COMPARISON OF CLASSIFICATION TECHNIQUES USED FOR CREDIT RISK ASSESSMENT IN FINANCIAL MODELING

T.N.Pandey *
D.Choudhury **
A.K.Jagdev ***
S.Dehuri ****

Abstract

Credit risk refers to the risk that a borrower will default on any type of debt by failing to make payments which it is obligated to do. Assessment of Credit Risk is very important for any type of financial institution for avoiding huge amount of losses that may be associated with any type of inappropriate credit approval decision. In this paper, we are going to compare different classification techniques used for credit risk assessment such as linear discriminant analysis, logistic regression, classification and regression tree, support vector machine, neural network and genetic algorithm.

***Index Terms*—Credit Risk; LDA; LR; CART; SVM; NN; GA**

* T.N.Pandey, is with the Institute of Technical Education and Research, SOA University, Bhubaneswar, India.

** D.Choudhury, is continuing his M.Tech in the department of C.S.E, I.T.E.R, SOA University,Bhubaneswar,India.

*** A.K.Jagadev is with the Computer Science and Engg .dept.,SOA University ,Bhubaneswar, India.

**** S.Dehuri is with the Department of System Engg.., Ajou University ,Republic of Korea.

I. INTRODUCTION

Financial modeling is the task of building an abstract representation of a financial decision making situation. Financial model is designed to represent in mathematical terms the relationships among the variables of a financial problem so that it can be used to answer what if questions or makes projection. Financial modeling is a general term that means different things to different users; the reference usually relates either to accounting and corporate finance applications, or to quantitative finance applications. While there has been some debate in the industry as to the nature of financial modeling - whether it is a tradecraft, such as welding, or a science - the task of financial modeling has been gaining acceptance and rigor over the years.

Objective of Financial Modeling

1) to demonstrate the size of the market opportunity
2) to explain the business model
3) to show the path to profitability
4) to quantify the investment requirement
5) to facilitate valuation of the business

Scope of Financial Modeling

1) Risk Analysis
2) Portfolio Management
3) Profitability Analysis
4) Sales Forecasting
5) Bond Rating

II. CREDIT RISK ASSESSMENT

Credit risk refers to the risk that a borrower will default on any type of debt by failing to make payments which it is obligated to do. Assessment of Credit Risk is very important for any type of financial institution for avoiding huge amount of losses that may be associated with any type of

inappropriate credit approval decision. In the case of frequent credit decisions like thousands, financial institution will not take any judgmental decision for every individual case manually, but it will try to adopt the automated credit scoring system to easier and accelerate the decision making process. So, here comes the concept of the "Credit Scoring Model". Credit Scoring is a method of measuring the risk incorporated with a potential customer by analyzing his data. Usually, in a credit scoring system, an applicant's data are assessed and evaluated, like his financial status, preceding past payments and company background to distinguish between a "good" and a "bad" applicant [1]. It is one of the earliest financial risk management tools developed [2]. Its significance is more highlighted because of recent financial crisis.

The benefits of credit scoring involve reducing the time needed in the loan approval process, saving cost average per loan, objectivity improvement which helps lenders ensure they are applying the same criteria to all borrowers [3] and easier supervising of existing accounts [4]. Development of credit scoring was started in the 1960s [5]. It has been widely studied in the areas of artificial intelligence, machine learning, and statistics.

There are basically two types of algorithms used for credit risk assessment i.e. conventional algorithms and bio-inspired algorithms.

III. CONVENTIONAL ALGORITHMS

A. *Linear Discriminant Analysis*

LDA proposed by Fisher [22] is the first classification algorithm applied in credit scoring.LDA has been the most commonly used statistical technique in constructing credit classification model due to its simplicity. LDA attempts to find a linear combination of predictor variables to classify loaners into various groups. LDA has been regarded as a data mining technique in handling classification problems which reduces the observed variables into a smaller number of dimensions that result in decreasing the number of features to be considered by the classifiers. Rather than directly eliminating irrelevant or redundant variables from the original feature space, LDA merely transform the original variables through linear combination into a new subset of

variables. Thus, the linear methods provide a new way of understanding the data, but they are not able to reduce the number of original features. The LDA can be expressed as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (1)$$

Where β_0 is called the "intercept" and β_1, β_2, β_3 and so on are called the "regression coefficients" of X_1, X_2, X_3 respectively.

