

Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA

Yugal kumar

Assistant Professor in CSE/IT Dept, Hindu College of Engineering, Industrial Area, Sonapat, Haryana, India.
yugalkumar.14@gmail.com

G. Sahoo

Professor in Dept. of Information Technology, Birla Institute of Technology, Mesra, Ranchi, Jhrakhand, India.
gsahoo@bitmesra.ac.in

Abstract— In the field of Machine learning & Data Mining, lot of work had been done to construct new classification techniques/ classifiers and lot of research is going on to construct further new classifiers with the help of nature inspired technique such as Genetic Algorithm, Ant Colony Optimization, Bee Colony Optimization, Neural Network, Particle Swarm Optimization etc. Many researchers provided comparative study/ analysis of classification techniques. But this paper deals with another form of analysis of classification techniques i.e. parametric and non parametric classifiers analysis. This paper identifies parametric & non parametric classifiers that are used in classification process and provides tree representation of these classifiers. For the analysis purpose, four classifiers are used in which two of them are parametric and rest of are non-parametric in nature.

Index Terms— Bayesian Net, Logistic Regression, Multi layer perceptron, and Navie Bayes

I. Introduction

In present era, research in every field has become inter disciplinary i.e. research is not bind with in boundaries but it include two or more disciplines to carry out the research and researcher of different field work together. Such type of interdisciplinary research field is Machine Learning that provides a common platform to the researcher from computer science, biology, mathematics, AI, Economics, Life Sciences Statistics and control theory to work together. Due to the research, development and application of machine learning algorithms, day by day new classifications methods/classifiers have been developed. A Large number of classification techniques were developed by the means of statistical method, logical method , symbolic method, perceptron based method, fuzzy

based concept, Neural network based and rule based etc. D. Dutton & G. Conroy et al [10] provided brief review of what Machine Learning algorithms has done. De Mantaras and Armengol also reported historical survey of logic and instance based learning classifiers [11]. So, it is quite difficult how to categories classification methods/ classifiers. The purpose of this paper is to build up relation between classifiers or provide a Skelton how classifiers can be categories under one name and examine the performance of classifiers by using WEKA tool. Figure 1 shows tree like representation of classifiers used in classification process. Root node of the tree defined by the term classification is further divided into supervised and unsupervised classification. Basically, unsupervised classification is known as clustering. Supervised technique can be defined as form of learning with the help of supervisor(learning through example). Unsupervised classification is just opposite to the supervised classification i.e. learning without supervisor(Learning from observations). Classification techniques consist of following steps to perform the mining process:-

1. Data Acquisition
2. Data preprocessing
3. Data Presentation
4. Decision Making
5. Performance Evaluation

II. Supervised Classification

This technique is used to arrange the objects into different categories with help of supervisor. Supervised classification can be divided into probabilistic based and geomatric based classification. Probabilistic based classifier involves density based function to classify instance. Density based function is further divided into parametric and non parametric

methods. Parametric classifier is based on the statistical probability distribution of each class. Non parametric classifiers are used in case of unknown density function and used to estimate the probability density function. So, there are large number of classifiers/methods come to existence that are used for classification task. Some of these are shown in figure 1 such as K-NN, Voted perceptron, naive bayes so on. Some of classifiers includes decision tree method that are performing a successive partitioning of cases until all subsets belong to a single class[12,13] and statistical & rough set approaches [7]. Neural Network inspired classifiers were also developed to perform classification task [14,15]. Some of the classifier are based on the fuzzy if then rule[16], linear discriminating model [17,18] and

integration of rule induction & lazy learning [19] is also developed. These classifiers are both parametric and non parametric. BayesNet, Naviebayes, Multilayer Perceptron and Logistic regression are used to evaluate the performance of parametric and non parametric classification techniques in this paper. To evaluate the performance of these classifier weka data mining tool is used. First reason of selecting this tool is free of cost availability under general public license. Second reason is to learn the concepts and application of data mining using this tool in real time. Weka is providing large number of different classifiers that are used in data mining task and analyze the output produced by these classifiers. Figure 2 shows the supervised classification process.

