Analysis of Credit Scoring Using Particle Swarm Optimization Algorithm under Logistic Regression Model

Ulfa Rahmani
Master Program of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Padjadjaran, Indonesia
Email: ulfarahmani91@gmail.com

Sukono, Riaman
Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Padjadjaran, Indonesia
Email: sukono@unpad.ac.id; gusriani99@gmail.com

Subiyanto
Department of Marine Science, Faculty of Fishery and Marine Science,
Universitas Padjadjaran, Indonesia.
Email: subiyanto@unpad.ac.id

Abdul Talib Bon
Department of Production and Operations,
University Tun Hussein Onn Malaysia, Malaysia
Email: talibon@gmail.com

Abstract

Banking is a financial institution that has very important role in an economic and trade activity that is useful for channeling funds in the form of loans to the public who need fresh funds for business in the hope of helping improve the people's economy. In the process of lending, bank are often faced with a risk known as credit risk or problem loans. Therefore, the credit rating was analyzed by using the Particle Swarm Optimization algorithm in the Logistic Regression model. It was determined by using feasibility parameters of prospective debtors based on past data variables held by prospective debtors. This risk can be overcome by using a scoring system that is owned by each bank. In this paper, the data used is data on financial service cooperatives in Indonesia. Of the eight factors were analyzed, there are six factors that have a significant effect on the risk of default, namely including the age of debtors, family dependents, the amount of savings, the value of collateral, given the credit limit, and the loan term.

Keywords:
Credit risk, credit scoring, logistic regression, particle swarm optimization algorithm

1. Introduction

Banking is a financial institution whose role is very important in an economic and trade activity [Tan, 2015]. One of the benefits of the bank is to channel funds in the form of loans to the public who need fresh funds for business or for consumer needs. So that it can be said that banks are the core of each country's financial system. In the implementation of credit, besides helping to improve the people's economy, banks also get a source of revenue in the form of profit sharing. However, in the process of bank lending it is often faced with a risk known as problem loans. One of the causes of problem loans is the failure of the bank in conducting a credit analysis of prospective borrowers. In determining the decision to give a loan depends on the models and algorithms that each financial service has to reduce risk in their operations [Turan, 2016]. Empirical studies [Doblas-Madrid et al, 2013; Kusi et al, 2017] shows that credit information sharing reduces adverse bank selection, moral danger and serves as a motivation for repayment of loans which apparently reduces bank bad debts and improves the quality of bank assets.

A good credit scoring model should be able to correctly rank the customers from low to high probabilities of default, which is the basis of approval strategies [Fang et al, 2018]. Studying credit risk in commercial banks from Pakistan using models Credit assessment used credit valuation techniques such as credit rating models for individuals, logistic regression and discriminant analysis [Semreen et al, 2012]. Studied the optimal credit growth assessment from the perspective of financial stability based on the functional relationship between the square deviation of quarterly credit growth from the desired level and changes in provision costs one year later [Jakubik et al, 2015]. Examines credit information sharing through private credit bureaus and public creditregistries and their effect on bank credit risk in low and high income countries in Africa [Kusi et al, 2017]. The method of approach has been carried out by researchers on credit valuation. However, the analysis of credit valuation has not been widely applied to financial services.

Therefore, in this paper a discussion on credit assessment for financial service cooperatives is expanded using the estimation of the Particle Swarm Optimization algorithm with the Logistic Regression model. As an object taken from one of the financial services in Indonesia. Credit assessment analysis is done to minimize the risk of credit or non-performing loans.

2. Data and Methodology

2.1 Data

In this study data was obtained from the Cooperative of Financial Services in the period 2001-2011. The data obtained consisted of 100 samples, which were divided into 2 categories, namely category 0 as a credit that was not problematic or said to be feasible amounting to $n_0 = 85$ and category 1 which was said to be problematic or non-feasible credit totaling $n_0 = 15$. The variables that influence the credit assessment analysis are 8 variables, including the age of debtors (X_1), family dependents (X_2), the amount of savings (X_3), the value of collateral (X_4), the amount of income per month (X_5), given the credit limit (X_6), take home pay (X_7), and the loan term (X_8) [Sukono et al, 2014].

