W.P.No. IT-11-07
August 2011

INDIAN INSTITUTE OF FOREIGN TRADE

Working Paper

An Empirical comparison of rule based classification techniques in medical databases

R.P.Datta
Sanjib Saha

W.P. No: IT-11-07

Working Paper Series

Aim

The main aim of the working paper series of IIFT is to help faculty members share their research findings with professional colleagues in the pre publication stage.

Submission

All faculty members of IIFT are eligible to submit working papers. Additionally any scholar who has presented her/his paper in any of the IIFT campuses in a seminar/conference will also be eligible to submit the paper as a working paper of IIFT.

Review Process

All working papers are refereed

Copyright Issues

Keys to the first two digits of the working paper numbers

GM: General Management
MA: Marketing Management
FI: Finance
IT: Information and Technology
QT: Quantitative Techniques
EC: Economics
LD: Trade Logistics and Documentation

Disclaimer

Views expressed in this working paper are those of the authors and not necessarily that of IIFT.

Printed and published by

Indian Institute of Foreign Trade

Delhi Centre

IIFT Bhawan, B-21, Qutab Institutional Area, New Delhi – 110016

Kolkata Centre

J1/14, EP & GP Block, Sector –V, Salt Lake, Kolkata - 700091

Contact

workingpapers@iift.ac.in

List of working papers of IIFT

See end of the document

Series Editor

Ranajoy Bhattacharyya

W.P. No: IT-11-07

An Empirical comparison of rule based classification techniques in medical databases

R.P.Datta*
Sanjib Saha**

Abstract

Classification techniques have been widely applied in the field of medical databases and have gained a lot of success. At present various classification algorithms are available in the literature and the problem of choosing the best method for a particular data set is faced by many researchers. In this paper, we apply five well-known rule based classification techniques, Decision Tree, JRIP, NNGE, PART and RIDOR, on different medical databases and compare their relative merits & demerits. Subsequently, we interpret their applicability to segment patients into groups.

Keywords: classification, datamining, knowledge discovery, and rule based classification

---

* Indian Institute of Foreign Trade, Kolkata Campus, J1/14, EP & GP Block, Sector V, Salt Lake , Kolkata – 700091, India. E-mail: rpdatta@iift.ac.in
** Tata Consultancy Services, TCS New Building, Block EP and GP, Plot B-1, Salt Lake, Sector 5, Kolkata - 700091
Readers should send their comments on this paper directly to the authors. An earlier version of this paper was presented at the 2nd International Congress on Pervasive Computing and Management 12-13, Dec 2009 at Sydney, Australia and published in the Conference Proceedings of the same conference.
The authors wish to thank The School of Education Technology, Jadavpur University, Kolkata, 700032, India, where part of the work was carried out.

An Empirical comparison of rule based classification techniques in medical databases

1. Introduction

In data mining and knowledge discovery it is an important task to construct fast and accurate classifiers for large data sets. Classification plays an important role in data mining. Some of the commonly used approaches for classification used to extract knowledge from data are statistical John G.H (1995), divide and conquer (Furnkranz, J. 1996) and covering (Cendrowska, J. 1987).

Many algorithms have been derived from these approaches, like Naïve Bayes (John G.H 1995), See5 (Quinlan 2011), C4.5 (Quinlan, J. R. 1993), PART (Frank, E. and Witten, I. 1998), Prism (Cendrowska, J. 1987) and IREP (Furnkranz, J. and Widmer, G. 1994). Sometimes a small subset of rules is found by traditional classification techniques and detailed rules that may play an important role are often missed (Pazzani, M., Mani, S., and Shankle, W. R. 1997).

Classification and Association rules play a major part in data mining. Classification is the process of dividing a dataset into mutually exclusive groups. Association rules give a process to find relationships among data items in a given dataset. The objective of this study is to try to provide some suggestions to the following questions faced by researchers, namely: (i) How does one choose the algorithm which is best suitable for the particular data set under consideration. (ii) How does one compare a particular algorithm with another as far as effectiveness is concerned.

The overall objective of this paper is to study the performance comparison of five different rule based classification techniques using medical databases to segment patients into groups. The detailed objective is to compare the performance of the following rule based classification techniques,- Decision Tree, JRIP, NNGE, PART, and RIDOR, based on the following criteria,- number of rules generated, classification accuracy, kappa statistic, mean absolute error, root mean squared error, relative absolute error and root relative squared error on medical databases and to find which algorithm performs best for which dataset. To this end we give a brief background of machine learning and the methods used in section 2, followed by results and discussions in section 3 and summarization and conclusions in section 4.