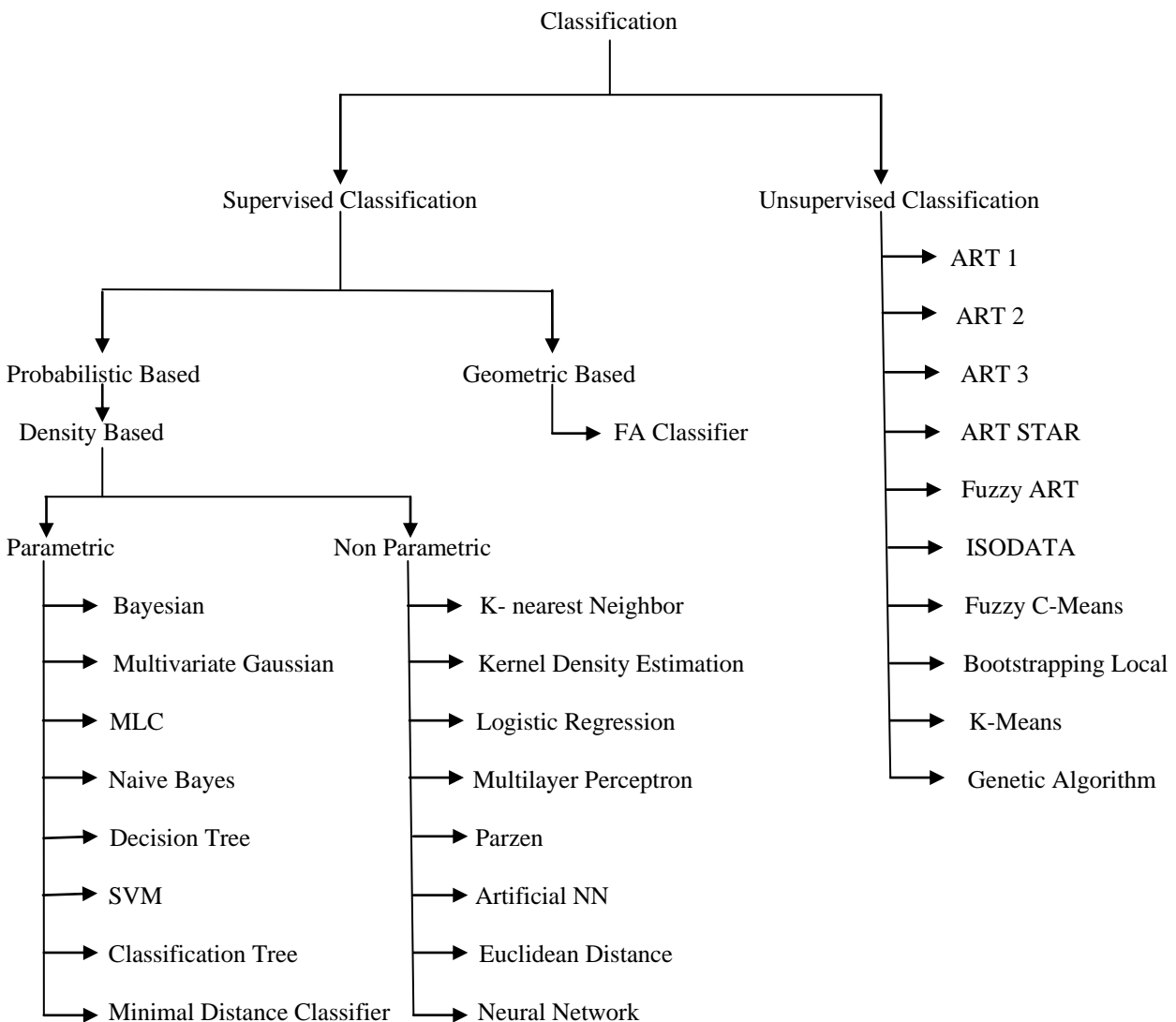


Figure1: Representation of Classification Classifiers in Tree Form

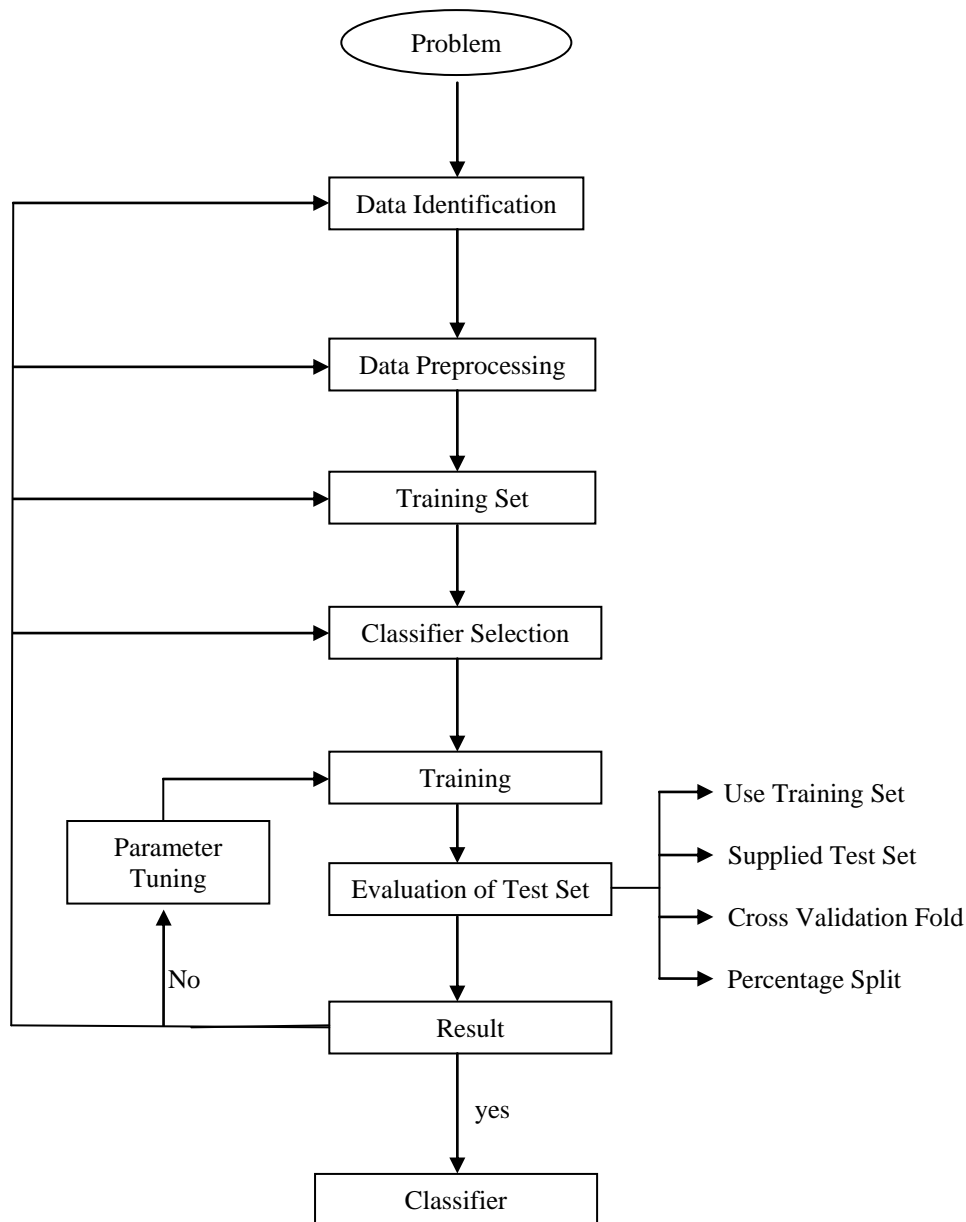


Figure 2: Process of Supervised Classification

2.1. Bayesian Classifier

Bayes classifier is an example of statistical-based learning scheme. In Bayesian classifier, it is assumed that all attribute have equal importance and value of one attribute does not provide the value of another attribute. In bayesian net, classification is done with the help of Bayes rule.

2.2. Navie Bayes Classifier

Bayes theorem with independent assumptions between predictors is core concept of Naive Bayes classifier. The simplest approach of Bayesian network is naive bayes in which all attribute of a dataset is independent to its class variable value. So, Naive Bayes classifier is a Bayesian network where the class

has no parents and each attribute has the class as its sole parent [4, 5]. Naive Bayes classifiers build the model easily with no complicated iterative parameter estimation. Due to this feature, naive bayes classifiers are useful for very large datasets. Despite its simplicity, Naive Bayes classifiers provide better results for complex real-world problems. But there is a problem with naive bayes classifier when it deals with a real world problem; it cannot provide an explicit model for its deployment. But recent study on analysis of the Bayesian classification problem has shown that there are some theoretical reasons for the unreasonable effectiveness of Naive Bayes classifiers [8]. On the other hand, naive bayes classifiers require small

amount of training data to predict the parameters for classification.

