Journal of
Applied & Computational Mathematics

Ghassempouri and Hoseini, J Appl Computat Math 2017, 7:1
DOI: 10.4172/2168-9679.1000381

Research Article Open Access

A Methodology for Calculating Customer Credit Score Based on Customer Lifetime Value Model

Ghassempouri M* and Hoseini SMS

Islamic Azad University, Tehran Province, Tehran, Iran

Abstract

Proper customer relationship management is among the facets that contribute to productivity at institutions. It is a requirement for customer relationship managers, especially at financial and credit institutions and at banks, to calculate and determine the customer's creditworthiness and credit score. The aim of this study is to present a solution for calculating the customers' value and their credit score without incurring the costs for collecting extra information. The primary source of data for this study is operation system database. Due to differences among operation systems, a comprehensive schema of the database is defined first. Only conventional indices and variables have been used in this schema, so that the presented solution can be generalized and will be applicable to most economic institutions. The calculation of the customer's creditworthiness is performed with regard to the three variables of the "recency" of contact, the "frequency" of transactions, and the "monetary" amount. The collected data is divided into the two populations of "good" and "bad" customers. Variables from those two populations that possess significant differences are identified using statistical methods. Those variables are used in determining the customer's credit score. Next, a solution is presented for comparing the efficiency of the models for the identification of the customer's credit score. We will test and compare two statistical methods, the Logistic Regression model and the Fisher Discriminant Analysis, and two soft computing methods, the Multilayer Perception Network and the Vector Machine for determining the customer's credit score. Additionally, a solution is offered for setting the number of layers and the number of neurons in the Multilayer Perception Network.

Keywords: Credit scoring; Customer lifetime value; Discriminant analysis; RFM model

Introduction

Breakthroughs in information and communication technology together with the globalization of economy have created new challenges for economic institutions in general and financial institutions and banks in particular [1-3]. Overcoming these challenges will not be possible without relying on scientific management. One of the foundations of scientific management is using mathematical models and soft computing to model and quantify phenomena [4-7]. Sound and accurate decision-making to resolve managerial problems would be impossible without the quantification of phenomena [8-10]. The problem of determining the customer's creditworthiness is one of these challenges that demand addressing particularly following the prevalence of credit cards and selling methods that rely on the customer's credit [11-15]. The information on the individuals and businesses credit scores is now a valuable asset produced and sold by certain credit rating agencies [16-20]. Many institutions prefer to calculate their own customers' credit scores relying on the data already existing at the institution [21-23]. This way, such institutions would avoid paying the costs pertaining to collecting extra data or buying entities' credit scores from credit rating agencies. This study is attempting to meet this need by presenting a plan that is practical, comprehensive, and feasible to implement at most institutions, and would also eliminate the need to pay extra costs [24,25].

Literature and Basic Concepts

The customer's credit score, his lifetime value, and the default risk from his likelihood of non-repayment are among the primary requirements in formulating sales strategy and policy. The perception that long term relationship with a customer translates into more profitability has been held since long. Based on this concept is raised the subject of customer lifetime value (CLV, or LTV). In classic resources, CLV has been defined as the net worth gained from the customer as a result of having relationship with him, with a part or a cluster of customers as a continuing, consistent, and interactive flow. In models for calculating CLV, the total net income gained in the past is considered together with the customer's current and future value. Questions that are brought up in calculating CLV are: What is the probability that the customer will buy the product or service again, as specified for each period? What is the approximate gross size of his purchase, as specified for each period? What is the cost of rendering services to the customer for his retention, as specified for each period? What is the cost of attracting a new customer (3). A variety of models have been introduced for calculating CLV, all of which in essence aim at predicting the customers' future behavior by observing their past conduct. If we represent the customer's transactions withθ, we will have past=f (θ) and future=f (θ). The aim of modeling is to determine the function *f*. Offering services in market is done based on a contract or without it and in continuous or discrete manner. Therefore, the relationship with the customer is considered in four different forms, and the modeling for the calculation of CLV is performed in correspondence with these four states. Studies show that, having the indices for the "recency" of contact and the "frequency" of purchase, it is possible to estimate CLV for each of these four states (Hughes, 2006). Recency and frequency are represented with the triplet (x, t_x, T). x Is the

***Corresponding author:** Ghassempouri M, Islamic Azad University, Tehran, Tehran Province, Tehran, Iran, Tel: 981132290331; E-mail: st_m_ghasempouri@azad.ac.ir