2. Background and Methods

A machine learning algorithm is an algorithm that can be implemented on a computer which can learn from past experience (observed instances) with respect to some category of tasks and some measure of performance (Mitchel, 1997). Machine learning methods are suitable for data from a variety of sources like transactional data, financial data, molecular biology related data etc. This is because the learning ability of the algorithm can construct classsifiers/hypotheses that can explain complex relationships in the data which are not visible otherwise. The classifiers or hypotheses thus constructed can be further verified by domain experts or subject matter specialists who can suggest some real lab experiments if needed to validate or refute the hypotheses.

Learning schemes in machine learning can be generally divided into two types: supervised learning where the output has been labeled apriori or the learner has some previous knowledge about the data; and unsupervised learning where no previous information is available to the learner about the data or the output. Some of the commonly performed tasks of the learner are to classify, characterize and cluster the input data as need be.

Knowledge discovery in databases (KDD), also frequently termed as data mining aims to find useful information from large collection of data. Data mining is the technique of extracting meaningful information from a large and mostly unorganized database. It is the process of performing automated extraction and generating predictive information from large databases. The discovered knowledge may be rules that describe properties of the data, patters that occur frequently and objects that are found to be in clusters in the database etc. (Heikki Mannila 1997).

Classification is one of the most common tasks in machine learning where given two or more different sets of example data, the learner needs to construct a classifier to distinguish between the different classes. Classification enables us to categorize records in a large database into predefined set of classes. The classes are defined before studying or examining records in the database. It also enables us to predict the future behaviour of that sample data. Classification can be looked upon as supervised learning.

Association rule enables us to establish association and relationships between large unclassified data items based on certain attributes and characteristics. Association rules define certain rules of associability between data items and then use those rules to establish relationships.

Comparison of classification techniques have been the subject of several previous studies. A comparison of rule based and association rule mining algorithms is dealt with in Majid M.M, Ali, A.B.M.S and Tickle, K.S. 2009. A comparison of Fuzzy based classification with neural network approaches for medical diagnosis is given in Herrman C, Halgamuge S.K, and Glesner M 1995. Again in the medical detection field the paper Frame AJ, Undrill PE, Cree MJ et al 1998, gives a comparison of computer based classification methods applied to the detection of micro aneurysms in ophthalmic fluoresce in angiograms. In a study by Nosofsky R.M., Gluck Mark A et al 1994, the authors have partially relocated and extended Shepard, Hovland, and Jenkins's 1961 classic study involving the task difficulty for learning six fundamental types of rule-based categorization problems. A comparison of various classification methods for predicting deception in Computer-Mediated Communication is presented in Zhou L, Burgoon J.K, Twitchell D.P. et al 2004.

2.1. Rule Based Classification

Rule based classification categorize records by using a collection of "if…then…" rules. Rule :(Condition) -> y where Condition is a conjunction of attributes and y is the class label e.g. (Blood Type=Warm) ^ (Lay Eggs=Yes) -> Birds. A rule r covers an instance x if the attributes of the instance satisfy the condition of the rule. Building classification rules are of two types- direct method and indirect method. Direct methods are those that extract rules directly from data e.g. RIPPER, (Cohen, 1995). Indirect methods are those that extract rules from other classification model like decision trees e.g. C4.5rules (Quinlan, 1993). In direct method we first grow a single rule (Rule Growing) then remove instances from this rule (Instance Elimination) after that prune the rule (Stopping Criterion and Rule Pruning) and then finally add rule to current rule set.

Advantages of rule-based classification are,

- it is as highly expressive as decision trees
- it is easy to interpret
- it is easy to generate
- it can classify new instances rapidly

Some commonly used Rule Based Classification Techniques are described briefly below:

2.1.1. Decision Table (DT)

Decision Table (Kohavi, 1995; Holmes, 1999) is an accurate method for numeric prediction from decision trees and it is an ordered set of If-Then rules that have the potential to be more compact and therefore more understandable than the decision trees.

2.1.2. JRIP

JRIP (Cohen, 1995) is a prepositional rule learner, i.e. Repeated Incremental Pruning to Produce Error Reduction (RIPPER). Initial rule set for each class is generated using IREP. The Minimum Description Length (MDL) based stopping condition is used. Once a rule set has been produced for each class, each rule is reconsidered and two variants are produced.