2.3. Multilayer Perceptron

The Multi Layer Perceptron (MLP) or Feed-forward network is a type of artificial neural network that consists of a non linear activation function in hidden layer [1]. MLP network provide nonlinear mapping between input and output vectors. Neural networks have two important functions i.e. pattern classifiers and nonlinear adaptive filters [6, 7]. A general framework of neural network consist of three layer architecture i.e. an input layer that define the input value, one or more hidden layers define the mathematical function and an output layer define final outcome of [3]. Each layer consists of a large number of neurons that are interconnected through weights. Each neuron has mathematical function (also known as activation function) that accepts input from previous layer and produced output for next layer. So, in neural networks the prediction is defined by the activation function [9].

2.4. Logistic regression

The term regression is defined as an analyzing or measuring the relation between a dependent variable and one or more independent variable. Regression can be defined by two types: Linear regression and logistic regression. Logistic regression is a generalization of linear regression [11]. It is basically used for estimating binary or multi-class dependent variables and the response variable is discrete, it cannot be modeled directly by linear regression i.e. discrete variable changed into continuous value. Logistic regression basically is used to classify the low dimensional data having non linear boundaries. It also provides the difference in the percentage of dependent variable and provides the rank of individual variable according to its importance. So, the main motto of Logistic regression is to determine the result of each variable correctly. Logistic regression is also known as logistic model/ logit model that provide categorical variable for target variable with two categories such as light or dark, slim/ healthy. Following Figure show the logistic regression implementation. Figure 3 shows logistic regression implementation.

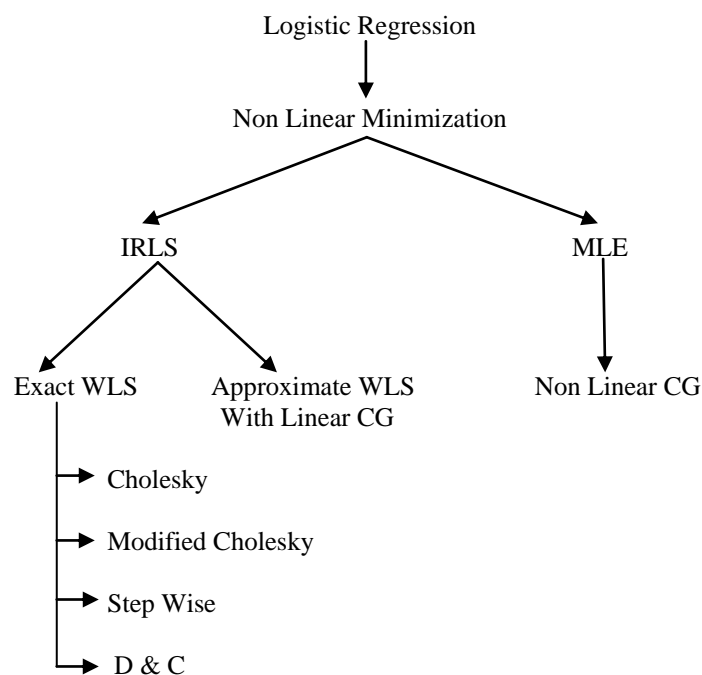


Figure 3: Logistic Regression Categorization.

III. Result & Discussion

To evaluate performance of parametric and non parametric classifier, four classifier are used i.e. Navie Bayes, Bayesian Net as parametric classifiers & Multi layer perceptron, Logistic regression as non parametric classifiers. Weka toolkit is used to analyze the performance of these classifiers. To analyze

performance of discussed classifiers following parameters are used:-

3.1. Kappa statistic

The Kappa Statistic can be defined as measuring degree of agreement between two sets of categorized data [22]. Kappa result varies between 0 to 1 intervals. Higher the value of Kappa means stronger the agreement/ bonding. If Kappa = 1, then there is perfect

agreement. If Kappa = 0, then there is no agreement. If values of Kappa statistics are varying in the range of 0.40 to 0.59 considered as moderate, 0.60 to 0.79 considered as substantial, and above 0.80 considered as outstanding [23].

