Credit Scoring for *Cooperative of Financial Services* Using Logistic Regression Estimated by Genetic Algorithm

Sukono¹, Asep Sholahuddin², Mustafa Mamat³,⁵ and Krishna Prafidya⁴

¹,²,⁴ Department of Mathematics, University of Padjadjaran, Indonesia
³ Research Fellow, Department of Mathematics, Universiti Malaysia Terengganu, Malaysia
⁵ Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin 22200 Besut, Terengganu, Malaysia

Abstract. One of the *Cooperative of Financial Services* is disbursed loans to debtors (members and prospective members). In lending (provision of credit) is likely to arise the problem, namely the possibility of debt default by the debtor. To anticipate the risk of default (credit risk), to prospective debtors applying for credit risk analysis was performed using credit scoring. In this paper the analysis of credit scoring is done using logistic regression model, which is estimated using genetic algorithms. As a numerical illustration, the method used to analyze the credit scoring on a *cooperative of financial services* in Indonesia. Of the eight factors were analyzed, it was only six factors that significantly influence to the risk of default. Six of these factors include: number of dependents, the amount of savings, the value of collateral, monthly income, credit limit is realized, and the loan repayment period.

Mathematical Subject Classification: 62H20, 62J99, 68T20

Keywords: credit risk, management risk, credit scoring, logistic regression, genetic algorithms
1 Introduction

Lending is an activity that dominates the cooperative of financial services efforts in its function as microfinance institutions. Credit transaction occurs when there is a desire, especially our customers who need additional funds to accelerate its business or for consumptive needs. In lending, the cooperative of financial services are often faced with a known risk with credit risk or problem loans. One factor is the cause of the problem loans for the failure of cooperative of financial services in performing credit analysis of prospective debtors (Thanh & Kleimeier [14]). These risks can be minimized by using the scoring system. This scoring model was built based on data belonging to the debtors good or bad category (Emel et al. [5]).

Credit scoring analysis can be performed using several approaches or models. Bartolozzi et al. [2], conducted a study on credit scoring in retail banking sector. In their research, Bartolozzi et al., using a logistic function (logit) for perform modeling of credit scoring. For estimate the parameters, Bartolozzi et al. using the simulation approach. Koh et al. [12], studied the formation of credit scoring models to financial institutions in the German bank. In conducted the establishment of credit scoring models, Koh, et al. using data mining approaches to the analysis. Wu [15] constructed credit scoring models using multiple linear regression models, and to obtain the model fitting is done by bootstrap confidence interval approach. Lahsasna, et al. [13], analyzed credit scoring models using soft computing methods. The main modeling issues are discussed especially from the data mining point of view. To model credit scoring in the commercial bank, Islam [11] applied artificial neural networks and genetic algorithms approaches. Comparing the results by the two approaches suggests not significantly different. Modeling of credit scoring had been done by many researchers using different approaches. However, analysis of credit scoring is carried out generally to the banking system. Analysis of the credit scoring on cooperative of financial services are still not widely applied.

In fact, operationally, the function of the cooperative of financial services as well as the bank's operations. Accordingly, in this paper we carried out the development of credit scoring for the cooperative of financial services. Analyses were performed using logistic regression model, which is estimated using genetic algorithms. Analysis of credit scoring on a cooperative of financial services is important in order to minimize credit risk or problem loans. The cooperative of financial services in Indonesia is considered as the object of study.

2 Material and Methodology

2.1 Material

The data used in this study was a Cooperative of Financial Services in period 2001-2011. The data consists of 100 samples, which $n_1 = 15$ contained as
Credit scoring for cooperative of financial services

47

much as a category of 1, meaning that there is a problem loans, and \( n_0 = 85 \) as a category of 0, meaning that there is not problem loans. Factors contributing to the problem loans are studied include 8 (eight), namely: 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 \((X_8)\). To the data, which will require normality test, as discussed below.

The normality test of data is done because the independence variable in data values fluctuating high and low. Striking differences in values can lead to bias in the data analysis, so it does not reflect the real situation. The normality test of data is done by using statistical software SPSS 17.0. Once the data is then used in the standardized assessment of logistic regression models. Furthermore, for programming genetic algorithm is done by using the software of Matlab 7.0.