Altman [26] collected 33 bankruptcy companies and 33 contrary healthy companies to construct a LDA credit scoring model and found that the linear discriminant credit scoring model performed very well, especially in short time period. Other studies [27], [28], [29], [30] also utilized LDA to develop credit scoring models for bank and credit card sectors.

Sustersic et al. [25] state that the weakness of the linear discriminant analysis is the assumption of a linear relationship between variables, which is usually nonlinear and the sensitivity to deviations from the multivariate normality assumption.

Advantages of LDA

1) dichotomous response variable.
2) easy to calculate
3) reduced error rates

Disadvantages of LDA

1) normality assumption on variables.
2) approximately equal variances in each group.
3) assumption on equivalent correlation patterns for groups.

B. *Logistic Regression*

LR is a type of predictive model (also known as logit model) in the field of statistical learning, which is used for binary classification when the target variable is a categorical variable with two categories - for example true or false, active or inactive, success or failure, purchase product or doesn't purchase product etc. LR makes use of predictor variables; variables may be either numerical or categorical.

For example, the probability that a person has a heart attack in a specified time that might be predicted from the knowledge of person's age, sex and body mass index.

Logistic regression model is one of the most used methods in building credit scoring models. Logistic regression can fit various kinds of distribution functions such as Gamble, Poisson, and normal distributions [40]. In order to increase its accuracy and flexibility, several methods have been proposed to extend the traditional binary logistic regression model including multinomial logistic regression model [41] and logistic regression model for ordered categories [42].

Logistic regression is used extensively in the medical and social sciences as well as in marketing applications such as prediction of customer's propensity to purchase a product or cease a subscription. The response 'Y' of a subject can take one of two possible values, denoted by 1 and 0 (for example, Y=1 if a disease is present; otherwise Y=0). Let $X=(x_1, x_2,\ldots, x_n)$ be the vector of explanatory variables. The logistic regression model is used to explain the effects of the explanatory variables in the form of binary response.

$$Logit\{Pr(Y=1|x)\} = log\frac{Pr(Y=1|x)}{1-Pr(Y=1|x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (2)$$

Where β_0 is called the "intercept" and β_1, β_2, β_3 and so on are called the "regression coefficients" of x1, x2, x3 respectively.

The logistic function is given by

$$P = {}^{1}/_{1+e^{-logit(p)}} \quad (3)$$

A graph of the function is shown in Figure 1. The logistic function is useful because it can take an input any value from negative infinity to positive infinity, whereas the output is confined to values between 0 and 1.

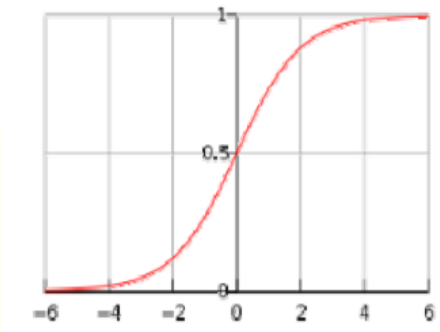

Fig.1. A graph of logistic regression function

Advantages of LR

1) Scores are interpretable in terms of log odds.
2) Constructed probabilities have chance of being meaningful.
3) It is modeled as a function directly rather than as ratio of two densities.
4) It is a good default tool to use when appropriate, especially, combined with feature creation and selection.

Disadvantages of LR

1) It invites to an over-interpretation of some parameters.
2) It requires large number of data points per predictor in order to achieve stable result.

C. *Classification and Regression Tree*

CART is developed by Breiman [21]. Classification and Regression trees (CART) is a nonparametric decision tree learning technique that produces either classification or regression trees, depending on the dependent variable. Classification trees are designed for dependent

variables that take a finite number of unordered values, with prediction error measured in terms of misclassification cost. Regression trees are for dependent variables that take continuous or ordered discrete values, with prediction error typically measured by the squared difference between the observed and predicted values.