2.1.3. NNGE

This technique is Nearest-neighbor-like algorithm (Martin, 2002) using non-nested generalized exemplars, which are hyper rectangles that can be viewed as if-then rules.

2.1.4. PART

The PART (Frank & Witten, 1998) technique avoids global optimization step used in C4.5rules and RIPPER. It generates an unrestricted decision list using basic separate and -conquer procedure. It builds a partial decision tree to obtain a rule. It uses C4.5's procedures to build a tree. It uses separate-and-conquer. It builds a partial C4.5 decision tree in every iteration and makes the "best" leaf into a rule.

2.1.5. RIDOR

RIpple DOwn Rule learner (Gaines & Compton, 1995) generates a default rule first and then the exceptions for the default rule with the least (weighted) error rate. Then it generates the "best" exceptions for each exception and iterates until pure. Thus it performs a tree-like expansion of exceptions. The exceptions are a set of rules that predict classes other than the default. IREP is used to generate the exceptions.

Association rule mining was first proposed by Agrawal, Imielinski and Swami (1993), consists of “Finding frequent patterns, associations, correlations or casual structure sets of items or objects in transaction database, relational database and other information repositories”. The application of association rule is ranging from business management, production control, and market analysis, to engineering design and science exploration. At present association rule mining is an important task of data mining and is used in market basket analysis that tries to find out the shopping behaviour of customers in the hope of finding patterns (Agrawal, R., Amielinski, T. and Swami, A. 1993). One of the most popular algorithms of finding association rules is Apriori (Agrawal & Srikant, 1994; Liu-Hsu & Ma, 1998).

Association Rule(R): Implication expressions of the form X -> Y [s, c], where X and Y are item sets. (X, Y subset of I) and X∩Y=empty.

Support(s): Fraction of transactions that contain both X and Y. Probability that a transaction contains XUY. i.e. P(XUY).

Confidence(c): Measures how often items in Y appear in transactions that contain X. Conditional probability that a transaction having X also contains Y. i.e. P(X|Y)=Support(XUY)/Support(X).

2.2. Different Measurement Criteria

Confusion Matrix

The general structure of n class confusion matrix is:

Actual Class		Predicted Class			
		A	B	C	N
	A	tpA	eAB	eAC	eAN
	B	eBA	tpB	eBC	eBN
	C	eCA	eCB	tpC	eCN
	N	eNA	eNB	eNC	tpN

Classification Accuracy = (tpA + tpB + tpC +…+ tpN)/ total instances

$$\text{Kappa Statistics} = \frac{(\text{Observed agreement} - \text{Chance agreement})}{(1 - \text{Chance agreement})}$$

Observed agreement = (tpA + tpB + tpC +…+ tpN)/ total instances

Let us define 2n variables A1, A2, B1, B2, C1, C2, …, N1, N2

A1 = (tpA + eAB + eAC +…+ eAN) / total instances
A2 = (tpA + eBA + eCA +…+ eNA) / total instances
B1 = (eBA + tpB + eBC +…+ eBN) / total instances
B2 = (eAB + tpB + eCB +…+ eNB) / total instances
C1 = (eCA + eCB + tpC +…+ eCN) / total instances
C2 = (eAC + eBC + tpC +…+ eNC) / total instances
N1 = (eNA + eNB + eNC +…+ eN(N-1) + tpN) / total instances

N2 = (eAN + eBN + eCN +…+ e(N-1)N + tpN) / total instances

Chance agreement = A1 x A2 + B1 x B2 + C1 x C2 +…+ N1 x N2

Actual target values: a1 a2 … an

Predicted target values: p1 p2 … pn

Mean Absolute Error = $\dfrac{|p1-a1|+\ldots+|pn-an|}{n}$

Root Mean Squared Error= $\dfrac{(p1-a1)2+\ldots+ (pn-an) 2}{n}$

Relative Absolute Error= $\dfrac{|p1-a1|+\ldots+|pn-an|}{|a-a1|+\ldots+|a-an|}$

Root Relative Squared Error= $\dfrac{(p1-a1)2+\ldots+(pn-an)2}{(a-a1)2+\ldots+(a-an)2}$

3. Experimental Results and Observations

In this paper, 11 medical datasets have been taken from UCI Machine Learning Repository (Blake & Merz, 2000) and all the datasets are in arff (attribute relation file format). The datasets have both continuous and discrete attributes and also contain missing values. In the classification, all attributes of the dataset have been first selected. Then cross validation of 10 folds have been chosen as test method, fold determines the amount of data used for pruning; one fold is used for pruning, the rest for growing the rules.