2.2 Binary Logistic Regression Model

According to Agresti [1], binary logistic regression used to estimate the influence of some explanatory variable \((X)\) to the response variable \((Y)\) that is binary or dichotomous. Said to be binary or dichotomous because it has two possible values that is 0 and 1. Form a binary logistic regression equation used in this study is

\[
\pi(X_i) = \frac{e^{\sum_{k=0}^{K} \beta_k X_{ik}}}{1 + e^{\sum_{k=0}^{K} \beta_k X_{ik}}} \quad ; \quad i = 1, \ldots, N
\]  

(1)

If the equation (1) is transformed by the natural logarithm, it will obtain the linear regression equation as follows:

\[
g(X_i) = \ln \left( \frac{\pi(X_i)}{1 - \pi(X_i)} \right) = \sum_{k=0}^{K} \beta_k X_{ik} \quad ; \quad i = 1, \ldots, N
\]  

(2)

2.3 Logistic Regression Parameter Estimation

Suppose that \( \beta = (\beta_0, \beta_1, \ldots, \beta_K) \) is a vector of logistic regression parameters. According to Czepiel [4], the goal of logistic regression is to estimate the parameters \( \beta_k \quad (k = 0,1,\ldots,K) \) which contribute to equation (2). Suppose there are \( k \) independence variables \( X_1, \ldots, X_K \), conditional density function of the dependence variable \( Y \) towards \( \beta \) opportunity to follow a Bernoulli distribution

\[
f(Y \mid \beta) = \prod_{i=1}^{N} \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i} \quad ; \quad Y_i = 0,1
\]  

(3)

Maximum likelihood estimators are the values that \( \beta \) can maximize likelihood function of equation (3).
Variable $Y_i$ is given the value 0 or 1 for each pair $(X_i, Y_i)$. If $Y_i = 1$ then the contribution to the likelihood function is $\pi(X_i)$, and if $Y_i = 0$ then the contribution is $1 - \pi(X_i)$. Thus contribute to the likelihood function of the pair $(X_i, Y_i)$ can be written as (Feelders [6]):

$$L(\beta) = \prod_{i=1}^{N} \pi_i^{Y_i} (1 - \pi_i)^{1-Y_i}; Y_i = 0,1.$$  

(4)

Substituting (1) into (4) given the equation

$$L(\beta) = \prod_{i=1}^{N} \left( e^{\sum_{k=0}^{K} \beta_k X_{ik}} \right)^{Y_i} \left( 1 + e^{\sum_{k=0}^{K} \beta_k X_{ik}} \right)^{-1}.$$  

(5)

If the left and right of equation (5) take the natural logarithm, then the equation is obtained as (Hans & Schreiner [8]):

$$\mathcal{L}(\beta) = \sum_{i=1}^{N} \left[ Y_i \sum_{k=0}^{K} \beta_k X_{ik} - \ln \left( 1 + e^{\sum_{k=0}^{K} \beta_k X_{ik}} \right) \right].$$  

(6)

Elements of the vector $\beta$ in equation (6) would then be estimated using genetic algorithms.

2.4 Genetic Algorithms

According to Hermanto (2007), genetic algorithm developed by Goldberg in 1989, is a computational algorithm that inspired Darwin's theory of evolution. In Darwin's theory of evolution, that the survival of an organism can be maintained through the process of reproduction, crossover, and mutation, and follow the rules that the strong are winning. Darwin's theory of evolution has adopted a computational algorithm to find the solution of a problem in a more "natural".

A solution generated by the genetic algorithm named chromosome, where a collection of chromosomes is called a population. A chromosome consists of compiler components are called genes, and its value can be any number of numeric, binary, symbols, or characters, depending on the issues to be resolved. Chromosomes will evolve continuously, are called generations. In each generation of chromosomes are evaluated the success rate of the value of the solution to a problem that was about to be solved (the objective function), using a measure called the fitness. To select a chromosome which will be retained for the next generation is called the selection process.