CART is a recursive partitioning method to be used both for regression and classification. CART is constructed by splitting subsets of the data set using all predictor variables
to create two child nodes repeatedly, beginning with the entire data set. The best predictor is chosen using a variety of impurity or diversity measures (Gini, twoing, ordered twoing and least-squared deviation). The goal is to produce subsets of the data which are as homogeneous as possible with respect to the target variable. In this study, we used measure of Gini impurity that used for categorical target variables.

Gini Impurity Measure:

The Gini index at node t, g(t), is defined as

$g(t) = \sum_{j \neq i} p(j|t)p(i|t)$ (4)

where i and j are categories of the target variable. The
equation for the Gini index can also be written as

$g(t) = 1 - \sum_j p^2(j|t)$ (5)

Thus, when the cases in a node are evenly distributed across the categories, the Gini index takes its maximum value of 1-(1/k), where k is the number of categories for the target variable. When all cases in the node belong to the same category, the Gini index equals 0.

If costs of misclassification are specified, the Gini index
is computed as

$g(t) = \sum_{j \neq i} C(i|j)p(j|t)p(i|t)$ (6)

where C(i|j) is the probability of misclassifying a category j
case as category i.

The Gini criterion function for split s at node t is defined

as

$$\emptyset(s,t) = g(t) - p_L g(t_L) - p_R g(t_R) \quad (7)$$

where p_L is the proportion of cases in t sent to the left child node, and p_R is the proportion sent to the right child node.

The split s is chosen to maximize the value of U(s, t). This value is reported as the improvement in the tree by Breiman [21].

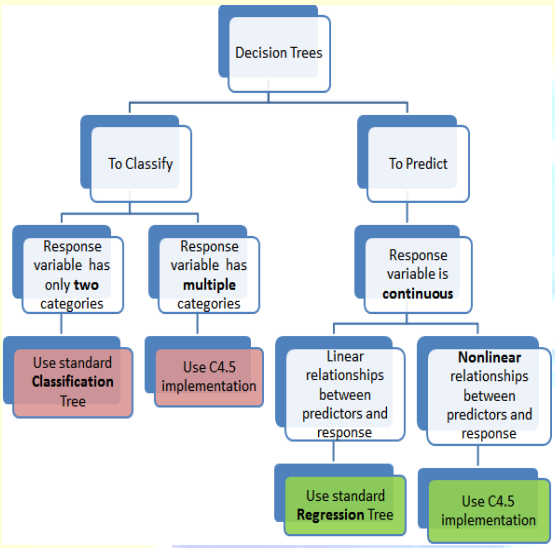

Fig.2. Classification of decision trees

Advantages of CART

1) It is nonparametric.
2) It does not require variables to be selected in advance.
3) It can easily handle outliers.
4) It has no assumptions and computationally fast.
5) It is flexible and has an ability to adjust in time.

Disadvantages of CART

1) It may have unstable decision trees.
2) It splits only by one variable at a time.

3) It does not vary under a monotone transformation of independent variables.

D. *Support Vector Machine*

Support vector machines (SVM) were first suggested by Vapnik [24]. SVM technique is a classification technique that has proven its performance in many fields, such as text categorization, credit risk, and bankruptcy prediction [32]. The strength of this technique lies with its capability to model nonlinearity and resulting in complex mathematical models. SVMs are used to find an optimal hyper-plane which maximizes the margin between itself and the nearest training examples in the new high-dimensional space and minimizes the expected generalization error.

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input

data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other.