After that the particular classification technique (e.g. Decision Tree) would be chosen and also the parameters would be specified such as search method greedy stepwise, seed value 1 i.e. seed is used for randomizing the data, prune value True i.e. whether pruning is performed, debug value False i.e. whether debug information is output to the console, confidence factor 0.25 i.e. confidence factor used for pruning (smaller values giver more pruning). In the association rules, only the discrete attributes of the dataset would be selected first. Then the particular association technique (e.g. Apriori) would be chosen and also the parameters would be specified as minimum support value, minimum confidence value, number of rules, and class index value –1 i.e. the last attribute is taken as class attribute. The analysis of these rule based classification techniques have been done by same criteria on number of rules generated, classification accuracy, kappa statistic, relative absolute error and root relative squared error. Analysis of association rules from the individual datasets has been done by same criteria on number of best rules, different rules and their corresponding confidence values. In this paper, the WEKA version 3.5.7 (Witten Ian H. & Frank Eibe, 2000 & 2005) framework is used for experimentations.

The Dataset used in the experiments are described in Tables 1 and 2.

Table 1: Characteristics of the Data Sets used

Dataset	Instances	Classes
Audiology	226	24
Breast Cancer	286	2
Breast Cancer-w	699	2
Colic	368	2
Diabetes	768	2
Heart-c	303	2
Heart Statlog	270	2
Hepatitis	155	2
Hypothyroid	3772	4
Lymph	148	4
Primary Tumor	339	22

Table 2: More characteristics of the Data Sets used

Dataset	Continuous Attributes	Discrete Attributes
Audiology	0	70
Breast Cancer	0	10
Breast Cancer-w	9	1
Colic	7	16
Diabetes	8	1
Heart-c	6	8
Heart Statlog	13	1
Hepatitis	6	14

Table 2: More characteristics of the Data Sets used (contd)

Hypothyroid	7	23
Lymph	3	16
Primary Tumor	0	18

The experimental Results for Rule Based Classification are given in the following tables 3 through Table 9.

Table 3: Number of Rules obtained from Different Data Sets

Dataset	**Rule Based Classification**				
	Number of Rules				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	27	20	46	21	105
Breast Cancer	25	3	105	20	3
Breast Cancer-w	23	6	271	10	4
Colic	32	4	109	9	4
Diabetes	32	4	280	13	4
Heart-c	17	4	77	19	6
Heart Statlog	16	5	102	24	6
Hepatitis	28	4	24	8	2
Hypothyroid	76	5	39	11	11
Lymph	18	6	33	13	10
Primary Tumor	65	7	165	43	162

Table 4: Classification Accuracy obtained from Different Data Sets

Dataset	**Rule Based Classification**				
	Classification Accuracy (%)				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	69.47	73.00	71.24	78.32	73.00
Breast Cancer	74.83	70.97	65.03	71.23	70.10
Breast Cancer-w	95.42	95.42	96.00	93.84	95.85
Colic	83.69	84.24	80.43	84.78	83.69
Diabetes	71.09	76.04	74.00	75.26	75.00
Heart-c	76.24	81.52	80.86	79.86	79.54
Heart Statlog	82.96	78.88	78.15	73.33	78.15
Hepatitis	81.93	78.06	84.51	84.51	78.71
Hypothyroid	99.33	99.33	98.70	99.41	99.44
Lymph	78.38	77.70	78.38	76.35	85.13
Primary Tumor	39.82	39.23	40.70	40.70	37.19

Table 5: Kappa Statistics Calculation for Different Data Sets

Dataset	**Rule Based Classification**				
	Kappa Statistic				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	0.63	0.68	0.66	0.74	0.68
Breast Cancer	0.27	0.24	0.12	0.2	0.18
Breast Cancer-w	0.89	0.89	0.91	0.86	0.90
Colic	0.64	0.65	0.56	0.66	0.63
Diabetes	0.33	0.45	0.41	0.44	0.42
Heart-c	0.52	0.63	0.61	0.59	0.58
Heart Statlog	0.65	0.56	0.55	0.46	0.55
Hepatitis	0.36	0.26	0.43	0.54	0.19
Hypothyroid	0.95	0.95	0.90	0.96	0.96
Lymph	0.58	0.57	0.58	0.55	0.71
Primary Tumor	0.30	0.24	0.32	0.32	0.29