The new chromosomes are called offspring, formed by mating between chromosomes in a generation is called crossover. The number of chromosomes in the population who are determined based crossover parameter is called crossover_rate. The mechanism of changes in the constituent elements of living things as a result of natural factors is called a mutation. Mutations are represented as the change in the value of one or more genes in the chromosome
by a random value. The number of genes that have mutations in the population is determined by a parameter called \textit{mutation\_rate}. After several generations, the value will be generated chromosomes genes is converges to a certain value. These genes are the best solution generated by the genetic algorithm for the problems solved.

An application of this research is the genetic algorithm where it is used to estimate logistic regression model parameters, ie, used in maximizing the log likelihood function is expressed as equation (6). The maximum solution of equation (6) using genetic algorithms can be found by the following steps:

1) Establishment of chromosome. Since the sought value is $\beta_k$ ($k = 0,1,\ldots, K$), then used as a parameter $\beta_k$ forming chromosome genes. The parameter $\beta_k$ thresholds are members of real numbers.

2) Initialization. Initialization process is done by providing an initial value of genes with random values corresponding restrictions set.

3) Evaluation of chromosome. The problems were solved is to determine the value of $\beta_k$ in equation (6), therefore the objective function as a chromosome that can be used is equation (6).

4) Selection chromosome. The selection process is done by making chromosome which has an objective function value is small, has likely chosen a large or have a high probability. In this case, the fitness function can be used $fitness = 1/(1 + objectif\_function)$, plus 1 divider necessary to avoid the possibility of division by zero. While it can be used to find the probability formula $P[i] = fitness[i] / total\_fitness$. For the selection process can be used roulette wheel (random number generator), it is necessary to look first cumulative probability $C[k]$. Having calculated the value of the cumulative probability and random numbers generated $R$ in the range $[0, 1]$. If $R[k] < C[1]$ then select chromosome 1 as the parent, except that select k chromosomes as the parent, provided that $C[k - 1] < R < C[k]$. Do as much of the population.

5) Crossover. After the selection process, the next process is the process of crossover. Methods used one of them is a one-cut-point, which randomly selects one position in the parent chromosome, then exchanged genes. Chromosome which is used as the parent chosen at random, and the number of chromosomes that undergo crossover influenced by parameters \textit{crossover\_rate} ($\rho_c$). Suppose specified crossover probability of 25%, it is expected that within a generation there is a 50% chromosome from one generation undergo crossover.

6) Mutation. The number of chromosomes that have mutations in a population is determined by the parameters \textit{mutation\_rate}. The process of mutation is done by replacing a randomly selected genes with a new value obtained randomly. The process is as follows: first, the total length calculated previously existing genes in a population. In this case the total length of the
gene is \( \text{total}_\text{gen} = (\text{number of genes in a chromosome}) \times (\text{total population}) \). To select the position of the mutated genes is done by generating a random number between 1 to integer \( \text{total}_\text{gen} \). If the random number generated is less than the variable \( \text{mutation}_\text{rate} (\rho_m) \), then select the position as a sub-chromosome mutation. After the mutation process, meaning it has completed one of iteration of the genetic algorithm, also called a generation. This process will be repeated until a predetermined number of generations and ultimately be acquired chromosome as the optimum objective function.

2.5 Parameter Significance Tests

In testing the significance of the estimator parameter logistic regression models will be used several statistical tests, namely: Likelihood Ratio Test, Wald Test, Hosmer & Lemeshow Test, and \( R^2 \).