3.2. Mean Absolute Error (MAE)

Mean absolute error can be defined as sum of absolute errors divided by number of predictions. It is measure set of predicted value to actual value i.e. how close a predicted model to actual model.

3.3. ROC Area

ROC provides comparison between predicted and actual target values in a classification. It describes the performance of a model with complete range of classification thresholds or in other words, it has been used for model comparisons. ROC area varies between 0 to 1 intervals [21]. By default classification threshold for binary classification is .5. When the probability of a prediction is 50% or more, the model predicts that class. Changes in the classification threshold affects the predictions made by the model; if the threshold for predicting the positive class is changed from 0.4 to 0.7 then fewer positive predictions will be made. This will affect the distribution of values in the confusion matrix.

3.4. Root Mean Square Error (RMSE)

Root mean square error is defined as square root of sum of squares error divided number of predictions. It is measure the differences between values predicted by a model and the values actually observed. Small value of RMSE means better accuracy of model. So, minimum of RMSE & MAE better is prediction and accuracy.

3.5. Data Set Description

In this paper, Breast feeding dataset has been used to evaluate the performance of classifiers. This dataset

consist of 927 first-born children whose mothers choose to breast feed for newly born children up to three month. Data set consist of 927 no. of instances and 10 attributes. Initially data is given in .txt format. Detailed description of dataset as given below:

Title: Breast Feeding Data

Reference: National Longitudinal Survey of Youth Handbook the Ohio State University, 1995.

1. Data Set Characteristics: Multivariate
2. Number of Instances: 927
3. Area: Life
4. Attribute Characteristics: Real
5. Number of Attributes: 10
6. Details of Attributes
 - Duration of breast feeding, weeks
 - Indicator of completed breast feeding (1=yes, 0=no)
 - Race of mother (1=white, 2=black, 3=other)
 - Mother in poverty (1=yes, 0=no)
 - Mother smoked at birth of child (1=yes, 0=no)
 - Mother used alcohol at birth of child (1=yes, 0=no)
 - Age of mother at birth of child
 - Year of birth
 - Education level of mother (years of school)
 - Prenatal care after 3rd month (1=yes, 0=no)
7. Associated Tasks: Classification
8. Missing Values? N/A
9. Distinct: 49
10. Unique: 10
11. Source: Data set has been taken from the book titled as Survival Analysis Techniques for Censored and Truncated Data published by Springer

Table 1: Shows comparison of parametric & non parametric classifiers using two test methods

Classifier	Use Training Set Method				10 Cross Fold Method			
	Kappa statistic	MAE	ROC Area	RMSE	Kappa statistic	MAE	ROC Area	RMSE
Navie Bayes	0.2724	0.2463	0.746	0.3577	0.1165	0.2674	0.653	0.3845
Multi Layer Perceptron	0.9144	0.0303	0.92	0.1516	0.045	0.2646	0.556	0.4684
Logistic Regression	0.2232	0.2397	0.787	0.3464	0.1165	0.2672	0.622	0.3901
Bayesian Net	0.2724	0.2463	0.746	0.3577	0.1165	0.2674	0.653	0.3845

The table 1 shows comparisons between parametric as well as nonparametric classifiers with the help of two test option i.e. use training method and 10 cross fold method. In this paper four parameters are used to analyze the performance of classifier as well as to find out which method is better. From the table 1 statistics, it is clear that the use training set method has better performance than 10 cross fold method.

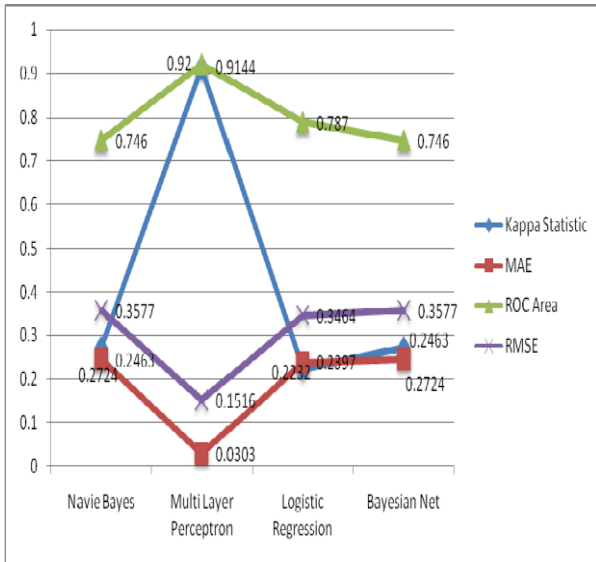


Figure 4: Comparison of Classifiers using Use Training Set Method.

