

# Comparative Evaluation of Predictive Modeling Techniques on Credit Card Data

Ravinder Singh and Rinkle Rani Aggarwal

**Abstract**—Credit Scoring studies are very important for any financial house. Both traditional statistical and modern data mining/machine learning tools have been evaluated in the credit scoring problem. But very few of the studies facilitate the comparison of majority of the commonly employed tools in single comprehensive study. All the tools such as LDA (Linear Discriminant Analysis), SVM (support vector machines), Kernel density estimation, LR (logistic regression), GP (genetic programming), K neighborhood, which are available in SAS enterprise miner 6.2. The results revealed that support vector machine and genetic programming are superior tools for the purpose of classifying the loan applicant as their misclassification rates were least as compared to others.

**Index Terms**—Credit risk, credit scoring, machine learning, predictive modeling

## I. INTRODUCTION

Financial crimes are increasing at enormous rate every year and financial institutions must adopt methods to safeguard their reputation and their customers. The use of statistical methods to address these problems faces many challenges [9]. Interestingly financial crimes are uncommon events that lead to extreme class imbalances. Criminals deliberately attempt to hide their usage patterns and quickly change their strategies over time, making the process of fraud detection sophisticated. Also sometimes legal constraints and investigations delays make it impossible to actually verify suspected crimes in a timely manner.

Credit scoring methods are statistical tools employed by various banks and other financial institutions, marketing and advertisement companies to estimate the probability whether the loan applicant could be categorized as potential defaulter or not [20]. Basically, credit scoring aims to classify the dependent variable with respect to the response variables. Banks collect the information about the applicant from various sources such as historical data, questionnaires and interviews. This aims at collecting all demographic details such as income, age, sex, type of loan, nationality, job, and income pertaining to the applicant.

The accurate prediction of consumer credit risk is indispensable for lending organizations. Credit scoring is a common technique that helps financial institutions to evaluate the likelihood for a credit applicant to be a potential defaulter on the financial obligation and decide

whether to grant credit or not. The precise judgment of the creditworthiness of applicants allows financial institutions to increase the volume of granted credit while minimizing possible losses. The increasing number of potential applicants has driven the need for the development of sophisticated fraud detection techniques that automate the credit approval procedure.

Earlier credit scoring was restricted to statistical techniques such as discrimination between several groups in a data sample. History reveals that credit scoring came into existence in early thirties of the nineteenth century when some financial institution decided to classify their applicants with respect to default status. In late 1950's the first automated system was used to develop predictive model of the applicants based on their historical data.

After the card holders default, several data collection procedures are undertaken. These procedures are expensive relative to the size of most loans. They are often useless when the card holder has gone bankrupt. Therefore, it is significant for the card issuers to identify card holder type at early stage in order to minimize lending to risky customers before the default occurs. It means that it becomes necessary to maximize the 'True positives' (TP). The true positive rate (TRP) is known as sensitivity and the true negative rate (TNR) is sometimes called specificity.

## II. DATA MINING/PREDICTIVE MODELING TOOLS

The classification techniques can be classified into parametric and non-parametric problems. Traditional statistical methods are parametric in nature. Parametric methods are based upon the assumptions of normally distributed population and estimate the parameters of the distributions to solve the problem [5]. Non-parametric methods, on the other hand, make no assumptions about the specific distributions involved, and are therefore distribution-free [15]. Parametric techniques encompass linear regression, generalized linear regression, logistic regression and discriminant analysis. On the other hand neural networks, decision tree, genetic algorithms and k-nearest neighbor techniques fall in to the category of non-parametric methods.