Received October 09, 2017; **Accepted** December 20, 2017; **Published** December 27, 2017

Citation: Ghassempouri M, Hoseini SMS (2017) A Methodology for Calculating Customer Credit Score Based on Customer Lifetime Value Model. J Appl Computat Math 7: 381. doi: 10.4172/2168-9679.1000381

number of transactions in interval (0,T], and t_x is the time of transaction x. The probability that a customer with the attribute (x, t_x, T) will still be active in time T is represented by $P(\text{active} \mid x, t_x, T)$, and the number of expected transactions in the future with mathematical expectation is represented by $E[X(T, T+t) \mid x, t_x, T]$. Authors have calculated this expression in their studies. The RFM (Recency, Frequency, Monetary) model was introduced by Cullinan [5]. The popularity of this model is due to its simplicity of use and the access to input data. This way, organizations can use their transactions in this model. In this model, the three factors of the regency of contact, the frequency of purchase or contact, and the monetary amount of purchase or exchange in a linear relation are the basis for calculation. Due to the popularity of this model, ways to further expand and solidify its theoretic foundations are still under study. The present study will show that, having the three factors of the recency of contact, the frequency of purchase, and the monetary amount of purchase in a given time period, we can calculate the remaining CLV. Despite the popularity and reception that the RFM model has enjoyed, it suffers from some shortcomings, too. Using the results from this model may lead to weak decisions, it does not allocate resources efficiently, and is not capable of predicting the customer's future status. Models other than RFM have also been introduced. These models try to predict the customer's future status and cash flow. Also, models have been presented that adopt a statistical approach. NBD (Negative Binomial Distribution) is one of these statistical models that work for continuous and repetitive purchase mode. Such models offer a good statistical distribution for function. Rating customers' creditworthiness is a measurement and evaluation system that uses the data from the past behavior and present specifications of the individuals that had received loans before. This system is in fact a tool for assessing and judging about the right amount of loan to be made to the customer based on his personal data and his past credit history. Just like any other approach in modeling, rating is only a simplification of a complex phenomenon in the real world, and at its best, it is just an approximation of the risk. The aims of rating are: controlling the choice of risk, converting the risk of non-repayment into appropriate cost, managing credit reduction, assessing new loan plans, shortening the time required for approval, making sure about the soundness of the credit rules and their consistent execution, improving goal setting for amendment purposes, making financial policies proportionate to specific risk limits, improving the request for the repayment of installments with the aim to reduce the costs of delay in installment payment. Rating customers' creditworthiness and classifying the loans according to the customers' credit score will result in a reduction in financial risk. To take control of the loan making process, the two variables of minimum deposit and maximum loan allowed are incorporated. This way, the amount of deposit and the maximum amount of permitted loan for a given customer will be determined after his credit score has been calculated. Methods like multivariate regression for determining function *f* are presented in the relation

$f: R \rightarrow \{\text{Higher Score, Medium Score, Low Score}\}$.

In this equation, the *R* domain includes the two variables of the customer's income and the customer's debt to income ratio. Different approaches exist to customers' credit rating. Some of these are: rating and scoring based on the customer's creditworthiness evaluation, behavioral rating, rating and scoring for the receipt of installments, rating and scoring customers to identify abuse and fraud. In the creditworthiness evaluation approach, the customer's income or salary, the number of people under his sponsorship, his residence, and other such variables are the criteria for determining the customer's credit. The behavioral approach relies on historical data from the past. For example, the customer's financial activities, number of transactions, account balance, frequency and quantity of loan installment maturities, and the lifetime of his account are input variables for predicting the probability of his delinquency in paying the installments. In rating for installment receipt approach, the customer's income and his delays in repayments are among the variables in decision making. Usually, the goal is a timely and proper identification and action proportionate to the customer's credit and repayment delays. In the abuse identification approach, the customers are rated based on their probability of abuse. Different methods and algorithms have been used for rating. The modeling methods have been classified in four categories: statistical modeling, expert system based modeling, modeling with artificial intelligence techniques, and soft computing. In credit rating based on experts' view, factors such as the customer's quality of loyalty, his ability to repay, the value of the collateral, loan purpose or type, loan period, and its installment maturities are taken into consideration. This method is heavily dependent on the expert's experience and opinion, and is also time consuming and prone to error. The efficiency of different rating methods is compared. The results indicate that soft computing approaches, especially the hybrid ones, are more accurate than statistical methods. Still, neural network and soft computing methods lack clarity. In other words, they are not based on reasoning and are unable to precisely explain the vague and uncertain if customer loss, customer life cycle, customer value, and the policies for attracting and rating. To solve this problem, it seems that studies are needed to develop the hybrid methods and clustering methods. Another approach to achieve clarity is to merge neural network with fuzzy system. Research on customer rating is usually conducted by computer experts, hence generally lacking real financial and market considerations. In fact, customer rating is aiming at financial risk reduction, and the customer's value and profitability and his life time value are hardly accounted for. The practical outcomes of using customer rating will be retaining customers are not taken into account. In most statistical methods, it is a required condition for the data distribution to be normal, but this condition is not necessary in logistic regression approach.