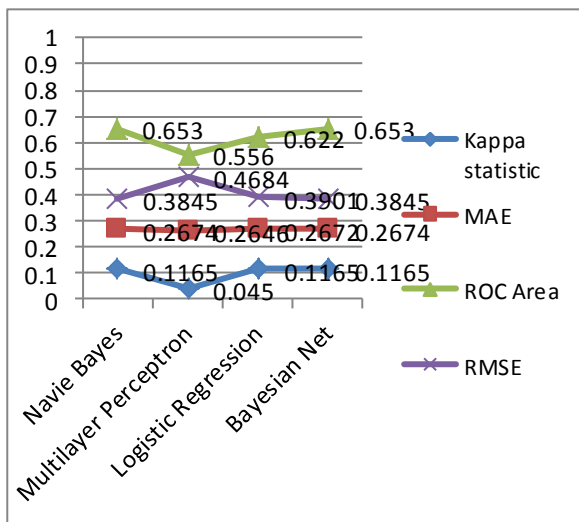


Figure 5: Comparison of Classifiers using 10 Cross Fold Method

Figure 4 & 5 shows the comparison of parametric/nonparametric classifiers with the help of two test method i.e. use training set method and 10 cross fold method. These figures also state that use training set method has better performance than 10 cross fold method. So, it is not easy to say that which techniques i.e. parametric or non parametric provide better result. But, there are several reasons that shows use training method has better than 10 cross fold method. Firstly, the value of RMSE and MAE i.e. minimum value of RMSE and MAE better is prediction.

So, use training set method has minimum value of RMSE and MAE as compare to 10 cross fold method. Secondly the value of kappa statistic, it provides poor agreement of predictions with actual class in both cases but comparing above discussed method; use training set method has better value than 10 cross fold method. Third reason is the value of ROC i.e. Higher value of ROC means higher positive predictions that can affect confuse matrix that provides the value of sensitivity, specification etc. Analyzing the performance of individual classifiers rather than model; multilayer perceptron has best performance among all these classifiers by using use training set method i.e. maximum value of Kappa statistic i.e. 0.91, minimum value of RMSE and MAE i.e. 0.03 & 0.15 and maximum value of ROC i.e. 0.92 while logistic regression has poor performance. The performance of navie bayes and Bayesian net is same. But in case of 10 cross fold method, performance of navie bayes, Bayesian net and logistic regression is almost same. Using this method multilayer perceptron shows poor performance among all these method i.e. kappa statistic value (0.04) almost null predication of actual class, value of RMSE & MAE (0.2646 & 0.46) and value of ROC (0.543).

IV. Conclusion

In this paper, we have investigated the performance of navie bayes, Bayesian net, multilayer perceptron and logistic regression. Navie bayes and Bayesian net classifiers derived from classical bayes theorem while multilayer perceptron based on the neural network technique that has been trained by back propagation method. Logistic regression is a statistical method that are use to classify the data. In these classifiers, multilayer perceptron and logistic regression classified as non parametric classifier of supervised classification technique while Bayesian net and navie bayes classified as parametric classifier of supervised classification. So in this paper, we try to sketch out comparative analysis of parametric and non parametric classifiers with the help of two test method i.e. use training set method and 10 cross fold method. Statistics shows that multilayer perceptron provide the best performance when we are used use training set method to build the model to classify the data and logistic regression classifier provide poor performance. But, multilayer perceptron has provided poor performance when 10 cross fold method was used and all other classifiers has provided almost equal performance.

References

[1] Sarle, Warren S. (1994), "Neural Networks and Statistical Models," Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, pp 1-13.