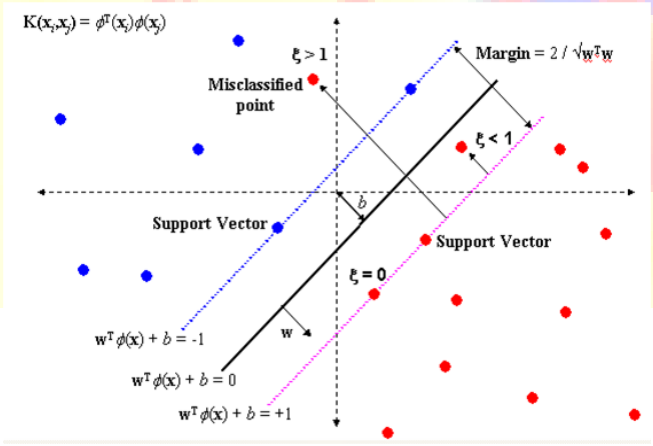

Fig.3. support vector machine

An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Let D be a training set formed by l pattern s_i . Each pattern is a couple of values ($\mathbf{x}_i$, y_i) where $\mathbf{x}_i \in R^l$ and $y_i \in \{-1,1\}$ where i=1,….,l. The patterns with output +1 are called positive patterns, while the others are called negative patterns. The points x belonging to the hyper plane must satisfy w.x+b = 0, where w is normal to the hyper plane and b is the intercept. The vectors that are not on this hyper plane is defined by w.x+b≠0.

The decision function f(x) is given by

$$f(x) = sgn(w.x + b) \quad (8)$$

An optimal hyper plane is located where the margin between two classes of interest is maximized and the error is minimized. To compute the optimal hyper plane, *the* following optimization problem has to be solved:

Minimization: $\frac{1}{2}\|w^2\|$

Subject to: $y_i((w.x_i) + b) - 1 \geq 0$ (9)

The margin of the hyper plane is $\frac{2}{\|w\|}$.

The constrained optimization in Eq. (9) is solved by the method of Lagrange multipliers. The equivalent optimization problem becomes,

Maximize: $\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j (x_i . x_j)$

Subject to: $\sum_{i=1}^{l} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, for i=1,2,..,l (10)

where $\alpha_i \geq 0$ are the Lagrange multipliers.

The constant $0 < C < \infty$, called the penalty value or C value, is a regularization parameter. It defines the trade-off between the number of misclassification in the training data and the maximization of margin.

Nonlinear transformation function $\emptyset$ maps the data into a higher dimensional space. There exists a function k, called a kernel function, such that, $k(x_i, x_j) \equiv \emptyset(x_i) . \emptyset(x_j)$. The optimization problem then becomes,

Maximize: $\sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, x_j)$

Subject to: $\sum_{i=1}^{l} \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$, for i=1,2,..,l (11)

Advantages of SVM

1) SVM is used in the situation of finite sample data. It aims to get the optimal solution based on the present information rather than the optimal value when the number of sample tends to be infinite.
2) The algorithm is finally transformed into the optimization of quadratic program. Theoretically, it will get a global optimization value, which solves the unavoidable local optimization problem while using neural network.
3) The algorithm performs a nonlinear mapping from the original data space into some high dimension feature space, in which it constructs a linear discriminant function to replace the nonlinear functions in the original data space. This special character assures that SVM has good generalization ability.

Disadvantages of SVM

1) SVM is a binary classifier. To do a multi-class classification, pair-wise classifications can be used (one class against all others, for all classes).
2) Computationally expensive and thus runs slow.

We have seen that the techniques that are used in conventional algorithms are not suitable for credit risk assessment due to several limitations.

IV. BIO-INSPIRED ALGORITHMS

Biologically inspired algorithms or bio-inspired algorithms are a class of algorithms that imitate specific phenomena from nature. Bio-inspired algorithms are usually bottom-up, decentralized approaches that specify a simple set of conditions and rules and attempt to solve a complex problem by iteratively applying these rules [36]. Such algorithms tend to be adaptive, reactive and distributed [37].

TABLE I. COMPARISON BETWEEN BIO-INSPIRED ALGORITHMS AND CONVENTIONAL ALGORITHMS

Criteria	Bio-inspired Algorithms	Conventional Algorithms
Flexibility	Strength through flexibility, or strength in numbers	Start with a fixed size or population in mind and hence are not very flexible
Performance	Work well even when the task is poorly defined	Reach a saturation limit in their performance
Scalability	Scalability is not really a challenge	Scalable, but only to a certain degree
Flexibility in decision making	Tend to find the alternate best available solution	Depends on programmer's understanding of the

		program
Improvement scope and innovation	Largely unexplored field	Conventional algorithms are optimized and developed almost to their limits

Bio-inspired algorithms depend heavily on component behaviour. They take a bottom-up decentralized approach to solving any problem. They are called computationally intelligent with respect to the field of artificial intelligence. This is because the system is not told how to achieve an overall goal. Instead, through iterative individual component behaviour, the system produces an emergent, overall behaviour. This emergent behaviour is then utilized for solving the problem.