**Likelihood Ratio Test.** To test the significance of the parameters estimator on the whole, to use statistical

\[
\hat{G} = 2 \left[ \sum_{i=1}^{N} Y_i \ln \hat{\pi}_i + \sum_{i=1}^{N} (1-Y_i)\ln(1-\hat{\pi}_i) - n_1 \ln n_1 - n_0 \ln n_0 + N \ln N \right].
\]  

Hypothesis for the likelihood ratio test is \( H_0 : \hat{\beta}_0 = \hat{\beta}_1 = \cdots = \hat{\beta}_k = 0 \), with alternative \( H_1 : \exists \beta_0 \neq \hat{\beta}_1 \neq \cdots \neq \hat{\beta}_k \neq 0 (k = 0,1,\ldots,K) \). Because the \( \hat{G} \) statistic is asymptotically distributed Chi-Square, the test criteria used are: Reject \( H_0 \) if \( \hat{G} > \chi^2_{(1-\alpha)(df)} \), otherwise accept \( H_0 \) if \( \hat{G} \leq \chi^2_{(1-\alpha)(df)} \). Where \( \alpha \) is the significance level established, and \( df = m-1 \) with \( m \) the number of model parameters (Bolton [3]).

**Wald Test.** According to Hosmer & Lameshow [9], to test the significance of the parameter \( \hat{\beta}_k (k = 0,1,\ldots,K) \) can be used individually Wald test. Wald test \( Z \) statistic is used, where the \( Z \) statistic follows the standard normal distribution. The value of the \( Z \) statistic is

\[
\hat{Z} = \left( \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right) ; k = 0,1,\ldots,K,
\]  

where \( \hat{\beta}_k \) is estimator parameter of \( \beta_k \), and \( SE(\hat{\beta}_k) \) standard error of \( \hat{\beta}_k \).

Hypothesis for Wald test is \( H_0 : \hat{\beta}_k = 0 \), with alternative \( H_1 : \hat{\beta}_k \neq 0 (k = 0,1,\ldots,K) \). Test criteria are rejected \( H_0 \) when \( Z < Z_{\frac{1}{2}}(\alpha) \) or \( Z > Z_{\frac{1}{2}}(1-\alpha) \), the opposite received \( H_0 \) when \( Z_{\frac{1}{2}}(\alpha) \leq Z \leq Z_{\frac{1}{2}}(1-\alpha) \). Where \( Z_{\frac{1}{2}}(\alpha) \) is the percentile of the standard normal distribution with a significance level of \( \alpha \).
Credit scoring for cooperative of financial services

Hosmer & Lemeshow Test. Hosmer & Lameshow test known as a test of logistic regression models fit to the data. Statistic test of Hosmer & Lemeshow is

\[
\hat{C} = \sum_{k=1}^{g} \left( \frac{O_k - n_k \bar{\pi}_k}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \right) \quad \text{or} \quad P_{Value} = \Pr(\hat{C})
\]  

(9)

where \( O_k = \sum_{j=1}^{n_k} Y_j \) and \( \bar{\pi}_k = \sum_{j=1}^{n_k} (m_j \bar{\pi}_j / n_k) \).

The hypotheses are 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.

Hosmer & Lemeshow following the Chi-Square distribution with degrees of freedom \( df = (g - 2) \), with a general \( g = 10 \). Test criteria used were: Reject \( H_0 \) if \( \hat{C} > \chi^2_{(1-\alpha)(g)} \), otherwise accept \( H_0 \) when \( \hat{C} \leq \chi^2_{(1-\alpha)(g)} \) where \( \alpha \) the significance level established (Hosmer & Lemeshow [9]).

R-Square. According to Hosmer and Lemeshow [9], the value of \( R^2 \) in logistic regression analysis showed strong relationships between independence variables and dependence variables. Statistic of \( R^2 \) can be determined using the formula

\[
R^2 = 1 - \exp \left( -\frac{\hat{L}}{N} \right),
\]

(10)

where the \( \hat{L} \) value of the log likelihood of the model, and \( N \) the number of data. If \( R^2 \to 1 \), then the relationship between the independent variable and the independent variable is strong. Conversely, if \( R^2 \to 0 \), then relationship is weak.

3 Results and Analysis

The data were analyzed as described in section 2.1. For multivariate analysis, before the data is used to further analysis is necessary to normality test. The purpose of normality test in multivariate analysis is want to know whether distribution of the data (following or close to) normal distribution. Because good data for multivariate analysis, is the data that has a pattern like a normal distribution (not skewed to the right or to the left). Normality test of data is done using SPSS version 17.0. Once the data are normally distributed, then this data is used to estimate parameters of binary logistic models.
3.1 Results

Binary logistic regression parameter estimates performed with the aim to determine the estimator vector \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_8) \) which can maximize the log likelihood function in equation (6). Estimation is done using a genetic algorithm, where the steps were described in section 2.4. Parameter estimation using genetic algorithms is done by using Matlab 7.0. The parameter estimates and their standard error values of each parameter estimator are given in Table 1.