One of the challenges faced with in rating is optimal default point definition for defining bad customer. Different factors such as laws and the economic institution's requirements are involved in defining bad customer. According to the definition by the Basle Committee on Banking Supervision, a delay of more than 90 days in paying installments would render the customer as gone bad. At some other banks, delay in paying even one installment is enough basis to regard the customer as gone bad. Also, it is possible to simultaneously have several criteria for bad customer definition. This study tries to use random forest tree algorithm to obtain an optimal default for bad customer definition.

Aims of the Study

As shown in Literature Review, customer credit rating requires collecting personal and financial data on the customers. This is a costly attempt which in some cases violates the customer's legal rights, too. Studies show that the customer's legal rights are often neglected in data collection, the customer information is collected without their consent, and the collected data is usually more than what is needed for customer rating. Regarding their application and benefits, modeling for calculating the customer's lifetime value and his creditworthiness is a need even for small institutions. Unfortunately, data collection and modeling is costly. Therefore, it seems that studies need to be conducted to come up with low cost models that would be able to cover the two

subjects of customer LTV and credit rating using minimum attributes. So, the first question is how to collect the data required for modeling LTV and credit rating with minimum cost, and the second is what method is the right method for calculating the credit score and how we can identify this right method. Here, we first present a methodology for extracting data from the institution's operation system database. We will show that the required data can be extracted without the need for extra costs such as filling out forms or customer interview and just by a general and conventional schema of the operation system database. After calculating the customer value based on the classic RFM model, we will answer the question of whether this model can be used for customer credit rating too. Then, an appropriate methodology will be presented for comparing discriminant analysis methods in customer credit rating. Four discriminant analysis methods will be compared and ordered according to their efficiency.

Theoretic Foundations of this Study

In this study, statistical methods and soft computing are used for customer credit rating. Of statistical methods, logistic regression and Fisher discriminant analysis, and of soft computing methods, multilayer perception network and support vector machines are examined and compared. Their mathematical foundations are briefly presented.

Multilayer perception network

In multilayer perception network, each layer is made up of the matrix of weights W, bias vector b, and output vector Y. The output from the first hidden layer is defined as $Y_1=f(W_1.X+b_1)$ and the output from the second layer is defined as $Y_2=f(W_2.X+b_2)$. Function $f(.)$ is the activation function, and is usually chosen of sigmoid or tangent sigmoid type. The chosen number of layers in most applications is three. If the input vector has m elements, each element from the input vector will feed forward n neurons with the activation function output. The activation function outputs will linearly mix with the weights and create the network output. Propagating the input vector to the output vector, the network will generate its own output. The generated output is compared with the real output. The difference between the two outputs is called error. The error value is back-propagated into the network as changes in weights and bias. If the error reduces to the specified value, the learning process will stop.

Support vector machine

This method separates with a hyperplane the data that have been grouped in binary form. The method of separating is in a way that the separation area between the hyper plane and the sample training data is maximized. In action, the data will not be completely separated by the hyper plane into two groups. Part of the data will be placed on the wrong side. Therefore, simultaneous with the increase in the separation area, the addition of the error from placing some of the data on the wrong side should be minimized. The C parameter as the error penalty coefficient in the model controls the relative cost of these objectives. The training data placed on the separation area or on the wrong side are called support vectors; as such vectors suffice for determining the separating hyper plane. In algebraic language, the ordered couples $\{(x_i,y_i),i=1,\ldots,l\}$ are available as training data where x_i R^n and $y_i \in\{1,-1\}$. The data have been classified into two sets by the y_i value. The separation area stands between the two planes $W^Tx+b=1$ and $W^Tx+b=-1$. In nonlinear situations, the input data are projected onto another Euclidean space by the function $z_i=(x_i)$. If we represent the error value for each input vector as ε_i, the support vector machine (SVM) is the optimal answer for the following problem [24]:

$$Min_{w,\,b,\,\varepsilon}\ \frac{1}{2}W^TW + C\sum_{i=1}^{l}\varepsilon_i$$

Discriminant analysis with fisher method

In his studies, Fisher introduces a linear function for partitioning of a set into two groups [11]. If X is representative of the set of random variables X_i, so that $X=(X_1,X_2,\ldots X_p)$, and if we write the linear combination of random variables X_i, as the Y relation $Y=w_1X_1+w_2X_2+\ldots+w_3X_p$ where p is the number of attributes, then one criterion for partitioning is the Y mean for the two groups. If we represent these two groups by B and G one solution is to examine the difference between $E(Y|B)$ *and* $E(Y|G)$ to find values for w_i s that maximize this difference. However, the condition $\sum_i w_i = 1$ should be met for the Y relation to be linear. The fundamental drawback of this method is that, despite the presence of two different means, the data distribution may be in a way that it cannot be partitioned into two groups. To solve this problem, Fisher assumes that the sample variances of these two groups are identical. Then, he introduces the relation M as the separation criterion:

$$M = \frac{distance\ between\ sample\ means\ of\ \ two\ groups}{\left(sample\ variance\ of\ \ each\ group\right)^{\frac{1}{2}}} \tag{1}$$

Dividing the means distance by the square root of the variance, the separation criterion will become independent of the scale. So, changing Y into cY would not change the size of M.

Suppose that the sample means for the groups B and G are m_B and m_G, respectively. Also, S is the common and sample variance. The separating distance M will be:

$$M = w^T.\frac{m_G - m_B}{\left(w^T.S.w\right)^{\frac{1}{2}}} \tag{2}$$

The number of attributes in each group is equal to p. Therefore, m_B, m_G,and w are vectors for p entries. And we have:

$$\mathrm{E}\left(\mathrm{Y}|\mathrm{G}\right) = w.\mathrm{m}_{\mathrm{G}}^{\mathrm{T}} \tag{3}$$

$$\mathrm{E}\left(\mathrm{Y}|\mathrm{B}\right) = \mathrm{w}.\mathrm{m}_{\mathrm{B}}^{\mathrm{T}} \tag{4}$$

$$\mathrm{Var}\left(\mathrm{Y}\right) = \mathrm{w}.\mathrm{S}.\mathrm{w}^{\mathrm{T}} \tag{5}$$

We will put in relation M and will equal it to 0 after differentiating with respect to w:

$$\frac{m_G - m_B}{\left(w.S.w^T\right)^{\frac{1}{2}}} - \frac{\left(w.\left(m_G - m_B\right)^T\right)\left(Sw^T\right)}{\left(w.S.w^T\right)^{\frac{3}{2}}} = 0 \tag{6}$$

$$\left(m_G - m_B\right)\left(w.S.w^T\right) = \left(S.w^T\right)\left(w.\left(m_G - m_B\right)^T\right) \tag{7}$$

The second derivative of the relation M is positive. Also, the expression $\frac{w.S.w^T}{\left(w.\left(m_G - m_B\right)^T\right)}$ is a scalar. So, the minimum point will be as:

$$w^T((S^{(-1)}(m_G - m_B)^T) \tag{8}$$

In this approach, only the means and variance have been used. Therefore, the existence of a normal distribution was not a requirement and this method will be applicable to any statistical distribution. In fact, $w.x=c$ is the line perpendicular to the line that connects the means of the two groups. Then, $w.x=c$ is deemed as the separating line. If c is defined as the cutoff point, then c should be chosen in the distance between the means of the two groups.

Discriminant analysis with logistic regression

The weakness of the linear regression in discriminant analysis is the large size of the difference in the range of the two sides of the regression equation. If the left hand side of the relation is a function of probability, a better outcome will be achieved, because the likelihood that different points from the right hand side of the equation are projected onto one point will be reduced. Therefore, a function of p_i is written on the left hand side of the regression equation. In logistic regression, this function is recommended to be a logarithmic function, so that the regression equation will be written as eqn. (9):

$$log\left(\frac{p_i}{1-p_i}\right) = w_0 + x_{i1}w_1 + x_{i2}w_2 + \ldots + x_{ip}w_p \quad for\ all\ i=1\ldots.n \tag{9}$$