Table 6: Mean Absolute Error for Different Data Sets

Dataset	**Rule Based Classification**				
	Mean Absolute Error				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	0.06	0.03	0.02	0.02	0.02
Breast Cancer	0.37	0.38	0.35	0.36	0.29
Breast Cancer-w	0.08	0.06	0.04	0.06	0.04
Colic	0.26	0.23	0.19	0.24	0.16
Diabetes	0.34	0.34	0.26	0.31	0.25
Heart-c	0.16	0.11	0.07	0.09	0.08
Heart Statlog	0.27	0.29	0.21	0.27	0.22
Hepatitis	0.26	0.26	0.15	0.18	0.21
Hypothyroid	0.02	0	0	0	0
Lymph	0.20	0.14	0.10	0.13	0.07
Primary Tumor	0.07	0.06	0.05	0.06	0.05

Table 7: Root Mean Squared Error for Different Data Sets

Dataset	**Rule Based Classification**				
	Root Mean Squared Error				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	0.16	0.13	0.15	0.12	0.15
Breast Cancer	0.44	0.45	0.60	0.47	0.54
Breast Cancer-w	0.18	0.20	0.20	0.22	0.20
Colic	0.36	0.36	0.44	0.35	0.40
Diabetes	0.42	0.42	0.51	0.41	0.50
Heart-c	0.26	0.24	0.28	0.26	0.28
Heart Statlog	0.37	0.41	0.46	0.49	0.46
Hepatitis	0.36	0.41	0.39	0.36	0.46

Table 7: Root Mean Squared Error for Different Data Sets (contd)

Hypothyroid	0.07	0.05	0.08	0.05	0.05
Lymph	0.29	0.31	0.32	0.33	0.27
Primary Tumor	0.19	0.19	0.23	0.19	0.23

Table 8: Relative Absolute Error for Different Data Sets

Dataset	**Rule Based Classification Relative Absolute Error (%)**				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	85.28	41.37	33.04	30.08	31.01
Breast Cancer	89.45	90.78	83.56	87.22	69.36
Breast Cancer-w	18.56	13.67	8.86	15.16	9.18
Colic	57.34	50.27	42.00	50.87	35.00
Diabetes	75.41	75.23	57.29	68.22	55.00
Heart-c	81.03	52.47	38.06	45.95	40.79
Heart Statlog	55.23	58.60	44.24	55.97	44.24
Hepatitis	80.31	78.56	46.89	56.18	64.47
Hypothyroid	29.48	6.56	8.91	4.78	3.81
Lymph	76.08	52.74	40.31	48.69	27.71
Primary Tumor	97.04	85.47	55.32	75.98	70.28

Table 9: Root Relative Squared Error for Different Data Sets

Dataset	**Rule Based Classification Root Relative Squared Error (%)**				
	DT	**JRIP**	**NNGE**	**PART**	**RIDOR**
Audiology	85.74	70.44	81.70	64.76	79.15
Breast Cancer	96.07	98.32	129.37	104.18	117.86
Breast Cancer-w	39.28	42.54	42.11	47.00	42.85
Colic	75.61	76.41	91.67	72.47	83.64
Diabetes	89.17	88.93	107.06	87.04	104.90
Heart-c	84.50	76.71	87.83	84.00	90.81
Heart Statlog	75.16	83.15	94.07	99.24	94.07
Hepatitis	89.46	101.79	97.17	88.94	113.94
Hypothyroid	40.27	29.10	42.32	26.42	27.70
Lymph	81.59	85.34	90.28	89.80	74.86
Primary Tumor	96.47	95.69	115.41	97.04	118.81

For Audiology dataset, the PART algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms since the Audiology dataset has large number of attributes (more than equal to 20).

For Breast Cancer dataset, the Decision Tree algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms as the Breast Cancer dataset has only discrete attributes.

For Breast Cancer-w dataset, the NNGE & Decision Tree algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms.

For Colic dataset, the PART algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms as the Colic dataset has large number of attributes (more than equal to 20).

For Diabetes dataset, the JRIP & Decision Tree algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms. This could be because the Diabetes dataset has all continuous attributes except the class attribute.

For Heart-c dataset, the JRIP algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms.

