PROFILING OF INTERNET BANKING USERS IN INDIA USING INTELLIGENT TECHNIQUES

V. Ravi **Mahil Carr** **N. Vidya Sagar**

Internet banking is a new delivery channel for banks in India. The Internet banking channel is both an informative and a transactional medium. However, Internet banking has not been popularly adopted in India as expected. The objective of this paper is to find the profiles of Internet banking users as well as non-users using intelligent techniques. This study investigates and identifies potential customers based on profiles of existing users. The profiles may be used to target and attract potential customers to adopt Internet banking. Significant determining variables that influence adoption of Internet banking were identified from the literature especially from the theory of reasoned action, theory of planned behavior, technology acceptance model and diffusion of innovations theory. Likert scale responses were collected from a sample of users and non-users of Internet Banking using a questionnaire. The resultant data set was analyzed using statistical and intelligent techniques like Classification and Regression Trees (CART), Support Vector Machines (SVM), Neural Networks and Logistic Regression and classification models were built. This paper compares these four predictive models for their accuracy and usefulness. CART turns out to be a good predictive model since it also provides rules for identifying potential Internet banking users, apart from performing feature selection.

INTRODUCTION

Internet banking is a new channel for the distribution of financial services through the Internet and the World Wide Web infrastructure. The Internet is now being considered as a strategic weapon – a competitive advantage that will revolutionize the way banks operate, deliver, and compete against one another. The Internet promises a revolution in retail banking of monumental proportions providing customers with new levels of convenience and flexibility. Internet banking allows customers to perform a wide range of banking transactions electronically via the bank's Web site. When first introduced, Internet banking was used mainly as an information presentation medium in which banks marketed their products and services on their Web sites. With the development of asynchronous technologies and secured electronic transaction technologies, however, more banks have come forward

Journal of Services Research, Volume 6, Number 2 (October 2006 - March 2007)

to use Internet banking both as a transactional as well as an informational medium. As a result, registered Internet banking users can now perform common banking transactions such as writing checks, paying bills, transferring funds, printing statements, and inquiring about account balances. Internet banking has evolved into a “one stop service and information unit” that promises great benefits to both banks and consumers. Internet banking would help banks present a potentially low cost alternative to brick and mortar branch banking.

From the consumers’ perspective, Internet banking provides a very convenient and effective approach to manage one’s finances as it is easily accessible 24 hours a day, and seven days a week. Besides, the information is current. For corporate customers, sophisticated cash management packages offered through Internet banking provides them with up to the minute information, allowing for timely funds management decisions. There are about four million computer users in India. The banking industry has tried to provide Internet banking services to its customers. However, customers have not adopted Internet banking in a big way in India.

Lack of trust in Internet banking has been identified as the key to the failure to adopt Internet banking (Araujo and Araujo, 2003; Castelfranchi and Tan 2001; Noterberg et al., 2003). Several studies have attempted to model technology adoption. In this paper, based on the antecedents to trust we identify significant variables that influence the adoption of Internet banking. After identifying key variables, we operationalize them in the form of a questionnaire. After collecting responses from a group of Internet banking users and non-users, four predictive classification models were built using intelligent techniques such as Classification and Regression Trees (CART 2005), Logistic Regression (Garson, 1998) Support Vector Machines (http://www.research.microsoft.com/users/jplatt/svm.html) and Neural Networks (Rummelhart and McClelland, 1986). The profile developed of users and non-users of Internet banking may be used to target segments of potential customers.

The rest of this paper is organized follows. The next section reviews existing trust models to identify key variables that affect adoption of Internet banking. Then the various methodologies that were used, the data collection and analysis via intelligent techniques are discussed. The results are followed by conclusions.

REVIEW OF LITERATURE

Theories such as Theory of Reasoned Action (TRA) (Lin and Wu, 2002; Gefem et al., 2003), Theory of Planned Behavior (TPB) (Matheison, 1991), Technology Acceptance Model (TAM) (Araujo and Araujo, 2003; Noteberg et al. 2003; Gefen et al., 2003, Matheison, 1991; Malhotra and Galleta, 1999) and Diffusion of Innovation Theory (DIT) (Loch et al; 2003) provide insight into usage of Internet banking. Significant variables from these theories are drawn to build predictive classification models. This work differs from Confirmatory Factor Analysis (Garson, 2005), which tries to validate the designed model and its reliability. The variables derived from established theories mentioned above were used to make a questionnaire to collect data from Internet banking users and non-users. The variables are outlined and discussed below:

Beliefs

Belief is a representational mental state that takes the form of a propositional attitude. It can be defined as any cognitive content that is held true. It means that an individual is convinced of the truth of a statement or allegation. This is derived from TRA (Lin and Wu 2002).