So, the probability value will be obtained via eqn. (10):

$$p_i = \frac{e^{w.x}}{1+e^{w.x}} \tag{10}$$

In general, if the vector values y in equation $y=w_0+w^T{}_X$ assume only 0 and 1 so that we have $y_{i\in}\{0,1\}$, then to discriminate and separate the data sets into two classes, the probabilities $P(y=0|x)$ *and* $P(y=1|x)$ need to be calculated.

$$P(y=1\,|\,x) = \frac{1}{1+exp\left(-\left(w_0+w^t x\right)\right)} \tag{11}$$

$$P(y=0\,|\,x) = \frac{exp\left(-\left(w_0+w^T x\right)\right)}{1+exp\left(-\left(w_0+w^T x\right)\right)} \tag{12}$$

The Methodology of the Study

The steps for data collection methodology are: Creating a database subschema of the operation system. Define a diagram of the relation between the entities in the database. Extracting data with SQL commands and recording them in a table. Standardizing the variable values. Identifying and separating the independent and dependent variables. Identifying the distribution of the variables using the Anderson–Darling test. Identifying and determining the outliers. If the variable distribution is normal, the Dixon-Grubbs' test is used. If the variable distribution is not normal, the box plot is used. Determining the index for classifying the customers into two populations of "good" and "bad". Analyzing and determining the significant difference between the two populations of "good" and "bad" customers. If the variable distribution is normal, the analysis of variance is used. If the variable distribution is not normal, the Kruskal-Wallis H test or the Mann–Whitney U test are used. Narrowing down the variables via clustering method and drawing dendrogram. The RFM model is developed and implemented. That is, the creditworthiness of each customer for each fiscal period is calculated. Also, to identify and predict the general trend of changes in customer creditworthiness, at institution level, the means for the creditworthiness of the customers for each fiscal period is calculated. Using time series and the customer's creditworthiness for previous fiscal periods, we will be able to predict the customer's creditworthiness for the coming fiscal periods. Therefore, the creditworthiness of the ith customer for the jth fiscal period is defined by $V_{i,j}$ in eqn. (13). In this relation, $R_{i,j}$ is the recency of contact for ith customer in the jth fiscal period and $F_{i,j}$ and $M_{i,j}$ are the number of transactions and the mean deposit of the ith customer for the jth fiscal period, respectively.

$$V_{ij} = Wr^{*}R_{ij} + Wf^{*}F_{ij} + Wm^{*}M_{ij} \tag{13}$$

If n_j is the number of the customers of the institution in the jth fiscal period, the mean creditworthiness of the customers for the jth fiscal period is defined by n_j as in eqn. (14).

$$w_j = \frac{\sum_{i=1}^{n_j} V_{ij}}{n_j} \tag{14}$$

If the number of past fiscal periods is m, and T is defined as a time series system for predicting, then $V_{i,m+1}$ is the creditworthiness of the ith customer and V_{m+1} is the mean creditworthiness of the customers for the future fiscal period, so that these values are obtained via eqns. (15) and (16).

$$V_{i,m+1} = T(V_1, V_2, \ldots, V_m) \tag{15}$$

$$V_{i,m+1} = T(V_1, V_2, \ldots, V_m) \tag{16}$$

The variable $V_{i,m+1}$ represents the future creditworthiness of the ith customer. The variable V_{m+1} shows the future creditworthiness of all customers. For those customers who have received loans, and whose goodness and badness have been determined by examining the way they repay the loan installments, the values for $V_{i,j}$ s are extracted and the V_B and V_G is obtained. So that:

$$V_G = \left\{V_{i,j}\,|\,\forall i\ \ Good = 1\ j = 1\ldots m\right\} \tag{17}$$

$$V_B = \left\{V_{i,j}\,|\,\forall i\ \ Good = 0\ j = 1\ldots m\right\} \tag{18}$$

If the two populations V_B and V_G are different, then it is inferred that the histories of the good and bad customers differ. In case there is a significant difference between these two populations, there will be a credit scoring model based on the RFM model. To choose the right model for customer credit rating, the efficiency of the models is compared using the Kolmogorov index. The number of layers and neurons in the MLP model is determined by design of experiments. The steps for executing the credit scoring methodology are: Calculating the logistic regression equation. Calculating the Kolmogorov index, and the error for the logistic model. Calculating the Fisher equations. Calculating the Kolmogorov index, and the error for the Fisher model. Choosing the number of layers and neurons for the MLP model by design of experiments. Executing the steps for training and testing the neural network at different levels of the design of experiments. Determining the right number of levels and neurons using the regression equation of the design of experiments. Executing the steps for training and testing for the right neural network model. Calculating the Kolmogorov index and the error for the neural network model. Calculating the discriminant equation for the vector machine method. Calculating the Kolmogorov index for the vector machine model. Comparing the means of the Kolmogorov index for the models using the analysis of variance and determining the most appropriate model.