### A. Linear Discriminant Analysis (LDA)

It is a popular classification technique was developed by Fisher. It focused on optimally separating two groups by on the basis of a function, which provides the maximum distance between their respective means [8]. The linear discriminant function used to map input values into output or target variable is:

Manuscript received April 12, 2011; revised June, 22, 2011  
Authors are with Department of Computer Science & Engineering,  
Thapar University, Patiala, 47004 India (email:  
ravinder1bhatia@gmail.com, raggarwal@thapar.edu).

$$Y_i = a_1 X_1 + a_2 X_2 + \dots + a_n X_n \quad (1)$$

where  $Y$  represents target variable and  $a_1, a_2, \dots, a_n$  indicates the attributes or the input values. In more technical terms, the discriminant analysis problem can be stated as: let P1 (good) and P2 (bad) denote the two mutually exclusive strata of the population and let a customer coming either from P1 or from P2 be represented by the random vector  $X=(X_1, X_2, \dots, X_p)$  of explanatory variables. It is assumed here that the rule of discrimination is based on a sample containing both response and explanatory variables randomly chosen from the entire population [2].

The advantages of LDA are its simplicity and that it can be easily estimated. Due to parametric nature, it requires the assumption of normality to be maintained. However, Reichert et al., (1983) proposed that the non-normality of credit information does not provide a limitation for the empirical utilization of the method. However some of the problem observed

- Many times, due to the qualitative nature of the data may not follow normal distribution, so the parametric nature of discriminant analysis is violated.
- The linear discriminant model assumes equality between the variance covariance matrices across all groups. Failure to adjust for inequality in the variance covariance matrices makes inefficient use of the available credit information and can severely distort the classification results.
- Prior probabilities should be incorporated to adjust for classification function adopted by the model [18]. Failure to incorporate the prior probabilities in the classification process will not minimize prediction errors and frequently has severe disruptive effect between group classification results.

### B. Logistic Regression (LR)

Logistic regression model is a commonly employed statistical tool that predicts the membership of an object among two or more groups. It is constrained by the fact that the target or the response variable should be binary in nature [3], [12]. It provides a powerful technique analogous to multiple regression and anova for continuous responses. Reference 8 shows that the likelihood function of mutually independent variables  $y_1, y_2, y_3, \dots, y_n$  with binary responses is a member of exponential family with  $(\log(\prod_1/(1-\prod_1)), \dots, \log(\prod_n/(1-\prod_n)))$  as a canonical parameter. In technical terms the relationship between a canonical parameter and the vector of explanatory variables  $x$  is stated as:

$$\text{Log}(\prod_i(1-\prod_i)=x\beta_i \quad (2)$$

This linear relationship between the logarithm of odds and the vector of explanatory variables results in a nonlinear relationship between the probability of  $y$  equals 1 and the vector of explanatory variables [8].

$$\prod_i \exp(x\beta_i / (1 + \exp(x\beta_i)) \quad (3)$$

Logistic regression is best suited for dealing with classification problems since the computed outputs can be expressed in terms of probabilities.

### C. Decision Trees

They are also one of the commonly used tools for performing classification tasks. Decision tree learning is a method for approximating discrete valued target function, in which the learned function is represented by a decision tree[15]. The classification process begins by sorting down the tree from the root node, which provides the classification of the instance. Using appropriate search strategies, decision trees explore the attributes or the input values to generate understandable rules-like relationships with respect to the target variable. Search heuristics basically adopt recursive partitioning algorithms to split the original data into finer and finer subsets, or clusters. The objective is to find the optimum number of splits and determine the node to maximize the information gain. The algorithm tries to minimize the number of splits, as the output would be easier to understand. The most predictive variable is placed at the top node of the tree. The operation of decision trees is based on the ID3 or C4.5 algorithms. The algorithms make the clusters at the node gradually purer by progressively reducing disorder (impurity) in the original data set. Disorder and impurity can be measured by the well-established measures of entropy and information gain borrowed from information theory.

The central choice pertaining to tree splitting algorithm is selecting which attribute to test at each node of the tree. A quantitative measure called information gain measures the extent how well a given attribute separates the training examples according to their target classification.

One of the most significant advantages of decision trees is the fact that knowledge can be extracted and represented in the form of classification (if-then) rules. Each rule represents a unique path from the root to each leaf. Over fitting the training data is an important issue in decision tree learning. Because the training examples are only a sample space of all possible instances it is possible to add branches to the tree that improve performance on the training examples while decreasing performance instances outside this set.

