other classifiers which can be expensive. Advantage of naïve base is that it can estimate the parameters that are necessary for classification with a small number of training data.

2.2 Multi level Perceptron –

Multilayer perceptron(MLP) consists of at least three layers of nodes. MLP is a class of feed forward artificial neural network. Back propagation technique which is a supervised learning technique is utilized for training. MLP are sometimes referred as "vanilla" neural networks mostly when they have a single hidden layer. The network is a network of simple processing elements which work together to produce a complex output. A multilayer feed forward network has input layer, one or more hidden layer and an output layer.

2.3 J48–

J48 is an open source java implementation of c4.5 decision tree algorithm. The implementation of particular learning algorithms encapsulated in a class and is dependent on other classes for some functionalities in WEKA. Each time when j48 executes in java virtual machine an instance of that class is created by allocating memory for building and storing a decision tree classifier. Programs which are large a broken down into more than one class. For building a decision tree the j84 classifier does not contain any code but it has references to instance that do most of the work.

[1] MCA V Semester, St Aloysius College, AIMIT, Mangalore, Karnataka, India
[2] MCA V Semester, St Aloysius College, AIMIT, Mangalore, Karnataka, India
[3] Asso. Professor, St Aloysius College, AIMIT, Mangalore, Karnataka, India

3. EXPERIMENTAL RESULT

3.1 Simulation enviroment

To compare and get the values of different parameters we have use a tool called WEKA. Waikato environment for Knowledge analysis (WEKA) was developed in University of Waikato in New Zealand [10]. It is basically written on java which is a suite of machine learning software [10]. It is provided with free license under GNU General Public License. It contains virtualization tools and algorithms for data analysis and predictive modelling. Data mining tasks such as clustering, data preprocessing, regression, visualization, classification and feature selection re supported by WEKA. All the techniques are predicted on the assumption that the data is available in file or relation where fixed number of attributes describe each data point. It also provides access to SQL database using java database connectivity which can process and return the result of the query. It does not have the capacity for multi relational data mining but other software can be used to a collection of linked database table to single table which can be processed. Sequence modelling if an important area that is currently not covered by algorithms.

In this paper we have taken three data sets the detail of each data set is shown in Table 1 the datasets taken are from UCI Machine learning repository.

Table -1 Details of 3 Data Sets

Data sets	Instances	Attributes	No, of Classes	Type
Breast Cancer	286	10	2	nominal
Glass	214	10	7	numeric
Diabetes	768	9	2	numeric

The Breast cancer data is provided by oncology Institute that appeared in machine learning literature.it is described by 9 instances in which some are linear and some are nominal [9].

Diabetes data was obtained from an automatic electronic recording device and paper records. Paper records have fictitious uniform recording times but electronic records are more realistic [9].

The Glass dataset is used to determine the type of class. At the crime scene glass left can be used as evidence if it can be identified properly [9].

Table -2 Accuracy on Breast Cancer

	ACCURACY ON BREAST CANCER			
Sr. No	Parameters	Naïve Bayes	MLP	J48
1	TP Rate	0.717	0.647	0.755
2	FP Rate	0.446	0.489	0.524
3	Precision	0.704	0.648	0.752
4	Recall	0.717	0.647	0.755
5	F-Measure	0.708	0.647	0.713
6	MCC	0.288	0.158	0.339
7	ROC Area	0.701	0.623	0.584

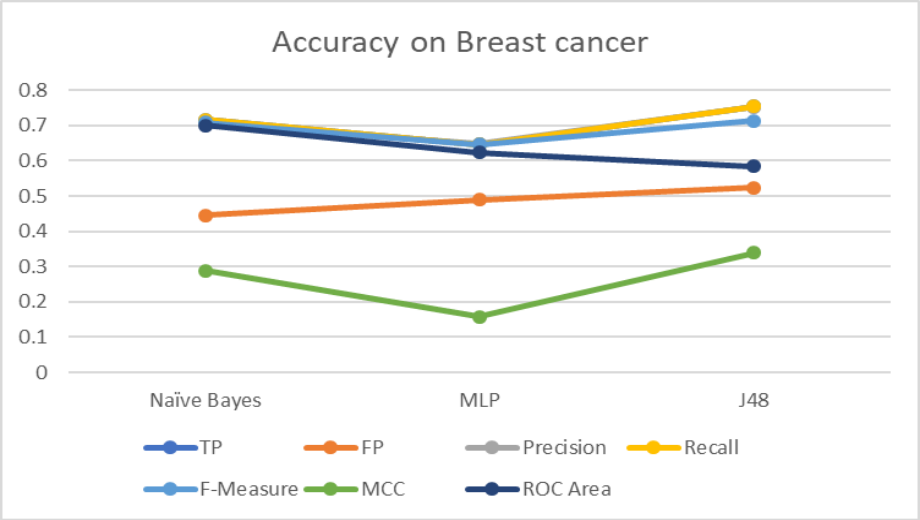

Figure 1. Accuracy chart on Breast Cancer

From the above table and fig we can see that all the parameters of J48 have higher accuracy accept ROC area where Naïve Bayes hav the highest followed by MLP and J48.If we see Naïve Bayes and MLP Naïve bayes have higher accuracy except in FP rate.