Implementation and Case Study

This case study is conducted in financial and credit institutions environment. The statistical population for this study has been extracted from the databases of four financial institutions. In the operation system database of the institution, the subschema with entity diagram is defined as Figure 1. The savings account specifications, the loan specifications, and the account transactions are stored in the ACC table, the LOAN table, and the DETAILS table, respectively. There is one loan for each customer and there are some transactions for each deposit account or loan account. Data are extracted from this subschema using the SQL language.

After standardization and elimination of the out-of-range data, the result is stored in a table. About 70% of the data will be used in building, and the remaining 30% will be incorporated in testing the models. The statistic of the tests is calculated for the study variables

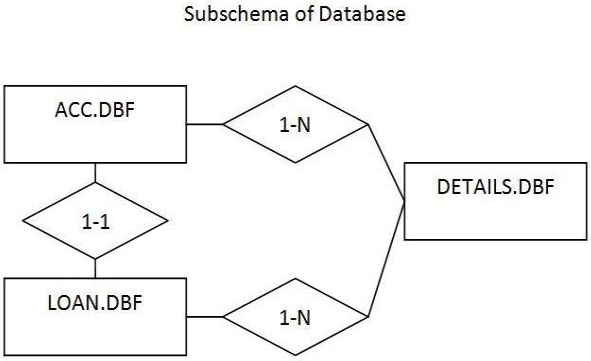

Figure 1: Subschema of Database.

Difference Between Variables in Two Populations , Good and Bad Customers					
Row	**Variable**	**Description**	**Mean for good**	**Mean for bad**	**H statistic value**
1	R	Recency of contact	77.82	37.56	254.23
2	NUMPAID	Number of paid installments	7	12	160.44
3	LIFETIME	Duration of cooperation	10	8	60.46
4	F	Number of transactions	1.170	1.110	55.80
5	M	Account balance	1.030	1.030	19.32
6	AGE	Age	43	39	17.98
7	AMT	Loan amount	5,000,000	8,000,000	14.45
8	PART	Installment amount	250,000	400,000	12.26

Table 1: Difference between variables in two populations, good and bad, customers.

using the Minitab software. In case the statistic of the Anderson test is above 0.576, distribution of data is not normal and zero hypothesis at the significance level of α=15% rejects. The results indicate that the distribution of the variables is not normal. To analyze the difference between the two populations, the Kruskal-Wallis H test is conducted. The Good variable, as the factor in the test assumes 1 and 0 values for good customers with no delay in repayment and bad customers, respectively. In Table 1, there are significant differences in the study variables for good and bad customers. We narrow down the number of variables through clustering. Here, correlation coefficient is defined as the criterion of similarity. Therefore, the entry (i,j) in the matrix of distances will be defined by relation $d_{ij}=1-|r_{ij}|$. With a similarity size of 70%, the AMT and PART variables go into the same cluster. Therefore, one of them can be eliminated.

The execution of the proposed methodology for data collection and extraction resulted in the applicability of the variables in rows 1 to 7 of Table 1 for determining the customer value and his credit score. The calculation of CLV is done using eqn. (14). Now, the good and bad customer populations with the CLV mean index are compared during the fiscal periods. The one-way analysis of variance test shows that the two populations have significant differences.

Now, with this difference, the RFM variables can be used to determine the customer's credit score. Four discriminant approaches will be discussed and their efficiencies will be compared. Each of these methods is tested several times using different data. Their indices of efficiency will be compared, and the best method will be selected. The output from the model is the customer's credit score. Good or bad customer is defined by cutoff point. If we show the cutoff point with s , and the obtained probability for the jth customer as P_i, then we have:

$$Pi \le s \rightarrow \{\text{customer i is Bad}\} \quad (19)$$

$$Pi > s \rightarrow \{\text{customer i is Good}\} \quad (20)$$

We show the model's efficiency index with Kolmogorov-Smirnov (KS) in eqn. (21). The KS efficiency index is in fact the maximum distance between the two probabilities.