Subjective Norms

The degree, to which people think that others who are important to them think they should perform the behavior. People often choose their family and friends to step into such new things as they have faith in them first and then the new concept. This variable is an important variable from TPB (Matheison, 1991).

Perceived Ease of Use

Perceived ease of use (PEOU) refers to the degree to which a person believes that using a particular system would be free of effort. According to TAM, the PEOU is one of the main variables influencing the trust on any newly adopted technology (Araujo and Araujo, 2003, Noteberg et al. 2003, Gefen et al. 2003, Matheison, 1991, Malhotra and Galleta, 1999).

Perceived Usefulness

Perceived Usefulness (PU) is defined as the degree to which a person believes that using a particular technology will enhance his performance.

The PU is also an important variable from TAM (Araujo and Araujo, 2003; Noteberg et al. 2003; Gefen et al., 2003; Matheison, 1991; Malhotra and Galleta, 1999). Perceived Usefulness has been confirmed as an important variable that influences user technology acceptance and therefore has received a great deal of attention from prior researchers.

Trust in the Bank

To have confidence or faith in the bank providing the Internet banking service (Kim and Prabhakar, 2005).

Security

Security is the state of being free from dangers like theft or loosing money and information (Gefen et al., 2003). The need of Security for the safe guard of ones money and information should be made available and take necessary precautions against all possibilities of frauds and loosing information and property.

Attitude

Attitude is a complex mental state involving Beliefs and feelings and values and dispositions to act in certain ways towards accepting something. People can also be 'ambivalent' towards a target, meaning that they simultaneously possess a positive and a negative attitude towards it. This variable is extracted from TPB (Matheison, 1991) and TAM (Araujo and Araujo, 2003; Noteberg et al.; 2003, Gefen et al., 2003; Matheison, 1991; Malhotra and Galleta, 1999).

Intention

It is the course of action that intends to flow towards an objective. It is an anticipated outcome that is intended or that guides ones planned actions. It is a decision or wish to do something specific with determination. The intention is an important variable in the TPB (Mathieson, 1991) and TAM (Araujo and Araujo, 2003; Noteberg et al., 2003; Gefen et al., 2003; Matheison, 1991; Malhotra and Galleta, 1999).

METHODOLOGY

A survey was conducted using the questionnaire, given to 165 individuals with minimum educational background of graduation and falling under age groups of 23 years to 52 years. This group comprises 103 Internet banking users and 62 non-users. This data set is analyzed with data mining approaches viz., CART, Logistic Regression, SVM and Neural Networks to obtain the overall percentage prediction accuracy.

The questionnaire is formed using Likert scale (Babbie, 1994) that presents a set of attitude statements. Here respondents were asked to express agreement or disagreement on a five-point scale. Each degree of agreement is given a numerical value from one to five. Thus a total numerical value was calculated from all the responses. In the questionnaire, the variables were operationalized into 58 related questions. The questionnaire was pretested. This data was preprocessed to arrive at a single score for each variable by taking the average of all its related questions. Thus a dataset of 165 respondents and 8 variables was obtained.

Data Analysis

The predictive classification models used in this research work were chosen from the areas of statistics, machine learning, statistical learning theory and the neural networks family. The preprocessed data set was split into training and test sets such that training set comprises 70% and test set 30% of data set. The training set was used to train the following four predictive models and their effectiveness was tested using the test data - Classification and Regression Trees (CART), Logistic Regression, Support Vector Machines (SVM) and Neural Networks.

CART is very versatile and a successful decision tree building technique. It automatically searches for important relationships and uncovers hidden structure even in highly complex data. It is being used increasingly in medical, marketing, environmental, banking and commercial applications. Our work on CART is carried out using the Salford Systems CART 5.0 (CART 2005). Logistic Regression is chosen, as it is the most basic non-linear classification model coming from statistical family. SVM is chosen from statistical learning theory because of its sound theoretical basis and wide recognition in diverse disciplines. Finally, a multi layer perception is chosen from neural networks family because of its wide applicability and robust nature.

RESULTS

CART Results

Table 1 gives the prediction success of the test sample of 50 respondents. This table gives percentage of correctness for non-users and users of Internet banking as actual class 0 and 1 respectively. The percentage of correctness of non-users of Internet banking is 12 upon 19 i.e. 63.16% and for Internet banking users for 24 upon 31 i.e., 77.42%. The average success percentage of both non-users and users of Internet banking is 70.29%.