Table -3 Accuracy on Glass

	ACCURACY ON GLASS			
Sr. No	Parameters	Naïve Bayes	MLP	J48
1	TP Rate	0.486	0.678	0.668
2	FP Rate	0.188	0.141	0.13
3	Precision	0.496	0.671	0.67
4	Recall	0.486	0.678	0.668
5	F-Measure	0.453	0.659	0.668
6	MCC	0.297	0.537	0.539
7	ROC Area	0.762	0.847	0.807

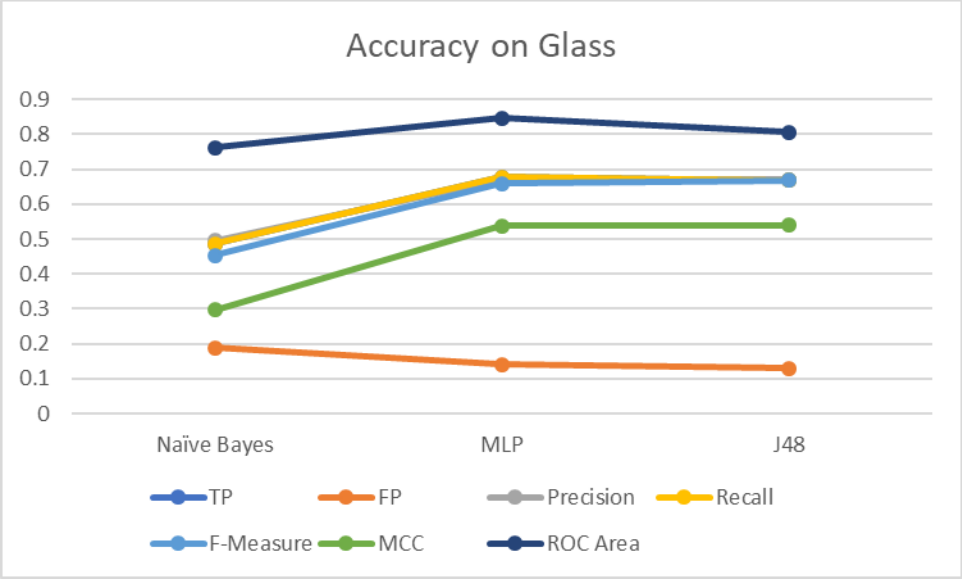

Figure 2. Accuracy chart on Glass

In Glass data set between MLP and J48 have almost equal accuracy measures except ROC measure where MLP has higher accuracy measure. In Naïve Bayes and j48 has higher accuracy measure exceptin FpP rate. Naïve bayes has higher accura only at FP rate compared to MLP

Table -4 Accuracy on Diabetes

	ACCURACY ON DIABETES			
Sr. No	Parameters	Naïve Bayes	MLP	J48
1	TP Rate	0.763	0.754	0.738
2	FP Rate	0.307	0.314	0.327
3	Precision	0.759	0.75	0.735
4	Recall	0.763	0.754	0.738
5	F-Measure	0.76	0.751	0.736
6	MCC	0.468	0.449	0.417
7	ROC Area	0.819	0.793	0.715

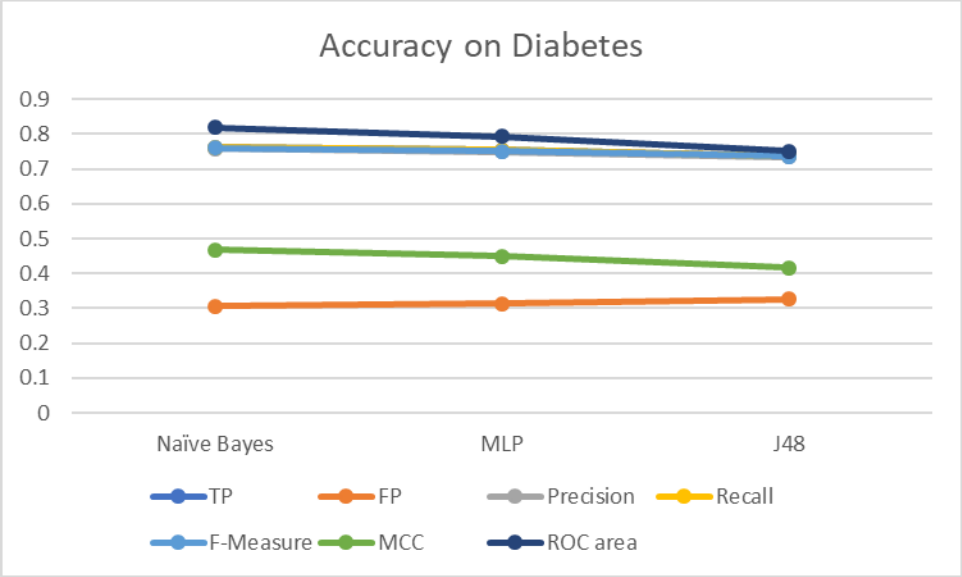

Figure 3. Accuracy chart on Diabetes

On diabetes dat set it almost the same but it shows that naïve Bayes have higher accuracy then MLP followed by J48 except in FP rate wher J48 has higher accuracy.