$$KS = Max|P_G - P_B| \quad (21)$$

In this relation, P_G is the correct identification of a good customer and P_B is the correct identification of a bad customer. These probabilities will change as the cutoff point changes. The logistic regression equation is calculated using Minitab software. As seen in eqn. (22), the R and F coefficients are positive and the M coefficient is negative. The goodness probability of the customer is obtained by eqn. (23).

$$y' = 1.8296 + 0.002716R + 0.05867F - 0.02009M \quad (22)$$

$$P(1) = \frac{\exp(y')}{(1+exp(y'))} \quad (23)$$

In MATLAB environment, the efficiency index is calculated. To make sure about the soundness of the model performance and also to compare the efficiency of the models, the model efficiency index for 9 datasets is obtained. The results from the program execution are presented in Table 2.

The discriminant analysis with Fisher method is executed in SPSS environment. The Fisher linear functions are presented in eqns. (24 and 25) for bad and good customers, respectively.

$$y_B = -3.231 + 0.072R - 0.023F + 0.103M \text{ for Good=0} \quad (24)$$

$$y_G = -3.566 + 0.076R + 0.013F + 0.08M \text{ for Good=1} \quad (25)$$

The discriminant function is selected as y=min (y_B, y_G). The efficiency index for 9 datasets is calculated and recorded in Table 3.

The multilayer perception neural network is implemented using MATLAB software. The network is of feed forward type with Hyperbolic tangent activation function and updating network weights as error back propagation with Levenberg-Marquart (LM) algorithm. To select the number of hidden layers and the number of neurons in each layer, design of experiments executed in Minitab environment with two factors in two levels. The regression equation of the test is obtained as eqn. (26).

$$KS = 76.18 + 0.695*Neurons + 6.69*Layers - 0.339*Neurons*Layers \quad (26)$$

To determine an appropriate model, we use regression eqn. (26). The conclusion is that a model with one hidden layer and 20 neurons is better. Test result is recorded in Table 4.

To execute the algorithm for the vector machine, the model in question needs to be converted into a quadratic equation. To that end, good and bad customers are defined with indices 1 and -1, respectively. Vector Y shows the good and bad customers with these indices. Then, the good and bad customers' data are placed in Class A and Class B, respectively. The discriminant equation is obtained as eqn. (27).

$$Y = w_1R + w_2F + w_3M + b = 212R + 124F + 256M - 10046 \quad (27)$$

The Kolmogorov indices are calculated for 9 test samples and are recorded in Table 5. The test results are compared using one-way analysis of variance. The Tukey method has been chosen for testing the means difference. In comparison to other methods, the neural network method shows a higher efficiency.

Efficiency Indexes for Logistic Model									
Test No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
KS	90.35	90.04	90.09	89.52	90.10	89.51	90.11	90.15	89.93

Table 2: Efficiency indexes for logistic model.

Efficiency Indexes for Logistic Model									
Test No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
KS	89.42	90.30	89.98	89.88	89.86	90.23	90.24	89.94	90.60

Table 3: Efficiency indexes for fisher model.

Efficiency Indexes for Logistic Model									
Test No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
KS	90.0532	89.7793	90.3263	89.9628	89.8013	90.2519	89.8439	90.157	89.7775

Table 4: Efficiency indexes for MLP model.

Efficiency Indexes for Logistic Model									
Test No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
KS	89.7775	90.157	89.8439	90.2519	89.8013	89.9628	90.3263	89.7793	90.0532

Table 5: Efficiency indexes for support vector machine model.

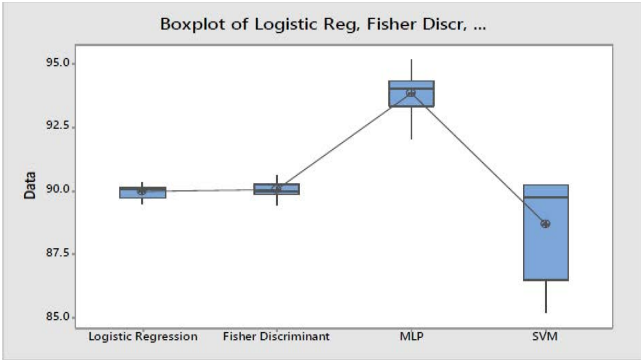

Figure 2: Logistic regression and Fischer methods.

No significant difference in efficiency is observed among the logistic, Fisher, and vector machine methods. Despite the superiority of the neural network method, as seen in Figure 2, the logistic regression and Fischer methods are more stable against different data.