After the data is obtained, the first step is to do a normalization test with the aim to determine the data that has been collected normally distributed or not. Normality tests are carried out to avoid striking differences in values that can cause bias in data analysis. This test is carried out using the help of SPSS statistical software. Data that has been normalized is then carried out standard assessment with a logistic regression model. Next, the Particle Swarm Optimization programming is done using Matlab software.

2.2 Logistic Regression Model

Logistic regression is a prediction model used in classification aimed at determining a causal relationship between a variable and other variables [Bekhet et al, 2014]. The variable "cause" can be called an explanatory variable an independent variable (X). While the "effect" variable is the variable that is affected the dependent variable (Y) [Agresti, 2002].

Regression Analysis or called logistic model or logit model is one part of Regression Analysis which is used to analyze data with response variables consisting of qualitative data [Hosmer, 2013]. Logistic Regression Analysis is used to predict the probability of an event occurring, by matching the data to the logit function. This method is a general linear model used for binomial regression. Logistic regression does not require the assumption of normality, heteroscedasticity, and autocorrelation, because the dependent variable is dichotomic / binary. Variables that only have two categories, namely the category that states the success event 0 and the category that states the failure event 1. The parameters of the Binary Logistic Regression model are estimated by the Maximum Likelihood (MLE) method which is then solved by the Particle Swarm Optimization algorithm. The binary logistic regression equation used in this study are:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_i xi)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \ldots + \beta_i x_{ii})} \qquad i = 1,2,\ldots,N \tag{1}$$

with $\pi(x_i)$ is the chance of a successful event with probability values $0 \le \pi(x_i) \le 1$ and β_i are parameter values with $i = 1,2,\ldots,p$. $\pi(x_i)$ is a non-linear function, so it needs to be transformed into a logit form to obtain linear functions so that the relationship between independent variables and non-independent variables can be seen. Transformation of logit $\pi(x_i)$ is done, so the simpler equation is obtained, namely:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = (\beta_0 + \beta_1 x_1 + \ldots + \beta_i x_i) \tag{2}$$

The essential assumptions of logistic regression are independence between the successive observations and the existence of a linear relationship between logit (x) and the predictors $X_1, X_2, \ldots, X_i$. One of the necessary considerations before applying the logistic regression model is to determine whether the relationship between the independent variable and the probability of the event changes its sense or direction, or not.

2.3 Estimation of Logistic Regression Parameters

The purpose of logistic regression is to estimate the parameter $\beta_i(i = 0,1,\ldots,p)$ which has an effect on equation (2) [Sohn et al, 2016]. Suppose there are independent variables $X_1, X_2, \ldots, X_i$, the conditional density function Y to β follows the Bernoulli Distribution as follows [Montrenko et al, 2014]:

$$f(y|\beta) = \prod_{i=1}^{N} \pi_i^{\,y_i}(1-\pi_i)^{1-y_i} \qquad y_i = 0,1 \tag{3}$$

The Y_i variable is given a code 0 and 1 for each pair $(X_i Y_i)$. If $Y_i = 1$, the contribution to the likelihood function is $\pi(X_i)$, and if $Y_i = 0$ then the contribution to the likelihood function is $1 - \pi(X_i)$, where $\pi(X_i)$ denotes the value of π (x) on x_i. So that the contribution to the likelihood function of the pair (x_i, y_i) can be written as follows [Feelders, 2000]:

$$L(\beta) = \prod_{i=1}^{N} \pi_i^{\,y_i}(1-\pi_i)^{1-y_i} \qquad y_i = 0,1 \tag{4}$$

Substitute equation (1) to equation (4)

$$L(\beta) = \prod_{i=1}^{N} \left(e^{\sum_{i=0}^{p}\beta_i x_{ii}}\right)^{Y_i} \left(1 + e^{\sum_{i=0}^{p}\beta_i x_{ii}}\right)^{-1} \tag{5}$$

from equation (5) is logically fed with natural logarithms, so the log likelihood function is obtained as follows [Dellin et al, 2005]:

$$l(\beta) = \sum_{i=1}^{N} \left\{ y_i \sum_{i=0}^{p} \beta_i x_{ii} - \ln\left(1 + e^{\sum_{i=0}^{p}\beta_i x_{ii}}\right) \right\} \tag{6}$$

The vector element β in equation (6) is estimated to use Particle Swarm Optimization.