Table 1: Test Sample Prediction Success Table

Actual Class	Total Cases	Percent Correct	0 N=19	1 N=31
0	19	63.158	12	7
1	31	77.419	7	24

Table 2 gives the variable importance of all the variables used in the questionnaire for profiling the customers of Internet banking. This is as proposed by most methods in CART namely, the Gini Reduction, Symmetric Gini, Class probability and Twoing.

Table 2: Variable Importance Factors

Variable	Score	
INTENTION	100.00	
BELIEFS	94.57	
SN	92.60	
TIB	69.02	
ATTITUDE	56.09	
PU	40.16	
SECURITY	39.76	
PEOU	30.90	

CART performed feature selection and accordingly Intention is the most important variables while Beliefs, Subjective Norms, Trust in Bank, Attitude, Perceived Usefulness, Security and Perceived Ease of Use followed in that order. So, this states that Intention, Beliefs and Subjective Norms are the most important variables in customers' point of view according to the collected empirical data.

Table 3 provides 17 rules derived from the 17 terminal nodes of the decision tree constructed with entropy method in CART. Out of these, 9 rules profiled the Internet banking users and rest profiled as non-users.

Table 3: Rules for 17 Nodes of Test Data

Node	Rule(s) Framed	Result
1	If *ATTITUDE* ………………….is 2	An Internet banking user
2	If *BELIEFS* ……………………is 2 or 4 And *SUBJECTIVE NORMS* …….is 3 or 4 And *INTENTION* ………………...is 2 or 4 or 5 And *ATTITUDE* ………………….is 3 or 4	An Internet banking user
3.	If *TRUST IN BANK*……………is 2 or 4 And *BELIEFS* ……………………is 3 And *SUBJECTIVE NORMS* …….is 3 or 4 And *INTENTION* ………………...is 2 or 4 or 5 And *ATTITUDE* ………………….is 3 or 4	An Internet banking user
4.	If *ATTITUDE* ………………….is 4 And *TRUST IN BANK* …………...is 3 And *BELIEFS* ……………………is 3 And *SUBJECTIVE NORMS* …….is 3 or 4 And *INTENTION* ………………...is 2 or 4 or 5	An Internet banking user
5.	If *PERCEIVED EASE OF USE* is 2 or 4 And *ATTITUDE* ………………….is 3 And *TRUST IN BANK*..…………..is 3 And *BELIEFS* …………………….is 3 And *SUBJECTIVE NORMS* ……..is 3 or 4 And *INTENTION* ………………....is 2 or 4 or 5	An Internet banking user
6.	If *PERCEIVED EASE OF USE* is 3 And *ATTITUDE* ………………….is 3 And *TRUST IN BANK* ……….….is 3 And *BELIEFS* …………………….is 3 And *SUBJECTIVE NORMS* ……..is 3 or 4 And *INTENTION* ………………....is 2 or 4 or 5	Not an Internet banking user
7.	If *TRUST IN BANK* …………….is 2 or 4 And *BELIEFS* ……………….….is 3 or 4 And *PERCEIVED EASE OF USE* is 3 or 4 And *SUBJECTIVE NORMS* ……..is 2 And *INTENTION* ……………….....is 2 or 4 or 5 And *ATTITUDE* ……………….….is 3 or 4	An Internet banking user
8.	If *TRUST IN BANK* ……………is 3 And *BELIEFS* …………………….is 3 or 4 And *PERCEIVED EASE OF USE* is 2 or 3 And *SUBJECTIVE NORMS* ……..is 2 And *INTENTION* ………………....is 2 or 4 or 5 And *ATTITUDE* ……….……….is 3 or 4	Not an Internet banking user
9.	If *BELIEFS* …………………….is 2 And *PERCEIVED EASE OF USE* is 2 or 3 And *SUBJECTIVE NORMS* …….is 2 And *INTENTION* ………………...is 2 or 4 or 5 And *ATTITUDE* ………………….is 3 or 4	Not an Internet banking user
10.	If *PERCEIVED EASE OF USE* is 4 And *SUBJECTIVE NORMS* …….is 2 And *INTENTION* ………………..is 2 or 4 or 5 And *ATTITUDE* ………………….is 3 or 4	Not an Internet banking user