Results of the Study

Executing the data collection methodology, statistical analysis, and customer value calculation, the following results have been obtained:

- Extraction and identification of the right variables for determining the customer value and his credit score from the operation system database without incurring extra costs for collecting data from the customers.
- Making sure about the existence of significant difference between the two populations of bad and good customers based on the obtained variables.
- Calculating the customer value and determining the trend of changes in their value.
- Predicting the customer value for future fiscal periods.
- Identifying significant difference in customer value between the two populations of bad and good customers.

Executing the methodology for customer credit rating, the following results were obtained in addition to presenting a method for comparing and selecting the rating model:

- If the multilayer perception neural network is set correctly, this model will be more efficient than other models in rating.
- Design of experiments can be used to set the number of hidden layers and the number of neurons in the multilayer perception model.
- The results from testing the statistical models show that this method has less variance compared to soft computing methods.

In this study, a method was proposed for calculating the customer score based on the customer value model. We also showed how the efficiency of the models are evaluated and compared. It is evident that managerial considerations will be influential in the final decision. Considering the fact that statistical methods are explainable and have less variance in the output, despite the superior efficiency of the multilayer perception model, statistical methods may be preferred.

References

1. Andrea B, Marisa B (2007) Credit Scoring for Microenterprise Lenders. Microenterprise Fund for Innovation.
2. Berger PD, Nasr NJ (1998) Customer lifetime value: Marketing models and Applications. Journal of Interactive Marketing.
3. Buttle Francis (2009) Customer Relationship Management. Elsevier.
4. Ricahrd C, Jiang W (1999) A Stochastic RFM Model . Journal of Interactive Marketing 13: 2-12.
5. Cullinan GJ (1978) Picking them by their batting averages: Recency-frequency-monetary .Marketing Association.
6. Hosmer WD, Lemeshow S, Sturdivant XR (2013) Logistic Regression Models for the Analysis of Correlated Data . Applied Logistic Regression.
7. Ehrenberg ASC (1959) The Pattern of Consumer Purchases. Journal of the Royal Statistical Society. Series C 8: 26-41.
8. Liran E, Mark J, Jonathan L (2013) Then impact of credit scoring on consumer lending. The RAND Journal of Economics 44: 249-274.
9. Peter F, Hardie B (2005) The Value of Simple Models in New Product Forecasting and Custome-Base Analysis. Applied Stochastic Models in Business and Industry .
10. FDIC (2007) Scoring and Modeling : Division of Supervison and Consumer Protection.

11. Fisher RA (1936) The use of multiple measurments in taxonomic problems. Anals of Human Genetics, pp. 179-188.
12. Alfred F (2014) Acquiring Profitable Customers with Credit Scoring Models . Capital Service.
13. Terry H (2013) Default definition selection for credit scoring . Artificial Intelligence Research.
14. Arthur MH (2006) Strategic Database Marketing .
15. Huseyin I, Bora A (2010) A comparisson of data mining techniques for credit scoring in banking: A managerial perspective. Journal of Bussiness Economics and Management 10: 233-240 .
16. Meike K, Barbara K, Meints M (2008) Profiling of Customers and Consumers-Customers Loyalty Programs and Practices.
17. Adel L (2008) Credit Scoring Models Using Soft Computing Methods: A Survey The International Arab Journal of Information Technology 7.
18. Miglautsch J (2002) Application of RFM principles: What to do with 1-1-1 customers? Journal of Database Marketing & Customer Strategy Management 9: 319-324.
19. Reichheld FF, Sasser WE (1990) Zero defections: Quality comes to services 68: 105-111.
20. Sadatrasoul SM, Gholamian M, Siami M, Hajimohammadi Z (2013) Credit scoring in bank and financial institution via data mining techniques a literature review. Journal of AI and Data Mining 1: 119-129.
21. David CS, Robert P (1994) An Industrial Purchase Process Application . Marketing Science.
22. Nazmul S , Hojjat A (2013) Computational Intelligence Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing.
23. Lyn CT, David BE, Jonathan NC (2002) Credit Scoring and Its Application: SIAM.
24. Vapnik V (1998) Statistical Learing Theory. IEEE Transactions on Neural Networks 10: 988-999
25. Zhong Y, Xiao-LL (2012) An Overview of Personal Credit Scoring Techniques and Future Work. International Journal of Intelligence Science 2: 181-189.