Table -5 Accuracy Measures of Naïve Bayes, MLP and J48

Sr No	Data Set	Naïve Bayes	MLP	J48
1	Breast Cancer	71.678	64.6853	75.524
2	Glass	48.598	67.757	66.8224
3	Diabetes	76.3021	75.3906	73.8281

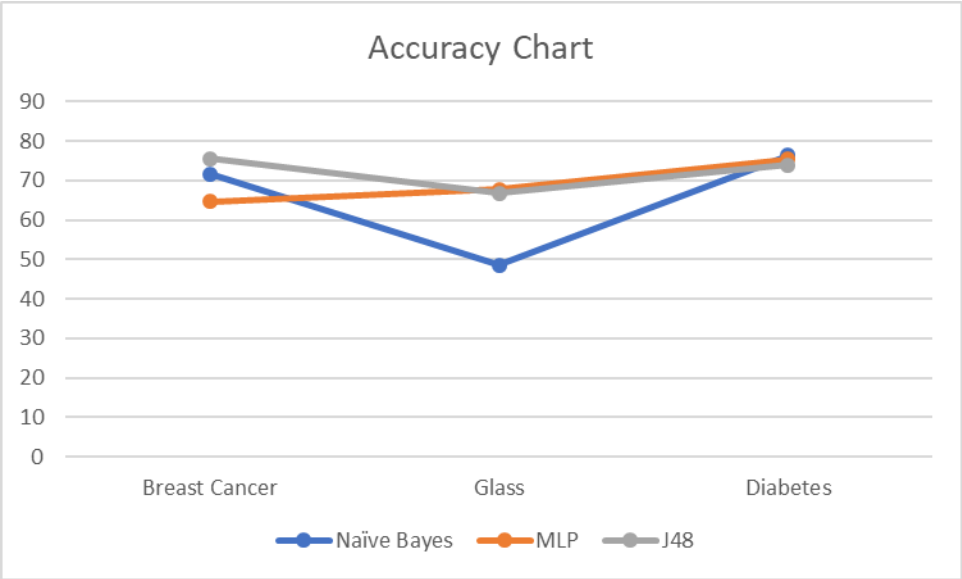

Figure 4. Accuracy chart of Breast cancer, glass and J48

From the above fig we can clearly see that J48 gives the best accuray with this three datasets followed by MLP and J48.we can say that J48 is the best among Naïve Byes,MLP and J48 for this three data sets.but if we see only two data sets glass and diabetes MLP is better than J48

4. CONCLUSION

In this paper we evaluated the performance in terms of classification accuracy of Naïve Bayes, Multilayer Perceptron and J48 algorithm taking various measures like TP rate, FP rate, Precision, Recall, F-measure, MMC and ROC area. Accuracy has

been measured taking into consideration each data set. J48 gives the best accuracy for breast cancer data set. On diabetes and glass MLP and J48 are almost same but MLP has an edge over J48. Naïve Bayes is slightly better than MLP on diabetes data set.

5. REFERENCES

[1] Rich Caruana, Alexandru Niculescu-Mizil, "An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics" A revised version of this paper appeared in the Proceedings of ICML'06.

[2] Rohit Arora, Suman "Comparitive analysis of classification algorithms on different datasets using weka,". International Journal of Computer Applications (0975 – 8887).

[3] Pablo D. Robles-Granda and Ivan V. Belik, "A Comparison of Machine Learning Classifiers Applied to Financial Datasets," Proceedings of the World Congress on Engineering and Computer Science 2010 Vol I WCECS 2010, October 20-22, 2010, San Francisco, USA.

[4] Bharat Deshmukh, Ajay S. Patil2 & B.V. Pawar, "Comparison of Classification Algorithms using WEKA on Various Datasets" IJCSIT International Journal of Computer Science and Information Technology, Vol. 4, No. 2, December 2011.

[5] William Elazmeh "Machine Learning algorithms and methods inWeka"

[6] Aaditya Desai, Dr. Sunil Rai, "Analysis of Machine Learning Algorithms using WEKA" International Conference & Workshop on Recent Trends in Technology, (TCET) 2012.

[7] André Rodrigues OliveraI , Valter RoeslerII, Cirano IochpeII, Maria Inês SchmidtIII, Álvaro VigoIV, Sandhi Maria BarretoV , Bruce Bartholow DuncanII " Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes – ELSA-Brasil: accuracy study".Available: http://www.scielo.br/pdf/spmj/v135n3/1806-9460-spmj-135-03-00234.pdf

[8] Nigel Williams, Sebastian Zander, Grenville Armitage"A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow" ACM SIGCOMM Computer Communication Review Volume 36, Number 5, October 2006

[9] UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/datasets.html.

[10] Weka: http://www.cs.waikato.ac.nz/~ml/weka/

.