2.4 Particle Swarm Optimization

The PSO algorithm is one of the optimization algorithms that can be used for decision making. But it can also be used for disbursing lines. PSO is an optimization technique by continuously calculating prospective solutions using a quality reference. This algorithm optimizes problems by moving particles / prospective solutions in the problem space using certain functions for the position and velocity of the particles. Particle Swarm has two main operators: Speed updates and position updates. During each generation each particle is accelerated towards the previous best particle position and global best position. At each iteration, the new velocity value for each particle is calculated based on the current velocity, the distance from the previous best position, and the distance from the best global position. A new speed value is then used to calculate the next position of the particles in the search space. This process is then repeated several times, or until the minimum error is reached.

PSO simulates the behavior of group birds. Suppose the following scenario: a group of birds randomly looking for food in an area. There is only one piece of food in the area that is being sought. All birds don't know where the food is. But they know how far the food is in each iteration. So, what is the best strategy for finding food? What is effective is to follow the bird closest to food.

The PSO is initialized with a group of random particles (solutions) and then searches optimally by updating generations. In each iteration, each particle is updated by following two "best" values. The first is the best solution (fitness) that has been achieved so far, this value is called pbest. The "best" value obtained by each particle in the population is called the best global and is called the gbest because it does solid work in a challenging, high-dimensional, non-convex, continuous environment. When a particle takes part of the population as a neighbor topology, the best value is the best local and is called pbest [Kiran, 2017].

Below are two equations that form the PSO algorithm. As a head, "*k*" references the current iteration, therefore "$k+1$" implies the next iteration.

Particle position

$$x_{i+1}^{p} = x_{i}^{p} + v_{i+1}^{p} \tag{8}$$

Every particle that raises a point in the D-dimensional space that has a velocity value that allows flying through the problem space. Modification of the position of particles is done by using the information of the previous position and its current speed with the following equation [Qasim et al, 2018]:

$$v_{i+1}^{p} = w_i v_i^{p} + c_1 r_1 \left(p_i^{p} - x_i^{p}\right) + c_2 r_2 \left(p_i^{k} - x_i^{p}\right) \tag{9}$$

Where

x_i^p	: position of particles.
v_i^p	: particle speed.
p_i^p	: position of the best individual particles.
p_i^g	: position of the best bunch.
w_i	: constant inertia weight.
$c_1 c_2$	: cognitive and social parameters respectively.
$r_1 r_2$	: random number between 0 and 1.

From the particle velocity equation, two important groups appear:

1. Social terms: $c_2 r_2 \left(p_i^g - x_i^p\right)$.
2. Cognitive terms: $c_1 r_1 \left(p_i^p - x_i^p\right)$

Using these two simple equations, the basic structure of PSO routine flow is as follows:

A) Initialization

a. Set constants : $k_{\max}, w_i, c_1, c_2$.
b. Initialize the particle position randomly.
c. Initialize particle speed randomly.
d. Set $i = 1$ (calculate iterations).

B) Optimize

a. Evaluating cost functions f_i^p at each particle position x_i^p.
b. If $f_i^p \le f_{best}^p$ then $f_{best}^p \le f_i^p$ and $f_i^p \le x_i^p$.
c. If $f_i^p \le f_{best}^g$ then $f_{best}^g \le f_i^p$ and $p_i^g \le x_i^p$

d. If the condition stops fulfilling, proceed to C.
e. Update all particle speeds.
f. Update all particle positions.
g. Increase i.
h. Return to B (a).

C) End

The main concept behind PSO, as evidenced by the particle velocity equation above, is that there is a constant balance between three different forces that attract each particle:

a. Previous particle speed (inertia).
b. Distance from the position of the most famous particles (cognitive strength).
c. Distance from famous position herd (social power).

The parameters of particles (p_{best} and g_{best}) are updated at each iteration. Fitness function is calculated for each particle to find the best solution in these arch space. This solution value is stored and represents the pbest (position). The value gbest is a global optimal value for the whole population [Zouache et al, 2015].

2.5 Estimated Significance Test Parameters

Estimation is estimating certain characteristics of a particular population or it can be called the estimated population value (parameter) using the sample value. The way to draw conclusions about parameters relates to ways of estimating parameter prices. So, the unknown price of the actual parameter will be estimated based on sample statistics taken from the population concerned. In this paper, a significance test is carried out from a logistic model in order to see which significant independent variables are. The significance test of the logistic regression model in this paper was carried out with several statistical tests, namely: Likelihood Ratio Test, Wald Test, Hosmer & Lemeshow Test, and R-Square $\left(R^2\right)$.