Towards estimator the parameters is necessary to test the significance of the overall effect on the dependence variable \( \pi(X) \). Hypothesis test is \( H_0: \hat{\beta}_0 = \hat{\beta}_1 = ... = \hat{\beta}_8 = 0 \) with alternative \( H_1: \exists \hat{\beta}_0 \neq \hat{\beta}_1 \neq ... \neq \hat{\beta}_k \neq 0 \) \((k = 0,1, ..., 8)\). Tests performed using the log likelihood ratio test of equation (7). The results are also given in Table 1.

### Table 1. Parameter and Standard Error Estimators

<table>
<thead>
<tr>
<th>Coefficient Parameter of Variables ((X_i))</th>
<th>Estimator of Parameter ((\hat{\beta}_i))</th>
<th>Error Standard ( SE(\hat{\beta}_i) )</th>
<th>Ratio ((\hat{Z}_i)) (\frac{\hat{\beta}_i}{SE(\hat{\beta}_i)})</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.68076175</td>
<td>0.539</td>
<td></td>
<td>Significance</td>
</tr>
<tr>
<td>Age of debtors ((X_1))</td>
<td>0.08972099</td>
<td>0.520</td>
<td>0.17254036</td>
<td>No Significance</td>
</tr>
<tr>
<td>Family dependents ((X_2))</td>
<td>0.76596963</td>
<td>0.517</td>
<td>1.48156602</td>
<td>Significance</td>
</tr>
<tr>
<td>Amount of savings ((X_3))</td>
<td>0.75022570</td>
<td>654.0</td>
<td>1.14713410</td>
<td>Significance</td>
</tr>
<tr>
<td>Value of collateral ((X_4))</td>
<td>2.05312936</td>
<td>1.988</td>
<td>1.03276125</td>
<td>Significance</td>
</tr>
<tr>
<td>Income per month ((X_5))</td>
<td>-2.21395298</td>
<td>3.006</td>
<td>-0.73651130</td>
<td>Significance</td>
</tr>
<tr>
<td>Credit limit ((X_6))</td>
<td>-1.042501243</td>
<td>2.068</td>
<td>-0.60411085</td>
<td>Significance</td>
</tr>
<tr>
<td>Take Home Pay ((X_7))</td>
<td>0.44284592</td>
<td>3.082</td>
<td>0.14368784</td>
<td>No Significance</td>
</tr>
<tr>
<td>Loan term ((X_8))</td>
<td>0.91824481</td>
<td>0.654</td>
<td>1.40404405</td>
<td>Significance</td>
</tr>
</tbody>
</table>

Log Likelihood Statistic = -26.82226249994948

Calculations using SPSS version 17.0 generates log likelihood ratio statistic of \( \hat{G} = -26.82226249994948 \). This \( \hat{G} \) statistic is asymptotically distributed chi-square \( \chi^2 \) with degrees of freedom, \( df = 8 \). Using significance level, \( \alpha = 0.05 \), from the table of chi-square, \( \chi^2_{(1-0.05),(8)} = 2.7326 \). Since, \( \hat{G} > \chi^2_{(1-0.05),(8)} \), then the hypothesis \( H_0 \) is rejected, which means that the estimator \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_8) \) is significantly effect on \( \pi(X) \).