A. *Neural Network*

A neural network is a field of Artificial Intelligence which is inspired by human brain. It is used to predict outputs from a set of inputs by taking linear combination of input and then making nonlinear transformations of the linear combination
using activation function. Biological neural networks are made up of real biological neurons that are connected or functionally related in a nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis.

Gately [38] defined neural networks as "an artificial intelligence problem solving computer program that learns through a training process of trial and error". Therefore, neural networks building require a training process and the linear or nonlinear variables in the training procedure help distinguish variables for a better decision-making outcome. In the credit scoring area, neural networks can be distinguished from other statistical techniques.

Amari [39] gave an example to differentiate between regression models and neural networks models. In his discussion, he stated that to build an applicant score using regression models, the "inverse matrix" should be used, whilst in neural networks the "applicant's profile" is used to perceive those applicants relative scores. Also, using neural networks, if the outcomes are unacceptable, the estimated scores will be changed by the nets until they become acceptable or until having each applicant's optimal score.

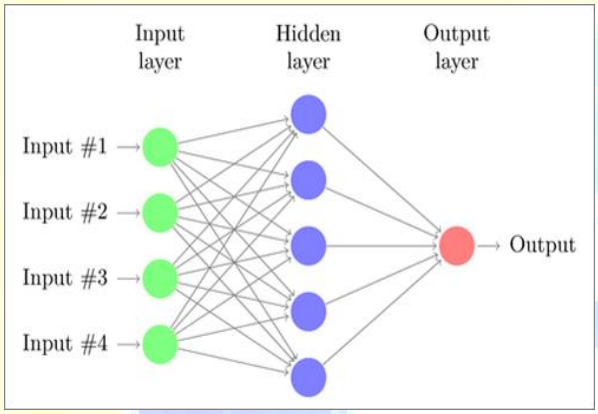

Fig.4. Neural Network

Advantages of NN

1) does not use pre-programmed knowledge base.
2) suited to analyze complex pattern.
3) have no restrictive assumptions.
4) can handle noisy data.
5) can overcome autocorrelation.
6) robust and flexible.

Disadvantages of NN

1) requires high quality data.

2) variables must be carefully selected a priori.
3) can have risk of overfitting.
4) requires long processing time.
5) requires large training sample.

B. *Genetic Algorithm*

Genetic algorithms (GA), a general adaptive optimization search methodology based on a direct analogy to Darwinian natural selection and genetics in biological systems, is a promising alternative to conventional heuristic methods. GA works with a set of candidate solutions called a population. Based on the Darwinian principle of "survival of the fittest", the GA obtains the optimal solution after a series of iterative computations. GA generates successive population of alternate solutions that are represented by a chromosome, i.e. a solution to the problem, until acceptable results are obtained. Associated with the characteristics of exploitation and exploration search, GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms.

Genetic algorithm is an efficient optimization procedure. The basic principle of the genetic algorithm is inspired by the mechanisms of biological evolution [25]. In a genetic algorithm, a population of strings (called chromosomes), which encode candidate solutions (called individuals, members, or phenotypes) to an optimization problem, evolves toward better solutions.