Test of Likelihood Ratio

Maximum Likelihood is a method that can be used to estimate a parameter in a regression in order to get an estimator for unknown parameters from a population with a maximum Likelihood function. Before forming a logistic regression model, a significant parameter test was first carried out. The first test conducted to test the effect of the role of parameters in the overall model, namely the hypothesis as follows: Maximum Likelihood is a method that can be used to estimate a parameter in a regression with the aim of obtaining an estimator for unknown parameters from a population with a maximum probability function. According to Hosmer and Lameshow (1989; 2013), the Likelihood Ratio test is to test the significance of all the coefficients of the independent variables in the model shown by the G statistics, whose equations are as follows:

$H_0 : \beta_1 = \beta_2 = \ldots \beta_i = 0$ (the model has no significant effect)

$H_1 : \exists \beta_1 \neq \beta_2 \neq \ldots \beta_i \neq 0$ (the model has a significant effect on the model)

Where $i = 0,1,\ldots,p$

According to Hosmer and Lameshow (1989), the Likelihood Ratio test is to test the significance of all the coefficients of the independent variables in the model shown by the G statistics, whose equations are as follows:

$$G = 2\left[\sum_{i=1}^{n} y_i \ln \hat{\pi}_i + \sum_{i=1}^{n}(1-y_i)\ln(1-\hat{\pi}_i) - n_1 \ln n_1 - n_0 \ln n_0 + n \ln n\right] \quad (9)$$

The hypothesis for the Likelihood ratio test is $H_0 : \beta_0 = \beta_1 = \ldots = \beta_i = 0$, with alternatives $H_1 : \exists \beta_0 \neq \beta_1 \neq \ldots \neq \beta_i \neq 0 \quad (i = 0,1,\ldots,p)$. Because the statistic G follows the Chi-Square distribution with the degree of freedom equal to the number of independent variables. The criteria used are: if $G \geq \chi^2_{(1-\alpha)(df)}$

then H_0 rejected and if $G < \chi^2{}_{(1-\alpha)(df)}$ then H_0 be accepted. Where α is the significant level specified, and $df = m - 1$ with m number of model parameters.

Wald Test

According to Hosmer and Lemeshow (1989; 2013), to test the significance of $\beta_i (i = 0,1,\dots,p)$ parameters, the Wald test is used individually. The Wald test uses Z statistics, where this Z statistic follows the Raw Normal distribution. The Z statistics are:

$$Z = \frac{\beta_1}{SE(\beta_1)} \quad ; i = 0,1,\dots,p \tag{10}$$

Where β_1 is the estimator for parameters (β_1) and $SE(\beta_1)$= estimator of standard error for the coefficient β_1. The Wald test hypothesis is $H_0 : \beta_i = 0$. With alternatives $H_1 : \beta_i \neq 0 \quad (i = 0,1,\dots,p)$. The criteria used if $-Z_{\frac{1}{2}(1-\alpha)} < Z < Z_{\frac{1}{2}(1-\alpha)}$ then H_0 accepted and if $Z_{\frac{1}{2}(\alpha)} \leq Z \leq Z_{\frac{1}{2}(1-\alpha)}$ then H_0 rejected. Where $Z_{\frac{1}{2}(\alpha)}$ is the percentile of a standard normal distribution with level significance α.

Hosmer & Lemeshow Test

According to Hosmer and Lemeshow (1989; 2013), the Hosmer and Lemeshow test is known as the Logistic Regression Model compatibility test for data. The equation of this test is as follows:

$$C = \sum_{i=1}^{g} \frac{(o_i - n_i \bar{\pi}_i)^2}{n_i \bar{\pi}_i (1 - \bar{\pi}_i)} \tag{11}$$

The hypothesis used is as follows:

H_0 :there is no difference between the results of observations with the model used

H_1 :there is a difference between the results of observations with the model used

This Hosmer and Lemeshow test will follow the Chi-Square distribution with degrees of freedom $df = (g - 2)$. In general use $g = 10$. Test the criteria used, namely: H_0 rejected if $C > \chi^2{}_{(1-\alpha)(g)}$ and H_0 accepted if $C < \chi^2{}_{(1-\alpha)(g)}$. Where the α level of significance is determined.