Further tests carried out partially with respect to each parameter estimator, is intended to test the significance of the effect of each estimator for dependence variable \( \pi(X) \). Hypothesis test is \( H_0: \hat{\beta}_k = 0 \) with alternative \( H_1: \hat{\beta}_k \neq 0 \) \((k = 0,1, ..., 8)\). Tests carried out using the ratio \((Z)\) test or Wald
test equation (8), the calculation results of the \( \hat{Z} \) are given in Table 1. This \( \hat{Z} \) statistic is asymptotically normal distributed. Using \( \alpha = 0.05 \), 
\[
Z_{\frac{1}{2}(0.05)} = -0.27 \quad \text{and} \quad Z_{\frac{1}{2}(1-0.05)} = 0.27 .
\]
Since \(-0.27 \leq \hat{Z} \leq 0.27 \), then the \( H_0 \) is accepted, else \( H_0 \) is rejected. Based on the \( \hat{Z} \) statistics in Table 1, it is clear that the parameter estimator \( \hat{\beta}_1 \) and \( \hat{\beta}_7 \) were not significant, while the other is significant.

Because the parameter estimator of \( \hat{\beta}_1 \) and \( \hat{\beta}_7 \) are not significant, then the parameters are removed from the model. In other words, these parameters do not significantly influence the estimated models. With the issuance of the parameters \( \hat{\beta}_1 \) and \( \hat{\beta}_7 \), it is necessary to re-estimate without including variables \( X_1 \) and \( X_7 \). Because the variables \( X_1 \) and \( X_7 \) generate parameter coefficients \( \hat{\beta}_1 \) and \( \hat{\beta}_7 \) are not significant. The re-estimation is also performed with the genetic algorithm, and the results are given in Table 2.

<table>
<thead>
<tr>
<th>Coefficient Parameter of Variables ( (X_i) )</th>
<th>Estimator of Parameter ( (\hat{\beta}_i) )</th>
<th>Error Standard ( SE(\hat{\beta}_i) )</th>
<th>Ratio ( (\hat{Z}) ) ( \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} )</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.67391565</td>
<td>0.535</td>
<td>Significance</td>
<td></td>
</tr>
<tr>
<td>Family dependents ( (X_2) )</td>
<td>0.78627418</td>
<td>0.415</td>
<td>1.89463657</td>
<td>Significance</td>
</tr>
<tr>
<td>Amount of savings ( (X_3) )</td>
<td>0.72977551</td>
<td>0.650</td>
<td>1.12273155</td>
<td>Significance</td>
</tr>
<tr>
<td>Value of collateral ( (X_4) )</td>
<td>2.15363477</td>
<td>1.918</td>
<td>1.12285441</td>
<td>Significance</td>
</tr>
<tr>
<td>Income per month ( (X_5) )</td>
<td>-1.79684955</td>
<td>0.974</td>
<td>-1.84481473</td>
<td>Significance</td>
</tr>
<tr>
<td>Credit limit ( (X_6) )</td>
<td>-1.09647121</td>
<td>2.045</td>
<td>-0.53617174</td>
<td>Significance</td>
</tr>
<tr>
<td>Loan term ( (X_8) )</td>
<td>0.90983972</td>
<td>0.646</td>
<td>1.40842061</td>
<td>Significance</td>
</tr>
</tbody>
</table>

Log Likelihood = -26.84255798909220

Likelihood ratio test and Wald test has been performed in the same way as the previous estimate, and significant results is shown. After the Likelihood ratio test and Wald test, then performed a test of Hosmer & Lemeshow.

Hosmer & Lemeshow test for re-estimation results is conducted to analyze the suitability of logistic models with real data. Hypotheses for the Hosmer & Lemeshow test are as follows:

\( H_0 \) : There is no difference between the observations with the predictions of the model;

\( H_1 \) : There is a difference between the observations with the model predictions.

Statistical test of Hosmer & Lemeshow is given as equation (9). Test can also be performed using the statistical of \( P\_Value \), the test criteria, reject \( H_0 \)
if \( P\_Value \) is less than the significance level. In this test, obtained the magnitude of \( P\_Value \) is 0.392. When specified significance level of \( \alpha \) is 0.05, then obviously \( P\_Value \) is greater than the significance level. Therefore, the hypothesis \( H_0 \) is accepted, which means “there is no difference between the observations with the model estimator”.