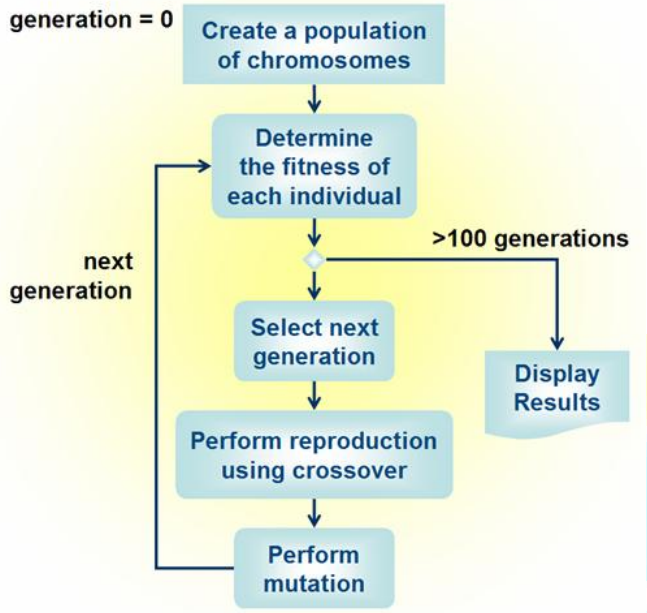

Fig.5. Working principle of Genetic Algorithm

Advantages of GA

1) It can solve every optimization problem which can be described with the chromosome encoding.
2) It solves problems with multiple solutions.
3) Since the genetic algorithm execution technique is not dependent on the error surface, we can solve multidimensional, nondifferential, continuous, and even nonparametrical problems.
4) Structural genetic algorithm gives us the possibility to solve the solution structure and solution parameter problems at the same time by means of genetic algorithm.
5) Genetic algorithm is a method which is very easy to understand and it practically does not demand the knowledge of mathematics.
6) Genetic algorithms are easily transferred to existing simulations and models.

Disadvantages of GA

1) Certain optimization problems (they are called variant problems) cannot be solved by means of genetic algorithms. This occurs due to poorly known fitness functions which generate bad chromosome blocks in spite of the fact that only good chromosome blocks cross-over.
2) There is no absolute assurance that a genetic algorithm will find a global optimum. It happens very often when the populations have a lot of subjects.

3) Like other artificial intelligence techniques, the genetic algorithm cannot assure constant optimization response times. Even more, the difference between the shortest and the longest optimization response time is much larger than with conventional gradient methods. This unfortunate genetic algorithm property limits the genetic algorithms' use in real time applications.
4) Genetic algorithm applications in controls which are performed in real time are limited because of random solutions and convergence, in other words this means that the entire population is improving, but this could not be said for an individual within this population. Therefore, it is unreasonable to use genetic algorithms for on-line controls in real systems without testing the first on a simulation model.

V. LITERATURE SURVEY

Practitioners and researchers have developed a variety of traditional statistical models and data mining tools for credit scoring, which involve linear discriminant models [15], logistic regression models [16], k-nearest neighbor models Henley [17], decision tree models [18], neural network models [9, 19, 14] and genetic programming models [20].

Desai et al. [9] investigated neural networks, linear discriminant analysis and logistic regression for scoring credit decision. They concluded that neural networks outperform linear discriminant analysis in classifying loan applicants into good and bad credits, and logistic regression is comparable to neural networks.

From the computational results made by Tam and Kiang [10], the neural network is most accurate in bank failure prediction, followed by linear discriminant analysis, logistic regression, decision trees, and k-nearest neighbor. In comparison with other techniques, they concluded that neural network models are more accurate, adaptive and robust.

Kim [11] compared the neural network approach with linear regression, discriminant analysis, logistic analysis, and a rule-based system for bond rating. They found that neural networks achieved better performance than other methods in terms of classification accuracy.

Huang et al. [12] compared SVMs with a back-propagation neural network to predict corporate credit ratings but find inconsequential differences in performance.

Zhang [13, 31] found that GA+SVM gives better result as compared to a pure SVM, back-propagation neural network (BPN), Genetic Programming (GP) and logistic regression (LR).

Desai et al. [9] found that GA approach was better than linear discriminant analysis, logistic regression and a variety of neural networks.

West [14] investigated the credit scoring accuracy of several neural networks. Results were benchmarked against traditional statistical methods such as linear discriminant analysis, logistic regression, k-nearest neighbor and decision trees.

Li et al. [23] found that SVMs outperform multilayer perceptrons for consumer credit data, but their results are also based on a small sample size.