R-Squared Test R^2

According to Hosmer and Lemeshow (1989; 2013), the value of R^2 in the Logistic Regression analysis shows the strong relationship between independent variables and free variables. For the value of R^2 it is:

$$R^2 = 1 - \exp\left[-\left(\frac{L^2}{n}\right)\right] \tag{12}$$

where: L = log Likelihood value of the model and n = amount of data. If $R^2 \to 1$, then the relationship between the independent variable and the dependent variable is strong and if $R^2 \to 0$ then the relationship is weak.

3. Results and Analysis

Data analysis is done in the manner as described in section 2.1. Before the data is carried out further, it is necessary to test the normality that applies to multivariate analysts. In the multivariate analysis the normality test aims to determine whether the data distribution is close to or follows a normal distribution. Data that has a pattern like a normal distribution is good data for multivariate analysis. Data normality test is done by using SPSS. Data can be used to estimate the parameters of a binary logistic model after data is normally distributed.

3.1 Results

The vector estimator $\beta = (\beta_{0,}\beta_1,\ldots,\beta_8)$ is determined using the parameter estimation of Binary Logistic Regression by maximizing the Likelihood function in equation (6). Estimates are performed using the Particle Swarm Optimization algorithm as described in section 2.4. Parameter estimation using Particle Swarm Optimization algorithm is done using Matlab, while for estimation and Standard Error (SE) value is do using SPSS. Parameter and Standrad Error estimation results are shown in Table 1.

Testing the significance aims to test parameter estimators that influence the dependency variable $\pi(X)$. Testing the significance aims to test parameter estimators that influence the dependency variable. To find out the significant factors on the model carried out by Wald test, namely to test the overall parameters to obtain the best value by minimizing several parameters. This step is done by matching a model that only contains significant variables. The test uses the (Z) ratio test that is using equation (10), with the hypothesis test used is $H_0 : \beta_i = 0$ with alternative $H_1 : \beta_i \neq 0$ $(i = 0,1,\ldots,8)$. Used $\alpha = 0.05, Z_{\frac{1}{2}(0.05)} = -0.20$ and $Z_{\frac{1}{2}(0.05)} = 0.20$. Since $-0.27 \leq Z \leq 0.27$, it H_0 is accepted, and H_0 the others are rejected. The results are as follows:

Table 1. Significant Variable Parameter Estimates

Parameter Coefficient (X_i)	Parameter Estimator (β_i)	Standard Error $SE(\beta_i)$	Ratio (Z) $\frac{\beta_i}{SE(\beta_i)}$	Significance
Age (X_1)	-0.896069	0.974	-0.919988706	Signifikance
Family dependents (X_2)	-0.118182	0.415	-0.284775903	Signifikance
The amount of savings (X_3)	0.382378	0.650	0.588273846	Signifikance
The value of collateral (X_4)	2.083042	1.918	1.086049009	Signifikance
Given the credit limit (X_6)	-0.477245	2.045	-0.233371638	Signifikance
The loan term (X_8)	0.791815	0.646	1.225719814	Signifikance

Maximum Likelihood value $\hat{\beta} = 30.0000$

Then, the results obtained β_5 and β_7 are not significant. Because of parameter estimation β_5 and β_7 not significant, this parameter is removed from the model because the parameters do not significantly affect the model. Therefore, it is necessary to re-estimate the model without including β_5 and β_7.

Based on the Wald test produced in Table 1, it can be seen that the factors that influence decision making significantly do not have a significant effect on the level $\alpha = 5\%$. The difference between the two models is Family dependents (X_5) and Take home pay (X_7) for the loan to be removed from the original model. The Likelihood ratio test that compares these two models is obtained using G definitions, which follow the Chi-Square distribution.