Furthermore, to see the strength of the relationship between independent variables and the dependent variable can be done based on the value of \( R^2 \). The value of \( R^2 \) statistic can be calculated using equation (10). The value of \( R^2 = 0.99925730 \), obtained from re-estimation suggests that the relationship between independence variables: a dependent family \( (X_2) \), the amount of savings \( (X_3) \), the value of collateral \( (X_4) \), the amount of income per month \( (X_5) \), given the credit limit \( (X_6) \), and loan term \( (X_8) \), with dependence variables of default probability \( \pi(X) \) is very strong. Based on re-estimation analysis in Table 2, then the estimated logistic regression has the equation:

\[
\hat{\pi}(X) = e^{-2.67391565 + 0.78627418X_2 + 0.72977551X_3 + 2.15363477X_4 - 1.79684955X_5 - 1.09647121X_6 + 0.90983972X_8}
\]

The logistic regression equation mentioned is the probability of default of borrowers in paying the loan installments.

3.2 Analysis

Estimator based on the probability of default, the feasibility of a prospective borrower's decision can be based on the prediction of credit risk (Scoring) are presented in Table-3. Scoring is widely used by financial institutions in general. **Illustration.** Prospective debtors D, age \( (X_1) \) 40 years; have dependents \( (X_2) \) 4 persons; the amount of savings in financial services cooperatives \( (X_3) \) of Rp 750,000; will give a guarantee \( (X_4) \) of Books Motor Vehicle Owner (BMVO) value of Rp 9,375,000; the amount of income per month \( (X_5) \) is Rp 3,500,000; credit limit \( (X_6) \) the proposed Rp 11,000,000; take home pay \( (X_7) \) amounting to Rp 2,800,000, and the loan term \( (X_8) \) for 24 months (or 2 years).

<table>
<thead>
<tr>
<th>Probability of default (Credit Risk)</th>
<th>Predicate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 ( &lt; \pi(X) \leq 0.49 )</td>
<td>A</td>
<td>Eligible</td>
</tr>
<tr>
<td>0.49 ( &lt; \pi(X) \leq 0.69 )</td>
<td>B</td>
<td>Just Eligible</td>
</tr>
<tr>
<td>0.69 ( &lt; \pi(X) &lt; 1.00 )</td>
<td>C</td>
<td>No Eligible</td>
</tr>
</tbody>
</table>

**Analysis of the case.** Based on the above information, can be predicted that the probability of default (credit risk) for a prospective borrower D is
\( \pi(X) = 0.55940566 \). Based on the predicate of credit risk in Table-3, that the prospective borrower D belonging to the predicate B (Just Eligible) to receive credit with the above information. Predicate B thus, the cooperative of financial services can make the decision to accept or reject the loan application submitted by the prospective borrower D. In the event of doubt as such, the cooperative of financial services may request additional information which makes it possible to raise the credit worthiness predicate. For example, by asking to increase the collateral value \( (X_4) \), and or lower the credit limit \( (X_6) \) is proposed, so that after the analysis of default probability calculation can be increased to a predicate A (Eligible).

Such examples of applying the analysis of credit scoring in cooperative of financial services, in doing the lending. Using credit scoring analysis thus expected to reduce the level of risk of default by the debtors, which can result in losses for cooperative of financial services. Thus credit risk analysis can be done using credit scoring models which are widely used by financial institutions in general. However, in order to have a standard of lending decisions strictly in accordance with the characteristics of the debtors owned, cooperative of financial services should establish a credit scoring models based on their own debtors.

4 Conclusion

Binary logistic regression model estimator parameters estimated by using a genetic algorithm, can be used for the analysis of credit scoring in cooperative of financial services. At a cooperative of financial services, of the eight factors that were analyzed, there were only six factors that significantly affect the incidence of the possibility of default. Estimator of default probability of borrowers are estimated via binary logistic regression model, then matched with the interval eligibility of borrowers. So that each debtor has a predicted predicate that describes the level of credit default risk. Based on these predicates, cooperative of financial services can take a decision received by each debtor. Analysis of credit scoring can be done using the scoring used by financial institutions in general, but it would be nice if a cooperative of financial services shall establish its own scoring debtors based on historical data.

References


Received: October 21, 2013