Oreski et al. [44] found that GA+NN model is significantly better in feature selection for classification as compared to some other techniques used for selecting features.

The predictive accuracy of different techniques based on German Dataset is shown in the table given below.

TABLE II. COMPARISON OF CLASSIFICATION ACCURACY OF DIFFERENT CREDIT SCORING METHODOLOGIES

Sl.No	Classifier	Accuracy(%)
1	LDA[33]	66.0
2	LR[33]	72.4
3	CART(t-test)[33]	68.9
4	C4.5[44]	72.4
5	SVM[45]	75.4
6	NN[33]	75.2
7	GA+SVM[34]	77.92
8	GP[35]	77.34
9	RBF[14]	75.63
10	GA+NN[43]	82.3

V. CONCLUSION

From the above comparison, we found that GA+NN method is giving better accuracy than all other methods. The future work is to extend the dataset or to propose different methods such as FLANN, CFLANN to obtain better accuracy.

REFERENCES

[1] Hancock, Simon, "Developing a More Accurate and Efficient Scorecard", *Credit Control*, 20, (8), 10-16, 1999.

[2] L.C. Thomas, D.B. Edelman, and J.N. Crook, "*Credit scoring and its applications*", Society for Industrial Mathematics, 2002.

[3] L.J. Mester, "What's the point of credit scoring?", *Business review*, vol. 3, 1997, pp. 3–16.

[4] Zhang Lei-lei, Hui Xiao-feng and Wang Lei, "Application of adaptive support vector machines method in credit scoring", *Management Science and Engineering, 2009. ICMSE 2009. International Conference on*, 2009, pp. 1410-1415.

[5] K. Leung and F. Cheong, "Consumer credit scoring using an artificial immune system algorithm", *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, 2007, pp. 3377-3384.

[6] SJ. Press, S. Wilson, "Choosing between logistic regression and discriminant analysis", *Journal of the American Statistical Association*, 73 (4)(1978) ,699-705.

[7] Agresti, A. (1990). Categorical data analysis. New York: Wiley. Ahn, B.S., Cho, S. S., & Kim, C. Y. (2000). "The integrated methodology of rough set theory and artificial neural network for business failure prediction", *Expert Systems with Applications*, 18(2),65-74.

[8] McCullagh, P. (1980). "Regression model for ordinal data," *Journal of the Royal Statistical Society*, Series B, 42(2), 109-142.

[9] Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). "A comparison of neural networks and linear scoring models in the credit union environment", *European Journal of Operational Research*, 95(1), 24–37.

[10] Tam, K. Y., & Kiang, M. Y. (1992). "Managerial applications of the neural networks: the case of bank failure prediction", *Management Science,* 38(7), 926–947

[11] J.W. Kim, "Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems", *Expert Systems* 10 (1993) 167–171.

[12] Huang, Z., Chen, H., Hsu, C., Chen, W. & Wu, S. 2004. "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study", *Decision Support Systems*, ***37***, 543-558.

[13] Zhang Lei-lei, Hui Xiao-feng, and Wang Lei, "Application of adaptive support vector machines method in credit scoring", *Management Science and Engineering, 2009.*

[14] West, D. (2000). "Neural network credit scoring models", *Computers and Operations Research*, 27(11–12), 1131–1152.

[15] Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models.
Journal of Business and Economic Statistics, 1(2), 101–114.

[16] Henley, W. E. (1995). Statistical aspects of credit scoring. Dissertation, The Open University, Milton Keynes, UK.

[17] Henley, W. E., & Hand, D. J. (1996). A k-nearest neighbor classifier for assessing consumer credit risk. Statistician, 44(1), 77–95.

[18] Davis, R. H., Edelman, D. B., & Gammerman, A. J. (1992). Machine learning algorithms for credit-card applications. Journal of Mathematics Applied in Business and Industry, 4, 43–51.

[19]Malhotra, R., & Malhotra, D. K. (2002). Differentiating between good credits and bad credits using neuro-fuzzy systems. European Journal of Operational Research, 136(1), 190–211.