### D. Support Vector Machines (SVM)

Support vector machines implement the principle of structure risk minimization by constructing an optimal hyper plane [19]. In case of two group classification, the equation of hyper plane for a given input vector  $w$  is given by:

$$wx+b=0 \quad (4)$$

At the same time the input vector  $w$  should satisfy

$$Y_i [(wx+ b)]-1 + \epsilon \quad (5)$$

are ideally suited for data sets characterized by large feature space.

SVM's do well in classifying nonlinear separable groups. They do not require large training datasets and training converges to a unique global solution. These features make SVM's suitable for applications such as credit card fraud detection [18], [22]. At the same time they suffer from many disadvantages. They are relatively more expensive to implement. The results are not easily interpretable, and many parameters of the algorithm must be specified, e.g.,

the type of kernel function used to transform the feature space and all its parameters.

#### E. Kernel Discrimination

A common approach to non parametric discriminant analysis with continuous or discrete nature of predictor variables is to substitute non parametric estimates of group conditional densities in the definition of bayes rule [14]. For a continuous p-dimensional feature vector x, a nonparametric estimate of the ith group conditional density  $f_i(x)$  provided by kernel method is:

$$F_i(x) = n_i^{-1} h_i^{-p} \sum k_p((x-x_{ij})/h_i) \quad (6)$$

where  $k_p$  is a kernel function and  $h_i$  is a smoothing parameter.

The simplest class of kernels consists of probability density function that satisfies:

$$\begin{aligned} K(x) &> 0 \\ \int_{\mathbb{R}^p} k(x) dx &= 1 \end{aligned} \quad (7)$$

#### F. Artificial Neural Network(ANN)

Neural network learning methods provide a robust approach to approximating real valued, discrete-valued and vector-valued target functions [15]. The motivation behind the neural networks comes from the fact that biological learning systems are built of very complex networks of interconnected neurons [3],[4]. Typically, feed-forward networks with only three layers (input, hidden, and output layers) are used in fraud detection. The input to the neural network is the vector of features. The output layer gives the probability of the activity being criminal, which is used as a suspicion score. Back propagation is commonly used algorithm for training neural networks. The weights are initialized with random values, which are then changed in the direction that minimizes training error [21].

Neural networks are attractive tools in financial crime detection for a few reasons [19].

- First, three-layer nets were shown to be capable of dealing with the highly skewed data that arise in this application.
  - Second, once they are trained, they can analyze new data very quickly, an attribute that is necessary when trying to catch fraudulent transactions in real time.
- However, neural networks also suffer from drawbacks:
- One major demerit is the need to select and adjust the structure of the network.
  - Also the choice of the number of hidden states must be specified be made to optimize learning and generalization.
  - Further, the performance of the classifier is very sensitive to the input vectors chosen, so significant attribute selection and preprocessing are necessary.

A particular type of ANN called multilayer perceptron (MLP) is especially suitable for classification. It consists of one input layer, one or more hidden layers and one output layer, each consisting of several neurons. Each neuron processes its inputs and generates one output value that is transmitted to the neurons in the subsequent layer [14]. Each

neuron in the input layer (indexed  $i = 1; 2; \dots; n$ ) delivers the value of one predictor (or the characteristics) from vector x. When considering default/non-default behaviour, one output neuron is satisfactory.

#### G. Instance-Based Learning Algorithms

They are characterized as lazy-learning algorithms. Learning in these algorithms consists of simply storing the presented training data. When a new query instance is encountered a set of similar related instances is retrieved from memory and used to classify the new query instance [15]. Such predictive tools can construct different approximation to the target function for each distinct query instance that must be classified.

Lazy-learning algorithms require less computation time during the training phase as compared to their counterparts such as decision trees, neural and Bayesian network but more computation time during the classification process [7]. One of the most straightforward instance-based learning algorithms is the nearest neighbor algorithm. The k-nearest neighbor classifier serves as an example of a non-parametric statistical approach. When given an unknown case, a K-NN classifier searches the pattern space for the k training cases that are similar to unknown cases. These k training cases are the "K-nearest neighbors" of the unknown cases.