Node	Rule(s) Framed	Result
11.	If *SUBJECTIVE NORMS* ...…...is 3 or 4 And *PERCEIVED USEFULNESS* is 4 And *ATTITUDE* ………………..….is 3 And *TRUST IN BANK* ………..….is 2 or 3 And *INTENTION* ………………...is 3	An Internet banking user
12.	If *SECURITY* ………………..…..is 3 And *SUBJECTIVE NORMS* ……..is 2 And *PERCEIVED USEFULNESS* is 4 And *ATTITUDE* ………………..….is 3 And *TRUST IN BANK* …………...is 2 or 3 And *INTENTION* ………………...is 3	An Internet banking user
13.	If *SECURITY* …………………...is 4 And *SUBJECTIVE NORMS* ……..is 2 And *PERCEIVED USEFULNESS* is 4 And *ATTITUDE* ………………….is 3 And *TRUST IN BANK* …………...is 2 or 3 And *INTENTION* ………………...is 3	Not an Internet banking user
14.	If *PERCEIVED USEFULNESS* is 3 or 5 And *ATTITUDE* …………………..is 3 And *TRUST IN BANK* ……………is 2 or 3 And *INTENTION* ………………...is 3	Not an Internet banking user
15.	If *SUBJECTIVE NORMS* …….is 4 And *ATTITUDE* ………………….is 4 And *TRUST IN BANK* …………...is 2 or 3 And *INTENTION* ………………...is 3	An Internet banking user
16.	If *SUBJECTIVE NORMS* ……..is 2 or 3 And *ATTITUDE* ………………….is 4 And *TRUST IN BANK* …………...is 2 or 3 And *INTENTION* ………………...is 3	Not an Internet banking user
17.	If *TRUST IN BANK* …………...is 4 And *INTENTION* ………………...is 3 And *ATTITUDE* ………………….is 3 or 4	Not an Internet banking user

Note: 5 → Very Strong; 4 → Strong; 3 → Moderate; 2 → Low and 1 → Very Low

Logistic Regression Results

Table 4 gives the classification table for the test data with an overall percentage of accuracy as 100% meaning that it gives the best accuracy for the empirical data collected. It classified 19 upon 19 non-users of Internet banking and 31 upon 31 as Internet banking users. For both the cases the cut off value is fixed at 0.500.

Table 4: Classification Table for the Test Data

Classification Table[a]

	Observed		Predicted TARGET 0	Predicted TARGET 1	Percentage Correct
Step 1	TARGET	0	19	0	100.0
		1	0	31	100.0
	Overall Percentage				100.0

a. The cut value is .500

SVM Results

Neucom's SVM (Support Vector Machines, (http://www.research.microsoft.com/users/jplatt/svm.html) was used to analyze the data set with two available SVM kernels namely Linear kernel and Polynomial kernel. For Linear kernel, we obtained 86.09% accuracy for train data and 70% accuracy for test data. For Polynomial kernel, we got 80% accuracy for train data and 78% accuracy for test data. The accuracies obtained using Polynomial kernel for the train and test data sets are very good at 80% and 78% respectively using the Neucom's SVM.

Neural Network Results

The current results were obtained using the back propagation algorithm implemented in Business Data Miner (BD Miner, 2005) . Here, we trained neural network first and then tested for various desired values by changing the values of Learning rate, Momentum and the Hidden Nodes. Usually, the values for the Learning rate and the Momentum vary from 0 to 1. So, the values are changed and tested for set of values {0, 0.1, 0.2, 0.3, …….0.9, 1.0}. The Hidden Node values were changed from 5 to 20 each time. Finally, the procedure was followed keeping any two parameters constant and changing the other at a time. For example, if the Learning Rate and Momentum are fixed at 0.1, then the number of Hidden Nodes will be changed. The default values for the Learning rate and Momentum are 0.1 and for Hidden nodes are 5. The percentage success for the train data obtained keeping the learning rate and momentum constant at 0.1 and 13 hidden nodes was 85.22%. But surprisingly, the percentage of success for the test data is 100%.