- [2] S.H.Musavi and M.Golabi (2008), "Application of Artificial Neural Networks in the River Water Quality Modeling: Karoon River,Iran", Journal Of Applied Sciences, Asian Network for Scientific Information, pp. 2324- 2328.
- [3] M.J. Diamantopoulou, V.Z. Antonopoulos and D.M. Papamichai (jan 2005), "The Use of a Neural Network Technique for the Prediction of Water Quality Parameters of Axios River in Northern Greece", Journal Of Operational Research, Springer-Verlag, pp. 115-125.
- [4] Buntine, W. (1991). 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
- [5] Daniel Grossman and Pedro Domingos (2004). Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood. In Press of Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.
- [6] D.Marquardt (1963), "An Algorithm for Least Squares Estimation of Non-Linear Parameter", J. Soc. Ind. Appl. Math., vol. 11 pp 431- 441.
- [7] L.Fausett (1994), "Fundamentals of Neural Networks Architecture.Algorithms and Applications", Pearson Prentice Hall, USA.
- [8] Ian h.Written and Eibe Frank.Data Mining Practical Machine Learning Tools and Techniques, Second Edition,Elsevier.
- [9] Janikow, C. Z. (1998). "Fuzzy decision trees: issues and methods." IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 28(1): 1-14.
- [10] Dutton,D. & Conroy, G. (1996),"A review of machine learning", Knowledge Engineering Review 12: 341-367.
- [11] De Mantaras & Armengol E. (1998),"Machine learning from example: Inductive and Lazy methods", Data & Knowledge Engineering 25: 99-123.
- [12] J. R. Quinlan (1986), "Induction of decision trees",Machine learning 1, pp 81-106.
- [13] J.R.Quinlan (1993),C 4.5: Programs for machine learning, morgan kaufmann,san francisco.
- [14] M. S. Hung, M. Hu, M. Shanker (2001),"Estimating breast cancer risks using neural network", International journal of operational research52, 1-10.
- [15] R. Setiono, L. C. K. Hui (1995), "Use of quasi-Newton method in a feed forward neural network construction algorithm", IEEE Trans. Neural Network 6(1) pp. no. 273-277.
- [16] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka (1995), "Selecting fuzzy if then rules for classification problems using genetic algorithm", IEEE Trans. Fuzzy System 3 (3) 260 – 270.
- [17] D. B. Fogel, E. C. Wason, E. M. Boughton, V. W. Porto, P. J. Angeline (1998), "Linear and Neural Model for classifying breast masses",IEEE Trans. On Medical imaging 17 (3) 485-488.
- [18] G. Fung, O. L. Mangasarian (Oct. 1999), "Semi supervised support vector machines for unlabeled data classification", Technical Report, Dept. of Computer science, University of Wisconsin.
- [19] C. H. Lee, D. G. Shin (1999), "A multi strategy approach to classification learning in database", Data Knowledge Engg. 31, 67-93.
- [20] http://www.mcw.edu/FileLibrary/Groups/Biostatistics/Publicfiles/DataFromSection/DataFromSectionTXT/Data_from_section_1.14.txt
- [21] Fawcett , T (2006),"An introduction to ROC analysis", Pattern Recognit Lett, Vol.No.27:861–874.
- [22] Melville, P.; Yang, S.M.; Saar-Tsechansky, M. & Mooney R (2005),"Active learning for probability estimation using Jensen-Shannon divergence". In Proceedings of the European Conference on Machine Learning (ECML), pages 268–279. Springer.
- [23] Landis, J.R.; & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 33: 159–174.

Authors Bibliography



G. Sahoo received his MSc in Mathematics from Utkal University in the year 1980 and PhD in the Area of Computational Mathematics from Indian Institute of Technology, Kharagpur in the year 1987. He has been associated with Birla Institute of Technology, Mesra, Ranchi, India since 1988, and currently, he is working as a

Professor and Head in the Department of Information Technology. His research interest includes theoretical computer science, parallel and distributed computing, cloud computing, evolutionary computing, information security, image processing and pattern recognition.



Mr. Yugal Kumar received his B.Tech in Information Technology from Maharishi Dayanand University, Rohtak, (India) in 2006 & M.Tech in Computer Engineering from Maharishi Dayanand University, Rohtak, India in 2009. His research interests include fuzzy

logic, computer network and data Mining & swarm intelligence system. At present, he has been worked as working as an Assistant Professor in Department of Computer Science and Engineering, Hindu College of Engineering, Sonapat, Haryana, India.