K-NN can also be useful when the dependent variable takes more than two values: high risk, medium risk and low risk. Generating the nearest-neighbor rule is very computationally intensive ( $O(n^2)$  process) and can take considerable computational time for large datasets [9]. K-NN also requires an equal number of good and bad sample cases for better performance.

The choice of k also affects the performance of the k-NN algorithm [15]. This can be determined experimentally. Starting with  $k=1$ , we use a test case to estimate the error rate of the classifier. This process is repeated each time by incrementing k to allow for one more neighbor. The K-value that gives the minimum error rate may be selected. In general, larger the number of training samples is, the larger the value of k will be.

One major weakness associated with instance based algorithms is that cost of classifying the new instances can be high. The reason for this is that all the computation is performed at classification time rather than when the training samples are first encountered. Second weakness pertaining to instance based techniques is that they typically consider all the attributes of the instances when attempting to retrieve similar training examples from memory.

#### H. Genetic Programming

Genetic programming (GP) is a machine learning tool motivated by analogy to biological evolution [15], [19]. Basically genetic programming searches for best candidate hypothesis. The best hypothesis refers to the one which optimizes a predefined numerical measure for the problem at hand, called the hypothesis fitness. On each iteration, all members of the population are evaluated according to fitness function. A new set is generated by probabilistically selecting the fit individuals from the current population.

A GP algorithm works on a population of individuals, each of which represents a potential solution to a problem.

In order to solve a problem using GP it is necessary to specify the following:

- The terminal set: A set of input variables or constants.
- The function set: A set of domain specific functions used in conjunction with the terminal set to construct potential solutions to a given problem. For symbolic regression this could consist of a set of basic mathematical functions, while Boolean and conditional operators could be included for classification problems.
- The fitness function: Fitness is a numeric value assigned to each member of a population to provide a measure of the appropriateness of a solution to the problem in question.
- The termination criterion: This is generally a predefined number of generations or an error tolerance on the fitness.

### III. LITERATURE SURVEY

Most of the attention in the literature focuses on fraud, particularly on credit card fraud. While there is some similarity among the various statistical methods, we can classify the predictive techniques into two broad classes as supervised learning and unsupervised learning. The performance of the different techniques is dependent on both quality and quantity of data.

One significant aspect of building credit scoring models is the selection of appropriate classification tool. This arises the need of an empirical comparison of various classification techniques in credit scoring. There are many traditional parametric classification techniques that have been applied to this type of problem, for example, discriminant analysis, linear regression, and logistic regression. Recently soft computing techniques such as, decision trees, neural networks, have been applied extensively.

Abdou compared two credit scoring neural architectures, probabilistic NN (PNN) and multilayer perceptron (MLP), with discriminant analysis, probit analysis and logistic regression [7]. Their results demonstrated that PNN and MLP perform better than other models. In the same domain, compared ANN with decision tree analysis and logistic regression for credit risk classification and they concluded that decision tree technique performs better than ANN (with 74.2% of accuracy) and ANN (with 73.4% of accuracy) performs better than logistic regression (with 71.1% of accuracy).

Beahamasy carried out comparative survey of various machine learning tools such as artificial neural network, expert systems and hybrid intelligent systems in finance applications [1]. The comparison was made pertaining to credit evaluation, portfolio management, and financial prediction and planning. The results depicted that modern machine learning tools are superior in terms of their classification rates as compared to traditional statistical tools.

In the same year, Angelini developed two NN systems, one with a standard feed-forward network and other one with special purpose architecture. The system was validated with real-world data, obtained from Italian small businesses. They show that NNs can be strong in learning and estimating the default tendency of a borrower if careful data

analysis, data preprocessing and proper training are performed.

West Compared the accuracy of credit scoring of five ANN models: multilayer perceptron, mixture-of-experts, radial basis function, learning vector quantization and fuzzy adaptive resonance [6]. His study was based on two real world data sets : Australian and German. He employed 10 fold cross validation for enhancing his predictive power. He reported both good credit and bad credit rates. He benchmarked the results against five other traditional methods including linear discriminant analysis, logistic regression, k nearest neighbor, kernel density estimation and decision trees. Results demonstrated that the multilayer perceptron may not be the most accurate ANN model and that both the combination-of-experts and radial basis function NN models should be considered for credit scoring applications. Also, between traditional methods, logistic regression is more accurate method and more accurate than NN models in average case.