For Heart Statlog dataset, the Decision Tree algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms as the Heart Statlog dataset has all continuous attributes except the class attribute.

For Hepatitis dataset, the PART algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms as the Hepatitis dataset has large number of attributes (more than 20).

For Hypothyroid dataset, both the PART & RIDOR algorithms give better result on classification accuracy, kappa statistic and minimum error than the other algorithms. The Hypothyroid dataset has large number of attributes (more than 20).

For Lymph dataset, the RIDOR algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms.

For Primary Tumor dataset, both the NNGE & Decision Tree algorithm gives better result on classification accuracy, kappa statistic and minimum error than the other algorithms. The Primary Tumor dataset has only discrete attributes.

4. Summary and Conclusions

In this paper, five well-known rule based classification techniques namely Decision tree, JRIP, NNGE, PART, and RIDOR are applied on 11 medical datasets. From the experimental results, (Table 6 to Table 9) we make a comparative study of these algorithms and their applicability on medical databases.

From the results and their interpretations in this paper, one can make the following empirical observations:

- If the medical dataset has both continuous & discrete attributes and number of attribute is more than equal to 20 then PART algorithm performs better than the others.
- If the medical dataset has only discrete attributes then Decision Tree algorithm performs better than others.
- If the medical dataset has all continuous attributes except the class attribute then Decision Tree & JRIP algorithm performs better than others.

The values of Mean absolute error, Root mean squared error, Relative absolute error and Root relative squared error for these entire rule based classification algorithms on the 11 medical datasets make a comparative study of these algorithms and suggest which one is better for all these datasets.

From the experimental results (Table 6 to Table 9) of this paper, we see that:

- RIDOR algorithm gives lowest mean absolute errors for 10 medical datasets out of 11 datasets.
- Decision Tree gives lowest root mean squared errors for 10 medical datasets out of 11 datasets.
- RIDOR algorithm gives lowest relative absolute errors for 10 medical datasets out of 11 datasets.
- Decision Tree & PART gives lowest root relative squared errors for 8 medical datasets out of 11 datasets.

References

Agrawal R. & Srikant R. (1994) "Fast Algorithms for Mining Association Rules in Large Databases", *20th International Conference on Very Large Data Bases*, pp. 478-499.

Agrawal Rakesh, Imielinski Tomasz & Swami Arun (1993) "Database Mining: A performance Perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol.5. pp. 914-925,.

Agrawal, R., Amielinski, T., and Swami, A. (1993). "Mining association rule between sets of items in large databases", *Proceeding of the 1993 ACM SIGMOD International Conference on Management of Data*, pp. 207-216, Washington, DC, May 26-28

Blake, C. & Merz, C. (2000) *UCI repository of machine learning databases*

Cendrowska, J. (1987). "PRISM: An algorithm for inducing modular rules", *International Journal of Man-Machine Studies,* Vol.27, No.4, pp.349-370.

Cohen William W. (1995*)* "Fast Effective Rule Induction", *Twelfth International Conference on Machine Learning*, 115-123.

Flach P. A. & Lachiche N. (1999). "Confirmation-Guided Discovery of first-order rules with Tertius", *Machine Learning*. 42:61-95.

Frank Eibe & Ian H. Witten (1998). "Generating Accurate Rule Sets Without Global Optimization", *Fifteenth International Conference on Machine Learning*, 144-151.

Frank, E. and Witten, I. (1998). "Generating Accurate Rule Sets Without Global Optimization", Shavlik, J., ed., *Machine Learning: Proceedings of the Fifteenth International Conference*, pp. 144-151, Madison, Wisconsin, Morgan Kaufmann, San Francisco.

Frame AJ, Undrill PE, Cree MY, Olson JA et al (1998) "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms", *Comput Biol Med 1998*, May 28(3):225-38.

Furnkranz, J. and Widmer, G. (1994). "Incremental Reduced Error Pruning", *Machine Learning: Proceedings of the 11th Annual Conference*, New Brunswick, New Jersey, Morgan Kaufmann.

Furnkranz, J. (1996) "Separate-and-conquer rule learning", *Technical Report TR-96-25, Austrian Research Institute for Artificial Intelligence, Vienna*

Gaines. Brian R. & Compton. Paul (1995), "Induction of Ripple-Down Rules Applied to Modeling Large Databases", *J. Intell. Inf. Syst.* 5(3):211-228.