The re-estimation results of the Hosmer & Lemeshow Test aim to analyze the suitability of the Logistics model with data. The Hosmer & Lemeshow statistical test is done by equation (11). The hypothesis used is:

H_0 : observations with predictions of the same model

H_1 : observations with different model predictions

The Hosmer & Lemeshow statistical test uses equation (11) or uses statistics $P-Value$, the test criteria used, are H_0 rejected if $P-Value$ less than a significant level and H_0 accepted if $P-Value$ it is greater than the significant level. Significant level used is $\alpha = 0.05$. In this study, $P - Value$ obtained is 0.317, therefore $P - Value > \alpha$ and H_0 the hypothesis is accepted, which means "there is no difference between observation and model estimator".

The next step is to examine the relationship between independent variables and dependent variables based on R^2 values. This step follows equation (12) which is to produce a R^2 value of 0.94174283, which is obtained from re-estimation data that shows the relationship between independent variables, namely age, family dependence, amount of savings, collateral value, credit limit, and loan period, that dependence on variables $\pi(X)$ very significant probability. Thus, the estimated logistic regression based on re-estimation has the following equation:

$$\pi(X) = e^{3.00000 - 0.919988706 X_1 - 0.284775903 X_2 + 0.588273846 X_3 + 1.086049009 X_4 = 0.233371638 X_6 + 1.225719814 X_8}$$

The Logistic Regression equation above is a default problem obtained from debtor data in the opinion of the borrower credit.

3.2 Analysis

Credit worthiness decisions are made using risk prediction (valuation) by considering the probability of defaulting on the prospective debtor. Here are the credit risk titles:

Table 2. Predicate Credit Risk

Probability of default (Credit risk)	Predicate	Description
$0.00 < \pi(X) \leq 0.49$	A	Decent
$0.49 < \pi(X) \leq 0.69$	B	Pretty Decent
$0.69 < \pi(X) \leq 1.00$	C	Not Decent

The credit risk predicate above can be used to analyze credit in financial service cooperatives. The analysis of credit scoring aims to reduce the risk of default on prospective borrowers which results in losses for financial services. Therefore, financial service cooperatives to avoid losses must form a credit assessment model based on their own debtors.

4. Conclusion

The Particle Swarm Optimization algorithm is used as a Logistic Regression estimator with the aim of analyzing credit in financial service cooperatives. In this study, it was conducted on a financial service. It consists of eight factors analyzed, but there are six factors that influence significantly the probability of failure. Estimated probability of default is estimated by the logistic regression model, which is then matched with the loan feasibility interval to obtain the title of each prospective debtor. Based on the predicate obtained, the prospective debtor will obtain a decision from a financial service whether or not the prospective debtor is granted a loan.

Acknowledgements

Acknowledgments are conveyed to the Rector, Director of Directorate of Research, Community Involvement and Innovation, and the Dean of Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran.

References

Agresti, A. (2002) Categorical Data Analysis, New York: John Wiley & Sons, Inc.,

Bekhet,H. A. Eletter, S. F. K.,(2014) Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. Review of Development Finance. Vol 4, Pages 20–28. http://dx.doi.org/10.1016/j.rdf.2014.03.002

Czepiel, S.A. (2005). Maximum Likelihood Estimation of Logistic Regression. Models: Theory and Implementation, (On Line)

Dellien, H., Schreiner, M. (2005). Credit Scoring, Banks, and Microfinance: Balancing High-Tech with High-Touch, Working Paper, Women's World Banking, 8 West 40th Street, 9th Floor, New York, NY 10018, U.S.A.

Doblas-Madrid, A, & Minetti, R. (2013). Sharing information in the credit market: Contract-level evidence from US firms. Journal of financial Economics, Vol 109(1), Pages 198-223.

D. Zouache, et all (2015). Quantum-inspired firefly algorithm with particle swarm optimization for discrete optimization problems, Soft Computing Vol 20. Pages 2781–2799.

Fang, F., & Chen, Y, (2018). A New Approach for Credit Scoring by Directly Maximizing the Kolmogorov-Smirnov Statistic. Computational Statistic and Data Analysis Vol 133 (2), Pages 180-194

Feelders, A.J, (2000). Credit Scoring and Reject Inference With Mixture. Models, Working Paper, Tilburg University, The Netherlands, Copyright © 2000 John Wiley & Sons, Ltd.

Hosmer, D.W., Lemeshow, S, Sturdivant, R.X, (2013). Applied Logistic Regression. John Wiley & Sons. ISBN: 978-0-470-58247-3

Hosmer, D.W., Lemeshow, S, Applied Logistic Regression, Canada : John Wiley & Sons, Inc., (1989).