Zhang and Huang compared three data mining techniques namely back propagation, genetic programming and SVM for evaluating classification rates pertaining to credit scoring [5]. Initially they calculated the accuracy of each technique on German and Australian credit card data sets, which are easily available from machine repository of university of California. They observed that back propagation and genetic programming are better on average than svm, but the classification accuracy of the latter was more stable as in each run the result produced was same. The results obtained from German credit card data were not that good as it contained large number of optimistic cases ( who paid back their loan). They also proposed a combined model of these three techniques, which yielded better results as compared to individual methods in isolation [15]. This combined model was based on the concept of majority voting. For one applicant, If there are two or three models with same classification rate A, then the customer is classified as case A. Otherwise, the classification result of the customer is the same as that of the model with highest accuracy.

Yeh and Yang proposed an optimal credit scoring model to reassess the default risk of credit card holders for credit

The suggested model assessed the credit risk in various stages: In the very first stage different training data subsets were generated by employing a data sampling technique bagging pertaining to data shortage. In the subsequent stage, these training subsets were used to create the different NN models [21]. In the third stage, the generated neural network models were again trained with different training datasets and accordingly the classification score and reliability value of neural classifier were estimated. In the fourth stage, feature selection was performed using appropriate technique. In the fifth stage, the reliability values of the selected NN models will be scaled into a unit interval by logistic transformation. Finally the selected NN ensemble members are joined to obtain final classification result by means of reliability measurement. The authors also haven used two credit datasets to verify the effectiveness of their proposed model. : One from was Japanese credit card screening and other was pertaining to UK corporation database. They compared their ensemble method with two other categories

of classification tools: single methods such as logistic regression, support vector machines and hybrid methods such as neuro-fuzzy systems and neuro-support vector machine systems.

Thomas explored the techniques for predicting the financial risk involved in lending loans to customers [4]. Hardgrave presented a comprehensive and systematic approach to developing an optimal architecture of a neural network model for evaluating the creditworthiness of loan applicants [13]. Zurada investigated data mining techniques for loan granting decisions and predicted defaulters [10].

Jun used support vector machine to classify the loan applicants. Firstly methods of classification using SVM were introduced and then subsequently, index entropy weight screening methods were employed to screen the index of a bank credit risk. The study also highlighted the role of SVM in multiclass classification. It also proposed an incremental learning method to study the unknown cases.

Khandani and kim applied machine learning techniques for the purpose of development of non parametric forecasting models of consumer credit risk [2]. A credit beraeau data of the time span from January 2005 to April 2009 was used to develop the various models. Moreover, the time-series patterns of estimated delinquency rates from the developed model revealed that over the course of the recent financial crisis, credit risk analytics may have important applications in forecasting systemic risk.

Hajek and Ojek explored the modelling possibilities of kernel-based approaches to a complex real-world problem, of corporate and municipal credit rating classification [16]. The model design incorporated various factors such as data pre-processing, the labelling of individual parameter vectors using expert knowledge, the design of various support vector machines with supervised learning as well as kernel-based approaches with semi-supervised learning. The results revealed that the rating classes assigned to bond issuers can be classified with high classification accuracy using a limited subset of input variables. This holds true for kernel-based approaches with both supervised and semi-supervised learning.

Belloite and Crook evaluated support vector machines on a large credit database [18]. They found that the SVM is competitive and can be used as the basis of feature selection methods to identify most significant features which are most effective in determining the probability of default in the data set.

The above literature review reveals that most of the studies have evaluated the classification rates of various data mining/predictive modeling tools. But very few studies have evaluated eight-nine data mining tools simultaneously for their comparison on multiple parameters such as misclassification rate, positive predictive power, negative predictive power, sensitivity, f-measure etc. In order to model the real word scenario in credit scoring problem, a replicated study that would evaluate the efficacy of eight tools needs to be performed.