Figure 1 gives the Percentage of Success as 100% at Learning Rates = {0.0; 0.2; 0.3; 0.6; 0.7; 0.8; 0.9; 1.0} keeping Momentum=0.1 as constant with 5 Hidden Nodes

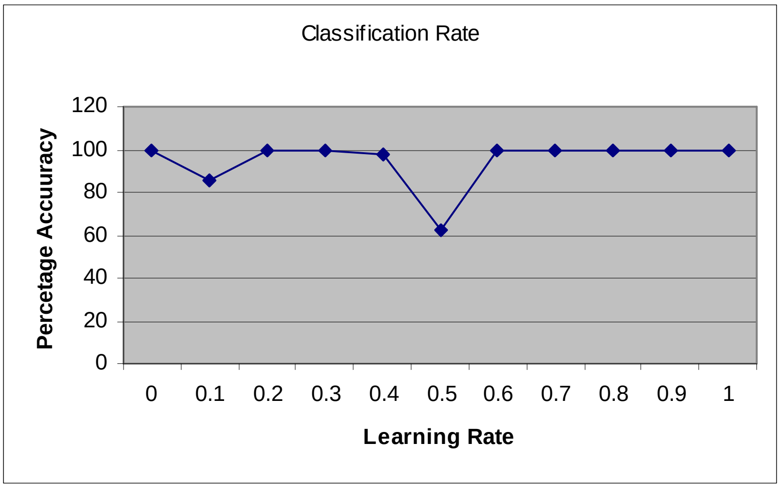

Figure 1: Classification Rate for Learning Rates

Figure 2 gives the Percentage of Success at 100% at Momentum = {0.1; 0.2; 0.3; 0.4; 0.5; 0.7; 0.8; 0.9; 1.0} keeping Learning Rate=0.1 at 5 Hidden Nodes.

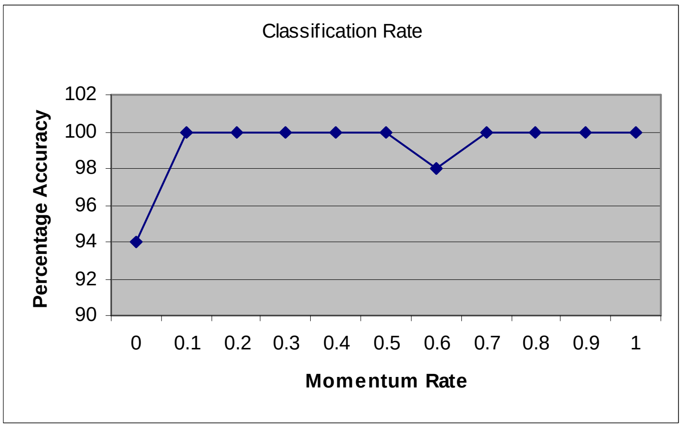

Figure 2: Classification Rate for Momentum Rates

Figure 3 gives the Percentage of Success is 86.67% at the following Hidden Nodes = {5; 6; 8; 10 12; 19; 20} keeping Momentum=0.1 and Learning Rate = 0.1 as constant.

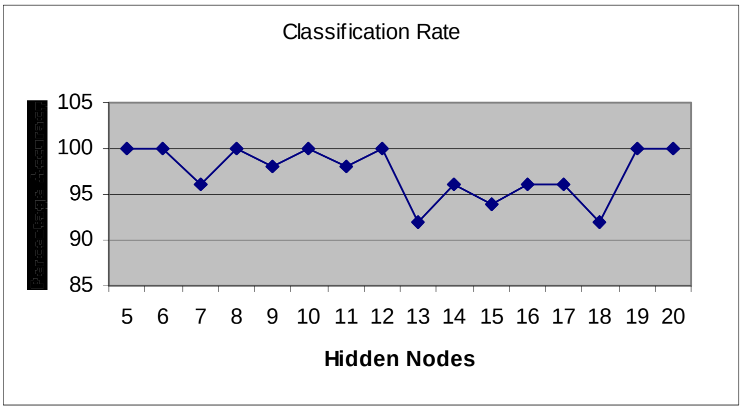

Figure 3: Classification Rate for Hidden Nodes

The results from Neural Networks gave 100% accuracy in classifying the test data on the Internet banking.

CONCLUSIONS

This research work on profiling Internet banking users was successful with the use of selected intelligent techniques viz. CART, Logistic Regression, SVM and Neural Networks. This work ranked variables according to the level of their influence on the usage of Internet banking using CART. The variables are ranked as Intention, Beliefs, Subjective Norms, Trust in the Bank, Attitude, Perceived Usefulness, Security and Perceived ease of use in that order. These results may probably vary when more empirical data is collected. While all the prediction models yielded an overall percentage prediction average, CART gave us variable importance factors and a set of “if-then” rules that can be used to profile both users and non-users of Internet banking. The Logistic Regression and the Neural Networks gave 100% overall accuracies for the test data. The CART and the SVM gave 70% and 78% accuracies respectively. Although Logistic Regression, SVM and Neural Networks gave better accuracies, CART is still chosen as the best intelligent method because it yielded other by-products like if-then classification rules. The classification rules provided here can be used to target potential customers of Internet banking and attract them by offering appropriate incentives.