Jakubik, P. Moinescu, B, (2015). Assessing optimal credit growth for an emerging banking system, Economic Systems http://dx.doi.org/10.1016/j.ecosys.2015.01.004

Kusi, B. A, & Ansah-Adu, K, (2015). Credit Information Sharing and its Impact on Access to Bank Credit across Income Bracket Groupings. International Journal, Vol 4(4).

Kusi, B. A, Komla, E. Ansah-Adu, Kwadjo, (2017). Bank Credit Risk and Credit Information Sharing in Africa: Does Credit Information Sharing Institutions and Context Matter?.Research in International Business and Finance. http://dx.doi.org/10.1016/j.ribaf.2017.07.047.

Kiran, M. S.,(2017), Particle swarm optimization with a new update mechanism, Appl. Soft Comput. Vol 60, pages 670–678.

Montrenko, A., Strijov, V, Weber, G, (2014). Sample Size Determination for Logistic Regression. Journal of Computational and Applied Mathematics 255 (2014) 743–752. http://dx.doi.org/10.1016/j.cam.2013.06.031.

Qasim, O. S., Algamal, Z. Y (2018). Feature Selection Using Partisle Swarm Optimization-based Logistic Regression Model. Chemometrics and Intelligent Laboratory Systems 182 (2018) 41–46. https://doi.org/10.1016/j.chemolab.2018.08.016

Samreen, A & Farheen Batul Zaidi, (2012). Design and Development of Credit Scoring Model for the Commercial banks of Pakistan: Forecasting Crediworthiness of Individual Borrowers,International Journal of Business and Social Science, Vol 3 No 17.

Sohn, S. Y., Kim, H. D, & Yoon, J.H. (2016) Techonology Credit Scoring Model with Fuzzy Logistic Regression. Applied Soft Computing, 43 (2016) 150-158 http://dx.doi.org/10.1016/j.asoc.2016.02.025.

Sukono, Sholahuddin, A, Mamat, M, Prafidya, K, (2014). Credit Scoring for Cooperative of Financial Services Using Logistic Regression Estimated by Genetic Algorithm. Journal for Theory and Applications. Vol. 8, no. 1-4, 2014

Tan, Y. (2015). The impact of risk and competition on bank profitability in China. J. Int. Finance. Markets. Inst. Money. doi: 10.1016/j.intfin.2015.09.003

Turan, H. (2016). The Weighting of Factors Affecting Credit Risk in Banking. Istanbul Conference of Economics and Finance, Vol 38 pages 49 – 53.

Wang, D, Zhang, Z., Bai, R., Mao, (2017). A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring, Journal of Computational and Applied Mathematics. http://dx.doi.org/10.1016/j.cam.2017.04.036.

Biographies

Ulfa Rahmani is a magister student in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, the field of applied mathematics, with a field of actuarial science.

Sukono is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Currently serves as Head of Master's Program in Mathematics, the field of applied mathematics, with a field of concentration of financial mathematics and actuarial sciences.

Riaman is a lecturer in the Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Her research focuses on Statistics, with a field of Insurance Mathematics

Subiyanto is a lecturer in the Department of Marine Science, Faculty of Fishery and Marine Science, Universitas Padjadjaran. He received his Ph.D in School of Ocean Engineering from Universiti Malaysia Terengganu (UMT), Malaysia in 2017. His research focuses on applied mathematics, numerical analysis and computational science.

Abdul Talib Bon is a professor of Production and Operations Management in the Faculty of Technology Management and Business at the Universiti Tun Hussein Onn Malaysia since 1999. He has a PhD in Computer Science, which he obtained from the Universite de La Rochelle, France in the year 2008. His doctoral thesis was on topic Process Quality Improvement on Beltline Moulding Manufacturing. He studied Business Administration in the Universiti Kebangsaan Malaysia for which he was awarded the MBA in the year 1998. He’s bachelor degree and diploma in Mechanical Engineering which his obtained from the Universiti Teknologi Malaysia. He received his postgraduate certificate in Mechatronics and Robotics from Carlisle, United Kingdom in 1997. He had published more 150 International Proceedings and International Journals and 8 books. He is a member of MSORSM, IIF, IEOM, IIE, INFORMS, TAM and MIM.