#### IV. DATA SET EMPLOYED

This study is carried out on Australian credit card data. The data set is available from machine learning repository

of University of California[14]. It contains 690 observations with 14 predictor variables and one response variable. These 14 predictor variables are divided into 8 categorical and six continuous variables. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. This dataset contains good mixture of both continuous and nominal attributes.

#### V. EMPIRICAL ANALYSIS

Comparative study was carried out on Australian data set. The eight techniques are assessed using the k-fold cross validation. But unfortunately since kernel density estimation and k-neighbourhood does not directly facilitate 10-fold cross validation, the parameters used to judge the efficacy of these techniques could not be calculated. So leave one cross validation approach is adopted in order to allow the comparison of more tools. Since leave one out cross validation is a special case of k-fold cross-validation when in the latter one the value of k is set to number of original cases. So the techniques are also assessed using 688 fold cross validation. Kernel density estimation and k-neighbourhood does not support k-fold cross validation. So the corresponding results can not be evaluated for them. All these parametric and non-parametric techniques are evaluated on parameters such as misclassification rate, positive predictive value(ppu), negative predictive value(npu), specificity and sensitivity. In case of LDA and logistic regression equal prior probability and misclassification costs were assumed. Support vector machine employed RBF kernel for non-linear mapping and produced 517 support vectors. A mlp model consisting of three layers and using logistic function for hidden and output layer is used. In case of non parametric techniques the value of K was adjusted by trying out various values of k.

TABLE1: COMPARATIVE EVALUATION USING 10 FOLD CROSS VALIDATION

	Misclassification rate	PPU	NPU	Sensitivity	Specificity
LDA	15.23	79.82	89.49	87.91	87.25
LR	13.7	78.80	76.80		
MLP	14.07	83.64	87.83	84.97	86.68
Decision Tree	14.36	78.67	93.29	92.80	79.90
GP	15.23	79.47	89.49	88.56	81.72
SVM	13.20	83.49	89.67	87.58	86.16
Kernel density estimation	*	*	*	*	*
Kneighbourhood	*	*	*	*	*

\* option not supported

#### VI. RESULTS/SUMMARY

The tables 1 and 2 depict that in case of both 10 fold and 688 fold crossvalidation, genetic programming yielded highest misclassification rates for classifying good cases in to bad and bad cases in to good. In case of leave-one –out

crossvalidation, MLP model emerged to be clear winner as it yielded the lower classification rate. In general Soft computing techniques such as decision tree and neural network were found to be superior as compared to statistical tools such as LDA and LR. Non parametric techniques did not performed well. The increase in value of k led to decrease in misclassification rate.

TABLE II: COMPARATIVE EVALUATION USING 688 FOLD CROSS VALIDATION OR LEAVE ONE OUT CROSSVALIDATION

	Misclassification rate	PPU	NP U	Sensitivity	Specificity
LDA	15.96	84.78	9511	94.44	86.42
LR	13.7	82.82	89.72	*	*
MLP	12.91	85.11	88.68	85.95	87.99
Decision tree	14.36	78.67	79.90	92.81	79.90
GP	16.54	77.76	89.91	88.89	79.11
SVM	13.49	82.97	89.23	87.58	85.64
Kernel density estimation	19.72	80.08	87.89	82.8	87.8
Kneighbour hood	18.23	79.08	85.02	84.50	82.20

\*The parameters cannot be calculated

REFERENCES

[1] A. Baharamzee. "Comparative survey of artificial Intelligence applications: artificial neural networks, expert systems, hybrid intelligence systems" Springer 2010

[2] A. E. Kim and A. J. Lo "Consumer credit-risk models via machine-learning algorithms" Journal of Banking and Finance 34 (11), pp. 2767-2787 2010

[3] C. Bolton. "Logistic regression and its application in credit Scoring" University of Pretoria 2009

[4] C. Thomas. "A survey of credit and behavioral scoring: forecasting financial risks of lending to customers," International Computer Journal of Forecasting" vol. 16, no. 2, pp. 149-172, 2000