Han J. & Kamber M. (2001), "Data Mining: Concepts and Techniques", *Morgan Kaufmann, San Francisco, CA*.

Heikki Mannila (1997), "Methods and Problems in Data Mining", *Proceeding of International Conference on Database Theory, Delphi, Greece, January 1997,* F Afrati and P. Kolaitis (ed.), Springer Verlag

Herrmann Christoph S, Halgamuge Saman K, and Glesner M (1995), "Comparison of Fuzzy Rule Based Classification with Neural Network Approaches for Medical Diagnosis", *European Congress on Fuzzy and Intelligent Technologies (EUFIT) 1995*

Holmes Geoffrey, Hall Mark & Frank Eibe (1999), "Generating rule sets from model trees" *Proc 12th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, pp. 1-12. Springer*.

John G.H. and Langley,P (1995), "Estimating Continuous Distributions in Bayesian Classifiers", *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence.* pp. 338-345. Morgan Kaufmann, San Mateo

Kohavi Ron (1995), "The Power of Decision Tables", *8th European Conference on Machine Learning*, pp. 174-189.

Liu Bing, Hsu Wynne & Ma Yiming (1998), "Integrating Classification and Association Rule Mining", *Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 80-86.

Martin Brent (1995), "Instance-Based learning: Nearest Neighbor With Generalization*", Hamilton, New Zealand.*

Mazid, M.M; Ali, A.B.M.S, Tickle, K.S.(2009), "A comparison between Rule based and Association Rule Mining Algorithms" *Third International Conference on Network Security 2009, NSS'09*.

Mitchell, T. (1997), "Machine Learning", *McGraw-Hill*

Nosofsky Robert M, Gluck Mark A et al (1961), "Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961)" *Memory and Cognition 1994, 22(3),* pp. 352-369.

Pazzani, M., Mani, S., and Shankle, W. R. (1997), "Beyond Concise and colorful: learning intelligible rules*", KDD-97*.

Quinlan, J. R. See5.0 (http://www.rulequest.com) viewed on April 2011.

Quinlan J. R. (1993). C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers*

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. *San Mateo, CA: Morgan Kaufmann, San Francisco.*

Sylvain Roy (2002), "Nearest Neighbor With Generalization" *Christchurch, New Zealand*.

Sarawagi S., Thomas S. & Agrawal R. (1998), "Integrating association rule mining with relational database systems: Alternatives and implications*" SIGMOD*.

Scheffer Tobias (2001), "Finding Association Rules That Trade Support Optimally against Confidence", *5th European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 424-435.

Witten Ian H. & Frank Eibe (2000), "Data Mining Practical Machine Learning Tools and Techniques with java Implementations"*, MAorgan Kaufmann Publishers.*

Witten Ian H. & Frank Eibe (2005), "Data Mining: Practical machine learning tools and techniques", *2nd Edition. Morgan Kaufmann, San Francisco*.

Zhou Lina, Burgoon Judee K, Twitchell Douglas P. at al (2004), "A comparison of Classification Methods for Predicting Deception in Computer –Mediated Communication", *Journal of Management Information Systems* Vol. 20, No. 4, Spring, 2004

List of working papers of IIFT

Sinha, Deepankar (2010), "Multi-Dimensional Approach to Management of Port Life Cycle: The Case of Major Ports in India" Working Paper No: LD-10-01, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/01.pdf

Raychaudhuri, Bibek and Chakraborty, Debottam (2010), "Export Potential at the State Level: A Case Study of Karnataka", Working Paper No: EC-10-02, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/02.pdf

Nag, Biswajit (2011), "Comprehensive Economic Partnership Agreement Between India and Sri Lanka: Where Does it Lead?", Working Paper No: EC-11-03, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/03.pdf

Sinha, Deepankar (2011), "Container Yard Capacity Planning: A Causal Approach" Working Paper No: LD-11-04, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/04.pdf

Rastogi, K. Siddhartha (2011), "Welfare Assessment of SPS Standards: An Empirical Study of Indo-US Mango Trade Dispute", Working Paper No: EC-11-05, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/05.pdf

Nag, Biswajit and Sikdar, Chandrima (2011), "Welfare Implications of India-ASEAN FTA: An Analysis using GTAP Model", Working Paper No: EC-11-06, Indian Institute of Foreign Trade, New Delhi and Kolkata. This paper can be downloaded from http://cc.iift.ac.in/research/Docs/WP/06.pdf