REFERENCES

A Neuro-computing Decision Support Environment (online) (cited 11 May, 2005) Available from <http://www.theneucom.com>.

Araujo, I. and Araujo, I. (2003) 'Developing trust in internet commerce', *Proceedings of the Centre for Advanced Studies Conference on Collaborative Research*, October 6-9, Toronto, Ontario, pp. 1– 15.

Babbie, E.R. (1994) *The Practice of Social Research*, (7th Edn.) Publishing Company, Belmont, California, Wadsworth.

BDMiner – An in-house Data Mining tool developed at IDRBT, Hyderabad, 2005.

Breimanm, L. Friedman, J., Olshen R., and Stone C (1984) Classification and Regression Trees, Wadsworth, Pacific Grove.

CART 5.0, Evaluation Copy, Salfords Systems Inc., San Diego, California (online) (cited 11 May, 2005), Available from <http://www.salford-systems.com/>.

Castelfranchi, C. & Tan, Y.H. (2001) 'The role of trust and deception in virtual societies', *Proceedings of the 34th Hawaii International Conference on System Sciences (HICSS-34), January 3-6, Island of Maui***,** vol. 7, pp.7011.

Garson, G.D. T*opics in Multivariate Analysis*, (online) (cited18 May, 2005) Available from <http://ertr.tamu.edu/showlink.cfm?linkid=813&catid=24>.

Garson, G.D. (1998) *Logistic regression*, (online) (cited 11 May, 2005) Available from <http://www2.chass.ncsu.edu/garson/pa765 logistic.htm>.

Gefen, D., Karahanna, E and Straub, D.W. (2003) 'Inexperience and experience with online stores: the importance of TAM and trust', *IEEE Transactions on Engineering Management,* 50:3, 307-321.

Hendersona, R. and Divett, M.J. (2003) 'Perceived usefulness, ease of use and electronic supermarket use', *International Journal of Human-Computer Studies***,** 59:3, 383-395.

Kim, K. and Prabhakar, B. 'Initial Trust, perceived risk and the adoption of internet banking', (online) (cited 18 July, 2005, Available from: <http://www.nr.no/~abie/ Papers/ 00RIP11.pdf.>.

Lin, C.S. and Wu, S. (2002) 'Exploring the impact of online service quality on portal site usage', *Proceedings of the 35th Hawaii International Conference on System Sciences (HICSS-35),* January 7-10, Island of Hawaii, Vol. 7, p.206.

Loch, K.D., Straub, D.W. and Kamel, S. (2003) 'Diffusing the Internet in the arab world: the role of social norms and technological culturation', *IEEE Transactions on Engineering Management*, 50:1, 45-63.

Malhotra, Y. and Galleta, D.F. (1999) 'Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation', *Proceedings of the 32nd Hawaii International Conference on System Sciences,* January 5-8, IEEE, pp. 1006.

Mathieson (1991) 'Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior', *Information Systems Research*, 2, 173-191.

Noteberg, A.E., Christiaanse & Wallage, P. 2003, 'Consumer trust in electronic channels', *e-Service Journal*, 2:2, 46-68.

Rumelhart, D.E. and McClelland, JL (1986), *Parallel D*istributed *Processing: Exploration in the Microstructure of Cognition*, vol. 1 & 2, Cambridge, MA., MIT Press.

Support Vector Machines, (online) (cited 21 May, 2005), Available from: <http://www.research.microsoft.com/users/jplatt/svm.html>.

Teltzrow, M. Gunther, O. and Pohle, C. (2003) 'Analyzing consumer behavior at retailers with hybrid distribution channels – A trust perspective', *Proceedings of the 5th International Conference on Electronic Commerce ACM, ICE - Pittsburgh*, pp. 422 – 428.

V. Ravi, is Assistant Professor, Institute for Development and Research in Banking Technology, Castle Hills, Masab Tank, Hyderabad, Andhra Pradesh.

Mahil Carr, is Assistant Professor, Institute for Development and Research in Banking Technology, Castle, Masab Tank, Hyderabad, Andhra Pradesh.

N. Vidya Sagar, Institute for Development and Research in Banking Technology, Castle Hills, Masab Tank, Hyderabad, Andhra Pradesh.