[5] D. Zhang, H. Huang, Q. Chen and Y. Jiang. "Comparison of credit scoring models" Third international conference of Natural Computation" 2007

[6] D. West. "Neural network credit scoring model" Computational Operation Research vol. 27 pp. 1131-1152, 2000

[7] H Abdou and J Pointon , E-Masry. "Neural nets versus conventional techniques in credit scoring in Egyptian banking." Expert System Application " vol. 35, no. 3, pp. 1275- 1292 , 2008

[8] H. Brighton and C. Mellish. "Advances in instance selection for instance-based learning algorithms." Data Mining and Knowledge Discovery vol. 6, pp. 153-172, 2002

[9] J. Hand and W. Henley. "Statistical classification methods in consumer credit scoring," Computer Journal of the Royal Statistical Society Series a Statistics in Society" vol. 160, no. 3, pp. 523-541, 1997.

[10] J. Zurada. and M. Zurada. "How Secure are Good Loans: Validating Loan-Granting Decisions and Predicting Default Rates on Consumer

Loans" The Review of Business Information Systems" vol. 6, no 3, pp. 65-83, 2002

[11] K. Cuh, W.C Tan, C.P Goh (2004). "Credit scoring using data mining techniques" Singapore Manage Rev 26(2):25-47

[12] Lee TH, Jung SC. "Forecasting credit worthiness: logistic vs. artificial neural net." Journal of Business Forecasting vol. 18, no 4, pp. 28-30, 1999/2000

[13] L.W. Glorfeld, B.C. Hardgrave, "An improved Method for developing neural networks: the Case of evaluating commercial loan credit worthiness," computer & operations Research, vol. 23, no. 10, pp. 933-944, 1996.

[14] Machine Learning repository, " Machine Learning Datasets", <http://www.uci.ml/>

[15] M.J.A, Berry, and Linoff, G.S." Data Mining Techniques for Marketing, Sales, and Customer Support", John Wiley & Sons, Inc, 1997

[16] P. Hajek, V. Olej "Credit Rating Modelling by Kernel-Based Approaches with Supervised and Semi-Supervised Learning " Neural Computing and Applications , pp. 1-13 2010

[17] SAS Institute, "Data Mining Using Enterprise Miner Software: A Case Study Approach", 1st edition, <http://www.sas.com>

[18] T. Bellotti, J. Crook, "Support Vector Machine For credit Scoring and Discovery of Significant Features. " Expert Systems with Applications vol36 no 2 , pp. 3302-3308 2009

[19] T. M. Mitchell "machine learning" International edition Mcgraw Hill, 1997

[20] W. E. Henley "Statistical aspects of credit scoring" Phd. Dissertation, The Open University, Milton Keynes, UK, 1995.

[21] Y. Lu, S Wang, "Credit risk assessment with a multistage neural network ensemble learning approach." Expert System Application vol. 34, no. 2, pp. 1434-1444, 2008

[22] Z Huang, H Chu, C-J Hsu, W-H Chen, Wu S " Credit rating analysis with support vector machines and neural networks: A market comparative study. Decision Support System" vol. 37, pp. 543-558, 2004



**Ravinder Singh** is a student of masters in software engineering at Thapar University Patiala. Currently he is pursuing thesis in macro designing and comparative evaluation of predictive modeling techniques in credit scoring. He has published two research posters in SAS data mining conferences M2009 and M2010. His areas of interests include predictive modeling, data mining, machine learning, credit scoring, SAS macro programming and semantic testing



**Dr. Rinkle Rani Aggarwal** is working as Assistant Professor in Computer Science and Engineering Department, Thapar University, Patiala since 2000. She has done her Post graduation from BITS, Pilani and Ph.D. from Punjabi University, Patiala in the area of parallel computing. She has more than 14 years of teaching experience. She has supervised 20 M.Tech. Dissertations and contributed 30 articles in Conferences and 21 papers in Research Journals.

Her areas of interest are Parallel Computing, Grid Computing and Algorithms. She is member of professional bodies: ISTE, IAENG, IACSIT & WASET and ICGST.