[20] Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Building credit scoring models using genetic programming. Expert Systems with Applications, 29(1), 41–47.

[21] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and regression trees. Wadsworth, Belmont CA.

[22] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, **7**, 1936 pp. 179–188

[23] Li ST, Shiue W, Huang MH (2006). The evaluation of consumer loans using support vector machines. *Expert Systems with Applications* 30(4):772-782.

[24] Vapnik, V. N. (1995). The nature of statistical learning theory. New York: Springer.

[25] Sustersic, M., Mramor, D., & Zupan, J. (2009). Consumer credit scoring models with limited data. Expert Systems with Applications, 36, 4736–4744.\

[26] E. L. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, The Journal of Finance, 23 (4) (1968) 589-609

[27] M. Bardos, Detecting the risk of company failure at the Banque de France, Journal of Banking and Finance, 22 (1998) 1405-1419

[28] R. A. Eisenbeis, Problems in applying discriminant analysis in credit scoring models, Journal of Banking and Finance, 2, 205-219

[29] P. Falbo, Credit scoring by enlarged discriminant analysis, OMEGA, 19 (4) (1991) 275-289

[30] B. J. Grablowsky and W. K. Talley, Probit and discriminant factors for classifying credit applicants: a comparison, Journal of Economics and Business, 33 (1981) 254-261

[31] Defu Zhang, M. Hifi, Qingshan Chen, and Weiguo Ye, "A Hybrid Credit Scoring Model Based on Genetic Programming and Support Vector Machines," *Natural Computation, 2008.*

[32] Amir F. Atiya, Senior member, IEEE, .Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results., *IEEE Transactions on Neural Networks, Vol. 12, 2001*

[33] P. Yao, "Hybrid Classifier Using Neighborhood Rough Set and SVM for Credit Scoring," *Business Intelligence and Financial Engineering, International Conference on*, Los Alamitos, CA, USA: IEEE Computer Society, 2009, pp. 138-142.

[34] Cheng-Lung Huang, Mu-Chen Chen, Chieh-Jen Wang, "Credit scoring with a data mining approach based on support vector machines." *Expert Systems with Applications,* (2006),doi: 10.10 16/j.eswa.2006.07.007

[35] Chomg-Shyong Ong, Jih-Jeng Huang, Gwo-Hshiung Tzeng, "Building credit scoring models using genetic programming." *Expert Systems with Applications ,29 (2005) ,41-47*

[36] Ding, J (2009), *Advances in Network Management*, Taylor & Francis, p. 164

[37] Rocha, L. (2011) "I485/H485/I585: Biologically Inspired Computing", http://informatics.indiana.edu/rocha/i-bic/. August 31, 2011 [accessed 11 Sep 2011]

[38] Gately, E. 1996. Neural Networks for Financial Forecasting: Top Techniques for Designing and Applying the Latest Trading Systems. New York: John Wiley & Sons, Inc.

[39] Al Amari, A. 2002. The credit evaluation process and the role of credit scoring: A case study of Qatar. Ph.D. Thesis, University College Dublin.

[40] SJ. Press, S. Wilson, Choosing between logistic regression and discriminant analysis, *Journal of the American Statistical Association*, 73 (4) (1978) ,699-705.

[41] Agresti, A. (1990). Categorical data analysis. New York: Wiley. Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, 18(2),65-74.

[42] McCullagh, P. (1980). Regression model for ordinal data. *Journal of the Royal Statistical Society*, Series B, 42(2), 109-142.

[43] Stjepan Oreski, Dijana Oreski, Goran Oreski , "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment ". *Expert Systems with Applications 39(2012) 12605-12617*

[45] Y. Lan, D. Janssens, G. Chen, and G. Wets, "Improving associative classification by incorporating novel interestingness measures," Expert Systems with Applications, vol. 31, pp. 184–192, 2006.

[45] Feng Chia Li, Peng-Kai Wang , Gwo-En Wang, "Comparison of the Primitive Classifiers with Extreme Learning Machine in Credit Scoring," IEEE ,*Industrial Engineering and Engineering Management*, 2009, 685-688.